



KIRGIZISTAN-TÜRKIYE MANAS ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
İŞLETME ANABİLİM DALI

BAĞIMSIZ DENETİMDE MAKİNE ÖĞRENMESİ YAKLAŞIMI:  
BORSA İSTANBUL A.Ş. ÖRNEĞİ

FATİH FAYDALI

DOKTORA TEZİ

EKİM – 2023



**KIRGIZİSTAN – TÜRKİYE MANAS ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
İŞLETME ANABİLİM DALI**

**BAĞIMSIZ DENETİMDE MAKİNE ÖĞRENMESİ YAKLAŞIMI: BORSA  
İSTANBUL A.Ş. ÖRNEĞİ**

**HAZIRLAYAN  
FATİH FAYDALI**

**DANIŞMAN  
DOÇ. DR. BİLAL SOLAK**

**DOKTORA TEZİ**

**EKİM- 2023**

## KABUL VE ONAY

Doç. Dr. Bilal SOLAK danışmanlığında Fatih FAYDALI tarafından hazırlanan “BAĞIMSIZ DENETİMDE MAKİNE ÖĞRENMESİ YAKLAŞIMI: BORSA İSTANBUL A.Ş. ÖRNEĞİ” adlı bu çalışma, jürimiz tarafından Kırgızistan - Türkiye Manas Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalında Doktora tezi olarak OYBİRLİĞİ / OYÇOKLUĞU ile kabul edilmiştir.

...../...../.....

Jüri	Unvanı, Adı Soyadı	Kurumu	Onay	İmza
Başkan				
Üye (Danışman)				
Üye				
Üye				
Üye				

## ONAY

Bu tezin kabulü Enstitü Yönetim Kurulunun ..... tarih ve ..... sayılı kararı ile onaylanmıştır.

...../...../.....

Prof. Dr. Kadir ARDIÇ  
Sosyal Bilimler Enstitüsü Müdürü

## ETİK BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun olarak sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, yararlanılan eserlere eksiksiz atıf yapıldığını, tezde yer alan verilerin hiçbir üniversitede herhangi bir tez çalışmasında kullanılmadığını ve tez yazım kurallarına uygun hazırlandığını beyan ederim.

Fatih FAYDALI

## ÖNSÖZ

İnsan benzeri karar alabilen makinelerin hayatımızın her alanına tesir ettiğini günümüzde denetimin de bu dönüşümden etkilenmemesi mümkün değildir. Bu çalışma kapsamında işletmenin sürekliliğinin değerlendirilmesinde makine öğrenmesi metotlarının kullanımı hakkında aydınlatıcı bilgiler verilmiş ve metotların etkinliğinin ortaya koyulması maksadı ile dokuz farklı model eğitilmiş ve değerlendirilmiştir. İşletmenin sürekliliğinin değerlendirilmesinde makine öğrenmesi metotlarının kullanımının daha etkin sonuçlar elde etmede önemli bir rol oynayacağı düşünülmektedir. Çalışma işletmenin sürekliliği değerlendirilirken denetçinin yaralanabileceği bir kılavuz niteliğinde tasarlanmış ve sunulmuştur. Çalışmanın denetim literatürüne katkı sağlayacağı düşünülmektedir.

Son olarak tez çalışmamın tüm kritik aşamalarında tarafıma yol gösteren ve her daim desteğini hissettiren danışmanım Doç. Dr. Bilal SOLAK hocama ve çalışmamın her aşamasında benden desteği esirgemeyen Prof. Dr. Kadir ARDIÇ hocama ve yaşamımın her anında benden desteğini esirgememiş olan annem Rukiye FAYDALI ve babam Beyhan FAYDALI başta olmak üzere tüm aileme teşekkürü bir borç bilirim.

Fatih FAYDALI

## İÇİNDEKİLER

ÖNSÖZ .....	i
İÇİNDEKİLER .....	ii
SİMGELER VE KISALTMALAR .....	vi
TABLolar LİSTESİ .....	vii
ŞEKİLLER LİSTESİ .....	ix
ÖZET .....	1
KIRGIZCA ÖZET (МАЗМУНУ) .....	3
RUSÇA ÖZET (АННОТАЦИЯ) .....	65
İNGİLİZCE ÖZET (SUMMARY).....	67
GİRİŞ .....	69
BİRİNCİ BÖLÜM .....	73
1.1. Denetim Kavramı.....	73
1.2. Muhasebe Denetiminin Tanımı ve Özellikleri.....	74
1.3. Denetim Türleri.....	76
1.3.1. Amacına Göre Denetim Türleri .....	76
1.3.1.1. Mali Tablolar Denetimi .....	76
1.3.1.2. Uygunluk Denetimi.....	77
1.3.1.3. Faaliyet Denetimi.....	77
1.3.2. Yapılış Nedenine Göre Denetim Türleri.....	78
1.3.2.1. Yasal (Zorunlu) Denetim .....	78
1.3.2.2. İsteğe Bağlı Denetim .....	78
1.3.3. Kapsamına Göre Denetim Türleri.....	78
1.3.3.1. Genel Denetim .....	78
1.3.3.2. Özel Denetim .....	79
1.3.4. Denetçinin Statüsüne Göre Denetim Türleri .....	79
1.3.4.1. Bağımsız Denetim.....	79
1.3.4.2. Kamu Denetimi.....	80
1.3.4.3. İç Denetim.....	80
1.4. Denetçi Kavramı ve Denetçi Türleri.....	81
1.4.1. İç Denetçi.....	81
1.4.2. Bağımsız Denetçi.....	81

1.4.3. Kamu Denetçisi.....	82
1.5. Genel Kabul Görmüş Denetim Standartları.....	82
1.5.1. Genel Standartlar .....	83
1.5.1.1. Mesleki Eğitim ve Deneyim .....	83
1.5.1.2. Bağımsızlık .....	84
1.5.1.3. Gerekli Mesleki Özen .....	84
1.5.2. Çalışma Alanı Standartları.....	85
1.5.2.1. Denetimin Planlanması ve Gözetim .....	85
1.5.2.2. İç Kontrol Sisteminin İncelenmesi ve Değerlendirilmesi.....	86
1.5.2.3. Yeterli Sayı ve Kalitede Kanıt Toplama.....	87
1.5.3. Raporlama Standartları .....	87
1.5.3.1. Genel Kabul Görmüş Muhasebe İlkelerine Uygunluk .....	87
1.5.3.2. Devamlılık .....	88
1.5.3.3. Açıklayıcı Bilgiler.....	88
1.5.3.4. Denetçi Görüşünün Beyanı .....	88
1.6. Uluslararası Denetim Standartları.....	89
1.7. Bağımsız Denetim Standardı 570 – İşletmenin Sürekliliği Standardı .....	90
1.7.1. İşletmenin Sürekliliği Esası .....	92
1.7.2. İşletmenin Sürekliliği ile İlgili İşletme Yönetimi ve Bağımsız Denetçinin Sorumlulukları .....	93
1.7.2.1. İşletme Yönetiminin Sorumluluğu.....	93
1.7.2.2. Bağımsız Denetçinin Sorumluluğu .....	96
1.7.3. İşletmenin Sürekliliği Varsayımı Üzerinde Ciddi Şüpheler Doğuracak Olay ve Durumlar .....	97
1.7.4. İşletmenin Sürekliliğine Dair İşletme Yönetiminin Yaptığı Değerlendirmenin Gözden Geçirilmesi .....	99
1.7.5 İşletmenin Sürekliliğine Dair İşletme Yönetiminin Yaptığı Değerlendirme Sonrası Dönem.....	100
<b>İKİNCİ BÖLÜM</b> .....	102
2.1. Makine Öğrenmesi.....	102
2.2. Öğrenme Stratejileri.....	106
2.2.1. Gözetimli Öğrenme.....	107
2.2.2. Gözetimsiz Öğrenme .....	109
2.2.3. Pekiştirmeli Öğrenme .....	109

2.3. Makine Öğrenmesi Süreci .....	110
2.4. Veri, Veri Seti ve Veri Ön İşleme Kavramları .....	112
2.4.1. Veri ve Veri Seti Kavramları .....	114
2.4.2. Gürültü .....	117
2.4.3. Örtük Değer ve Aykırı Değer Kavramları .....	117
2.4.4. Eksik Gözlem .....	118
2.4.5. Normalizasyon .....	119
2.4.6. Aşırı Uyum ve Yetersiz Uyum Sorunları .....	120
2.4.7. Öznitelik Seçimi .....	120
2.4.8. Çapraz Doğrulama .....	123
2.5. Makine Öğrenmesi Algoritmaları .....	124
2.5.1. Lojistik Regresyon .....	124
2.5.2. Karar Ağaçları .....	126
2.5.3. Destek Vektör Makineleri .....	130
2.5.4. Topluluk Öğrenmesi (Ensemble Learning) .....	132
2.5.4.1. Bootstrap Aggregating (Bagging) .....	133
2.5.4.1.1. Rastgele Orman Algoritması (Random Forest) .....	134
2.5.4.2. Boosting .....	135
2.5.4.2.1. Adaboost Algoritması .....	135
2.5.4.2.2. Gradient Boosting .....	137
2.5.4.2.3. Extreme Gradient Boosting (XGBoost) .....	138
2.6. Model Değerleme .....	138
<b>ÜÇÜNCÜ BÖLÜM</b> .....	143
3.1. Süreklilik Tahmini .....	143
3.2. İşletmenin Sürekliliğinin Tahmininde İstatistiksel Metotlardan Faydalanılan Çalışmalar .....	145
3.3. İşletmenin Sürekliliğinin Tahmininde Makine Öğrenmesi Metotlarından Faydalanılan Çalışmalar .....	149
<b>DÖRDÜNCÜ BÖLÜM</b> .....	169
4.1. İşletmenin Sürekliliğinin Tahminine Yönelik Borsa İstanbul Aş. İmalat Sektöründe Bir Uygulama .....	169
4.2. Araştırmanın Amacı ve Kapsamı .....	169
4.3. Araştırmanın Önemi .....	169

4.4. Araştırmanın Kısıtları .....	177
4.5. Araştırmanın Yaygın Etkisi .....	177
4.6. Araştırmanın Yöntemi .....	178
4.7. Çalışmada Kullanılan Veri Seti .....	180
4.8. Veri Setinde Yer Alan Mali Oranlar .....	181
4.9. R Studio Sürümü ve Kullanılan Kütüphaneler .....	185
4.10. Veri Ön İşleme Süreci.....	186
4.10.1. Eksik Verilerin Tamamlanması .....	186
4.10.2. Normalizasyon .....	188
4.10.3. Veri Setinin Dengelenmesi .....	189
4.10.4 Öznitelik Seçimi .....	189
4.11. Araştırmanın Bulguları .....	192
4.11.2. Eğitilen Modellerin Sonuçları.....	193
4.11.2.1. Lojistik Regresyon Modeli .....	194
4.11.2.2. Destek Vektör Makinesi Modelleri.....	198
4.11.2.2.1. Doğrusal Destek Vektör Makinesi Modeli .....	198
4.11.2.2.2. Polinomsal Destek Vektör Makinesi Modeli.....	201
4.11.2.2.3 Radyal Destek Vektör Makinesi Modeli .....	204
4.11.2.3. Sınıflama ve Regresyon Ağacı Modeli .....	208
4.11.2.4. Rastgele Orman Modeli .....	212
4.11.2.5. Adaboost Modeli.....	216
4.11.2.6. Gradient Boosting Modeli .....	219
4.11.2.7. XGBoost Modeli.....	222
SONUÇ VE ÖNERİLER.....	227
<b>KAYNAKLAR</b> .....	235
<b>ÖZGEÇMİŞ</b> .....	267

## SİMGELER VE KISALTMALAR

Adaboost	: Adaptive Boosting
Bagging	: Bootstrap Aggregating
BDS	: Bağımsız Denetim Standardı
DVM	: Destek Vektör Makineleri
LVQ	: Learning Vector Quantization
RPA	: Robotic Process Automation
SOM	: Self-Organizing Map
WRDS	: Wharton Research Data Services
XGboost	: Extreme Gradient Boosting
YSA	: Yapay Sinir Ağları

## TABLULAR LİSTESİ

<b>Tablo 1.1</b>	Uluslararası Bağımsız Denetim Standartları.....	90
<b>Tablo 4.1</b>	Veri Setinde Yer Alan Gözlem Yıllarının Sayısı.....	182
<b>Tablo 4.2</b>	Veri setinde Yer Alan Öznitelikler.....	183
<b>Tablo 4.3</b>	Veri Setinde Yer Alan Özniteliklerin İçerdiği Eksik Gözlem Sayıları.....	187
<b>Tablo 4.4</b>	Öznitelik Seçim İşlemi Sonrası Veri Setinde Yer Alan Öznitelikler.....	190
<b>Tablo 4.5</b>	Özniteliklere ait Özet İstatistikler.....	192
<b>Tablo 4.6</b>	Lojistik Regresyon Modeline Ait Karmaşıklık Matrisi.....	195
<b>Tablo 4.7</b>	Lojistik Regresyon Modeline ait Model Değerleme Metrikleri	195
<b>Tablo 4.8</b>	Doğrusal Destek Vektör Makineleri Modeline Ait Karmaşıklık Matrisi.....	199
<b>Tablo 4.9</b>	Doğrusal Destek Vektör Makinesi Modeline ait Model Değerleme Metrikleri.....	199
<b>Tablo 4.10</b>	Polinomsal Destek Vektör Makinelerine Ait Karmaşıklık Matrisi.....	202
<b>Tablo 4.11</b>	Polinomsal Destek Vektör Makinesi Modeline ait Model Değerleme Metrikleri.....	202
<b>Tablo 4.12</b>	Radyal Destek Vektör Makinesi Modeline Ait Karmaşıklık Matrisi.....	204
<b>Tablo 4.13</b>	Radyal Destek Vektör Makinesi Modeline Ait Model Değerleme Metrikleri.....	204

<b>Tablo 4.14</b>	Sınıflama ve Regresyon Ağacı Modeline Ait Karmaşıklık Matrisi.....	209
<b>Tablo 4.15</b>	Sınıflama ve Regresyon Ağacı Modeline Ait Model Değerleme Metrikleri.....	209
<b>Tablo 4.16</b>	Rastgele Orman Modeline Ait Karmaşıklık Matrisi.....	213
<b>Tablo 4.17</b>	Rastgele Orman Modeline Ait Model Değerleme Metrikleri....	214
<b>Tablo 4.18</b>	Adaboost Modeline Ait Karmaşıklık Matrisi.....	217
<b>Tablo 4.19</b>	Adaboost Modeline Ait Model Değerleme Metrikleri.....	217
<b>Tablo 4.20</b>	Gradient Boosting Modeline Ait Karmaşıklık Matrisi.....	220
<b>Tablo 4.21</b>	Gradient Boosting Modeline Ait Model Değerleme Metrikleri	220
<b>Tablo 4.22</b>	XGBoost Modeline Ait Karmaşıklık Matrisi.....	223
<b>Tablo 4.23</b>	XGBoost Modeline Ait Model Değerleme Metrikleri.....	223

## ŞEKİLLER LİSTESİ

Şekil 2.1	CRISP-DM Sürecinin Aşamaları.....	111
Şekil 2.2	Sigmoid Fonksiyon .....	125
Şekil 2.3	X ve Y Özniteliği ile Eğitilen Basit Bir Karar Ağacı.....	127
Şekil 2.4	Karmaşıklık Matrisi.....	139
Şekil 2.5	ROC Grafiği ve ROC Eğrisi Örneği.....	141
Şekil 4.1	Çalışmada Kullanılan Makine Öğrenmesi Süreci.....	179
Şekil 4.2	Veri Setinde Yer Alan Eksik Gözlem İçeren Öznitelikler.....	188
Şekil 4.3	Lojistik Regresyon Modele ait ROC Eğrisi ve AUC Skoru.....	196
Şekil 4.4	Doğrusal Destek Vektör Makineleri Parametre Optimizasyonu	199
Şekil 4.5	Doğrusal Destek Vektör Makinelerine ait ROC Eğrisi ve AUC Skoru.....	200
Şekil 4.6	Polinomsal Destek Vektör Makineleri Parametre Optimizasyonu.....	201
Şekil 4.7	Polinomsal Destek Vektör Makinesi Modeline Ait ROC Eğrisi ve AUC Skoru.....	203
Şekil 4.8	Radyal Destek Vektör Makinesi Modeline Ait ROC Eğrisi ve AUC Skoru.....	205
Şekil 4.9	Sınıflama ve Regresyon Ağacı Modeli Parametre Optimizasyonu.....	209
Şekil 4.10	Sınıflama ve Regresyon Ağacı Modeline Ait ROC Eğrisi ve AUC Skoru.....	210

<b>Şekil 4.11</b>	Rastgele Orman Modeli Parametre Optimizasyonu.....	213
<b>Şekil 4.12</b>	Rastgele Orman Modeline Ait ROC Eğrisi ve AUC Skoru.....	215
<b>Şekil 4.13</b>	Adaboost Modeli Parametre Optimizasyonu.....	217
<b>Şekil 4.14</b>	Adaboost Modeline Ait ROC Eğrisi ve AUC Skoru.....	218
<b>Şekil 4.15</b>	Gradient Boosting Modeline Ait ROC Eğrisi ve AUC Skoru....	221
<b>Şekil 4.16</b>	XGBoost Modeline Ait ROC Eğrisi ve AUC Skoru.....	224



## ÖZET

### **BAĞIMSIZ DENETİMDE MAKİNE ÖĞRENMESİ YAKLAŞIMI: BORSA İSTANBUL A.Ş. ÖRNEĞİ**

İşletmenin sürekliliğinin tahmini tüm işletme paydaşları için temel destekleyici bilgi kaynaklarında birisi olarak kabul edilmektedir. Paydaşlar açısından işletmenin süreklilik kabiliyeti hakkında doğru ve güvenilir bilgi sağlayabilen sınıflandırma ve tahmin modellerine sahip olmak çok önemlidir. Mali raporlama sistemleri, işletmelerin mali durumlarını değerlendirmek için gerek duyulan bilgileri sağlamaktadır. İstatistiksel metotlar 1960' lı yıllardan günümüze işletmenin sürekliliğinin tahmininde yaygın olarak kullanılmıştır. Son yıllarda yapılan çalışmalarda teknolojik gelişmelerin de hız kazanması ile katı varsayımlara sahip istatistiksel metotlara ikame olarak işletmenin süreklilik kabiliyetinin sınanmasında makine öğrenmesi metotlarında faydalanılmaya başlanmıştır. Bu çalışmada işletmenin sürekliliğinin tahmininde makine öğrenmesi metotlarının etkinliğinin sınanması amaçlanmaktadır. Bu amaçla çalışmada literatürde yaygın olarak kullanılan lojistik regresyon, destek vektör makineleri (doğrusal, polinomsal ve radyal tabanlı kernel çekirdekleri), sınıflama ve regresyon ağaçları, rastgele orman algoritması, adaboost, gradient boosting ve XGBoost metotları kullanılarak dokuz farklı model eğitilmiştir. Çalışmada Borsa İstanbul A.Ş. pay piyasası imalat ve alt sektörlerinde işlem görmekte olan işletmelere ait mali raporlardan faydalanılmıştır. Veri seti 2017 ile 2021 yılları arası kapsamaktadır. İlgili yıllar arasında işletmelerin Kamuyu Aydınlatma Platformunda duyurulmuş olan denetim raporları incelenmiş ve süreklilik görüşü almış olan işletmeler tespit edilmiştir. İşletmeler, yürütülen inceleme sonucu elde etiketlenmiştir. Böylece nihai veri setinde 43 süreklilik görüşü almış ve 788 süreklilik görüşü almamış gözlem yılını yer almıştır. Dengesiz veri setinin oluşturabileceği sorunların üstesinden gelmek amacı ile SMOTE metodundan faydalanılmış ve modellerin tahmin başarısını arttırmak maksadı ile genetik algoritmalar ile öznitelik seçimi stratejisi belirlenmiştir. Öznitelik seçimi öncesinde veri setinde, kapsamlı bir literatür taraması ile belirlenen 75 adet mali orana yer verilmiştir. Çalışmada eğitilen modellere ait doğruluk hassasiyet, seçicilik, tip-I hata, tip-II hata, kesinlik ve AUC skorları hesaplanmış ve modeller hesaplanan metrikler kullanılarak değerlendirilmiştir. Çalışmanın ampirik sonuçları incelendiğinde eğitilen tüm modeller içerisinde rastgele orman modelinin %100 doğruluk ile en başarılı sonucu ortaya koymuştur. Dokuz modelin hesaplanan metrikleri

incelendiğinde topluluk öğrenmesi modellerinin ve destek vektör makinesi modellerinin, sınıflama ve regresyon ağaçları ile lojistik regresyondan daha başarılı sonuçlar ortaya koyduğu görülmektedir. Çalışmanın ampirik sonuçları makine öğrenmesi modellerinin işletmenin sürekliliğinin tahminindeki etkinliği ortaya koymaktadır.

**Anahtar kelimeler:** Bağımsız Denetim, İşletmenin Sürekliliğinin Tahmini, Makine Öğrenmesi, Borsa İstanbul A.Ş.



## KIRGIZCA ÖZET (МАЗМУНУ)

### ЭРКИН АУДИТТЕГИ МАШИНАЛЫК ОКУУГА БОЛГОН ЖАКЫНДОО: "БОРСА ИСТАНБУЛ А.Ш" АКНУН МИСАЛЫНДА

#### 1. Аудит жана аудит түшүнүгү

Аудит түшүнүгү англис тилиндеги “Auditing” сөзүнүн эквиваленти. « Auditing » сөзү латын тилинен келип чыккан жана «угуу же угуп жатат» дегенди билдирген «audire» сөзүнүн негизинен келген түшүнүк. Сөздүн бул унгудан келип чыгышынын себеби бул жумушту аткаруу үчүн тандалган адамдар чындыкты табуу үчүн ишмердүүлүктү бирөөнү угуп ишке ашырууда жатат (Бозкурт 2018, 23). Адабиятта аудит түшүнүгүн аныктаган бирдиктүү аныктама жок. Изилдөөчүлөр аудитти ар кандай көз караштардан баалашкандыктан, окшош элементтерди көрсөтүп, ар кандай аныктамаларды беришкен. Адабияттардагы бир аныктама аудитти экономикалык ишмердүүлүккө жана окуяларга тиешелүү талаптардын канчалык деңгээлде алдын ала аныкталган чен белгилерге жооп берерин текшерүү жана натыйжалар жөнүндө кызыкдар тараптарга маалымат берүү үчүн далилдерди калыс чогултуу жана баалоо процесси катары аныктайт (Башпынар 2005, 35; Бозкурт 2018, 29; Гүредин 2014, 11; Кавал 2015, 3). Аудит экономикалык окуялар жөнүндө даярдалган маалыматтын тактыгын жана ишенимдүүлүгүн текшерүү процесси экендигин көрсөткөн аныктамадан тышкары, Аренс жана башкалар (2012) аудитти жүргүзүү үчүн аудитор ээ болушу керек болгон квалификацияга да кайрылат. Аныктама боюнча аудит маалымат менен алдын ала аныкталган чен белгилердин ортосундагы шайкештик даражасын аныктоо жана отчет берүү максатында маалыматтын далилдерин чогултуу жана баалоо процесси болуп саналат жана аудит компетенттүү жана көз карандысыз адам тарабынан жүргүзүлүшү керек. (Аренс, Элдер жана Бислей, 2012, 4-б.)

#### 2. Бизнесин үзгүлтүксүздүгү концепциясы

Акыркы жылдарда болуп өткөн глобалдык кризистер жана ишкердиктеги ийгиликсиздиктердин көбөйүшү Бизнесин үзгүлтүксүздүгүн болжолдоо бизнес жетекчилиги жана аудитор тарабынан кылдаттык менен бааланышы керектигин көрсөттү. Бизнесин үзгүлтүксүздүк концепциясы финансылык отчеттуулуктун негизги концепцияларынын бири болуп саналат жана ишкердик жакынкы келечекте өз ишин уланта алат деген божомолго негизделген. Ишкердикке

байланыштуу бардык кызыкчылык топтору бизнестин үзгүлтүксүздүгүн болжолдоо менен бизнестин финансылык абалы жөнүндө чечим кабыл алышат. Ишкана түзүлгөн учурдан тартып чексиз өмүргө ээ экендиги кабыл алынган. Бирок ишканалардын бар болушу алардын ишмердүүлүгүнүн үзгүлтүксүздүгүнө, өз убагында сарамжалдуу инвестицияларга, атаандаштыкка жетүү жөндөмдүүлүгүнө жана кардарлардын суроо-талаптарын жана керектөөлөрүн эң жакшы жол менен канааттандыруу мүмкүнчүлүгүнө жараша болот. Бул мааниде ишкана ээлери, өнөктөштөрү жана менежерлери бизнестин үзгүлтүксүздүгү жөнүндө маалымат берген кээ бир чен белгилерди, ошондой эле салттуу каржылык көрсөткүчтөрдү эске алуу менен каржылык натыйжаларды да, үзгүлтүксүздүк көрсөткүчтөрүн да өлчөй алышат. Чен белгилер ишкердиктин натыйжалуулугун өлчөө жана так чечимдерди кабыл алууга мүмкүндүк берет. Бизнестин үзгүлтүксүздүгүнө коркунуч келтириши мүмкүн болгон тобокелдик факторлорун алдын ала аныктоо ишкананын капыстан банкрот болуу тобокелдигине туш болушунун алдын алат. Азыркы гиператаандаштык чөйрөсүндө өтө маанилүү фактор катары өзгөчөлөнүп турган Бизнестин үзгүлтүксүздүгү бизнес кредиторлор, камсыздоочулар, инвесторлор жана көз карандысыз аудиторлор тарабынан эске алынат. Айрыкча, көз карандысыз аудит жагынан көз карандысыз аудиторлор бизнестин үзгүлтүксүздүгү көз карандысыз аудиторлордун аудитордук корутундуларды түзүү жоопкерчилигинде экендигин текшерүү үчүн маанилүү.

### **3. Машиналык окутуу**

Адамзат көптөн бери адам сыяктуу ой жүгүртүүгө жөндөмдүү машиналарды түзүүгө умтулуп келет. Бүгүнкү күндө машиналар өзүнүн технологиясы жана иштетүү жөндөмдүүлүгү менен адамдын акыл-эсинин белгилүү бир функцияларын аткара алат. Бирок, машиналар адам чийген чектерди кесип өтүү мүмкүнчүлүгүнө ээ эмес (Гопал 2018). Компьютерлер адамдын когнитивдик жөндөмдүүлүктөрүн талап кылган татаал тапшырмаларды аткара алабы деген маселе жасалма интеллект боюнча информатикада кеңири дисциплина катары классификацияланган. Адабиятта жасалма интеллекттин так аныктамасы жок. Жалпысынан алганда, жасалма интеллект окуу, кабыл алуу, ой жүгүртүү жана көйгөйлөрдү чечүү сыяктуу адамдын акылы менен байланышкан когнитивдик милдеттерди туураган компьютерлерди сүрөттөгөн термин катары аныкталат. Алгачкы жасалма

интеллекти изилдөөчүлөр компьютердик программалоо процесси жасалма интеллект маселесин чечүү үчүн колдонулушу мүмкүн деп ойлошкон. Бул божомол менен татаал тапшырмаларды аткарууга үйрөтүү үчүн компьютерлерге белгилүү бир шарттарда белгилүү чечимдерди кабыл алууга мүмкүндүк берүүчү жана көптөгөн шарттуу билдирүүлөрдү камтыган программаларды колдонууга болот. Жасалма интеллект маселелерин чечүүнүн альтернативалуу ыкмасы бул алгоритмдерди долбоорлоо, анда компьютер маселени чечүүдө өзүнүн мүмкүнчүлүктөрүн өнүктүрө алат. Сунушталган ыкма тажрыйбаны окутуу маалыматтар топтому аркылуу алгоритмге өткөрүүгө негизделген. Колдонулган алгоритмдердин дизайны биологиядан шыктанган окутуу машиналарынан баштап ыктымалдуулукка негизделген статистикалык үйрөнүү ыкмаларына чейин ар кандай стратегияларга негизделген. Жасалма интеллектте машиналык окутуу деген аталыш менен бир жармакташ чөйрө катары көзгө урунган бул тармак 1980-жылдардан бери көптөгөн изилдөөлөрдүн предмети болгон. Машиналык окутуунун табияты символикалык ыкмалардын кемчиликтеринен жабыркап калуудан сактайт. Символикалык ыкмалардан айырмаланып, берилиштерге негизделген машиналык окутуу алгоритмдери жасалма интеллект маселелерин чечүүдө программалоосуз математикалык моделдерди өндүрүү үчүн маалыматтарга басым жасайт (Jiang 2021).

Машиналык окутуу моделдеринде үлгү “ $n$ ” өзгөчөлүктөрдүн/атрибуттардын  $x_j; j = 1, 2, \dots, n$  туруктуу саны менен аныкталат жана ар бир өзгөчөлүк үлгү үчүн чыныгы сандык маанилерге ээ.  $x_j$  - маселеге тиешелүү вектордук мейкиндик жана элементи бар ар бир атрибут  $V_{x_j} \in \mathcal{R}$  топтомуна тиешелүү маанилерди камтыйт.  $S^{(i)}, x_j^{(i)} \in V_{x_j}$  берилиш үлгүсүн камсыз кылуучу  $\{x_1^{(i)}, \dots, x_n^{(i)}\}$  атрибутунун мааниси койду. Ар бир үлгү  $n$ -өлчөмдүү абал мейкиндигинде же гипотеза мейкиндигинде  $n$  сандык өзгөчөлүк менен көрсөтүлөт.

$$\mathbf{X} = [x_1 x_2 \dots x_n]^T \in \mathfrak{R}^n$$

### 3.1. Окутуу стратегиялары

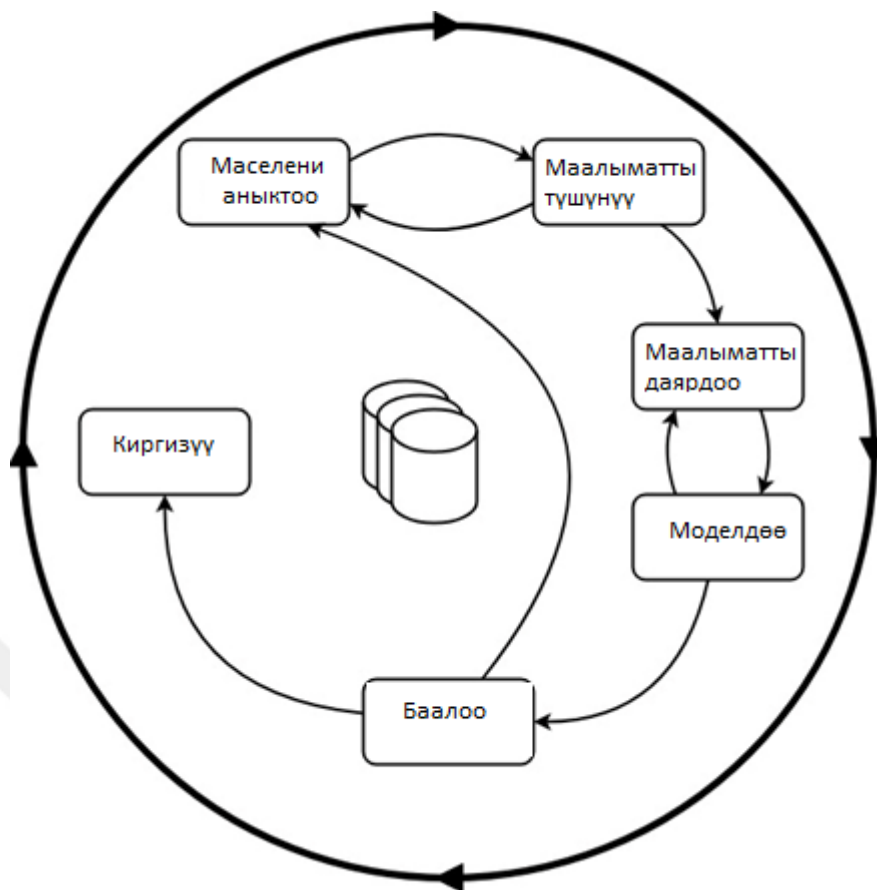
Берилиштер топтомундагы пайдалуу үлгүлөрдү автоматтык түрдө аныктай ала турган жана келечектеги кырдаалдарды болжолдоо үчүн бул үлгүлөрдү колдоно алган машиналык окутуу алгоритмдери негизинен “Көзөмөлдөгү окутуу” жана “Көзөмөлсүз окутуу” деп эки категорияга бөлүнөт (Abbott 2014, 5; James ж.б. 2013,

1). Бирок машиналык окутуунун башка түрлөрүнө караганда азыраак колдонулган бекемдөөчү окутуу деп аталган дагы бир үчүнчү типтеги машиналык окутуу ыкмасы бар (Murphy, 2012). Бекемдөөчү окутуу көзөмөлдөгү окутуу менен көзөмөлсүз окутуунун ортосундагы аралык болуп саналат. Бекемдөөчү окутууда жооп туура эмес болгондо, моделди окутуу процессинде кайра байланышты алат. Бирок моделге туура жоопту кантип табуу керектиги боюнча эч кандай белги берилбейт. Модель туура жооп тапканга чейин ар кандай мүмкүнчүлүктөрдү изилдейт жана аракет кылат (Marsland, 2015).

Көзөмөлдөнгөн окутууда окутуу маалыматтары киргизүү /чыгаруу векторлорунан турат. Көзөмөлдөнгөн маалыматтардагы ар бир киргизүү үчүн алгоритмдин көзөмөл сигналы катары колдоно ала турган чыгаруу вектору аныкталышы керек. Көзөмөлгө алынган окутуу ыкмалары - бул машиналык окутуу адабиятында кеңири талкууланган тармак жана алар, адатта, жетиштүү маалымат болгондо, жакшы иштешине кепилдик берет. Бирок алар киргизүү/чыгаруу жуптарын түзүү же маалымат топтомун белгилөө үчүн адамдын акыл-эсинин зарылдыгынан улам кымбат болушу мүмкүн. Көзөмөлсүз окутуу методдору окшош киргизүүлөрдү бардык мүмкүн болгон киргизүүлөрдү колдонгон сценарийге топтоого багытталган. Көзөмөлсүз окутууда маалыматтар белгиленбейт. Ошентип, чыгуу маалыматтарынын белгилениши жок болгон учурда, кайсы кириш маалыматтар окшош экенин аныктоо негизги көйгөй болуп саналат жана байкоо жүргүзүү үчүн колдоонун жоктугунан бир топ кыйыныраак. Бирок окутуу маалыматтарын чогултуу салыштырмалуу арзаныраак (Jiang 2021).

### **3.2. Машиналык окутуу процесси**

Берилиштерден билимди генерациялоо процесси ар кандай көндүмдөрдү жана билимдерди талап кылган чыгармачылык процесс. Бул процесс бизнес дүйнөсүндөгү көйгөйлөрдү түшүнүү жана аларды маалыматты чогултуу же машиналык окутуу тапшырмасына айландыруу үчүн базалык процессти талап кылат. CRISP-DM " Cross Industry Standard Processfor Data Mining " секторлорго жана колдонмо технологияларга карабастан маалыматтарды чогултуу жана машиналык окутуу колдонмолорунда айтылган көйгөйдү чечүү үчүн колдонула турган процесс моделин аныктайт (Larose жана Larose, 2015; Wirth жана Hipp 2000).



**Сүрөт 0.1** CRISP-DM процессинин этаптары ( Wirth жана Hipp 2000, 5)

CRIPS-DM процесси алты этаптан турат ( Wirth жана Hipp 2000, 5-7);

1. Маселени аныктоо: Процессин алгачкы баскычында машиналык окутуу долбоорунун максаттарын жана талаптарын түшүнүүгө, ал эми кийинки процессте бул түзүлгөн билимди машиналык окутуу ммаселесин аныктоого жана алдын ала долбоордун планына түзүүгө көңүл бурулат.

2. Маалыматты түшүнүү: Процессин экинчи этабы баштапкы маалыматтарды чогултуу менен башталат жана маалыматтар менен таанышуу, маалыматтардын сапатына байланыштуу көйгөйлөрдү аныктоо, маалыматтарды изилдөө жана жашыруун маалыматты ачуу үчүн гипотезаларды түзүү боюнча иш-аракеттер менен уланат. Маалыматты түшүнүү баскычы менен маселени аныктоо баскычынын ортосунда тыгыз байланыш бар. Маселени жана процессти калыптандыруу үчүн жок эле дегенде, бир аз маалыматты түшүнүү талап кылынат. Изилдөө маалыматтарын талдоо маалыматтарды таануу жана анын түзүмүн түшүнүү үчүн колдонулушу керек. Маалыматтын сапатына баа берилиши керек

жана зарыл болсо, ишке жарамдуу үлгүлөрдү камтый турган бөлүмдөр тандалышы абзел.

3. Маалыматты даярдоо: Бул баскыч чийки маалыматтардан акыркы маалымат топтомун түзүү үчүн зарыл болгон бардык иш-чараларды камтыйт. Бул этапта аткарылган иш-чаралардын конкреттүү ырааттуулугу жок. Анткени процесстин жүрүшүндө бир иш бир нече жолу аткарылышы мүмкүн. Маалыматтарды даярдоо этабында аткарылуучу милдеттердин арасында маалыматтарды тазалоо, функцияларды тандоо жана маалыматтарды трансформациялоо сыяктуу тапшырмалар бар. Бул баскычта чийки маалыматты моделдөөгө даяр кылуу үчүн бардык зарыл иш-чаралар жүзөгө ашырылат,

4. Моделдөө: Машиналык окутуунун ар кандай ыкмалары тандалып, үйрөтүлүп, алардын параметрлери оптималдаштырылат. Көп учурда маселени чечүү үчүн бир нече ыкмалар бар. Кээ бир ыкмалар белгилүү бир маалымат түрүн талап кылат жана көп учурда көйгөйлүү аймакка бирден ашык окшош ыкма колдонулат. Маалыматты даярдоо жана моделдөө этаптарынын ортосунда тыгыз байланыш бар. Моделдөө этабында маалыматтарда болгон көйгөйлөр жөнүндө идея түзүлөт. Техникалардын ичинен ылайыктуусун тандап, оптималдаштыруу керек. Маалыматтарды даярдоо фазасына кайтып келип, маалымат талаптарын же белгилүү бир ыкмалардын божомолдорун канааттандыруу үчүн моделдештирүү мүмкүн.

5. Балоо: Процесстин бул этабында, моделдөө баскычында үйрөтүлгөн бир же бир нече моделдер максаттарга жетип же жетпегендигин текшерүү үчүн кылдат бааланат.

6. Киргизүү (ишке киргизүү): Процесс моделди түзүү менен аяктайт. Модель муктаждыктарга ылайык түзүлүп, көрсөтүлүшү керек. Ушул себептен үйрөтүлө турган моделдерден чындап пайда алуу үчүн кандай аракеттердин жасалышы керектигин алдын ала түшүнүү абдан маанилүү. Акыркы моделдин ачыкка чыгышы процесс бүттү дегенди билдирбейт. Мындан тышкары бул моделди ишке ашыруу керек. Негизги ишке киргизүү отчетту түзүү болушу мүмкүн.

#### **4. Бизнесин үзгүлтүксүздүгүн болжолдоого багытталган “Борса Истанбул А.Ш.” акционердик коомунун өндүрүш секторунда иштелме**

Изилдөөнүн бул бөлүмүндө теориялык жактан түшүндүрүлгөн машиналык окутуу алгоритмдери Борса Истанбул А.Ш.” акционердик коомунун өндүрүш секторунда иштеген ишканалардын маалымат топтомун колдонуу менен үйрөтүлүп, натыйжалары кеңири талкууланат.

##### **4.1. Изилдөөнүн максаты жана көлөмү**

Изилдөөнүн максаты көз карандысыз аудит процессинде ишкана жетекчилиги жана көз карандысыз аудитордун бизнестин үзгүлтүксүздүгүн жүргүзүү үчүн жооптуу болгон баалоо процессинде машиналык окутуу алгоритмдеринин натыйжалуулугун аныктоо болуп саналат. Буга байланыштуу бул иште “Борса Истанбул А.Ш.” акционердик коомунун үлүш рыногунда өндүрүш өнөр жайы жана анын бөлүмчөлөрүндө ишмердүүлүк кылган 16 ишканага тиешелүү 47 байкоо жылы жана үзгүлтүксүздүк корутундусу берилбеген 198 ишканага тиешелүү 958 байкоо жылы каралат. Изилдөө 2017-жылдан 2021-жылга чейинки мезгилде жарыяланган каржылык отчетторду камтыйт. Үзгүлтүксүздүктү болжолдоо максатында логистикалык регрессия, классификация жана регрессия дарагы, таяныч векторлорунун машинасы (сызыктуу, полиномдук жана радиалдык негиздеги ядро функциялары), кокус токой алгоритми, “adaboost” алгоритми, “gradient boosting” алгоритми жана “extreme gradient boosting” алгоритми он катмарлуу кайчылаш текшерүү процедурасын колдонуу менен окутулду. Ошентип ишмердүүлүк (ишкана) үзгүлтүксүздүгүн болжолдоодо эң жогорку болжолдоо күчкө ээ моделди табуу максаты коюлган.

Изилдөөдө колдонулган финансылык катыштар Datastream маалымат базасынан алынган. Моделдерди окутууда R программалоо тили жана R Studio колдонулган.

##### **4.2. Изилдөөнүн маанилүүлүгү**

Ааламдашуу бардык секторлордогу ишканалардын ортосундагы атаандаштыкты күчөтөт. Атаандаштыктын күчөшү бизнеске басым жасап, алардын үзгүлтүксүздүгүн сактап калуу жөндөмдүүлүгүнө карата көбүрөөк белгисиздикти жаратууда. Бул процесс бизнеске жана бизнестин кызыкдар тараптарына барган сайын көбүрөөк басым жасап, аларды революциялык жаңы маалыматтык технологиялар менен аман калуу жолдорун издөөгө мажбурлоодо. Бизнес

функцияларына инновациялык системаларды кошпогон ишканалар процессти эффективдүү башкара алышпайт жана акыры ийгиликсиздикке дуушар болушат (Li жана Sun 2011a). Көпчүлүк учурларда бизнестин үзгүлтүксүздүгүн жоготуу ишкананын ичиндеги ар кандай факторлордун натыйжасы болуп саналат, алар ишкананын финансылык абалында чагылдырылат. Ишкананын ишмердүүлүк секторундагы белгисиздик, корпоративдик башкаруунун адекваттуу эместиги, инновациянын жоктугу, бизнестин башкаруудагы эффективдүүлүктүн жоктугу сыяктуу көптөгөн факторлор бизнестин үзгүлтүксүздүгүн жоготууга, демек банкроттукка алып келет жана уюштуруу максаттарына жетишүүгө тоскоол болот. (Gordini 2014). Ишкананын үзгүлтүксүз иштөө жөндөмдүүлүгүн сактай албай калышына алып келген көптөгөн факторлор бар. Булардын эң негизгилеринин арасында жогорку пайыздык чендер, экономикалык рецессиядан келип чыккан пайданын жоголушу, жогорку карыз запастары, сектордун өзгөчөлүктөрү, ишкананы каржылык кыйынчылыкка алып келиши мүмкүн болгон укуктук ченемдерди атаса болот (du Jardin 2015). Начар башкаруу менен коштолгон каржылык жоготуулардын топтолушу бизнестин үзгүлтүксүз иштешине терс таасирин тийгизет жана бизнестин кызыкдар тараптары катышкан сценарийде чоң экономикалык чыгымдарга алып келиши мүмкүн. Ошондуктан бизнестин үзгүлтүксүз иштөө жөндөмдүүлүгү жөнүндө эрте эскертүүчү системаларды иштеп чыгуу маанилүү маселеге айланууда жана бизнестин үзгүлтүксүздүгүн болжолдоо үчүн көптөгөн моделдер академиктер жана бизнестеги кызыкдар тараптар тарабынан иштелип чыгууда (Alam, Gao, жана Jones 2021; Murugan Anandarajan, Lee, жана Anandarajan 2001).

Бизнестин үзгүлтүксүздүгүн баалоо үчүн ийгиликтүү моделдерди иштеп чыгуу бизнес инвесторлору, кредиторлор, кызматкерлер, камсыздоочулар жана башка тармактын катышуучулары сыяктуу бизнестин кызыкдар тараптары үчүн абдан маанилүү маалыматтык ролду ойнойт (Alfaro ж.б. 2008). Аудит процессинде ар кандай максаттар үчүн маалымат аналитикасын жана машиналык окутуу ыкмаларын колдонууну мисал кылган көптөгөн изилдөөлөр бар (Amani жана Fadlalla 2017). Аудит учурунда текшерилген бизнестин ишин улантуу мүмкүнчүлүгү аудитор баалоого тийиш болгон маселелердин бири болуп саналат. Үзгүлтүксүздүк тобокелдиги адабиятта актуалдуу маселе болуп саналат, анткени ал бизнестин кызыкдар тараптарынын чечимдерине чоң таасирин тийгизет. Бул

контекстте ишкананын үзгүлтүксүз иштөө жөндөмдүүлүгү жөнүндө аудитордун көз карашын колдоо үчүн ар кандай ыкмалар жана моделдер иштелип чыккан жана үзгүлтүксүздүк тобокелдигин баалоо боюнча кеңири адабияттар бар. (Cecchini ж.б. 2010; Chye Koh жана Kee Low 2004a; Fallahpour, Lakvan, жана Zadeh 2017b; Fernández-Gámez vd. 2020; Gepp, Kumar, жана Bhattacharya 2010b; S. Y. Kim 2011b; S. Y. Kim жана Urpeja 2014; Li жана Sun 2011b; Liang ж.б. 2020a; Martens ж.б. 2008; Olson, Delen, жана Meng 2012a; Sánchez-Serrano ж.б. 2020; J. Sun ж.б. 2020; Véganzones жана Séverin 2018; Volkov, Benoit, жана Van den Poel 2017; Desheng(Dash) Wu, Liang, жана Yang 2008a; Xu жана Wang 2009a; Yeh, Chi, жана Lin 2014b; Lukason жана Camacho-Miñano 2019; Jayasekera 2018; Alam, Gao, жана Jones 2021; S. Jones жана Hensher 2008).

Аудитордун субъект жөнүндө чечим кабыл алуу процесси субъективдүү мүнөзгө ээ жана аудитор бул процессте профессионалдуу ой-пикирин көрсөтүүгө милдеттүү. Адабияттарда көз карандысыз аудитордун ишкананын үзгүлтүксүз иштеши жөнүндө чечиминин детерминанттары боюнча жүргүзүлгөн изилдөөлөр бар. Аудиторлор ишкананын үзгүлтүксүз ишмердүүлүгүн улантуу мүмкүнчүлүгүнүн айрым негизги көрсөткүчтөрүн өлчөйт (Mutchler 1984). Чечим кабыл алуу процессинде аудитор ишкананын операцияларынан түшкөн акча каражаттарынын агымы, учурдагы коэффициент, ишкананын таза наркы, узак мөөнөттүү карыздар, жалпы карыздар жана таза пайда сыяктуу негизги финансылык коэффициенттерди карайт. Адабиятта көрсөтүлгөн финансылык көрсөткүчтөр аудитордун акыркы чечимине көп таасир этет (Mutchler 1985). Мындан тышкары аудитордун ишкананын үзгүлтүксүздүгү тууралуу чечим чыгаруу процессин изилдеген иштерде ар кандай каржылык катыштары да процессте колдонуулары көрсөтүп турат (Reynolds жана Francis 2000; DeFond, Raghunandan, жана Subramanyam 2002; Krishnan жана Wang 2015).

#### **4.3. Изилдөөнүн чектөөлөрү**

Изилдөөнүн чектөөсү төмөнкүдөй;

Изилдөөнүн биринчи чектөөсү ал камтыган мезгилге байланыштуу. Изилдөө 2017-2021-жылдар аралыгында “Борса Истанбул А.Ш.” акционердик коому менен ишмердүүлүк жүргүзгөн ишканаларды камтыйт. BDS-570 Бизнес үзгүлтүксүздүк стандарты 2017-жылдан баштап жүргүзүлө турган аудит изилдөөлөрүндө ишке

ашырыла баштады. Изилдөөдө колдонулган маалымат топтому 2022-жылга карата түзүлгөн. Саналган эки себеп изилдөө камтыган мезгилди аныктоо үчүн негиз болгон,

Изилдөөнүн экинчи чектөөсү маалыматтар топтомундагы ишканалар иштеген секторго байланыштуу. Изилдөөдө колдонулган маалымат топтомунда “Борса Истанбул А.Ш.” акционердик коомунда өндүрүш жана бөлүмдөрүндө ишмердүүлүк кылган компаниялар гана камтылган. Изилдөөдө колдонула турган маалыматтар топтомундагы финансылык катыштар ортосундагы бир тектүүлүктү камсыз кылуу жана жүргүзүлө турган анализдерде так натыйжаларды алуу үчүн маалымат топтому өндүрүш ишканалары менен гана чектелет.

Изилдөөдөгү үчүнчү чектөө ишканалар кайсы мамлекетте иштеп жатканына байланыштуу. Изилдөөгө Түркияда иштеген ишканалар гана киргизилген.

Изилдөөдөгү төртүнчү чектөө колдонулган методдорго байланыштуу. Бул контекстте ишкананын үзгүлтүксүздүгүн болжолдоо үчүн бир гана машиналык окутуу алгоритмдери колдонулган. Изилдөөдө колдонулган көзөмөлдөнүүчү окутуу алгоритмдери логистикалык регрессия, классификация жана регрессия дарактары, таяныч векторлорунун машинасы (сызыктуу, полиномдук жана радиалдык негизде ядро функциялары), кокус токой алгоритми, “adaboost” алгоритми, “gradient boosting” алгоритми жана “extreme gradient boosting” алгоритмдери. Изилдөөдө классификациялоо алгоритмдери гана колдонулган. Изилдөөдөгү моделдер параметрдик ыкмалар эмес жана параметрдик ыкмалардагыдай 3 божомолдор жок.

Изилдөөдөгү бешинчи чектөө колдонулган касиеттерге байланыштуу. Аудитордун ишкананын үзгүлтүксүздүгүн баалоосу көптөгөн факторлорду бир эле учурда текшерүүнү талап кылат. Бирок бул изилдөөдө процеске байланыштуу структураланган маалыматтар жана финансылык катыштар гана колдонулган. Финансылык катыштар адабияттарда үзгүлтүксүздүктү баалоо үчүн кеңири колдонулганы көрүнүп турат. Колдонулган финансылык көрсөткүчтөр ишканалардын жылдык финансылык отчетторунан алынат.

#### **4.4. Изилдөөлөрдүн кеңири таралган таасири**

Машиналык окутуу ыкмалары бизнес үзгүлтүксүздүгүн болжолдоо боюнча адабиятта 1990-жылдардан бери кеңири колдонулуп келет. Адабиятта ар кайсы

өлкөлөрдө ар кандай маалымат топтомдору менен жүргүзүлгөн изилдөөлөр машиналык окутуу ыкмалары ишкананын үзгүлтүксүздүгүн баалоодо ийгиликтүү натыйжаларды берерин көрсөтүп турат. Жүргүзүлгөн изилдөөдө адабиятта кеңири колдонулган машиналык окутуу ыкмаларынын натыйжалуулугу бизнестин үзгүлтүксүздүгүн баалоодо бизнестин башкаруусу, аудиторлор жана бизнестин кызыкдар тараптары тарабынан текшерилет. Бул контекстте изилдөө Түркиядагы машиналык окутуу жана аудит адабиятына өзгөчө салым кошот. Изилдөөнүн тизмеленгендерден тышкары жалпы таасирлери төмөнкүлөр болуп саналат;

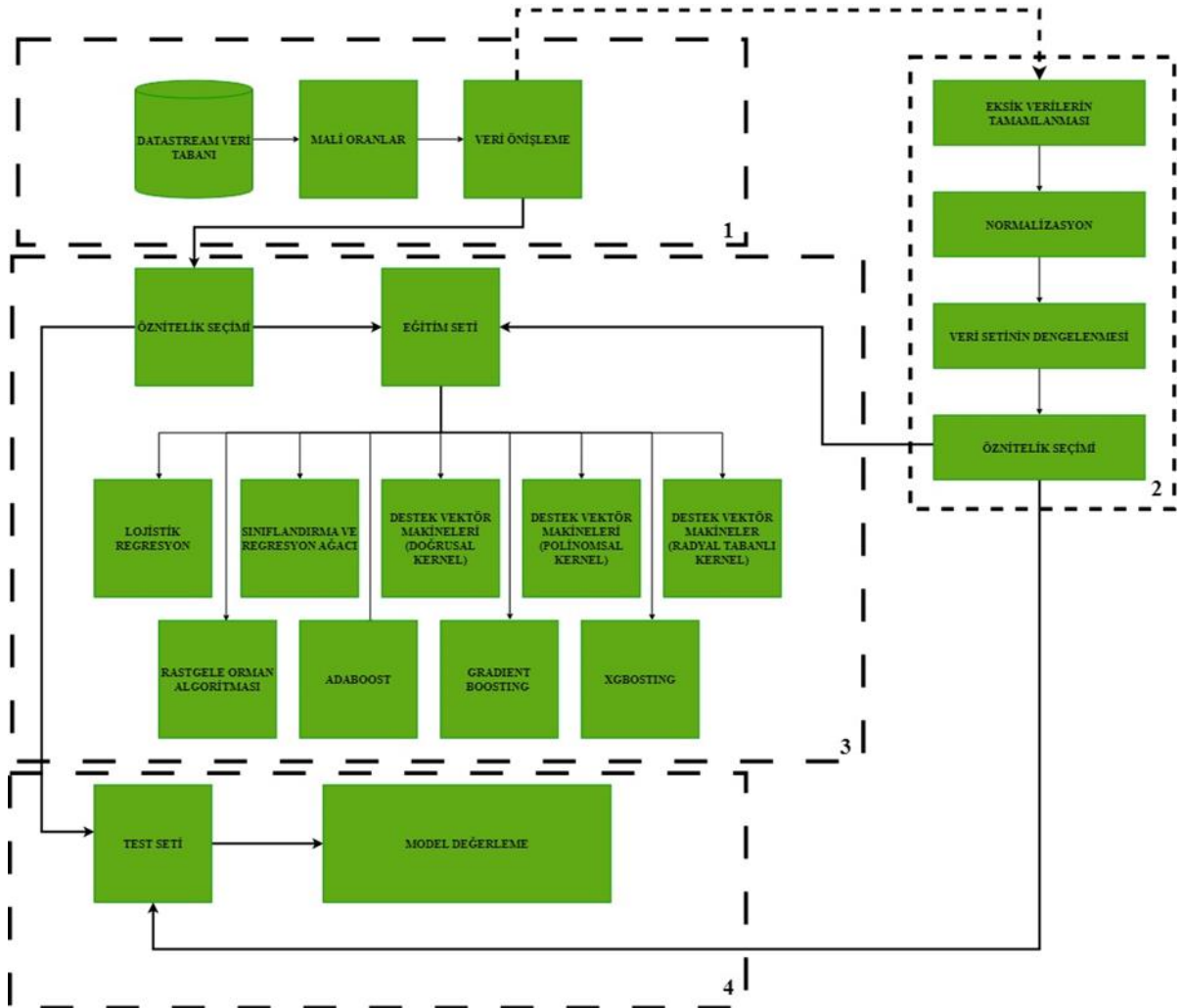
Иштелип чыга турган моделди аудитордун чечим кабыл алуу процессин колдоо үчүн көз карандысыз аудит программаларында колдонуу пландаштырылууда. Кошумчалай кетсек, сунушталып жаткан модель ишкананын ресурстарын пландаштыруу программасында жана жогорку деңгээлдеги чечимдерди колдоо программалык камсыздоосунда жогорку даражалуу менежерлердин чечим кабыл алуу процесстеринде колдонулушу мүмкүн деп болжолдонууда, алар бүгүнкү күндө бизнесте кеңири колдонулат. Сунушталган моделди ишкананын финансылык абалы жөнүндө чечим кабыл алуу процессинде байланышы бар тараптар (мамлекеттик мекемелер, кредиттик мекемелер, инвесторлор жана өнөктөштөр) колдонсо болот.

Изилдөөнүн камтыган чөйрөсүн кеңейтүү менен машиналык окутуу ыкмалары жана табигый тилди иштетүү алгоритмдери колдонулган реалдуу убакыт режиминде аудитордук чечимдерди колдоо программалык камсыздоо долбоору үчүн негиз болот деп эсептелет. Аудитордук чечимдерди колдоо программалык камсыздоосу реалдуу убакытта XBRL таксономияларын колдонуу менен каржылык отчеттуулукту талдоо аркылуу аудитордук иш-аракетти камсыз кылат. Ошентип бизнестин башкарууга байланыштуу тараптардын маалымат керектөөлөрү үзгүлтүксүз камсыз кылынат.

#### **4.5. Изилдөө ыкмалары**

Ишкананын үзгүлтүксүздүгүн баалоодо машиналык окутуу ыкмаларынын натыйжалуулугун өлчөөнү максат кылган бул изилдөөдө стандарттуу машиналык окутуу процесси колдонулган. Процесс 4.1-сүрөттө кеңири көрсөтүлгөн. Колдонулган процесс төрт этаптан турат: маалыматтарды алуу (1), маалыматтарды алдын ала иштетүү (2), моделдөө (3) жана баалоо (4). Процессинин биринчи этабы

Datastream маалымат базасынан чийки маалыматтарды алуу менен башталат. Чийки маалыматтарды алгандан кийин маалымат топтомуна киргизиле турган финансылык катыштар Ms Excel программасында эсептелип, R Studio чөйрөсүнө өткөрүлүп берилди. Маалыматтарды алдын ала иштетүү, моделдөө жана баалоо этаптары бүтүндөй R Studio чөйрөсүндө ишке ашырылган жана колдонмо бөлүмдө кенири түшүндүрүлгөн.



**Сүрөт 4.1. Изилдөөдө колдонулган машиналык окутуу процесси**

Маалыматтарды R studio чөйрөсүнө өткөрүп бергенден кийин биринчи кезекте, изилдене турган маалыматтардын түзүмүн жакшыраак таануу максатында маалыматтардын негизги көрсөткүчтөрү каралып, бул жол менен маалыматтарды алдын ала иштетүү стадиясында зарыл болгон процесстерди аныктоого багытталган. Маалыматтарды алдын ала иштетүү стадиясында маалыматтарды моделдөөгө даярдоо үчүн жетишпеген маалыматтарды толтуруу, нормалдаштыруу,

маалымат топтомун теңдөө жана өзгөчөлүктөрдү тандоо процесстери ишке ашырылган. Бул баскычтан кийин моделдөө баскычы башталат. Моделдөө баскычында логистикалык регрессия, классификация жана регрессия дарагы, таяныч векторлорунун машинасы (сызыктуу, полиномдук жана радиалдык негизде ядро функциялары), кокус токой алгоритмдери, adaboost, gradient boosting, жана XGBoost алгоритмдери окутулуп, окутулган моделдер баалоо этабында адабиятта кеңири колдонулган чен белгилер аркылуу бааланып, натыйжалар баяндалып, талкууланды.

#### **4.6. Изилдөөдө колдонулган маалыматтар топтому**

BDS-570 Бизнес үзгүлтүксүздүк стандарты 01.01.2017 же андан кийин баштала турган эсептик мезгилдердин аудитине карата колдонулат (Коомдук көзөмөл органы 2020, 2). Изилдөөдө стандарттын ишке ашырылган күн маалыматтарды аныктоо үчүн маалымдама катары кабыл алынган. Изилдөөдө колдонулган маалыматтар топтому 03.08.2022-жылы түзүлгөн. Бул күнү 2022-жылга карата каржылык отчеттуулук коомчулукка ачыкталган эмес. Бул чектөөлөр астында бизнестин үзгүлтүксүздүгүн болжолдоону максат кылган бул изилдөөдө маалымат топтомунун 2017-2021-жылдар аралыгындагы мезгилди камтый турганы аныкталды. Натыйжада маалымат топтому 2017, 2018, 2019, 2020 жана 2021-жылдары ачык жарыяланган финансылык отчеттордун маалыматтарын камтыды. Маалымат топтому “Борса Истанбул А.Ш.” акционердик коомунун үлүштөр рыногундагы өндүрүш тармагында жана анын суб-секторлорунда ишмердүүлүк кылган ишканаларды камтыйт. Финансы институттарынын ар кандай каржылык түзүмдөрүнөн улам изилдөөдө колдонула турган маалыматтар өндүрүш компаниялары менен гана чектелет. “Борса Истанбул А.Ш.” акционердик коомунун үлүштөр рыногундагы өндүрүш тармагында жана анын суб-секторлорунда 2023-жылдын май айына карата 209 ишкана ишмердүүлүк кылат. Бирок маалымат топтому түзүлгөн учурда 201 ишкана болгон.

Байкоо жылдарын белгилөө максатында өндүрүш тармагында ишмердүүлүк кылган ишканалардын Коомдук агартуу платформасынын веб-сайтында коомчулукка жарыяланган аудитордук отчеттор каралып чыкты. Буга байланыштуу көз карандысыз аудитордук отчетто ишкананын үзгүлтүксүз иштешин улантуу жөндөмдүүлүгүнө байланыштуу олуттуу белгисиздикке

шилтеме берген жалпысынан 16 ишканага таандык жалпысынан 47 жылга байкоо жүргүзүлгөн. Бул ишканалардан айырмаланып, 198 ишканага тиешелүү жана ишкананын үзгүлтүксүздүгүнө эч кандай шилтеме берилбестен 958 байкоо жылы бекем деп кабыл алынган. Изилдөөдө үзгүлтүксүздүк көрүнүшүнө ээ болгон байкоо жылдары шилтеме класс катары кабыл алынган. Демек тиешелүү класс "1" менен белгиленет. Үзгүлтүксүз көрүнүшкө ээ болбогон байкоо жылдары "0" деп белгиленген. Изилдөөдөгү маалыматтар топтомунда ишканага тиешелүү 2017-2021-жылдар аралыгындагы бир жылдан ашык болгон мезгил камтылган жана ал байкоо жылы деп аталган. Бизнесин үзгүлтүксүздүгүн болжолдоого багытталган изилдөөлөрдө үзгүлтүксүздүк көрүнүшүнө ээ ишканалардын аз санынан улам маалыматтар топтомунун көлөмүн көбөйтүү үчүн бир ишканага таандык бирден көр жылга маалымат топтомуна орун берүү кеңири таралган ыкма болуп саналат. Адабиятка кирген бул ыкма да изилдөөдө кабыл алынган (Alfaro ж.б. 2008).

**Таблица 4.1** Маалыматтар топтомундагы байкоо жылдардын саны

	2017	2018	2019	2020	2021	Бардыгы
Үзгүлтүксүздүк көрүнүшүнө ээ ишканалар	9	7	11	10	6	43
Үзгүлтүксүздүк көрүнүшүнө ээ болбогон ишканалар	151	159	157	158	163	788
Бардыгы	160	166	168	168	169	831

Маалыматтар топтомундагы финансылык катыштарды эсептеп чыккандан кийин жетишпеген маалыматтарды толуктоо үчүн кандайдыр бир иш-аракет жасалганга чейин көпчүлүк өзгөчөлүктөр жок болгон байкоо жылдары топтомдон чыгарылды. Ошентип, үзгүлтүксүздүк көрүнүшүн алган 4 байкоо жылы жана үзгүлтүксүздүк көрүнүшүн албаган 170 байкоо жылы маалымат топтомунан чыгарылды. Бул процессте маалыматтар топтомунан алынып салынган байкоолордун жылдары жетишпеген байкоолордун 100%дан 61%ке чейинкисин камтыйт. 4.1-таблицада маалыматтарды алдын ала иштетүү жана моделдөө этаптарында колдонулган маалыматтар топтомундагы байкоо жылдарынын саны камтылган. Маалыматтар топтому 2017-жылдан баштап жалпысынан 160 байкоо жылын камтыйт. Алардын ичинен бсы үзгүлтүксүздүк көрүнүшүн алса, 151и үзгүлтүксүздүк көрүнүшүн алган эмес. Маалыматтар топтому 2018-жылдан баштап жалпысынан 166 байкоо жылын камтыйт. Алардын 7си үзгүлтүксүздүк көрүнүшүн алса, 159у

үзгүлтүксүздүк көрүнүшүн алган эмес. Маалыматтар топтому 2019-жылдан баштап жалпысынан 168 байкоо жылын камтыйт. Анын ичинен 11и үзгүлтүксүздүк көрүнүшүн алса, 157си үзгүлтүксүздүк көрүнүшүн алган эмес. Маалымат топтомунда 2020-жылдан баштап жалпысынан 168 байкоо жылы бар. Алардын ичинен 10у үзгүлтүксүздүк көрүнүшүн алса, 158и үзгүлтүксүздүк көрүнүшүн алган эмес. Маалымат топтому 2021-жылдан баштап жалпысынан 169 байкоо жылын камтыйт. Алардын ичинен бсы үзгүлтүксүздүк көрүнүшүн алса, 163ү үзгүлтүксүздүк көрүнүшүн алган эмес. Маалымат топтомунда көрсөтүлгөн жылдар үчүн үзгүлтүксүздүк көрүнүшүн алган 43 байкоо жылы болсо, үзгүлтүксүздүк көрүнүшүн албаган жалпы 788 байкоо жылы бар.

#### **4.7. Маалыматтар топтомундагы каржылык катыштар**

Бизнестин үзгүлтүксүздүгүн болжолдоодо кандай өзгөчөлүктөр колдонула турганы жана саны кандай болот деген абдан маанилүү суроо (E. I. Altman ж.б. 2020). Bellovary, Giacominо жана Akers (2007) үзгүлтүксүздүктү болжолдоо адабиятында 752 түрдүү өзгөчөлүктөрдү аныктаган. Бул контексте өзгөчөлүктөрдү тандоо чоң мааниге ээ (du Jardin 2009). Адабиятта маалымат топтомуна кире турган өзгөчөлүктөрдү кантип тандоо керектиги боюнча консенсус жок. BDS-570 бизнес үзгүлтүксүздүгү стандартында ишкананын финансылык абалына байланыштуу, ишкана ишмердүүлүгүнө байланыштуу жана башка абалдарга байланыштуу болуп үч негизги тема менен каралат, бизнестин үзгүлтүксүз иштөө жөндөмдүүлүгүнө олуттуу шек келтире турган окуялар же шарттар жөнүндө шилтемелер жасалат (Коомдук көзөмөл органы 2020, 13-15). Изилдөөдө тиешелүү шилтемелерди эске алуу менен деталдуу жана ар тараптуу адабиятка сереп жүргүзүлүп, биринчи этапта 75 каржылык катыш аныкталган. Бул контексте кылдат адабияттарга сереп кылуу менен адабияттарда өзгөчөлүктөрдү тандоо көп колдонулган жана маанилүү натыйжаларды берген финансылык катыштар өзгөчөлүк катары аныкталган (M.-Y. Chen 2011b). Маалыматтар топтомунда орун алган өзгөчөлүктөр 4.2-таблицасында көрсөтүлгөн. 4.2-таблицадагы өзгөчөлүктөр өзгөчөлүктү тандоого чейинки процессти билдирет. Өзгөчөлүк тандалгандан кийин маалымат топтомунан алынган өзгөчөлүктөр жана моделдөө стадиясына которулган өзгөчөлүктөр тиешелүү бөлүмдө көрсөтүлөт. (Alfaro ж.б. 2008; Zięba, Tomczak, жана

Tomczak 2016; Zhou, Lai, жана Yen 2014; 2012; Yeh, Chi, жана Lin 2014a; Alam, Gao, жана Jones 2021; Ahn жана Kim 2009; Z. Zhang ж.б. 2022; Xiao ж.б. 2012; Chandra, Ravi, жана Bose 2009; Cheng, Chan, жана Sheu 2019; Yang, You, жана Ji 2011; Murugan Anandarajan, Lee, жана Anandarajan 2001; W.-S. Chen жана Du 2009; Zhao ж.б. 2022; L. Wang жана Wu 2017; L. Zhang ж.б. 2013; Zeng ж.б. 2020; Xu жана Wang 2009b; Yeh, Chi, жана Hsu 2010; Zhou, Lu, жана Fujita 2015; Desheng Wu, Ma, жана Olson 2022; Barboza, Kimura, жана Altman 2017; E. I. Altman ж.б. 2020; Zhou, Si, жана Fujita 2017; G. Wang ж.б. 2020; Desheng(Dash) Wu, Liang, жана Yang 2008b; Bunyaminu жана Issah 2012; Y.-P. Huang жана Yen 2019; Olson, Delen, жана Meng 2012b; Zhou, Lai, жана Yen 2014; Li, Andina, жана Sun 2013; Mselmi, Lahiani, жана Hamza 2017; Shin, Lee, жана Kim 2005; J. Sun ж.б. 2017; G. Wang, Chen, жана Chu 2018; X. Du ж.б. 2020; Liang ж.б. 2020b; Pham Vo Ninh, Do Thanh, жана Vo Hong 2018).

Изилдөөдө колдонулган бардык маалыматтар Datastream маалымат базасынан алынган. 4.2-таблицада өзгөчөлүктүн аталышы жана эсептөөдө колдонулган формула камтылган. Datastream маалымат базасындагы ар бир маалымат уникалдуу кодду камтыйт. Мисалы, ишканалардын жалпы активдери "WC02999" коду жана " Total Assets" деген аталыш менен маалымат базасында турат. 4.2-таблицада өзгөчөлүктөрдү эсептөө үчүн колдонулган формулалар Datastream коду жана аты менен бирге берилген.

#### **4.8. R Studio версиясы жана колдонулган китепканалар (жыйнактар)**

Изилдөөдө каржылык катыштарды же өзгөчөлүктөрдү эсептөө Microsoft Excel программасын колдонуу менен Datastream маалымат базасынан алынган маалыматтардын жардамы менен жасалган. Эсептөөдөн кийин аткарылган бардык операциялар R программалоо тили жана R Studio чөйрөсүндө аткарылган. 2000-жылы жарыялангандан бери R программалоо тили көптөгөн изилдөөчүлөрдүн жана илимпоздорунун салымы менен ар кандай максаттар үчүн даярдалган жалпысынан 19834 пакетти колдонуучуларга сунуштайт («R болгон пакеттер»). Изилдөөдө колдонулган R программалоо тилинин версиясы, R Studio версиясы жана колдонулган пакеттер тууралуу маалымат төмөндөгүдөй;

- R программалоо тилинин версиясы “4.1.3”,

- R Studio версиясы: “RStudio 2023.03.0+386 "Cherry Blossom" Release (3c53477afb13ab959aeb5b34df1f10c237b256c3, 2023-03-09) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) RStudio/2023.03.0+386 Chrome/108.0.5359.179 Electron/22.0.3 Safari/537.36”,
- “adabag” (Alfaro, Gámez, жана García 2013),
- “Amelia” (Honaker, King, жана Blackwell 2011),
- “caret” (Kuhn 2008),
- “rattle” (G. Williams 2011),
- “naniar” (Tierney жана Cook 2023),
- “gbm” (Greenwell ж.б. 2022),
- “kernlab” (Karatzoglou ж.б. 2004),
- “mice” (Buuren жана Groothuis-Oudshoorn 2011),
- “ranger” (Wright жана Ziegler 2017),
- “readxl” (Wickham жана Bryan 2023),
- “rpart” (Therneau жана Atkinson 2022),
- “unbalanced” (Dal Pozzolo, Caelen, жана Bontempi 2015),
- “Tidyverse” (Wickham ж.б. 2019),
- “xgboost” (T. Chen ж.б. 2022),
- “e1071” (D. Meyer ж.б. 2023),
- “dplyr” (Wickham ж.б. 2023),
- “plyr” (Wickham 2011),
- “pROC” (Robin ж.б. 2011).

#### **4.9. Маалыматтарды алдын ала иштетүү процесси**

Машиналык окутуу процессинде жалпылоо көрсөткүчтөрү жогору моделдерди окутууда колдонулган маалымат топтомунун сапаты абдан маанилүү роль ойнойт (Hagendorff 2021). Машиналык окутуу колдонмолорунда моделдөө үчүн даяр маалыматтарды алуу мүмкүн эмес. Маалыматтарды моделдөө үчүн даяр формага айландыруу үчүн ар кандай процесстерди жүргүзүү керек. Негизинен чийки маалыматтар жетишпеген маалыматтарды толтуруу жана өзгөчөлүктөрдү тандоо сыяктуу алдын ала иштетүү стадиясында иштетилет. Колдонула турган операциялар кээ бир учурларда берилиштердин структурасына байланыштуу керек

болсо, алгоритмди тандоо чийки маалыматтар боюнча ар кандай операцияларды талап кылышы мүмкүн. Бул контекстте маалыматтарды алдын ала иштетүү стадиясында жетишпеген маалыматтарды толтуруу, нормалдаштыруу, маалымат топтомун тең салмактоо жана өзгөчөлүктү тандоо сыяктуу төрт процесс аткарылды. Маалыматтарды алдын ала иштетүү стадиясындагы бардык иш-аракеттер R Studio чөйрөсүндө аткарылган. Маалыматтар топтомундагы жетишпеген маалыматтар тиешелүү өзгөчөлүктүн орточо маанисин колдонуу менен аяктады. Маалыматтар топтомун нормалдаштыруу үчүн Min-Max ыкмасы колдонулган. Маалыматтар топтомун тең салмактоо үчүн SMOTE ыкмасы колдонулган. Өзгөчөлүктөрдү тандоо үчүн генетикалык алгоритмдер колдонулган.

#### 4.10. Изилдөөнүн натыйжалары

##### 4.10.1. Кыскача статистика

4.5-таблицада маалыматтарды алдын ала иштетүү процессине чейинки маалыматтар топтомунун жалпы статистикасы камтылган. Демек ар бир өзгөчөлүктүн минималдуу, 1-чейрек, медиана, орточо, 3- чейрек, жана максималдуу маанилери таблицанда көрсөтүлгөн.

**4.5-таблица.** Өзгөчөлүктөрдүн кыскача статистикасы

Минималдуу	1-чейрек	Медиана	Орточо	3-чейрек	Максимум
3.356	5.218	5.712	5.744	6.195	8.101
-64.190	3.498	8.375	9.823	14.790	137.590
-1572.320	1.955	12.885	9.822	27.872	1235.580
0.000000	0.006818	0.030080	0.069425	0.098177	0.573386
0.000	0.026	0.122	24.014	0.430	16075.750
0.00000	0.03178	0.10180	0.15499	0.23732	0.93659
0.00000	0.03671	0.13784	0.37524	0.38150	18.36564
0.030	0.560	0.840	1.172	1.295	22.130
0.030	1.055	1.360	1.838	2.085	25.280
-17.03541	0.02567	0.17065	0.19645	0.39356	33.12458
-4.93436	0.02331	0.14911	0.14627	0.31423	0.82743
0.007215	0.242073	0.451378	0.556205	0.712355	6.211675
0.02222	0.42581	0.58447	0.56732	0.71307	0.99682
0.00318	0.28693	0.41553	0.43268	0.57419	0.97778
0.1764	1.3888	1.7234	2.4596	2.8076	47.3371
-172.826	1.587	2.448	3.116	3.587	112.745
0.0000	0.1357	0.2762	0.2824	0.4108	0.8659
-44.06480	0.02234	0.20385	0.45246	0.56072	30.77772
0.0000	0.1052	0.2783	0.2879	0.4174	2.8934
-102.2142	0.1674	0.6739	1.1360	1.4049	74.6958
0.07626	0.63149	0.76659	0.74947	0.88559	2.03128
0.07626	0.63149	0.76659	0.74947	0.88559	2.03128
-127.5247	0.4189	0.9694	1.5458	1.8389	111.5564
0.0402	0.3095	0.4820	0.6385	0.6904	23.7492
0.02113	0.35617	0.58023	0.56858	0.72007	5.66884

0.000	1.612	2.139	10.779	3.508	1954.400
-173.8262	0.5392	1.3872	2.0799	2.5402	111.7450
-20.508370	-0.000998	0.147441	0.066281	0.305491	0.867798
-67.4	0.6	1.3	488.5	4.1	325163.7
0.2223	2.9733	4.1772	5.4470	5.8097	121.5943
0.260	2.790	4.200	7.231	6.588	450.480
-3231.040	1.300	3.315	2.238	8.333	1134.273
0.01758	1.13573	1.52764	1.63386	1.89946	6.42559
0.0100	0.5900	0.8000	0.8851	1.0700	4.1600
0.2092	3.0928	4.6717	6.7695	7.4771	120.4895
-1.6269	0.2660	0.4409	0.5988	0.7392	12.6144
0.06363	0.95587	1.40759	1.84364	2.22186	16.82756
-3.30829	0.04288	0.10504	0.11051	0.17905	3.23090
-0.75551	0.03429	0.08739	0.09559	0.14793	0.91592
-149.280	1.076	3.163	47.400	10.772	7836.529
-436.020	0.790	5.800	4.444	12.520	321.830
-6.78826	0.01330	0.10050	0.08922	0.22185	1.75118
-3.79177	0.01577	0.13490	0.25001	0.33111	7.18284
-965.50	0.01	0.16	182.62	0.61	134442.33
-9.9570	0.0245	0.2470	1.5689	0.9980	191.5950
-8.844	0.005	0.318	1.495	1.154	113.037
-3.923	1.101	2.257	7.609	6.613	355.046
-11.0300	0.0785	0.4850	2.0585	1.5050	152.2700
-17.1644	-0.1063	0.1605	1.3350	1.0078	100.1159
-4.54582	0.02618	0.08081	0.08249	0.14658	0.84775
-915.33	0.07	0.29	213.31	0.82	159467.00
-13.59722	0.06464	0.21702	0.34358	0.43095	15.70046
-3.57320	0.05242	0.15233	0.26961	0.31172	12.93803
-10.3445	0.0600	0.2089	0.2681	0.3769	21.1218
-474.360	3.460	9.870	8.872	17.235	492.440
-4.65947	-0.02392	0.03838	0.03716	0.10927	0.81965
-1357.83	-0.07	0.13	66.51	0.51	49523.33
-5.77801	-0.06156	0.09397	0.20064	0.32072	14.21341
-5.30059	-0.04258	0.07023	0.15641	0.22126	11.71261
-1487.429	-0.708	1.474	18.876	6.856	6427.235
-34.59136	-0.02931	0.04962	-0.01717	0.12814	0.97052
-577.5424	-0.2939	0.1634	-0.2811	0.6982	163.5534
-0.9724	0.1049	0.2483	0.3708	0.4660	13.0870
-1.000	-0.018	0.188	15.987	0.512	9024.000
-76.18182	-0.08067	0.17151	0.92675	0.54801	240.87097
-0.9070	0.1174	0.2707	0.4163	0.4896	19.2226
-285.5000	-1.2620	-0.2665	1.1397	0.9519	480.6000
-1.0000	-0.2879	0.2648	1.7952	1.3149	69.6283
-72.3013	0.0379	0.2058	0.4622	0.5078	108.8154
-46.0484	-0.7080	0.1887	0.8037	1.2441	84.9924
2369	146302	506700	2893384	1895463	98560000
0.0	1.3	3.1	1294.6	11.1	629078.3
0.00000	0.01512	0.03483	0.05329	0.06200	1.07955
-90.130	1.087	1.730	3.262	2.942	342.700
0.000	0.636	0.854	8.675	1.234	3775.808

#### 4.10.2. Окутулган моделдердин натыйжалары

Машиналык окутуу моделдеринин иштеши маселенин спецификалык мүнөздөмөлөрүнөн жана маалымат топтомунун түзүмүнөн көз каранды экендиги бир ыкманын эң мыкты ыкма катары чыгышына жол бербейт (Duéñez-Guzmán жана Vose 2013). Ушул себептен улам изилдөөдө бирден ашык машиналык окутуу

алгоритми окутулган жана адабиятта кеңири колдонулган метрикаларга ылайык билдирилген (M.-Y. Chen 2011b). Логистикалык регрессия, таяныч векторунун машиналары (сызыктуу, полиномдук жана радиалдык негиздеги ядро функциялары), классификация жана регрессия дарагы, кокус токой алгоритми, adaboost алгоритми, gradient boosting алгоритми жана extreme gradient boosting алгоритми эмпирикалык натыйжалары деталдуу көрсөтүлүп, талкууланат. Моделдерди баалоо максатында адабиятта кеңири колдонулган метрикалар пайдаланылган. (Alfaro ж.б. 2008; E. I. Altman ж.б. 2020; Murugan Anandarajan, Lee, жана Anandarajan 2001; Barboza, Kimura, жана Altman 2017; Bunyaminu жана Issah 2012; Chandra, Ravi, жана Bose 2009; W.-S. Chen жана Du 2009; Cheng, Chan, жана Sheu 2019; Cho, Hong, жана Ha 2010; G. Wang, Chen, жана Chu 2018; Cleofas-Sánchez ж.б. 2016; De Bock 2017; Fallahpour, Lakvan, жана Zadeh 2017b; Fedorova, Gilenko, жана Dovzhenko 2013; Yeh, Chi, жана Lin 2014b; Жана rikas ж.б. 2010; Zhou жана Lai 2017; Garcia 2022). Тесттик маалыматтар топтому менен ар бир модель менен жасалган божомолдордун натыйжаларына байланыштуу каталар матрицалары жана эсептелген тууралык, Сезимталдуулук, Тандалмалык, тактык, I жана II типтеги каталар жана AUC упайлары көрсөтүлдү.

Маалыматтарды алдын ала иштетүү стадиясынан кийин алынган маалыматтар топтому көптөгөн ар кандай изилдөөлөрдө колдонулган 80/20 катышына ылайык туш келди экиге бөлүнүп, окутуу жана тесттик маалымат топтомдору түзүлдү (Harris 2015; He, Zhang, жана Zhang 2018). Ошентип маалымат топтомунун 0,8и окутуу маалыматы жана 0,2си тесттик маалымат катары сакталат. Окутулган моделдердин болжолдуу ийгилигин камсыз кылуу үчүн адабиятта кеңири колдонулган жана натыйжалуулугу далилденген 10 катмарлуу кайчылаш текшерүү процедурасы колдонулган (Alfaro ж.б. 2008; Chandra, Ravi, жана Bose 2009; G. Wang, Chen, жана Chu 2018; Cleofas-Sánchez ж.б. 2016; De Bock 2017; Fallahpour, Lakvan, жана Zadeh 2017b).

#### **4.10.2.1. Логистикалык регрессия модели**

Логистикалык регрессия экилик же ырааттуу жооп өзгөрмө менен түшүндүрүүчү өзгөрмөлөр ортосундагы байланышты изилдөө үчүн колдонулган машиналык окутуу ыкмасы болуп саналат. Логистикалык регрессия көптөгөн изилдөөчүлөр тарабынан адабияттардагы ар кандай изилдөөлөрдө колдонулган (M.-Y. Chen

2011a; Chye Koh жана Kee Low 2004; Hua ж.б. 2007; G. Zhang ж.б. 1999; Chandra, Ravi, жана al. Bose 2009).

**4.6-таблица.** Логистикалык регрессия моделине таандык каталар матрицасы

Болжол	Шилтеме		
	147 (GP)	8 (YP)	155
	16 (YN)	146 (GN)	162
163	154		

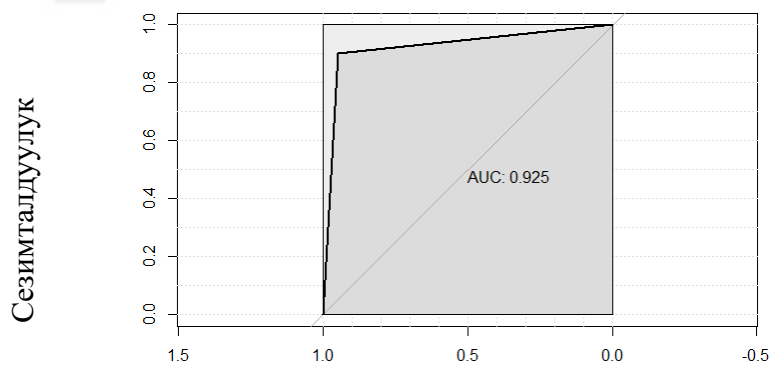
Логистикалык регрессиялык моделди окуткандан кийин тесттик маалыматтар топтому менен жасалган болжолдоонун натыйжаларын көрсөткөн каталар матрицасы 4.6-таблицада көрсөтүлгөн. Логистикалык регрессия моделинин тест маалымат топтомундагы жалпы 163 кооптуу байкоолордун ичинен 147 туура классификацияланган жана 16 туура эмес классификацияланган. Модель тобокелсиз болгон 154 байкоолордун ичинен 146 туура жана 8 туура эмес классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 155ин бизнестин үзгүлтүксүздүгү үчүн тобокелдүү жана 162син бизнестин үзгүлтүксүздүгүнө байланыштуу тобокелчиликсиз деп классификациялаган.

**4.7-таблица.** Логистикалык регрессия моделинин моделдик баалоо метрикалары

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9243	0.9018	0.9480	0.0519	0.0981	0.9483

Логистикалык регрессиялык моделдин моделдик баалоо метрикалары 4.7-таблицада көрсөтүлгөн. Логистикалык регрессия моделинин тууралык метрикасы 0,9243 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык логистикалык регрессия модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 92,43% туура болжолдогон. Моделдин сезимталдуулук метрикасы 0,9018 жана Тандалмалык метрикасы 0,9480 деп эсептелген. Сезимталдуулук көрсөткүчү үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрикасы үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Аныкталган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 90,18% туура болжолдогондугун, ал эми аныкталган Тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок

байкоолордун 94,80% туура болжолдоосун көрсөтөт. Сезимталдуулук жана Тандалмалык метрикаларынын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I-типтеги катасы 0,0519. I-типтеги ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы 0,0981. II типтеги ката – үзгүлтүксүз болуу тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Каталар матрицасы аркылуу эсептелген акыркы метрика тактык болуп саналат. Тактык метрикасы моделдин шилтеме классын канчалык так болжолдонгонунун көрсөткүчү. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган болжолдонгон байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигин алып келээринин өлчөмү. Моделдин тактык метрикасы 0,9483. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 94,83% үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

#### 4.3-сүрөт. Логистикалык регрессия моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.3-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы

ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокелди бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 92,5% ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дөн жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

Логистикалык регрессия ыкмасы ишкана үзгүлтүксүздүгү боюнча адабиятта статистикалык мезгилде көптөгөн изилдөөлөрдө үзгүлтүксүздүк тобокелдигин өлчөө максатында негизги ыкма катары жана машиналык окутуу ыкмаларынын статистикалык техникалары орун алган 2000-жылдардан баштап сунушталган моделди салыштыруу үчүн кеңири колдонулуп келет. Хiao жана башкалардын (2012) 276 байкоону камтыган маалымат топтому менен жүргүзүлгөн изилдөөлөрүндө логистикалык регрессия модели ишкана тобокелдүү деп эсептелгенге чейин t-2 жана t-3 мезгилине тиешелүү маалымат топтомдору менен окутулган. Логистикалык регрессия моделинин тууралык метрикасы t-2 мезгилинде 0,8346 жана логистикалык регрессия моделинин тууралык метрикасы ү t-3 мезгилинде 0,6400 деп эсептелген. Изилдөөнүн жыйынтыктары ишкананын тобокелдүү деп кабыл алынган мезгилге жакындаган сайын болжолдун ийгилиги жогорулайт экенин көрсөттү. L. Zhang жана башкалардын (2013) 92 ишкананын маалымат топтому менен жүргүзгөн жана финансылык катыштарды гана колдонгон изилдөөлөрүндө логистикалык регрессиялык моделдин тууралык метрикасы 0,9048 деп эсептелген. Zhou, Lai, жана Yen (2014) тарабынан даярдалган 7728 байкоодон турган маалымат топтому менен изилдөөдө логистикалык регрессиялык моделдин тууралык метрикасы 0,7399, сезимталдуулук 0,7828 жана Тандалмалык 0,6971 деп табылган. С.-F. Tsai (2014) Австралия, Германия, Жапония, Банкроттук жана UScompetiton маалымат топтомдору менен беш түрдүү логистикалык регрессиянын моделдерин окуткан. Австралиялык маалымат топтому менен окутулган моделдин тактык метрикасы 0,8714, I типтеги ката 0,1245 жана II типтеги ката 0,1330 деп эсептелген. Германиянын маалымат топтому менен окутулган моделдин тактык метрикасы 0,7722, I типтеги ката 0,0981 жана II типтеги ката 0,5137 деп эсептелген. Жапониянын маалымат топтому менен окутулган моделдин тактык метрикасы 0,8574, I типтеги ката 0,1171 жана II типтеги ката 0,1632 деп эсептелген. Банкроттук маалыматтары менен окутулган

моделдин тактык метрикасы 0,8291, I типтеги ката 0,1386 жана II типтеги ката 0,2078 деп эсептелген. UCcompetition маалымат топтому менен окутулган моделдин тактык метрикасы 0,8224, I типтеги ката 0,2293 жана II типтеги ката 0,1255 деп эсептелген. Veganzones жана Séverin (2018) 1500 байкоодон жана каржылык катыштан турган маалымат топтому менен окуткан логистикалык регрессия моделинин сезимталдуулук метрикасы 0,500, Тандалмалык метрикасы 0,965 жана AUC упайы 0,771 деп аныкталган. W. Zhang, He, жана Zhang (2019) изилдөөсүндө беш түрдүү маалымат топтомун колдонушкан. Австралиялык маалымат топтому менен окутулган логистикалык регрессия модели 0,8616 тууралык метрикасына жана 0,9255 AUC баллына ээ болгон, Германиянын маалымат топтому менен окутулган логистикалык регрессия модели 0,7317 тууралык метрикасына жана 0,7684 AUC баллына ээ болгон, Жапониянын маалымат топтому менен окутулган логистикалык регрессия моделинин тууралык метрикасы 0,8507 жана AUC баллы 0,9263 деп эсептелген, PPDai маалымат топтому менен окутулган логистикалык регрессия моделинин тууралык метрикасы 0,8713, AUC баллы 0,6506 жана. GMSC маалымат топтому менен окутулган логистикалык регрессия моделинин тууралык метрикасы 0,9313 жана AUC баллы 0,6892 болгон. Z. Zhang жана башкалардын (2022) изилдөөлөрүндө 2007-2020-жылдар аралыгындагы 2260 байкоону камтыган маалыматтар топтому менен окутулган логистикалык регрессия моделинин тууралык метрикасы 0,699, сезимталдуулук метрикасы 0,684, тактык метрикасы 0,705 жана AUC баллы 0,754. Изилдөөнүн маалыматтар топтомуна финансылык катыштар гана киргизилген. Бөлүштүрүлгөн метрикалары менен моделдер алар катышкан изилдөөлөрдө базалык классификатор жана сунушталган моделдер үчүн салыштыруу куралы катары колдонулган. Адабиятта орун алган ар кандай изилдөөлөрдө окутулган логистикалык регрессия моделдеринин эсептелген баалоо чен белгилери менен жүргүзүлгөн иштеги логистикалык регрессия моделинин модели салыштырылганда иштеги окутулган логистикалык регрессия моделинин жакшыраак болжолдоо жөндөмүнө ээ экени көрүнөт.

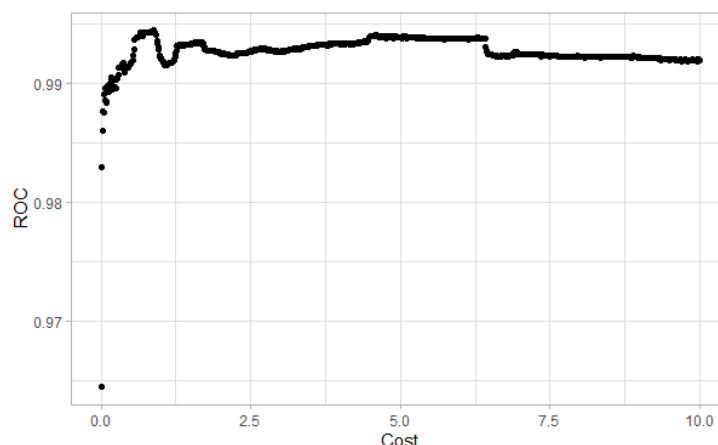
#### **4.10.2.2. Таяныч векторлорунун машинасы моделдери**

Таяныч векторлорунун машинасы бинардык классификация маселелери үчүн кеңири колдонулган ийгиликтүү машиналык окутуу ыкмасы (Ding, Song жана Zen

2008; J. Sun ж.б. 2017; 2021; J. Sun жана Li 2012). Ыкма маалымат топтомундагы байкоолорду жана кээ бир сызыктуу эмес кирүү векторлорун жогорку өлчөмдүү өзгөчөлүк мейкиндигине жылдыруу үчүн сызыктуу ыкманы колдонот. Жаңы байкоо мейкиндигинде түзүлгөн сызыктуу модель баштапкы байкоо мейкиндигинин чыныгы сызыктуу эмес чечим чектерин билдирет. Параметрлерди издөө жана оптималдаштыруу алгоритмдерди окутууда абдан маанилүү. Бирок адабиятта эң так ядро функциясын тандоо жана параметрди оптималдаштыруу боюнча жалпы ыкма жок. Ядро функцияларынын арасында радиалдык негиз функциясын колдонуу боюнча жалпы макулдашуу бар (Вае 2012). Ядро функцияларын салыштырган изилдөөчүлөр ядронун полиномдук функциясын окутуу радиалдык ядролук функцияга караганда кымбатыраак жана начар натыйжаларды берет деп белгилеп келишет (Ding, Song, жана Zen 2008; Z. Huang ж.б. 2004). Изилдөөдө үч түрдүү ядро функциясы, сызыктуу, полиномдук жана радиалдык негизде колдонулган.

#### 4.10.2.2.1. Сызыктуу таяныч векторлорунун машинасы модели

Сызыктуу ядро функциясын колдонгон вектордук машиналарды колдогон жалгыз параметр бул  $C$  параметри. Параметр моделдин туура эмес классификациядан качышы керектигин аныктайт. Параметр 0,001ден 10го чейинки диапазондо текшерилип, ар бир тестте 0,01ге көбөйүп, ROC баллынын өлчөгүн колдонуп, эң жакшы ROC баллын берген маани тандалды. Демек  $C$  параметри 0,871 болгондо 0,9944 мыкты ROC баллы алынган.



**4.4-сүрөт.** Сызыктуу таяныч векторлорунун машинасы параметрин оптималдаштыруу ( $Cost=C$ )

4.4-сүрөттө параметрди оптималдаштыруу процессинде сыналган маанилер жана эсептелген ROC баллдары менен графика көрсөтүлгөн.

**4.8-таблица.** Сызыктуу колдоонун вектордук машиналар моделинин ката матрицасы

Болжол	Шилтеме		
	163 (GP)	8 (YP)	171
	0 (YN)	146 (GN)	146
163	154		

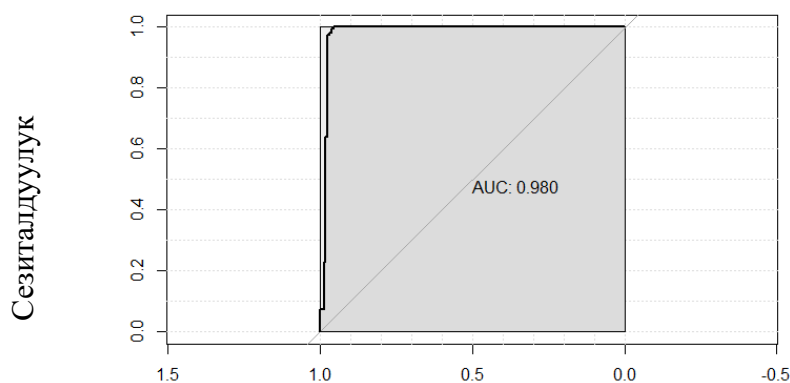
Сызыктуу колдоонун вектордук машина моделин окуткандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.8-таблицада көрсөтүлгөн. Модель тесттик маалымат топтомундагы бардык 163 тобокелдүү байкоону туура классификациялаган. Модель 154 тобокелдүү эмес байкоолордун ичинен 146 туура жана 8 туура эмес байкоону классификациялаган. Сыноодогу маалыматтар топтомундагы 317 байкоонун 171и ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 146сы ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

**4.9-таблица.** Сызыктуу колдоонун вектордук машина моделинин моделдик баалоо метрикалары

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9748	1	0.9480	0.0519	0	0.9532

Сызыктуу таяныч векторлорунун машинасы моделинин моделдик баалоо метрикасы 4.9-таблицада көрсөтүлгөн. Сызыктуу колдоонун вектордук машина моделинин тууралык метрикасы 0,9748 деп эсептелет. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык сызыктуу колдоонун вектордук машиналар модели тесттик маалымат топтому менен жасалган болжолдоолордун 97,48% туура болжолдогон. Моделдин сезимталдуулук метрикасы 1 жана тандалмалык метрикасы 0,9480 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Аныкталган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдигин алып жүрүүчү бардык байкоолорду туура

болжолдонгонун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 94,80% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүз тобокелдик менен байкоолорду да, үзгүлтүксүз тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Модель өзгөчө үзгүлтүксүздүк коркунучу бар байкоолорду туура классификациялоого жөндөмдүү. Моделдин тип I катасы 0,0519. Тип I ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун кооптуу катары классификациясы. II типтеги ката – үзгүлтүксүз тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрикасы болуп саналат. Тактык метрикасы модель шилтеме классын канчалык так болжолдоп жатканын көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0,9532. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 95,32% чындыгында үзгүлтүксүздүк тобокелдигин алып келерин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1- Тандалмалык

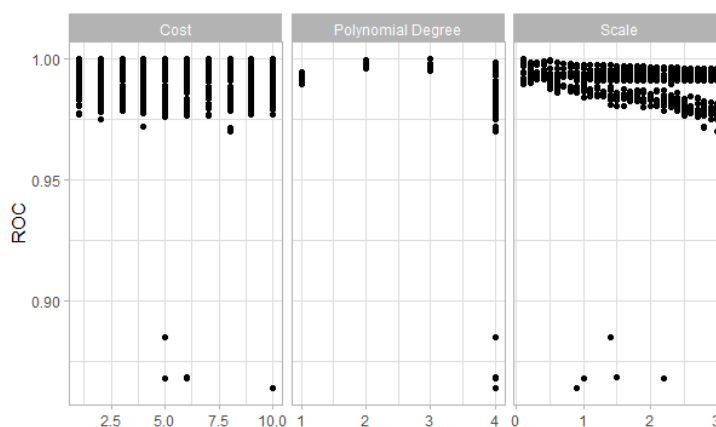
**4.5-сүрөт.** Сызыктуу таяныч векторлорунун машинасы үчүн ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.5-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокелдиги

бар жана үзгүлтүксүздүк тобокелдигисиз байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда, AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Моделдин эсептелген AUC баллы үзгүлтүксүздүк тобокели менен байкоолорду жана 98% ыктымалдык менен үзгүлтүксүздүк тобокелдигисиз байкоолорду туура айырмалоо мүмкүнчүлүгүнө ээ. Адабият изилденгенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC баллы 0,90дон жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

#### 4.10.2.2.2. Полиномдук таяныч векторлорунун машинасы модели

Полиномдук ядро функциясын колдонгон таяныч векторлорунун машиналары үч түрдүү параметрге ээ: C, scale жана degree. C параметри моделдин туура эмес классификациялоодон алдын алуу деңгээлин аныктайт. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 10го чейинки диапазондо сыналды. Degree параметри функциянын даражасын аныктайт. Параметр ар бир тестте 1ге көбөйүп, 1ден 4кө чейинки диапазондо сыналды. Scale моделдеги нормалдаштыруу параметри болуп саналат. Параметр ар бир тестте 0,1ге көбөйүп, 0,1ден 3кө чейинки диапазондо сыналды. Параметрди оптималдаштыруу ROC баллынын өлчөгүн колдонуу менен ишке ашырылган жана ROC баллынын эң жакшы маанилери тандалган. Буга байланыштуу C параметри 3, degree параметри 3 жана scale параметри 0,1 болгон шартта 0,9998 мыкты ROC баллы алынган.



**4.6-сүрөт.** Полиномдук таяныч векторлор машиналары параметрин оптималдаштыруу (Cost = C Полиномдук Degree=degree, Scale=scale)

4.6-сүрөттө параметрди оптималдаштыруу процессинде сыналган маанилер жана эсептелген ROC баллдары менен график көрсөтүлгөн.

**Таблица 4.10.** Полиномдук таяныч векторлорунун машиналарынын ката матрицасы

	Шилтеме		
Болжол	162 (GP)	3 (YP)	165
	1 (YN)	151 (GN)	152
	163	154	

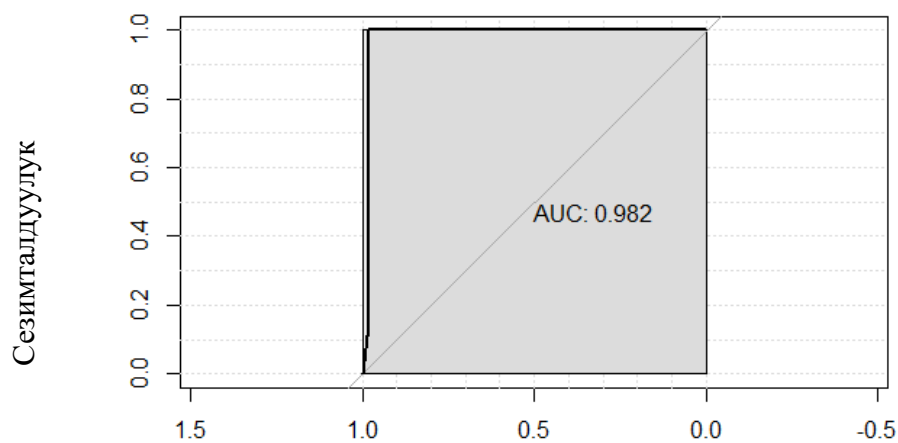
Полиномдук таяныч векторлор машиналары моделин окуткандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.10-таблицада көрсөтүлгөн. Полиномдук таяныч векторлор машиналары моделинин тест маалымат топтомундагы жалпы 163 кооптуу байкоолордун ичинен 162 туура жана 1 туура эмес деп классификацияланган. Модель 154 тобокелчиликсиз байкоонун 151ин туура, 3үн туура эмес деп классификациялаган. Ал тесттик маалыматтар топтомундагы 317 байкоонун 165ин ишкананын үзгүлтүксүздүгү жагынан тобокелдүү жана 152си ишкананын үзгүлтүксүздүгүнө байланыштуу тобокелчиликсиз деп классификацияланган.

**4.11-таблица.** Полиномдук таяныч векторлорунун машиналары моделинин моделдик баалоо метрикасы

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9873	0.9938	0.9805	0.0194	0.0061	0.9818

4.11-таблицада деполиномдук таяныч векторлорунун машиналары моделинин моделдик баалоо көрсөткүчтөрү көрсөтүлгөн. Полиномдук таяныч векторлорунун машиналары моделинин тууралык метрикасы 0,9873 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык полиномдук таяныч векторлорунун машиналары модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 98,73% туура болжолдогон. Моделдин сезимталдуулук метрикасы 0,9938 жана тандалмалык метрикасы 0,9805 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук

метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 99,38% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 98,05% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I тип катасы 0,0194 болуп саналат. Моделдин I тип катасы - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы 0,0061. II типтеги ката – үзгүлтүксүздүк тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык көрсөткүчү модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0,9818. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 98,18% үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

**4.7-сүрөт.** Полиномдук таяныч векторлорунун машиналары моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.7-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 98.2% ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дөн жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

#### 4.10.2.2.3 Радиалдык таяныч векторлорунун машинасы модели

Радиалдык негизде ядро функциясын колдонгон таяныч векторлорунун машиналары эки параметрге ээ, сигма жана C. C параметри моделдин туура эмес классификациядан качуу деңгээлин аныктайт. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 200гө чейинки диапазондо сыналды. Сигма параметри ядронун туурасын аныктайт. Ядронун кеңдиги ядро функциясынын жайылышын көзөмөлдөйт. Параметр ар бир тестте 0,1 ге көбөйүп, 0,1ден 5ке чейинки диапазондо сыналды. Параметрди оптималдаштыруу ROC баллынын өлчөгүн колдонуу менен ишке ашырылган жана ROC упайынын эң жакшы маанилери тандалган. Демек 0,9999 мыкты ROC баллы C параметри 2 жана везигма параметри 0,1 болгондо алынган.

**4.12-таблица.** Радиалдык таяныч векторлорунун машинасы моделинин ката матрицасы

Болжол	Шилтеме		
	161 (GP)	0 (YP)	161
2 (YN)	154 (GN)	156	
163	154		

Радиалдык таяныч векторлорунун машинасы моделин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.12-таблицада көрсөтүлгөн. Радиалдык таяныч векторлорунун машинасы моделинин тест маалымат топтомундагы 163 кооптуу байкоолордун 161и туура классификацияланган жана 2си туура эмес классификацияланган. Модель 154 тобокели жок байкоолордун 154үн туура

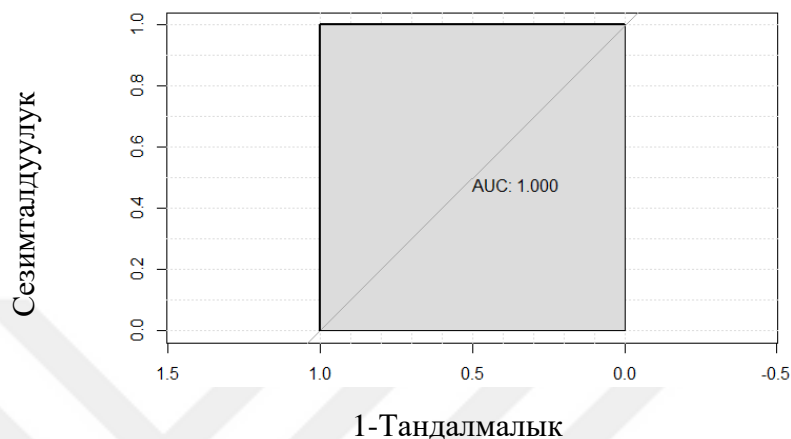
классификациялаган. Сыноодогу маалыматтар топтомундагы 317 байкоонун 171и ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 146сы ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

**4.13-таблица.** Радиалдык таяныч векторлорунун машинасы моделинин моделдик баалоо метрикалары

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9936	0.9877	1	0	0.0122	1

4.13-таблицада радиалдык таяныч векторлорунун машиналары моделинин моделдик баалоо көрсөткүчтөрү көрсөтүлгөн. Радиалдык таяныч векторлорунун машиналары моделинин тууралык метрикасы 0.9936 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык радиалдык таяныч векторлорунун машиналары модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 99.36%’ туура болжолдогон. Моделдин сезимталдуулук метрикасы 0.9877 жана тандалмалык метрикасы 1 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 98.77% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун бардыгын туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I тип катасы эсептелген эмес. I тип ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы 0.0122. II типтеги ката – үзгүлтүксүздүк тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык көрсөткүчү модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө

үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 1. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн бардыгы үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



#### 4.8-сүрөт. Радиалдык таяныч векторлорунун машиналары моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.8-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 100% ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дөн жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

Таяныч векторлорунун машиналары үзгүлтүксүздүктү баалоо адабиятында өзгөчө финансылык мүнөздөгү маалыматтар менен жалпылоонун жогорку көрсөткүчтөрүнө байланыштуу изилдөөлөрдө сунушталган моделди окутуу үчүн да, салыштыруу максаттары үчүн да кеңири колдонулат. Yeh, Chi, жана Hsu (2010) 114 байкоонун маалымат топтому менен жүргүзүлгөн изилдөөсүндө гаусстук радиалдык негизде ядронун өзөгүн колдонуу менен эки таяныч векторлорунун машинасы моделдерин окутушкан. Моделдердин биринчиси изилдөөдө кошумча

маалыматты алуу үчүн колдонулган иштөө чөйрөсүнүн анализинин жыйынтыгын камтыбайт. Моделдин тууралык метрикасы 0,8333, I типтеги ката 0,1579 жана II типтеги ката 0,1842 деп аныкталган. Экинчи таяныч векторлорунун машинасы модели иштөө чөйрөсүнүн анализинин жыйынтыгын камтыйт. Моделдин тууралык метрикасы 0,8684, I типтеги ката 0,1447 жана II типтеги ката 0,1053 деп эсептелген. Yang, You, жана Ji (2011) изилдөөлөрүндө 120 байкоо орун алган маалымат топтому менен окутулган радиалдык таяныч векторлорунун машинасы моделинин тууралык метрикасы 0,8667, сезимталдуулук метрикасы 0,8226, тандалмалык метрикасы 0,9138 жана тактык метрикасы 0,9170 катары эсептелген. Xiao жана башкалардын (2012) 276 байкоо орун алган маалымат топтому менен жүргүзүлгөн изилдөөлөрдө таяныч векторлорунун машинасы модели ишкана тобокелдүү деп эсептелгенге чейин t-2 жана t-3 мезгилине тиешелүү маалымат топтомдору менен окутулган. Таяныч векторлорунун машинасы моделинин тууралык метрикасы t-2 мезгилинде 0,8543 жана таяныч векторлорунун машинасы моделинин тууралык метрикасы t-3 мезгилинде 0,6440 катары эсептелген. Изилдөөнүн жыйынтыктары көрсөткөндөй, болжолдоонун ийгилиги ишкана тобокелдүү деп эсептелген мезгилге жакындаган сайын жогорулайт. L. Zhang жана башкалар (2013) тарабынан 92 ишкананын маалымат топтому менен жүргүзүлгөн изилдөөдө окутулган радиалдык таяныч векторлорунун машинасы моделинин тактык көрсөткүчү 0,9048 деп эсептелген. Wang, Ma, жана Yang (2014) изилдөөсүндө таяныч векторлорунун машинасы моделин окутуу Pietruszkiewicz (2008) жана Shmueli, Patel, жана Bruce (2010) тарабынан түзүлгөн эки башка маалымат топтомун колдонуу менен ишке ашырылган. Биринчи моделдин тууралык метрикасы 0,7221, I типтеги ката 0,2501 жана II типтеги ката 0,3019 деп аныкталган, экинчи моделдин тууралык метрикасы 0,7999, I типтеги ката 0,1819 жана II типтеги ката 0,2155. Zhou, Lai, жана Yen (2014) 6509 байкоо орун алган изилдөөлөрүндө окутулган генетикалык алгоритмдер менен өзгөчөлүктү тандоо колдонулган жана радиалдык негизде ядро функциясы колдонулган таяныч векторлорунун машинасы моделинин тууралык метрикасы 0,7627, сезимталдуулугу 0,7453 жана тандалмалык 0,7807 деп аныкталган. Yeh, Chi, жана Lin (2014) изилдөөлөрүндө 220 байкоо камтылган радиалдык негиздеги таяныч векторлорунун машинасы моделинин I типтеги катасы 0.1102 жана II типтеги ката 0,0374 деп эсептелген. Окутулган моделдин бардык метрикалары изилдөөдө

сунушталган модель менен салыштыруу максатында бөлүшүлгөн эмес. С.-F. Tsai, Hsu, жана Yen (2014) Австралиянын, Германиянын жана Жапониянын маалымат топтомдорун колдонуу менен үч түрдүү таяныч машина моделдерин окутушкан. Австралиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8563, Германиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,7568 жана Жапониянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8637. Ал Belfirst маалымат базасынан Volkov, Benoit, жана Van den Poel (2017) изилдөөлөрүндө 2011, 2012, 2013, 2014 жана 2015-жылдары үчүн 2007-2015-жылдарды камтыган чоң маалымат топтому менен беш түрдүү таяныч векторлорунун машиналары моделдерин окутушкан. 2011-жылга таандык эң ийгиликтүү моделдин тууралык метрикасы 0,7698 жана моделдин AUC баллы 0,8357. 2013-жылга таандык моделдин тууралык метрикасы 0,7717 жана AUC баллы 0,8352 болду. 2014-жылга таандык моделдин тууралык метрикасы 0,7646 жана моделдин AUC баллы 0,8231 деп аныкталган. 2015-жылдагы моделдин тууралык метрикасы 0,7705 жана моделдин AUC баллы 0,8344. П. Wang, Chen, жана Chu (2018) изилдөөсүндө 1726 ишканынын маалымат топтому колдонулуп, ишкана тобокелдүү деп классификацияланган жылдан 3.4 жана 5 жыл мурун маалыматтары бар полиномдук ядро өзөктөрдү колдонуу менен таяныч векторлорунун машиналары моделдерин окутушкан. Окутулган моделдердин үчөө SMOTE менен тең салмакталган маалыматтарды колдонуу менен окутулган. SMOTE жок моделдердин натыйжалары тиешелүүлүгүнө жараша тууралык метрикасы 0,7633, I типтеги ката 0,2267, II типтеги ката 0,3605, t-4 мезгилине таандык үлгүдөгү тууралык метрикасы 0,7672, I типтеги ката 0,2271 жана II типтеги ката 0,3034 жана t-5 мезгилине таандык моделдин тууралык метрикасы 0,7558, I типтеги ката 0,2341 жана II типтеги ката 0,3692 деп эсептелген. SMOTE аркылуу тең салмакталган маалымат топтому менен окутулган t-3 мезгил моделинин тууралык метрикасы 0,8806, I типтеги ката 0,1199 жана II типтеги ката 0,1132, t-4 мезгилине таандык моделдин тууралык 0,8649, I типтеги ката 0,1334, I типтеги ката 0,1334 II ката 0,1561 жана t-5 мезгилине таандык моделдин тууралык метрикасы 0,8413, I типтеги ката 0,1545 жана II типтеги ката 0,2107 болгон. Изилдөөнүн жыйынтыгы көрсөткөндөй, маалыматтар топтомун тең салмактоо моделдин ийгилигине өбөлгө түзөт. Veganzones жана Séverin (2018) 1500 байкоо жана финансылык катыштар топтому менен окутулган радиалдык негиздеги ядро

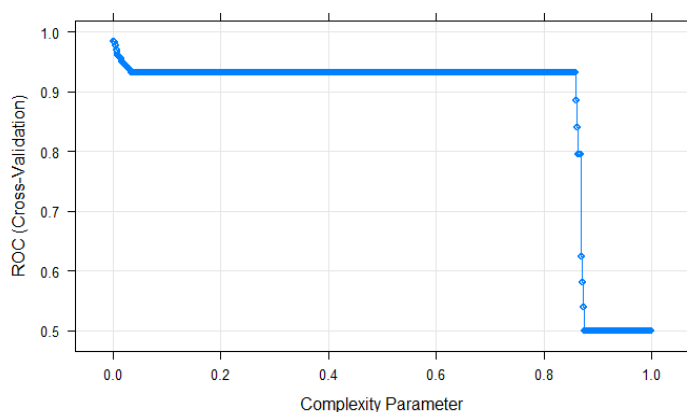
өзөгүн колдонуу менен таяныч векторлорунун машиналары моделинин сезимталдуулук метрикасы 0,725, тандалмалык метрикасы 0,913 жана AUC баллы 0,871 деп аныкталган. Zhang, He, жана Zhang (2019) Австралия, Германия, Япония, PPDa1 жана GMSC маалымат топтомдорун изилдөөдө колдонуу менен беш түрдүү таяныч векторлорунун машиналары моделдерин окутушту. Австралиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8507 жана AUC баллы 0,9183, Германиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,7483 жана AUC баллы 0,6060, Жапония маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8493 жана AUC баллы 0,8110, PPDa1 маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8709 жана AUC баллы 0,5424 жана GMSC маалымат топтому менен маалымат топтому менен окутулган моделдин тууралык метрикасы 0,9315 жана AUC баллы 0,6288 болуп эсептелген. G. Wang жана башкалар (2020) изилдөөсүндө 39 финансылык катышты колдонушкан. Изилдөөдө финансылык катыштарды гана камтыган маалыматтар топтому менен окутулган таяныч векторлорунун машиналары моделинин AUC баллы 0,8624 катары эсептелген. Z. Zhang жана башкалар (2022) негизги классификатор катары окутулган радиалдык таяныч векторлорунун машиналары моделинин тууралык метрикасы 0,701, моделдин сезимталдуулугу 0,701, тактыгы 0,705 жана AUC баллы 0,753 экенин аныктаган. Адабияттагы ар кандай изилдөөлөрдө окутулган таяныч векторлорунун машиналары моделдерин баалоо чен белгилери изилдөөдөгү таяныч векторлорунун машиналары моделдерин баалоо метрикалары менен салыштырылганда, изилдөөдө окутулган таяныч векторлорунун машиналары моделдеринин баалоосу жакшыраак экени көрүнүп турат. аткаруу.

#### **4.10.2.3. Классификация жана регрессия дарагынын модели**

Чечим дарактары классификация жана регрессия маселелерин чечүү үчүн кеңири колдонулган машиналык окутуу ыкмасы. Бизнес үзгүлтүксүздүгүн болжолдоо боюнча адабияттардагы көптөгөн изилдөөчүлөр чечим дарактарын колдонушкан. (Chye Koh жана Ki Low 2004b; Gepp, Kumar, жана Bhattacharya 2010a; M.-Y. Chen 2011a).

Классификация жана регрессия дарагынын моделине ээ болгон жалгыз параметр бул ср параметри. Параметр дарактарды түзүү процессинде дарактын татаалдыгы

(өлчөмү) менен тууралыктын ортосундагы тең салмактуулукту жөнгө салат. Параметр 0,00001ден 1ге чейинки диапазондо текшерилип, ар бир тестте 0,00001ге көбөйүп, ROC баллын өлчөгүн колдонуп, эң жакшы ROC баллын берген маани тандалды. Демек ср параметри 0,00179 болгондо, эң жакшы ROC баллы 0,9842 болгон.



**4.9-сүрөт.** Классификация жана регрессия дарагынын моделинин параметрин оптималдаштыруу (Complexity Parameter = Cp)

4.9-сүрөттө параметрди оптималдаштыруу процессинде сыналган баалуулуктар жана эсептелген ROC баллдары орун алган графика көрсөтүлгөн.

**4.14-таблица.** Классификация жана регрессия дарак моделинин ката матрицасы

Болжол	Шилтеме		
	158 (GP)	4 (YP)	162
5 (YN)	150 (GN)	155	
163	154		

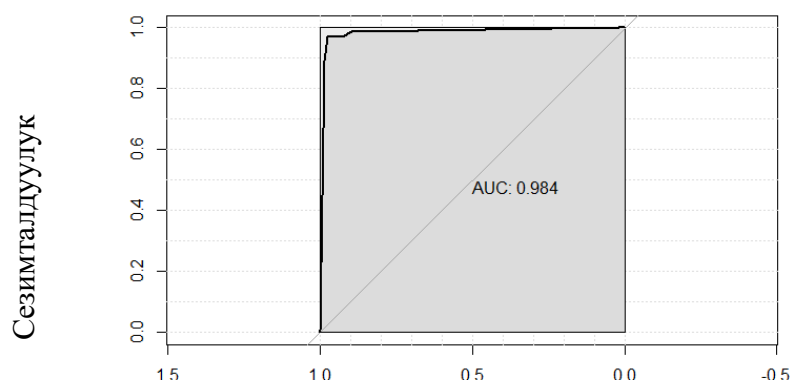
Классификация жана регрессия дарак моделинин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.14-таблицада көрсөтүлгөн. Классификация жана регрессия дарак моделинин тест маалымат топтомундагы 163 тобокелдүү байкоолордун 158и туура классификацияланган жана 5 туура эмес классификацияланган. Модель 154 тобокели жок байкоолордун 150үн туура жана 4 туура эмес классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 162си ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 155и ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

**4.15-таблица.** Классификация жана регрессия дарагынын моделинин моделдик баалоо метрикасы

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9716	0.9693	0.9740	0.0259	0.0306	0.9753

4.15-таблицада классификация жана регрессия дарагынын моделинин моделдик баалоо метрикалары көрсөтүлгөн. Классификация жана регрессия дарагынын моделинин тууралык метрикасы 0.9716 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык классификация жана регрессия дарагынын модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 97.16%’ туура болжолдогон. Моделдин сезимталдуулук метрикасы 0.9693 жана тандалмалык метрикасы 0.9740 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 96.93% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 97.40% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I тип катасы 0.0259. I тип ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы 0.0306. II типтеги ката – үзгүлтүксүздүк тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык көрсөткүчү модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0.9753. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 97.53% үзгүлтүксүздүк тобокелдигине ээ экенин

көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

#### 4.10-сүрөт. Классификация жана регрессия дарагынын моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.10-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 98.4 % ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дон жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

Ишкананын үзгүлтүксүздүгүн баалоо адабиятында ар кандай чечим дарагынын алгоритмдери колдонулган. Zhou, Lai, жана Yen (2014) 6509 байкоону камтыган изилдөөлөрүндө генетикалык алгоритмдер менен өзгөчөлүктү тандоону колдонгон C4.5 чечим дарагынын алгоритм моделинин тууралык метрикасы 0,5067, моделдин сезимталдуулугу 0,9276, ал эми тандалмалыгы 0,0858 деп табылган. С.-F. Tsai, Hsu, жана Yen (2014) Австралиянын, Германиянын жана Жапониянын маалымат топтомдорун колдонуу менен үч түрдүү чечим дарагынын моделдерин окутушкан. Австралиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8491, Германиянын маалымат топтому менен окутулган моделдин тууралык

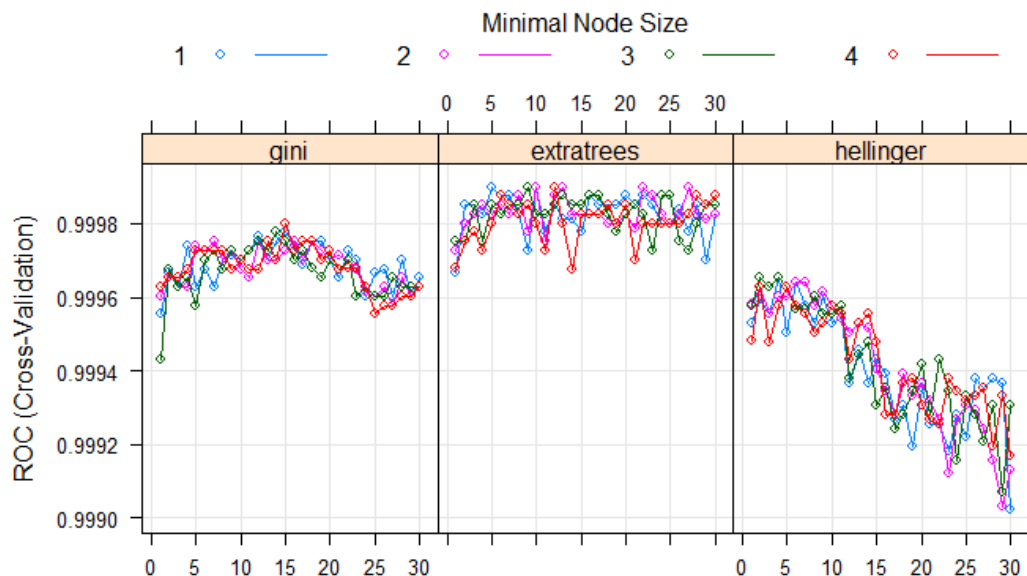
метрикасы 0,7377 жана Жапониянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8637.G болгон. Wang, Ma, жана Yang (2014) изилдөөсүндө чечим дарагынын моделдери Pietruszkiewicz (2008) жана Shmueli, Patel, жана Bruce (2010) тарабынан түзүлгөн эки башка маалымат топтомун колдонуу менен окутулган. Биринчи моделдин тууралык метрикасы 0,7163, I типтеги ката 0,2619 жана II типтеги ката 0,2324, экинчи моделдин тууралык метрикасы 0,7599, I типтеги ката 0,2583 жана II типтеги ката ката 0,2638 болгон. Yeh, Chi, жана Lin (2014) изилдөөлөрүндө 220 байкоону камтыган маалымат топтому менен окутулган C4.5 чечим дарагынын моделинин I типтеги катасы 0,2600 жана II типтеги катасы 0,0190 катары эсептелген. Окутулган моделдин бардык көрсөткүчтөрү изилдөөдө сунушталган модель менен салыштыруу үчүн бөлүшүлгөн эмес. C.-F. Tsai (2014) Австралия, Германия, Жапония, Банкроттук жана UScompetition маалымат топтомдору менен беш түрдүү чечим дарагынын (CART алгоритми) моделдерин окуткан. Австралиялык маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8611, I типтеги ката 0,1813 жана II типтеги ката 0,1028 деп эсептелген. Германиянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,7812, I типтеги ката 0,0952 жана II типтеги ката 0,4847 деп эсептелген. Жапониянын маалымат топтому менен окутулган моделдин тууралык метрикасы 0,9394, I типтеги ката 0,0612 жана II типтеги ката 0,0593 деп эсептелген. Банкроттук маалыматтары менен окутулган моделдин тууралык метрикасы 0,8593, I типтеги ката 0,1215 жана II типтеги ката 0,1641 деп эсептелген. UScompetition маалымат топтому менен окутулган моделдин тууралык метрикасы 0,8307, I типтеги ката 0,2035 жана II типтеги ката 0,1331 деп эсептелген. Zhang, He, жана Zhang (2019) изилдөөсүндө беш түрдүү маалымат топтомун колдонушкан. Австралиянын маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0,8145 жана AUC баллы 0,8124, Германиянын маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0,6633 жана AUC баллы 0,7753. Жапониянын маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0,8116 жана AUC баллы 0,9190. PPDai маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0.7816 жана AUC баллы 0.5793 жана GMSC маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0,8915 жана AUC метрикасы 0,6064 катары

эсептелген. Z. Zhang жана башкалар (2022) 2007-2020-жылдар аралыгындагы 2260 байкоону камтыган маалымат топтому менен окутулган чечим дарагынын моделинин тууралык метрикасы 0,799, сезимталдуулук метрикасы 0,811, тактык метрикасы 0,793 жана AUC баллы 0,879. Адабияттагы ар кандай изилдөөлөрдө окутулган чечим дарагынын моделдерин баалоо чен белгилери жүргүзүлгөн изилдөөдөгү чечим дарагынын моделинин баалоо көрсөткүчтөрү менен салыштырылганда, изилдөөдө окутулган чечим дарагынын моделинин баалоо көрсөткүчтөрү жакшыраак экени көрүнүп турат.

#### **4.10.2.4. Кокус токой модели**

Кокус токой алгоритми өнүккөн машиналык окутуу ыкмасы болуп саналат. Модель кокусунан түзүлгөн чечим дарактарынын коллекциясын колдонуу менен классификацияны түзөт. Кокус токой алгоритми башка алдыңкы ыкмалар менен катар үзгүлтүксүздүктү баалоо адабиятында практикалык мааниге ээ. Кокус токой алгоритми үзгүлтүксүздүктү баалоо изилдөөлөрүндө кеңири колдонулат (Chandra, Ravi, жана Bose 2009; S. Jones, Johnstone, жана Wilson 2017; Qian ж.б. 2022; Yeh, Chi, жана Lin 2014a; Barboza, Kimura, жана Altman 2017; Garcia 2022).

Кокус токой алгоритминин үч параметри бар: `mtry`, `splitrule` жана `min.node.size`. `Mtry` параметри ансамбль түзө турган чечим дарактарын түзүүдө ар бир бөлүмдө канча атрибут эске алынарын аныктайт. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 30га чейинки диапазондо сыналды. `Splitrule` параметри ансамблдин ар бир түйүнүндөгү эң жакшы бөлүүнү тандоо үчүн колдонулат. Параметрди тандоодо `gini`, `extratrees` жана `hellinger` бөлүү эрежелери сыналды. `Min.node.size` параметри чечим дарагынын генерация этабында алгоритмдин жүрүм-турумун көзөмөлдөө үчүн колдонулат. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 4кө чейинки диапазондо сыналды. Параметрди тандоо ROC баллынын өлчөгүн колдонуу менен жүргүзүлдү жана жакшы ROC балладрын берген параметрлер тандалды. Демек `mtry 5`, `splitrule extratrees` жана `min.node.size 1` болгондо, эң жакшы ROC баллы 0,9999 алынды.



**Сүрөт 4.11.** Кокус токой модели параметринин оптимизациясы

4.11-сүрөттө параметрди оптималдаштыруу процессинде сыналган маанилер жана эсептелген ROC баллдары менен графика көрсөтүлгөн.

**4.16-таблица.** Кокус токой моделинин ката матрицасы

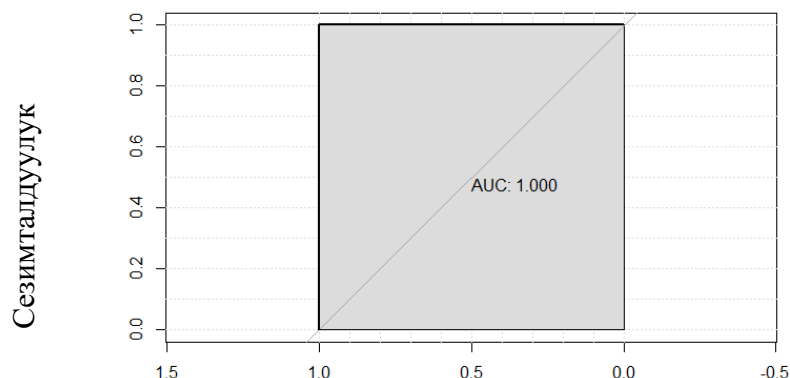
	Шилтеме		
<b>Болжол</b>	163 (GP)	0 (YP)	163
	0 (YN)	154 (GN)	154
	163	154	

Кокус токой моделинин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.16-таблицада көрсөтүлгөн. Кокус токой моделинин тест маалымат топтомундагы 163 тобокелдүү байкоолордун 63ү туура классификацияланган. Модель 154 тобокели жок байкоолордун 154үн туура классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 163ү ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 154ү ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган. Кокус токой модели тесттик маалыматтар топтому менен эн сонун болжолдоо ийгилигин көрсөттү.

**4.17-таблица.** Кокус токой моделинин моделдик баалоо көрсөткүчтөрү

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
<b>Нарк</b>	1	1	1	0	0	1

4.17-таблицада кокус токой моделинин моделдик баалоо метрикалары көрсөтүлгөн. Кокус токой моделинин тууралык метрикасы 1 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык кокус токой моделинин тесттик маалыматтар топтому менен жасалган болжолдоолордун 100% туура болжолдогон. Моделдин сезимталдуулук метрикасы 1 жана тандалмалык метрикасы 1 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 100% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 100% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Кокус токой модели I тип жана II типтеги катага ээ эмес. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык метрикасы модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 1. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 100% үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

#### 4.12-сүрөт. Кокус токой моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.12-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 100 % ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дон жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

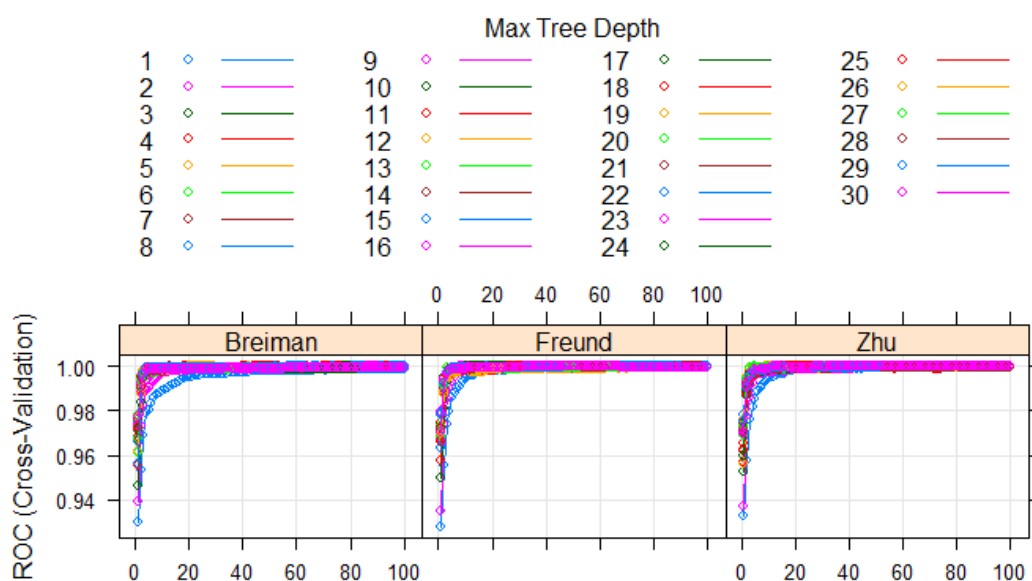
Bootstrap aggregating ыкмасы болгон кокус токой алгоритми адабиятта ар кандай изилдөөчүлөр тарабынан колдонулган. Volkov, Benoit, жана Van den Poel (2017) 2011, 2012, 2013, 2014 жана 2015-жылдар үчүн беш түрдүү кокус токой моделдерин 2007-жылдан 2015-жылга чейин Belfirst маалымат базасынан алынган чоң маалымат топтому менен окутушкан. 2011-жылга таандык эң ийгиликтүү моделдин тууралык метрикасы 0,7887 жана моделдин AUC баллы 0,8568 болгон. 2012-жылга таандык моделдин тууралык метрикасы 0,7789 жана моделдин AUC баллы 0,8458 болгон. 2013-жылдагы моделдин тууралык метрикасы 0,7874 жана моделдин AUC баллы 0,8521 болгон. 2014-жылга таандык моделдин тууралык метрикасы 0,7835 жана моделдин AUC баллы 0,8469 болгон. 2015-жылдагы моделдин тууралык метрикасы 0,7944, ал эми моделдин AUC баллы 0,8590. Vezanzones жана Séverin (2018) изилдөөсүндө 1500 байкоолордон турган, финансылык катыштардан

түзүлгөн маалымат топтому менен окутулган кокус токой моделинин сезимталдуулук метрикасы 0,571, тандалмалык метрикасы 0,918 жана AUC баллы 0,807 болгон. Jabeur ж.б. (2021) 133 байкоону камтыган жана 2017(Y), 2016(Y-1), 2015(Y-2) жана 2014(Y-3) маалыматтарын колдонгон изилдөөдө окутулган кокус токой моделдери боюнча 2016 жыл үчүн 0.9559 тууралык жана 0.939 AUC баллы, 2015 жыл үчүн 0.7308 жана 0.813 жана 2014 жыл үчүн 0.7436 жана 0.739 AUC баллына ээ. Qian жана башкалардын (2022) иштеринде 2001-2019 жылдар аралыгын жана 4167 байкоону камтыган маалымат топтому менен байкоонун тобокелин кабыл алган жылдан бир жыл мурунку берилиштер эске алынып окутулган кокус токой моделинин тууралык метрикасы 0,8499, сезимталдуулугу 0,8345, тактыгы 0,8608 жана AUC баллы 0,8248 деп эсептелген. Garcia (2022) тарабынан 2010-жылдын январынан 2018-жылдын декабрына чейинки мезгил үчүн 1824 ишканын 41933 кварталдык финансылык отчетун камтыган маалымат топтому боюнча окутулган кокус токой жана кокус токой топтору моделдеринин сезимталдуулук метрикасы ырааттуулугуна жараша 0.714 жана 1, тандалмалык метрикасы ырааттуулугуна жараша 0,978 жана 1, AUC баллы ырааттуулугуна жараша 0,846 жана 1. Z. Zhang жана башкалар (2022) 2007-2020-жылдар аралыгындагы 2260 байкоону камтыган маалымат топтому менен окуткан кокус токой моделинин тууралык метрикасы 0,832, сезимталдуулук метрикасы 0,811, тактык метрикасы 0,846 жана AUC баллы 0,893. Натыйжалары бөлүшүлгөн изилдөөлөрдүн арасында Гарсиа (2022) тарабынан иштелип чыккан кокус токой тобунун модели баалоо метрикалары боюнча башка маалыматтар топтому менен жүргүзүлгөн изилдөөдөгү кокус токой моделинин натыйжаларына так дал келет. Изилдөөдөгү моделдин натыйжалары натыйжалары бөлүшүлгөн башка изилдөөлөр менен салыштырылганда, иштелип чыккан моделдин болжолдоо ийгилиги жакшыраак экени көрүнүп турат.

#### **4.10.2.5. Adaboost модели**

Adaboost - Freund жана Schapire (1995) тарабынан түзүлгөн ансамблдик окутуу ыкмасы. Adaboost күчтүү классификаторду алуу үчүн базалык классификаторлордун ансамблин түзүүгө багытталган. Ыкма адабиятта кеңири колдонулат (Alfaro ж.б. 2008; Neo and Yang 2014; S. Y. Kim жана Urpeja 2014; J. Sun ж.б. 2017). Adaboost алгоритминин үч параметри бар:  $m_{final}$ ,  $maxdepth$  жана

coeflearn. Mfinal параметри моделдөөдө колдонула турган базалык классификаторлордун максималдуу санын жөнгө салат. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 100гө чейинки диапазондо сыналды. Maxdepth параметри моделдин терендигин белгилейт. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 30га чейинки диапазондо сыналды. Coeflearn параметри колдонула турган коэффициентти аныктайт. Параметр тандоодо Zhu, Breiman жана Freund методдору сыналган. Параметрди тандоо ROC баллынын өлчөгүн колдонуу менен жүргүзүлдү жана жакшы ROC баллын берген параметрлер тандалды. Демек mfinal 20, maxdepth 19 жана coeflearn Zhu учурунда 0,9999 мыкты ROC баллы алынды.



#### 4.13-сүрөт. Adaboost моделинин параметрин оптималдаштыруу

4.13-сүрөттө параметрди оптималдаштыруу процессинде сыналган маанилер жана эсептелген ROC баллдары менен графика көрсөтүлгөн.

#### 4.18-таблица. Adaboost моделинин ката матрицасы

		Шилтеме	
<b>Болжол</b>	163 (GP)	2 (YP)	165
	0 (YN)	152 (GN)	152
	163	154	

Adaboost моделинин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.18-таблицада көрсөтүлгөн. Adaboost моделинин тест маалымат топтомундагы 163 тобокелдүү байкоолордун 163ү классификацияланган. Модель 154 тобокели жок

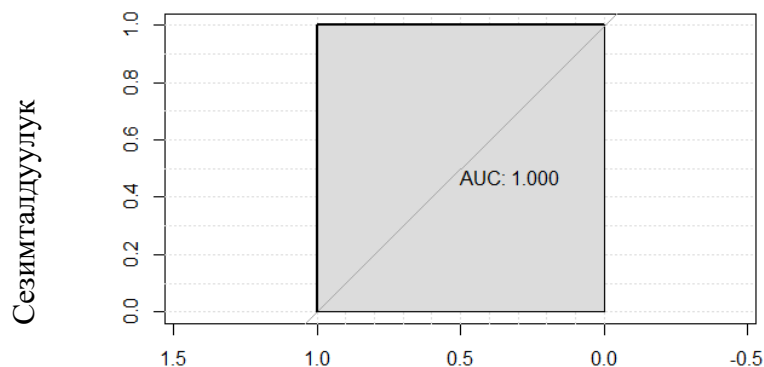
байкоолордун 152 туура классификациялаган жана 2 туура эмес классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 165и ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 152си ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

**4.19-таблица.** Adaboost моделинин моделдик баалоо метрикалары

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9936	1	0.9870	0.0129	0	0.9878

4.19-таблицада Adaboost моделинин моделдик баалоо метрикалары көрсөтүлгөн. Adaboost моделинин тууралык метрикасы 0.9936 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык Adaboost модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 99.36%’ туура болжолдогон. Моделдин сезимталдуулук метрикасы 1 жана тандалмалык метрикасы 0.9870 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 100% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 98.70% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I типтеги катасы 0.0129. I тип ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы эсептелбейт. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык метрикасы модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0.9878. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 98.78% үзгүлтүксүздүк тобокелдигине ээ экенин

көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

#### 4.14-сүрөт. Adaboost моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.14-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 100 % ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дөн жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

Boosting ыкмасы болгон adaboost алгоритми адабиятта даракка негизделген алгоритмдер жана ар кандай негизги классификаторлор менен үзгүлтүксүздүк тобокелдигин баалоо үчүн колдонулган. S. Y. Kim жана Urpeja (2014) тарабынан 1988-2011-жылдар аралыгында 826 байкоо менен жүргүзүлгөн изилдөөдө окутулган adaboost моделинин тууралык метрикасы 0,9769 жана AUC баллы 0,9880 болгон. Wang жана башкалар (2020) изилдөөсүндө 39 финансылык коэффициентти колдонушкан. Изилдөөдө атрибут катары финансылык катыштарды гана колдонуу менен окутулган adaboost моделинин AUC баллы 0,8862 деп эсептелген. X. Du жана башкалар (2020) 2017 жана 2018-жылдарды камтыган жана 7154 байкоо менен маалымат топтомун колдонуу аркылуу adaboost

моделинин сезимталдуулук метрикасы 0,9838 жана AUC баллы 0,9458 деп эсептешкен. Qian ж.б. (2022) изилдөөлөрүндө 2001-2019-жылдарды жана 4167 байкоону камтыган маалыматтар топтомун жана байкоонун тобокелдиги кабыл алынган жылдан бир жыл мурунку маалыматтарды эске алуу менен окутулган adaboost моделинин тууралык метрикасы 0,8410, сезимталдуулугу 0,8214, тактыгы 0,8548 жана AUC баллы 0,9182 болуп эсептелген. Адабияттагы ар кандай изилдөөлөрдө окутулган adaboost моделдерин баалоо чен белгилери менен жүргүзүлгөн изилдөөдөгү моделдин баалоо метрикалары менен салыштырылганда, изилдөөдө окутулган моделдин баалоо метрикалары жакшыраак экени көрүнүп турат.

#### **4.10.2.6. Gradient Boosting модели**

Gradient boosting машиналык окутууда эң жакшы эффективдүү ыкмалардын бири болуп эсептелет. Ыкма адабиятта ар кандай изилдөөлөрдө колдонулган (Son ж.б. 2019; Wyrobek жана Kluza 2019; Zięba, Tomczak, жана Tomczak 2016). Анын төрт параметри бар: Gradient Boostingn.trees, interaction.depth, shrinkage venminobsinnode. N.trees параметри моделдеги дарактардын санын жөнгө салат. Параметр 100дөн 2500гө чейинки диапазондо, аны ар бир сыноодо 100гө көбөйтүү жолу менен текшерилген. Interaction.depth параметри базалык классификатордун чечим дарактарынын тереңдигин белгилейт. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 10го чейинки диапазондо сыналды. Shrinkage ансамблдагы ар бир дарактын үлгүгө кошкон салымын жөнгө салат. Параметр 0,0005тен 0,002ге чейинки диапазондо сыналды, ар бир тестте 0,0001ге көбөйдү. N.minobsinnode дарак куруу процессинде жаңы түйүн түзүү үчүн талап кылынган байкоолордун минималдуу санын текшерет. Параметр ар бир тестте 1ди көбөйтүү менен 1ден 5ке чейинки диапазондо сыналды. Параметрди тандоо ROC баллынын өлчөгүн колдонуу менен жүргүзүлдү жана жакшы ROC баллын берген параметрлер тандалды. Демек мыкты ROC упай n.trees 1500, interaction.depth 10, shrinkage 0.0019 жана n.minobsinnode 3 учурунда алынган.

**4.20-таблица. Gradient Boosting моделинин ката матрицасы**

	Шилтеме		
Болжол	161 (GP)	3 (YP)	164
	2 (YN)	151 (GN)	154
	163	154	

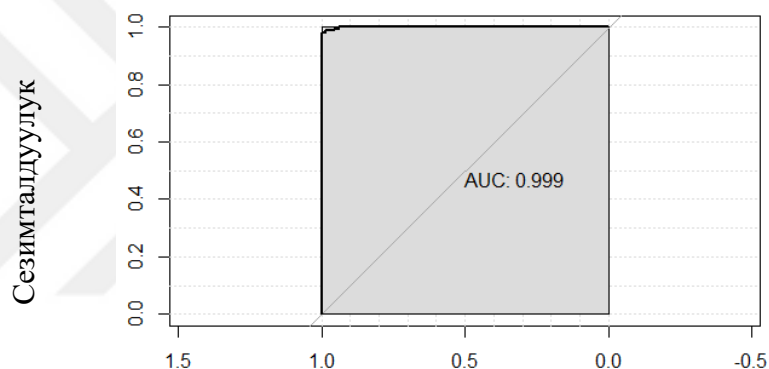
Gradient boosting моделинин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.20-таблицада көрсөтүлгөн. Gradient boosting моделинин тест маалымат топтомундагы 163 тобокелдүү байкоолордун 161 туура жана 2 туура эмес классификацияланган. Модель 154 тобокели жок байкоолордун 151 туура классификациялаган жана 3 туура эмес классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 164ү ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 154ү ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

**4.21-таблица. Gradient Boosting моделинин моделдик баалоо метрикалары**

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9842	0.9877	0.9806	0.0193	0.0122	0.9817

4.21-таблицада gradient boosting моделинин моделдик баалоо метрикалары көрсөтүлгөн. Gradient boosting моделинин тууралык метрикасы 0.9842 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык gradient boosting модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 98.42% туура болжолдогон. Моделдин сезимталдуулук метрикасы 0.9877 жана тандалмалык метрикасы 0.9806 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин үзгүлтүксүздүк тобокелдиги менен байкоолордун 98.77% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 98.06% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I типтеги катасы 0.0193. I тип ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары

классификациясы. Моделдин II типтеги катасы 0.0122. II типтеги ката – үзгүлтүксүздүк тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык метрикасы модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0.9817. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 98.17% үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



1-Тандалмалык

#### 4.15-сүрөт. Gradient Boosting моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.15-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 99.9 % ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дон жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

Gradient boosting алгоритми акыркы жылдары адабиятта кеңири колдонулуп келе жаткан ансамблдик окутуу ыкмасы. X. Du жана башкалар (2020) тарабынан 2017 жана 2018-жылдарды жана 7154 байкоону камтыган маалыматтар топтомун колдонуу менен окутулган gradient boosting моделдин сезимталдуулук метрикасы 0,9954 жана AUC баллы 0,9644 болгон. Jabeur ж.б. (2021) 133 байкоону камтыган жана 2017(Y), 2016(Y-1), 2015(Y-2) жана 2014(Y-3) жылдардагы маалыматтарды колдонгон изилдөөдө окутулган gradient boosting модели менен 2016 жыл үчүн 0,9231 тууралык жана 0.951 AUC баллы, 2015-жыл үчүн 0,7308 жана 0,733 AUC баллы, 2014-жыл үчүн 0,7607 жана 0,718 AUC баллы эсептелген. Qian ж.б. (2022) изилдөөлөрүндө 2001-2019 жылдар аралыгын жана 4167 байкоону камтыган маалыматтар топтому менен байкоонун тобокели кабыл алынган жылдан бир жыл мурунку берилиштерди эске алуу менен окутулган gradient boosting моделинин тууралык метрикасы 0,8509, сезимталдуулугу 0,8408, тактык 0,8580 жана AUC баллы 0,9274 деп эсептелген. Адабияттагы ар кандай изилдөөлөрдө окутулган gradient boosting моделдерин баалоо чен белгилери менен жүргүзүлгөн изилдөөдөгү моделдин баалоо метрикалары менен салыштырылганда, изилдөөдө окутулган моделдин баалоо метрикалары жакшыраак экени көрүнүп турат.

#### **4.10.2.7. XGBoost модели**

Extreme gradient boosting - бул T. Chen жана Guestrin (2016) тарабынан түзүлгөн gradient boosting ыкмасынын өркүндөтүлгөн версиясы. XGBoost жогорку болжолдоонун ийгилик потенциалынан улам кеңири колдонулган ыкма (Volkov, Benoit, жана Van den Poel 2017; Y.-P. Huang жана Yen 2019; Jabeur ж.б. 2021; Qian ж.б. 2022). Extreme gradient boosting алгоритминин жети параметри бар: nrounds, nrounds, max\_depth, eta, gamma, colsample\_bytree, min\_child\_weight жана subsample. Nrounds параметри - boosting итерациялардын саны, max\_depth параметри - базалык классификаторлор ээ боло турган максималдуу тереңдик, eta параметри - ар бир итерацияда модел өткөрө турган маалыматтардын көлөмү, gamma параметри - даракты түзүү процессинде жалбырак түйүнүндө бөлүү үчүн талап кылынган минималдуу жоготуу, colsample\_bytree параметри boosting учурунда ар бир даракты түзүп жатканда туш келди тандалып алынуучу өзгөчөлүктөрдүн үлүшү, min\_child\_weight - дарактын жалбырактарындагы байкоолордун минималдуу саны жана subsample параметри туш келди тандалып

алынган жана дарактарды түзүү учурунда камтыла турган байкоолорду белгилейт. Окутуу этабында параметрди издөө " caret" жыйнагынын (китепканасынын) автоматтык параметр издөө функциясынын жардамы менен жүргүзүлгөн. Параметрди тандоо ROC баллынын өлчөгүн колдонуу менен жүргүзүлдү жана жакшы ROC баллвн берген параметрлер тандалды. Демек nrounds 478, max\_depth 6, eta 0.2970209, gamma 0.5715507, colsample\_bytree 0.6139304, min\_child\_weight 5 жана subsample 0.7726817 болгондо мыкты ROC баллы 0,9988 алынган.

#### 4.22-таблица. XGBoost моделинин ката матрицасы

	Шилтеме		
Болжол	161 (GP)	3 (YP)	164
	2 (YN)	151 (GN)	154
	163	154	

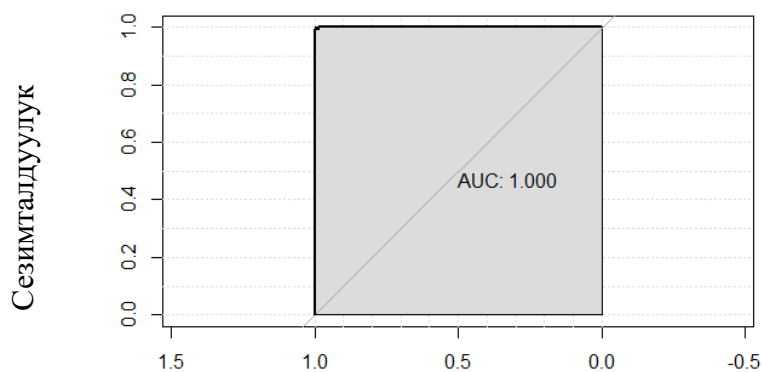
XGBoost моделинин окутулгандан кийин тесттик маалыматтар топтому менен жасалган божомолдун натыйжаларын көрсөткөн ката матрицасы 4.22-таблицада көрсөтүлгөн. XGBoost моделинин тест маалымат топтомундагы 163 тобокелдүү байкоолордун 161 туура жана 2 туура эмес классификацияланган. Модель 154 тобокели жок байкоолордун 151 туура классификациялаган жана 3 туура эмес классификациялаган. Тесттик маалыматтар топтомундагы 317 байкоонун 164ү ишкананын үзгүлтүксүз иштеши үчүн тобокелдүү жана 154ү ишкананын үзгүлтүксүз иштеши үчүн тобокелчиликсиз деп классификацияланган.

#### 4.23-таблица. XGBoost моделинин моделдик баалоо метрикалары

Метрика	Тууралык	Сезимталдуулук	Тандалмалык	Тип I ката	Тип II ката	Тактык
Нарк	0.9842	0.9877	0.9806	0.0193	0.0122	0.9817

4.23-таблицада XGBoost моделинин моделдик баалоо метрикалары көрсөтүлгөн. XGBoost моделинин тууралык метрикасы 0.9842 деп эсептелген. Тууралык метрикасы моделдин жалпы болжолдоо ийгилигин көрсөтөт. Эсептелген тууралык метрикасына ылайык XGBoost модели тесттик маалыматтар топтому менен жасалган болжолдоолордун 98.42% туура болжолдогон. Моделдин сезимталдуулук метрикасы 0.9877 жана тандалмалык метрикасы 0.9806 деп эсептелген. Сезимталдуулук метрикасы үзгүлтүксүздүк тобокелдигин алып жүргөн байкоолордун канчасы туура бааланганын көрсөтөт. Тандалмалык метрика үзгүлтүксүздүк тобокелдигин алып келбеген байкоолордун канчасы туура классификацияланганын көрсөтөт. Табылган сезимталдуулук метрикасы моделдин

үзгүлтүксүздүк тобокелдиги менен байкоолордун 98.77% туура болжолдогондугун, ал эми аныкталган тандалмалык метрикасы моделдин үзгүлтүксүздүк тобокелдиги жок байкоолордун 98.06% туура болжолдоосун көрсөтөт. Сезимталдуулук жана тандалмалык метрикалардын натыйжалары моделдин үзгүлтүксүздүк тобокелдиги менен байкоолорду да, үзгүлтүксүздүк тобокелдиги жок байкоолорду да туура классификациялоо үчүн жогорку жакындыкка ээ экенин көрсөтүп турат. Моделдин I типтеги катасы 0.0193. I тип ката - үзгүлтүксүздүк тобокелдигин алып келбеген байкоонун тобокелдүү катары классификациясы. Моделдин II типтеги катасы 0.0122. II типтеги ката – үзгүлтүксүздүк тобокелдигин алып жүргөн байкоонун тобокелдиксиз классификациясы. Ката матрицасы аркылуу эсептелген акыркы метрика тактык метрика болуп саналат. Тактык метрикасы модель шилтеме классын канчалык так болжолдой алат көрсөтөт. Башка сөз менен айтканда, тактык - бул үзгүлтүксүздүк тобокелдигин алып келе турган бааланган байкоолордун канчасы иш жүзүндө үзгүлтүксүздүк тобокелдигинин өлчөмү. Моделдин тактык метрикасы 0.9817. Моделдин тактык метрикасы үзгүлтүксүздүк тобокелдиги менен байкоолор катары болжолдогон үлгүлөрдүн 98.17% үзгүлтүксүздүк тобокелдигине ээ экенин көрсөтүп турат. Модель байкоону үзгүлтүксүздүк тобокелдиги катары классификациялоодо жогорку болжолдоо жөндөмүнө ээ.



#### 4.16-сүрөт. XGBoost моделинин ROC ийри сызыгы жана AUC баллы

Моделдин ROC ийри сызыгы жана AUC баллы 4.16-сүрөттө көрсөтүлгөн. ROC ийри сызыгы ыктымалдык ийри сызыгы, ал эми AUC баллы үзгүлтүксүздүк тобокели бар жана үзгүлтүксүздүк тобокели жок байкоолордун бөлүнүүчү өлчөмү

болуп саналат. Башка сөз менен айтканда AUC моделдин эки классты канчалык жакшы ажырата аларын көрсөтөт. Эсептелген AUC баллы моделдин үзгүлтүксүздүк тобокели бар байкоолор менен үзгүлтүксүздүк тобокелдиги жок байкоолорду 100% ыктымалдык менен туура айырмалоо мүмкүнчүлүгү бар экенин көрсөтүп турат. Адабиятты изилдегенде ROC баллы 0,50дөн жогору болушу керек жана моделдин ROC упайы 0,90дон жогору болушу керек деп кабыл алынат, бул модель кемчиликсиз иштегендигинин көрсөткүчү (Fawcett 2006).

XGBoost - акыркы жылдары адабиятта кеңири колдонулуп келе жаткан ансамблдик окутуу ыкмасы. Volkov, Benoit, жана Van den Poel (2017) 2011, 2012, 2013, 2014 жана 2015-жылдар үчүн беш түрдүү XGBoost моделин 2007-жылдан 2015-жылга чейин Belfirst маалымат базасынан алынган чоң маалымат топтому менен окутушкан. 2011-жылга таандык эң ийгиликтүү моделдин тууралык метрикасы 0.7869 жана моделдин AUC баллы 0.8506 болгон. 2012-жылга таандык моделдин тууралык метрикасы 0.7740 жана моделдин AUC баллы 0.8360 болгон. 2013-жылдагы моделдин тууралык метрикасы 0.7773 жана моделдин AUC баллы 0.8415 болгон. 2014-жылга таандык моделдин тууралык метрикасы 0.7729 жана моделдин AUC баллы 0.8337 болгон. 2015-жылдагы моделдин тууралык метрикасы 0.7824, ал эми моделдин AUC баллы 0.8461 деп эсептелген. Y.-P. Huang жана Yen (2019) тарабынан 2010-2016-жылдарды жана 64 байкоону камтыган маалыматтар топтому менен жүргүзүлгөн изилдөөдө окутулган XGBoost моделинин тууралык метрикасы 0,906, сезимталдуулук метрикасы 0,875, тандалмалык метрика 0,938, I типтеги ката 0.0310 жана II типтеги ката 0.0630. X. Du жана башкалар (2020) тарабынан 2017 жана 2018-жылдарды жана 7154 байкоону камтыган маалымат топтомун колдонуу менен Окутулган XGBoost моделинин сезимталдуулук метрикасы 0,9444 жана AUC баллы 0,9321 болгон. Jabeur жана башкалар (2021) тарабынан 133 байкоо орун алган жана 2017(Y), 2016(Y-1), 2015(Y-2) жана 2014(Y-3) жылдарындагы маалыматтар колдонулган изилдөөдө окутулган XGBoost модели менен 2016 -жыл үчүн тууралык метрикасы 0,9359 жана AUC баллы 0.931, 2015-жылга 0,8205 жана 0,807 жана 2014-жылга 0,8162 жана 0,715 катары эсептелген. Qian ж.б. (2022) изилдөөлөрүндө 2001-2019 жылдар аралыгын жана 4167 байкоону камтыган маалыматтар топтому менен байкоонун тобокели кабыл алынган жылдан бир жыл мурунку берилиштерди эске алуу менен окутулган XGBoost моделинин тууралык метрикасы 0.8542,

сезимталдуулугу 0.8397, тактык 0.8647 жана AUC баллы 0.9291 деп эсептелген. Garcia (2022) тарабынан 2010-жылдын январынан 2018-жылдын декабрына чейинки мезгил үчүн 1824 ишкананын 41933 кварталдык финансылык отчетун камтыган маалымат топтому боюнча окутулган XGBoost моделдеринин сезимталдуулук метрикасы 0.571, тандалмалык 0.989 жана AUC баллы 0.780 болгон. Z. Zhang жана башкалар (2022) 2007-2020-жылдар аралыгындагы 2260 байкоону камтыган маалымат топтому менен окуткан XGBoost моделинин тууралык метрикасы 0.835, сезимталдуулук метрикасы 0.808, тактык метрикасы 0.853 жана AUC баллы 0.913. Адабияттагы ар кандай изилдөөлөрдө окутулган XGBoost моделдерин баалоо чен белгилери менен жүргүзүлгөн изилдөөдөгү моделдин баалоо метрикалары менен салыштырылганда, изилдөөдө окутулган моделдин баалоо метрикалары жакшыраак экени көрүнүп турат.

## **5. Корутнуду жана сунуштар**

Аудит – бул экономикалык аракеттер же окуялар жөнүндө дооматтардын жана алдын ала аныкталган чен белгилердин шайкештик даражасын аныктоо максатында объективдүү түрдө бул дооматтар үчүн далилдерди өз алдынча алуу жана баалоо жана натыйжаларды кызыкдар адамдарга жеткирүүнүн системалуу процесси. Бухгалтердик эсептин негизги түшүнүктөрүнүн бири болгон бизнестин үзгүлтүксүздүк принциби бизнестин чексиз өмүрү бар жана жакынкы келечекте өз ишин улантат деген божомолду билдирет. Аудиттин эл аралык стандарттары бизнестин үзгүлтүксүздүгүн баалоо үчүн бизнестин башкарууну жана аудиторду жоопкерчиликке тартат. Жасалма интеллект технологиялары бизнестин бардык тармагына, ошондой эле көзөмөлгө таасирин тийгизди. Бул контекстте жасалма интеллект технологиялары аудит процессиндеги иш-аракеттердин контекстинде бааланып, эффективдүүлүктү жогорулатуу максатында ар кандай моделдер иштелип чыккан. Логистикалык регрессия, дискриминанттык анализ, байездик классификаторлор, чечим дарактары, таяныч векторлорунун машиналары, нейрон тармактары, кокус токой алгоритми, adaboost, gradient boosting жана extreme gradient boosting сыяктуу көзөмөлгө алынган окутуу ыкмаларынан тышкары, self-organizing maps жана k-каражаттар сыяктуу көзөмөлсүз машиналык окутуу ыкмалары да аудитте жасалма интеллект жана машиналык окутуу колдонмолорунда кеңири колдонулат.

Салттуу статистикалык методдор бизнестин үзгүлтүксүздүгүн баалоо максатында жүргүзүлгөн биринчи изилдөөлөрдө колдонулган (Chye Koh жана Moren Brown 1991; Doruch, Holthausen, жана Leftwich 1987). Beaver (1966) тарабынан жүргүзүлгөн бир өзгөрмөлүү статистикалык ыкмалар менен жана E. Altman (1968) тарабынан пайдаланылган көп дискриминанттык анализ менен изилдөөлөр бул чөйрөдө пионер болуп саналат. Beaver (1966) жана E. Altman (1968) тарабынан жүргүзүлгөн изилдөөлөрдөн кийинки жылдарда дискриминанттык анализ, логистикалык регрессия жана пробит анализи сыяктуу салттуу статистикалык ыкмалардан пайда алган көптөгөн изилдөөчүлөр адабиятка өз салымын кошушкан (Barnes 2004; Kida 1980; E. Altman 1968; Ohlson 1980; Zmijewski 1984). Бирок алар жөнөкөй түзүмгө, азыраак параметрлерге жана азыраак окутуу убактысын талап кылса да, салттуу методдор нормалдуулук, көз карандысыз өзгөрмөлөрдүн ортосундагы бирдей ковариация жана көз карандысыз өзгөрмөлөр ортосунда көп корреляция жок сыяктуу катаал жана чектөөчү божомолдор менен чектелет. Мындан тышкары каржылык маалыматтар менен иштөө көбүнчө божомолдордун бузулушуна алып келет (J. Sun ж.б. 2020; Yeh, Chi, жана Lin 2014a). Салттуу статистикалык ыкмалардын катаал божомолдоруна байланыштуу үзгүлтүксүздүктү баалоо адабияттарында машинанылык окутуу ыкмалары колдонула баштады (Z. Zhang ж.б. 2022). Чечим дарактары (Chye Koh жана Kee Low 2004a; Gepp, Kumar, жана Bhattacharya 2010b; M.-Y. Chen 2011a; Ruby жана Diamond 2013) жана таяныч векторлорунун машиналары (Hua ж.б. 2007; Ding, Song, жана Zen 2008; Yeh, Chi, жана Hsu 2010; S. Y. Kim 2011b; M.-Y. Chen 2012; Bae 2012; Hsieh, Hsiao, жана Yeh 2012; S.-C. Huang ж.б. 2012; Gordini 2014; Goo, Chi, жана Shen 2016; Fallahpour, Lakvan, жана Zadeh 2017b) жана ансамблди окутуу ыкмалары менен кокус токой алгоритми (Barboza, Kimura, жана Altman 2017; Terzi 2022; Garcia 2022), adaboost (Alfaro ж.б. 2008; Heo жана Yang 2014), gradient boosting (S. Jones 2017; E. I. Altman ж.б. 2020; X. Du ж.б. 2020; Jabeur ж.б. 2021) жана XGBoost (Zięba, Tomczak, жана Tomczak 2016; Y.-P. Huang жана Yen 2019) 1990-жылдардан баштап ар кайсы окумуштуулар тарабынан жүргүзүлгөн изилдөөлөрдө бизнестин үзгүлтүксүздүк тобокелдигин баалоого багытталган моделдерге киргизилген.

Бизнестин үзгүлтүксүздүгүн болжолдоого багытталган моделдердин негизги максаты бизнестин үзгүлтүксүздүгүн текшерүү үчүн бардык бизнестин кызыкдар

тараптары, өзгөчө бизнес жетекчиси жана аудитор тарабынан колдонула турган моделдерди иштеп чыгуу болуп саналат. Бул максатта адабиятта машиналык окутуу ыкмалары кеңири колдонулган. Ар кайсы өлкөлөрдүн маалымат топтомдору менен жүргүзүлгөн изилдөөлөрдө үзгүлтүксүздүктү баалоодо машиналык окутуу ыкмаларынын ырааттуулугу көрсөтүлдү. Бул изилдөөдө “Борса Истанбул А.Ш.” АКнын 2017, 2018, 2019, 2020 жана 2021-жылдар үчүн акция рыногунда, өндүрүштө жана субсекторлордо жалпы 201 бизнестин финансылык отчеттору каралып чыкты. Ишканаларга байкоо жүргүзгөн жылдарды белгилөө үчүн ишкананын үзгүлтүксүздүк корутундусун алган-албаганы эске алынган. Эгерде уюм тиешелүү жылы үзгүлтүксүз иш жүргүзүү жөнүндө корутунду алган болсо, байкоо жылы "1" деп белгиленет. Үзгүлтүксүз көрүнүшкө ээ болбогон байкоо жылдары "0" деп белгиленген. Изилдөөдө колдонулган маалымат топтомунда 16 ишканага тиешелүү 43 байкоо жылы жана 198 ишканага тиешелүү 788 байкоо жылы бар. Изилдөөдө үзгүлтүксүздүктү баалоодо кеңири колдонулган 75 каржылык катышы кеңири адабиятка сереп салуу менен тандалып алынган жана биринчи маалымат топтому түзүлгөн.

Маалыматтар топтому түзүлгөндөн кийин жетишпеген байкоолор өзгөчөлүктөрдүн орточо маанисин колдонуу менен аяктады жана маалыматтар топтому min-max нормалдаштыруу ыкмасы менен нормалдаштырылган. Изилдөөдө колдонулган маалымат топтому тең салмактуу эмес түзүлүшкө ээ. SMOTE - ашыкча тандап алуу ыкмасы тең салмактуу эмес маалыматтар топтомунун моделдин болжолдоо ийгилигине терс таасирин алдын алуу үчүн колдонулган. Маалыматтар топтому тең салмактагандан кийин «1» классында 814, «0» классында 774 байкоо болгон. «1» класстагы байкоолор үзгүлтүксүз тобокелин алып келет. «0» классындагы байкоолор үзгүлтүксүздүк тобокелин алып келбейт. Окутулган моделдердин болжолдоо ийгилигин жогорулатуу үчүн генетикалык алгоритмдер менен өзгөчөлүктөр тандалып алынган. Функцияны тандоодон кийин биринчи маалымат топтомундагы 75 өзгөчөлүктүн 58и тандалды. Генетикалык алгоритмдер менен жүргүзүлгөн тандоо фазасында ар бир итерациядагы ар кандай өзгөчөлүк айкалыштарынын болжолдоо ийгилиги кокус токой алгоритми менен өлчөнгөн.

Изилдөөдө жалпысынан тогуз түрдүү моделдер логистикалык регрессия, таяныч векторорунун машиналары (сызыктуу, полиномиялык, радиалдык негизде ядро

функциялары), классификация жана регрессия дарагы, кокус токой алгоритми, adaboost, gradient boosting жана XGBoost ыкмалары менен окутулду. Окутууда шилтеме (позитивдүү класс) класс “1” деп аныкталып, моделди баалоо этабында ушул чен белгиси эске алуу менен ката матрицасы түзүлүп, баалоо метрикалары эсептелген. Окутуунун алдында маалымат топтому 80-20 эрежеси боюнча туш келди бөлүнгөн. Маалыматтар топтомунан бөлүнгөн 80% маалымат окутуу максатында жана 20% тестирилөө үчүн колдонулган. Моделдерди окутууда 10 катмарлуу кайчылаш текшерүү процедурасы колдонулган.

Окутулган моделдер эсептелген тууралык метрикасы боюнча салыштырылганда эң ийгиликтүү натыйжа тесттик маалыматтар топтомундагы бардык байкоолорду туура классификациялаган кокус токой модели (100%) болду. Кокус токой моделинен кийин ырааттуулук менен радиалдык негизделген таяныч векторлорунун машиналары (99,36%) жана adaboost (99,36%), полиномдук таяныч векторлорунун машиналары (98,73%), gradient boosting (98,42%) жана XGBoost (98,42%), сызыктуу таяныч векторлорунун машиналары (97,48%), классификация жана регрессия дарагы (97,16%) жана логистикалык регрессия (92,43%) орун алды. Ансамблдик окутуу ыкмалары менен базалык классификаторлорду салыштырганда, моделдердин тууралык метрикасын эске алганда, ансамблдик моделдер менен таяныч векторлорунун машинасы моделдеринин ортосунда чоң айырма жок экени көрүнүп турат. Логистикалык регрессия модели тууралык метрикасы жагынан эң ийгиликсиз болуп саналат.

Моделдер эсептелген сезимталдык жана тандалмалык көрсөткүчтөр боюнча салыштырылганда эң ийгиликтүү модель кокус токой модели болуп саналат. Модель үзгүлтүксүздүк тобокели бар байкоолорду да, үзгүлтүксүздүк тобокели жок байкоолорду да классификациялоо үчүн эң сонун жөндөмгө ээ. Сезимталдуулук метрикасы жагынан кокус токой модели (1.0000) жана сызыктуу таяныч векторлорунун машинасы (1.0000), adaboost (1.0000) эң сонун алдын ала жакындыкка ээ. Моделдерден кийин полиномдук таяныч векторлорунун машинасы (0,9938), радиалдык негиздеги таяныч векторлорунун машинасы (0,9877), gradient boosting (0,9877) жана XGBoost (0,9877), классификация жана регрессия дарагы (0,9693) жана логистикалык регрессия (0,9018) турат. Моделдерди сезимталдуулук метрикасы боюнча салыштырганда ансамблдик моделдер менен таяныч векторлорунун машинасы моделдеринин ортосунда чоң айырма жок экени

көрүнүп турат. Логистикалык регрессиялык модель байкоолорду улантуу тобокелдиги менен классификациялоо үчүн эң төмөнкү жакындыкка ээ. Кокус токой модели (1000) жана радиалдык негиздеги таяныч векторлорунун машинасы модели (1000) үзгүлтүксүздүк тобокелин туудурбаган байкоолорду классификациялоо үчүн эң сонун жакындыкка ээ. Алардын моделдеринен кийин adaboost (0,9870), gradient boosting (0,9806) жана XGBoost (0,9806), полиномдук негиздеги таяныч векторлорунун машинасы (0,9805), классификация жана регрессия дарагы (0,9740), сызыктуу негиздеги таяныч векторлорунун машинасы (0,9480) жана логистикалык регрессия. Ансамблдик окутуу жана базалык классификаторлор тандалмалык метрикасы боюнча салыштырылганда, полиномдук жана радиалдык негиздеги моделдер ансамблдик моделдер менен жакын натыйжаларды көрсөткөнү көрүнүп турат. Логистикалык регрессия тандоо көрсөткүчү боюнча эң ийгиликсиз модель болуп саналат.

Кокус токой модели (1,000) жана радиалдык негиздеги таяныч векторлорунун машинасы модели (1,000) шилтеме классына болжолдоодо эң сонун натыйжаларды көрсөттү. Булардан кийин adaboost (0,9878), полиномдук таяныч векторлорунун машинасы (0,9818), gradient boosting (0,9817) жана XGBoost (0,9817), классификация жана регрессия дарактары (0,9753), сызыктуу таяныч векторлорунун машинасы (0,9532) жана логистикалык регрессия (0,9483) турат. Ансамбль моделдери менен радиалдык негиз менен полиномдуктун ортосунда тактык метрика жагынан чоң айырма жок. Логистикалык регрессия тактык метрика жагынан эң ийгиликсиз модель болуп саналат.

Моделдер AUC жагынан бааланганда, кокус токой, adaboost, XGBoost жана радиалдык негиздеги таяныч векторлорунун машинасы класстарды туура айырмалоонун 100% ыктымалдуулугуна ээ экени көрүнүп турат. Моделдерден кийин gradient boosting (0,9990), классификация жана регрессия дарагы (0,9840), полиномдук таяныч векторлорунун машинасы (0,9820), сызыктуу таяныч векторлорунун машинасы (0,9800) жана логистикалык регрессия (0,9250) турат. Моделдерди AUC баллдары боюнча салыштырганда, логистикалык регрессиядан башка бардык моделдер класстарды айырмалоонун эң аз 98% ыктымалдуулугуна ээ экени көрүнүп турат.

Изилдөөнүн эмпирикалык жыйынтыктары көрсөткөндөй, окутулган моделдердин ортосунда салыштыруу жүргүзүлгөндө, кокус токой модели эсептелген баалоо

көрсөткүчтөрүнө ылайык эң ийгиликтүү модель болуп саналат. Окутулган моделдердин ичинен болжолдоо жөндөмү жагынан эң ийгиликсиз модель логистикалык регрессия болуп саналат. Бул изилдөөдө даярдалган базалык классификаторлордун (таяныч векторлорунун машинасы, классификация жана регрессия дарагы жана логистикалык регрессия) арасында таяныч векторлорунун машинасы моделдери болжолдоочу көрсөткүчтөр боюнча болжолдоо ийгилигине ээ экенин көрсөтүп турат. Таяныч векторлорунун машинасы моделдери, өзгөчө радиалдык негиздеги модели, ансамблдик окутуу моделдери менен жакын болжолдоо натыйжаларына ээ. Алынган натыйжалар бизнес (ишкана) үзгүлтүксүздүгүн болжолдоодо машиналык окутуу моделдеринин натыйжалуулугун ачып берет.

Изилдөөнүн эмпирикалык натыйжалары айрыкча Түркияда машиналык окутуу жана көз карандысыз аудит боюнча изилдөөлөргө салым кошот. Изилдөөнүн жалпы таасирлери, тизмеленгендерден тышкары, төмөнкүлөр болуп саналат:

- Иштелип чыга турган моделди аудитордун чечим кабыл алуу процессин колдоо үчүн көз карандысыз аудит программаларында колдонуу пландаштырылууда. Кошумчалай кетсек, сунушталып жаткан модель ишкананын ресурстарын пландаштыруу программасында жана жогорку деңгээлдеги чечимдерди колдоо боюнча программалык камсыздоосунда жогорку даражалуу менежерлердин чечим кабыл алуу процесстеринде колдонулушу мүмкүн деп болжолдонууда, алар бүгүнкү күндө бизнесте кеңири колдонулат. Сунушталган моделди ишкананын финансылык абалы жөнүндө чечим кабыл алуу процессинде байланышы бар тараптар (мамлекеттик мекемелер, кредиттик мекемелер, инвесторлор жана өнөктөштөр) колдонсо болот.

- Изилдөөнүн көлөмүн кеңейтүү менен машиналык окутуу ыкмалары жана табигый тилди иштетүү алгоритмдери колдонулган реалдуу убакыт режиминде аудиттин чечимдерин колдоо программалык камсыздоо долбоору үчүн негиз түзүлөт деген ойлор бар. Аудиттик чечимдерди колдоо программалык камсыздоосу реалдуу убакытта XBRL таксономияларын колдонуу менен каржылык отчеттуулукту талдоо аркылуу аудитордук иш-аракетти камсыз кылат. Ошентип бизнесте башкарууга байланыштуу тараптардын маалымат керектөөлөрү үзгүлтүксүз камсыз кылынат.

Бизнести башкаруу жана аудитор үчүн бизнестин үзгүлтүксүз жүрүшүн алдын ала айтуу үчүн машиналык окутуу ыкмаларын колдонуунун пайдасы айдан ачык. Бирок моделдердин болжолдуу ийгилигине так процесс долбоору менен жетишүүгө болот. Бул контекстте келечектеги изилдөөлөрдө машиналык окутуу боюнча адабияттарындагы өнүгүүлөрдүн негизинде машинанылык окутуу процессинин ар бир этабын сынап көрүү алдын ала жакшыраак натыйжаларды көрсөткөн моделдердин өнүгүшүнө жол ачат деген тыянакка келсе болот. Маалыматтар машиналык окутуу процессинин эң маанилүү элементи болуп саналат. Изилдөөдө колдонулган маалымат топтому Түркия менен гана чектелген. Бирок келечектеги изилдөөлөрдө ар кандай маалымат топтомдору менен окутулган моделдердин бааланышы моделдердин глобалдашуусуна чоң салым кошо турганы көрүнүп турат. Мындан тышкары изилдөөдө финансылык катыштарды гана камтыган маалымат топтому колдонулган. Келечектеги изилдөөлөрдө каржылык катыштарды, ошондой эле корпоративдик башкаруу көрсөткүчтөрүн жана макроэкономикалык көрсөткүчтөрдү камтыган маалымат топтомдору менен моделдерди окутуу бул тармакка олуттуу салым кошот деп болжолдонууда. Машиналык окутуу изилдөөлөрүндө маалыматтар негизинен эки формада пайда болот. Структуралык маалыматтар менен жүргүзүлгөн изилдөөлөр адабиятта кеңири таралган. Акыркы жылдарда текстти интеллектуалдык анализи жана табигый тилди иштетүү технологияларындагы өнүгүүлөр түзүмдөлбөгөн тексттик маалыматтарды моделдерге киргизүүгө жол ачты. Бул контекстте түзүмдүк эмес форматтагы маалыматтарды камтыган моделдердин жайылтылышы бул тармактын өнүгүшүнө чоң салым кошот деген ой бар. Машиналык окутуу процессинде колдонулган ыкмалар дайыма өзгөрүп турат. Келечектеги изилдөөлөрдө машиналык окутуу адабиятындагы өнүгүүлөрдүн негизинде жаңы ыкмаларды сыноо үзгүлтүксүздүктү болжолдоо адабиятына маанилүү салым кошот деп айтсак болот.

**Ачкыч сөздөр:** Эркин Аудит, Машиналык Окуу, Бизнестин Үзгүлтүксүздүгүн Болжолдоо, Борса Истанбул А.Ш.

## RUSÇA ÖZET (АННОТАЦИЯ)

### ПОДХОД МАШИННОГО ОБУЧЕНИЯ В НЕЗАВИСИМОМ АУДИТЕ: НА ПРИМЕРЕ АО "БОРСА ИСТАНБУЛ А.Ш."

Развитие технологий искусственного интеллекта повлияло на аудит, а также на все элементы бизнеса. С начала 2000-х годов методы искусственного интеллекта и машинного обучения стали использоваться при оценке непрерывности деятельности предприятия, а также на разных этапах аудиторского процесса. Оценка непрерывности деятельности считается одним из основных источников вспомогательной информации для всех заинтересованных сторон. Для заинтересованных сторон очень важно иметь модели классификации и прогнозирования, которые могут предоставить точную и надежную информацию о непрерывности деятельности предприятия. Финансовые отчеты являются наиболее важными источниками, предоставляющими сведения, необходимые для оценки текущего состояния предприятия. Статистические методы широко используются для оценки непрерывности деятельности предприятия с 1960-х годов. Целью данного исследования является проверка эффективности методов машинного обучения при оценке непрерывности деятельности предприятия. Для этой цели было обучено девять различных моделей с использованием логистической регрессии, машин опорных векторов (линейные, полиномиальные и радиальные ядра), деревьев классификации и регрессии, алгоритма случайного леса, методов adaboost, gradient boosting и XGBoost, которые широко используются в литературе. В исследовании использовались финансовые отчеты предприятий, осуществляющую деятельность в производстве и подотраслях АО «Борса Истанбул А.Ш.». Набор данных охватывает период с 2017 по 2021 год. Были изучены аудиторские отчеты, опубликованные на Платформе публичного раскрытия информации предприятий за соответствующие годы, и определены предприятия, получившие заключение о непрерывности деятельности. В результате проведенной проверки предприятия были промаркированы вручную, в окончательном наборе данных было 43 года наблюдений с непрерывностью и 788 за отсутствием непрерывности. Перед выбором функций в набор данных были включены 75 финансовых коэффициентов, определенных в результате всестороннего обзора литературы. Были рассчитаны правильность, чувствительность, селективность,

ошибка типа I, ошибка типа II, точность и показатели AUC моделей, обученных в исследовании, и модели оценивались с использованием рассчитанных метрик. При рассмотрении эмпирических результатов исследования метрика точности модели случайного леса рассчитывается как 1,000 среди всех моделей. Это показывает, что модель случайного леса является наиболее успешной моделью среди всех моделей. Кроме того, когда исследуются рассчитанные метрики всех моделей, видно, что модели ансамблевого обучения и модели машин опорных векторов показывают более успешные результаты, чем деревья классификации и регрессии и логистическая регрессия. При рассмотрении эмпирических результатов исследования было установлено, что модели машинного обучения могут использоваться для оценки непрерывности деятельности предприятия.

**Ключевые слова:** Внешний Аудит, Прогнозирование Непрерывности Деятельности, Машинное Обучение, Борса Истанбул А.Ш.

## İNGİLİZCE ÖZET (SUMMARY)

### **MACHINE LEARNING APPROACH IN AUDITING: BORSA İSTANBUL A.Ş. CASE**

Prediction of going concern is accepted as one of the main supportive information sources for all stakeholders. It is very important for stakeholders to have classification and forecasting models that can provide accurate and reliable information about the going concern capability of the business. Financial reporting systems provide the information needed to evaluate the financial situation of businesses. Statistical methods have been widely used in the prediction of going concern since the 1960s. In recent years, with the acceleration of technological developments, machine learning methods have started to be used in testing the going concern of the enterprise as a substitute for statistical methods that have strict assumptions. This study aims to test the effectiveness of machine learning methods in the prediction of the going concern ability of companies. For this purpose, nine different machine learning models were trained namely logistic regression, support vector machines (linear, polynomial, and radial-based kernels), classification and regression trees, random forest algorithm, AdaBoost, gradient boosting, and XGBoost methods which are widely used in the literature. In the study, financial reports of the companies listed in the manufacturing and sub-sectors of the Borsa İstanbul A.Ş. have been used. The dataset covers between 2017 to 2021. The audit reports announced on the Public Disclosure Platform of the companies between the relevant years were examined and the companies that received the going concern opinion were determined. Businesses have been labeled as a result of the investigation carried out. Thus, the final data set included 43 observation years that received going concern opinion and 788 did not receive going concern opinion. In order to overcome the unbalanced dataset problem, SMOTE method was used and a feature selection strategy was determined with genetic algorithms in order to increase the prediction accuracy of the models. Before the feature selection, 75 financial ratios determined by a comprehensive literature review were included in the data set. The accuracy, sensitivity, selectivity, type-I error, type-II error, precision, and AUC scores of the models trained in the study were calculated and the models were evaluated using those calculated metrics. When the empirical results of the study were examined, it was revealed that the random forest model was the most successful among all the trained models with 100% accuracy. When the calculated metrics of the nine

models are examined, it is seen that ensemble learning models and support vector machine models show more successful results than classification regression trees, and logistic regression. The empirical results of the study reveal the effectiveness of machine learning models in predicting of going concern ability of companies.

**Keywords:** External Auditing, Going-Concern Prediction, Machine Learning, Borsa İstanbul A.Ş.



## GİRİŞ

Denetim, ekonomik olaylar ve eylemler ile ilgili iddialara ilişkin kanıtları nesnel olarak elde etmek ve değerlendirmek, bu iddialar ile önceden belirlenmiş kriterler arasındaki uygunluk derecesini tespit etmek ve sonuçları ilgili taraflara iletmeyi amaçlayan sistematik bir süreçtir (“A Statement of Basic Accounting Concepts” 1973).

İşletmenin sürekliliği varsayımı muhasebe teorisindeki en önemli kavramlardan birisidir (Sterling 1968). Süreklilik varsayımı aksi kararlaştırılmadıkça işletmenin yakın bir gelecekte tasfiye edilmeyeceği ve faaliyetlerinin durdurulmayacağını belirtmektedir (O’Regan 2004). İşletmelerin mali tabloları süreklilik varsayımının altında hazırlanmaktadır. Bu varsayımın ihlalinin mali tablolar üzerinde ciddi etkileri olabilir. İşletme yatırımcıları ve alacaklıları, işletmenin faaliyetlerinin devam edeceğini varsaymaları sebebi ile yatırım yapmakta ve borç vermektedir. Bu, meslek mensuplarının “işletmenin sürekliliği varsayımı” olarak aşına oldukları bir beklentidir. İşletmenin geçmişi ile ilgili bilgiler iflas halinde ise ya da iflas etmesi bekleniyor ise geleceğine ilişkin beklentilerin değerlendirilmesinde daha az yararlıdır. Bu durumda, odak işletmenin performansından işletmenin kaynakları ve yükümlülüklerine kaymaktadır. Bir işletmenin beklenen faaliyetten beklenen iflasa kayması mali raporlamanın amacını mutlaka değiştirmeyecektir. Ancak mali tablo unsurlarının ölçümleri de dahil olmak üzere bu hedeflerle ilgili bilgiler değişebilmektedir (Financial Accounting Standards Board 1978, 20).

İşletmenin sürekliliği yeteneği hakkında şüphe oluşturacak önemli bir belirsizlik bulunması işletmeyi, denetçiyi ve diğer mali tablo kullanıcılarını etkilemektedir. İşletmeler, mali sağlıklarına ve sürekliliklerine, sektördeki imajlarına ve bozulan durumlarının olumsuz etkilerine önem vermektedir. Özellikle işletme yatırımcılarının ve alacaklılarının işletmenin süreklilik kabiliyetinin, üzerinde direkt etkisi olan stratejik çıkarları bulunmaktadır.

Makine öğrenmesi bilgisayarların belirli bir görevde, zaman içerisinde deneyim yolu ile başarısını artırmasını sağlayan algoritmaların ve istatistiksel modellerin geliştirilmesine olanak sağlayan yapay zekanın geniş bir alanıdır. Bir öğrenme uzayındaki farklı hipotezlerin sınanmasını içeren yaklaşım, bir makinenin büyük ve karmaşık veri kümelerinden faydalı örüntüleri otonom olarak öğrenebilen ve böylece her senaryo için açık bir şekilde tekrar programlama yapmadan tahminde bulunma ve sınıflandırma yapma becerisine sahip olduğu varsayımına dayanmaktadır.

Çalışma dört bölümden oluşmaktadır. Birinci bölümde, denetim kavramı, muhasebe denetiminin tanımı ve özellikleri, denetim türleri, denetçi kavramı ve denetçi türleri, genel kabul görmüş denetim standartları kavramları ile uluslararası denetim standartları ve BDS- 570 işletmenin sürekliliği standardı hakkında bilgi verilmektedir. İkinci bölümde makine öğrenmesi kavramı, öğrenme stratejileri, gözetimli öğrenme, gözetimsiz öğrenme, pekiştirmeli öğrenme, makine öğrenmesi süreci, veri, veri seti ve veri ön işleme kavramları, gürültü, aykırı değer, normalizasyon, aşırı uyum ve yetersiz uyum, öznitelik seçimi, çapraz doğrulama kavramları hakkında bilgi verilmiş olup eğitilen makine öğrenmesi algoritmaları olan lojistik regresyon, destek vektör makineleri, karar ağaçları, adaboost, rastgele orman algoritması, gradient boosting ve extreme gradient boosting metotları hakkında bilgi verilmektedir. Üçüncü bölümünde işletmenin sürekliliğinin tahmini hakkında literatür taramasına yer verilmiştir. İşletmenin sürekliliği tahmini literatürü yaygın olarak istatistiksel dönem ve makine öğrenmesi metotlarının kullanıldığı dönem olarak iki aşamada ele alınmaktadır. Çalışmada bu yaklaşım benimsenmiş ve bölümde alana öncülük etmiş olan, istatistiksel metotlardan yararlanılmış olan çalışmalara ve makine öğrenmesi metotlarından yararlanılmış olan çalışmalara müstakil başlıklar altında yer verilmiştir. Literatür taramasında çalışmaların kullandıkları yöntemler, veri setleri, veri setlerinde yer verilen öznitelikler ve ampirik sonuçları hakkında bilgi verilmiştir. Çalışmanın dördüncü bölümünde araştırmanın amacı ve kapsamı, araştırmanın önemi, araştırmanın kısıtları, araştırmanın yaygın etkisi, araştırmanın yöntemi, çalışmada kullanılan veri seti ve öznitelikler, veri ön işleme süreci hakkında detaylı bilgiler paylaşılmıştır. Çalışmada eğitilen makine öğrenmesi modellerinin sonuçları detaylı bir şekilde raporlanmıştır. Çalışma sonucunda elde edilen bulguların yorumuna ve gelecek araştırmalar için önerilere sonuç ve öneriler bölümünde yer verilmiştir.

### **Çalışmanın Amacı**

Bu çalışmada makine öğrenmesi metotlarının işletmenin süreklilik kabiliyetinin tahminindeki etkinliğinin tespiti amaçlanmıştır. Bu amaçla, çalışmada lojistik regresyon, sınıflama ve regresyon ağaçları, destek vektör makineleri (doğrusal, polinomsal ve radyal tabanlı kernel çekirdekleri), rastgele orman algoritması, gradient boosting ve extreme gradient boosting algoritmaları kullanılarak dokuz farklı model eğitilmiş ve modellerin sonuçları karmaşıklık matrisi kullanılarak hesaplanan doğruluk, duyarlılık, seçicilik, kesinlik, tip-I hata ve tip-II hata metrikleri ile AUC skoru yardımı ile değerlendirilmiştir.

### **Çalışmanın Önemi ve Katkısı**

Denetimin temel amacı; işletmenin mali tablolarının işletmenin mali durumu ve faaliyet sonuçlarının tüm önemli yönleri ile gerçeğe uygun ve doğru bir biçimde gösterilip gösterilmediği hakkında görüş bildirilmesini sağlamaktır. Denetim sürecinde işletmenin süreklilik kabiliyetinin değerlendirmesi çok önemli bir yer tutmaktadır. Denetimde süreklilik kabiliyetinin değerlendirmesinin yasal zorunlulukları yanında yasal olmayan gereklilikleri de bulunmaktadır. İşletmenin sürekliliğinin tahmini mali raporlama ve mali analizde çok önemli bir yer tutmaktadır. Süreklilik kavramı mali tabloların temelini oluşturan, muhasebenin temel varsayımlarında birisidir ve işletmenin süreklilik kabiliyetinin kaybetmesinin farklı olumsuz etkileri bulunmaktadır.

İşletmenin süreklilik kabiliyetinin bulunmaması ilk olarak muhasebe politikaları üzerinde etkili olmaktadır. Bu bağlamda doğru tahmin modelleri geliştirilmesi mali tabloların doğruluğunun ve muhasebe standartlarına uyumun sağlanmasında önemli bir etkiye sahiptir. İşletme yatırımcıları ve borç verenler işletme ile ilgili karar verirken temel olarak işletmenin mali tablolarından faydalanmakta ve mali tabloların içerdiği bilgilere güvenmektedir. İşletmenin gelecekteki potansiyel bir iflas ile karşı karşıya olması ya da iflas riski taşıması yatırımcıların ve borç verenlerin kararları üzerinde etkilere sahiptir. Bu bağlamda tahmin doğruluğu yüksek modellerin geliştirilmesi yatırımcıların risk değerlendirme sürecine büyük katkı sağlayacaktır. İşletme yönetimi açısından işletmenin süreklilik riskinin belirlenmesi işletmenin geleceğini planlamada önem arz etmektedir. İşletme yönetiminin, süreklilik kabiliyeti hakkında bilgi sahibi olması potansiyel iflasların önüne geçebilecek proaktif eylemler almasına olanak tanıyacaktır.

İşletmenin süreklilik kabiliyeti hakkında önemli bir belirsizliğin mevcudiyetinin ortaya çıkarabileceği ekonomik kayıpların büyüklüğü ve bu kayıplardan etkilenen paydaşların sayısı çok büyüktür. İşletmelerin süreklilik riskini tahmin edebilmeleri, işletme ve işletme paydaşlarının potansiyel kayıpların önüne geçebilmelerine ve süreci en az zarar ile atlatabilmelerine olanak tanımaktadır. Bu bağlamda süreklilik kabiliyeti hakkında tahmin gücü yüksek modellerin geliştirilmesi, mikro düzeyde işletme paydaşların iktisadi kayıplar yaşamasının önüne geçebilecektir.

### **Çalışmanın Kapsamı**

Çalışma kapsamında kullanılan veri seti Borsa İstanbul A.Ş. pay piyasası imalat sanayi ve alt sektörlerinde işlem görmekte olan işletmeleri kapsamaktadır. Finans kuruluşlarının mali yapılarının farklı olması sebebi ile çalışmada kullanılacak olan veri seti sadece

imalat işletmeleri ile sınırlı tutulmuştur. Çalışmada gözlem yılları, bağımsız denetim raporunda işletmenin sürekliliğini devam ettirme kabiliyeti hakkında önemli bir belirsizliğe atıf yapıp yapılmadığına göre etiketlenmiştir. Veri seti 2017 ile 2021 yılları arasında toplam 201 işletmeye ait mali tabloları kapsamaktadır. Veri seti işletmenin sürekliliği açısından riskli olarak etiketlenmiş olan 16 işletmeye ait 47 gözlem yılına karşılık, işletmenin sürekliliği açısından risk taşımayan 198 işletmeye ait 958 gözlem yılından oluşmaktadır.

### **Çalışmanın Yöntemi**

Çalışmanın amacına bağlı kalınarak makine öğrenmesi metotlarının eğitimi maksadı ile işletmelere ait mali oranlar kullanılmıştır. BDS- 570 işletmenin sürekliliği standardında işletmenin mali durumu ile ilgili, işletme faaliyetleri ile ilgili ve diğer durumlar ile ilgili olmak üzere üç ana başlıkta atıf yapılan, işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin ciddi şüphe oluşturabilecek olaylar veya şartlar hakkında referanslar dikkate alınarak, kapsamlı bir literatür taraması ile ilk olarak 75 mali oran belirlenmiştir. İlk belirlenen mali oranlar arasından genetik algoritmaların kullanıldığı bir öznitelik seçim süreci ile 57 mali oran öznitelik olarak seçilmiştir. Veri setinde yer alan eksik gözlemler öznitelik ortalamaları kullanılarak tamamlanmıştır. Çalışmada min-max normalizasyon metodu kullanılarak normalizasyon yapılmıştır. Veri setinin dengelenmesi amacı ile SMOTE metodundan faydalanılmıştır. Makine öğrenmesi modellerinin eğitimine problemin tanımlanması ile başlanmıştır. Problem tanımlanması sonrasında veri elde etme işlemi yürütülmüştür. Veri ön işleme aşaması eksik verilerin tamamlanması, normalizasyon, veri setinin dengelenmesi ve öznitelik seçimi adımlarını içermektedir. Eğitilen modellere ait doğruluk, duyarlılık, seçicilik, tip-I hata, tip-II hata, AUC skoru ve kesinlik metrikleri hesaplanmıştır. Hesaplanan değerlendirme metrikleri kullanılarak hem çalışmada eğitilen modeller kendi aralarında karşılaştırılmış hem de literatürdeki çalışmalarda aynı algoritmalar kullanılarak eğitilen modeller ile karşılaştırılmıştır.

## BİRİNCİ BÖLÜM

Bu bölümde denetim kavramı ve özellikleri, denetim türleri, denetçi kavramı, denetçi türleri, uluslararası denetim standartları ve BDS-570 işletmenin sürekliliği standardı hakkında teorik bilgi verilmektedir.

### 1.1. Denetim Kavramı

Günümüzde işletmelerin ve paydaşlarının faaliyetlerini sürdürdükleri çevre koşulları değişim içerisindedir. İşletmeler gibi, onların faaliyetleri ile ilgilenen üçüncü kişi ve kurumlarda iş dünyasının karşılıklarına çıkardığı yeni uygulamaları benimsemekte, onlara uyum göstermektedir. Değişim, işletme içerisinde ortaya çıkan durumlardan kaynaklanabileceği gibi dış çevreye uyum sağlama amacı ile de kabullenilebilir. Her ne amaç ile olursa olsun değişimin en fazla etkilediği işletme fonksiyonlarında birisi muhasebedir. Değişim işletmelerin faaliyetlerini dolayısı ile muhasebe bilgi sisteminin sorumluluk sahasını değiştirmektedir. Muhasebe bilgi sisteminin ürettiği mali ve mali olmayan raporlar hem işletme içi hem de işletme dışı bilgi kullanıcıları açısından son derece önemlidir.

Mali tablolar, işletmenin mali durumu ve faaliyet sonuçlarına ilişkin bilgilerin ilgili taraflar ile paylaşılmasını sağlayan belgelerdir. Mali tablo kullanıcılarının temel amacı, işletme hakkındaki mali ve mali olmayan olaylar hakkında doğru ve güvenilir yargılar üretebilmektir. Bu nedenle, mali tabloların içerdiği bilgilerin doğruluk ve güvenilirlik derecesi, mali tablo kullanıcıları için hayati öneme sahiptir. Mali tablolarda sunulan bilgilerin doğruluğu ve güvenilirliği ile mali tablo kullanıcısının üreteceği yargının kalitesi arasında doğrudan bir ilişki mevcuttur (Demir 2010, 114).

Mali tabloları hazırlayan taraflar ile bilgi kullanıcıları arasındaki amaç farklılıkları nedeni ile ortaya çıkabilecek olan çıkar çatışmaları mali tabloların içerdiği bilginin doğruluğu ve güvenilirliğini etkileyebilmektedir. Denetim, mali tabloların içerdiği bilginin doğruluğu ve güvenilirliğini inceleyen ve bu konuda makul güvence sağlayan bir faaliyetler bütünüdür.

İktisadi unsurlar, yaralanacakları bilginin kalitesini artırmak amacıyla denetim hizmetine ihtiyaç duymaktadır. Denetim sürecinin başından sonuna kadar, denetim hizmeti alan işletmenin ilk talebi denetim kalitesinin yüksek olmasıdır (Arruñada 1999, 13).

Kaliteli finansal raporlamada denetim hizmetlerinin önemi bulunmaktadır. Denetim sonucunda verilen bağımsız denetim yargısı çok büyük bir etkiye sahip olan iktisadi bir

unsurdur. Muhasebe ve denetim sistemleri arasında amaç ve yöntem bağlamında farklılıkla bulunmasına rağmen, birbirileri ile ilişkili yapılarıdır. Muhasebe tarafından üretilen bilgilerin doğru, tarafsız ve güvenilir olması işletme ve ilişkili tarafların doğru kararlar alabilmesi için büyük önem taşımaktadır (Demir 2015, 342).

## 1.2. Muhasebe Denetiminin Tanımı ve Özellikleri

Denetim kavramı, dilimizde, İngilizce dilindeki “auditing” kelimesinin karşılığıdır. “Auditing” kelimesi köken itibari ile Latince ve “işitme ya da dinleme” anlamına sahip olan “audire” kelimesi kökünden gelmektedir. Kelimenin bu kökten türemiş olmasının sebebi, faaliyeti icra etmek için seçilen kişilerin, doğruları bulmaya yönelik yürüttükleri faaliyetlerini dinleyerek yapmalarıdır (Bozkurt 2018, 23).

Literatürde denetim kavramını tanımlayan tek bir tanım bulunmamaktadır. Araştırmacılar, denetimi farklı bakış açıları ile değerlendirdikleri için benzer unsurları işaret eden farklı tanımlar yapmışlardır. Literatürde yer alan tanımlardan birisi denetimi, iktisadi faaliyet ve olaylara ilişkin iddiaların, önceden saptanmış ölçütlere uygunluk derecesini araştırmak ve sonuçları ilgi duyanlara bildirmek amacıyla tarafsızca kanıt toplayan ve bu kanıtları değerleyen sistematik bir süreç olarak ifade etmektedir (Başpınar 2005, 35; Bozkurt 2018, 29; Güredin 2014, 11; Kaval vd. 2015, 3). Denetimin, iktisadi olaylar hakkında üretilen bilginin doğruluğunu ve güvenilirliğini sınamaya yönelik bir süreç olduğunu işaret eden tanımının yanında, Arens vd., (2012), denetimi icra edecek yetkilinin taşınması gerektiği niteliğine de atıfta bulunmaktadır. Tanıma göre, denetim, bilgi ile önceden saptanmış kriterler arasındaki uygunluk derecesini belirlemek ve raporlamak maksadı ile, bilgi hakkında kanıt toplama ve değerlendirme sürecidir ve denetim, yetkin ve bağımsız bir kişi tarafından yapılmalıdır. (Arens, Elder, ve Beasley 2012, 4)

Denetim tanımlarından yola çıkarak denetimin özellikleri şu şekilde sıralanabilir (Arens, Elder, ve Beasley 2012, 4; Başpınar 2005, 35-36; Bozkurt 2018, 29-31; Murat Erdoğan 2002, 56; Güredin 2014, 11; Kaval vd. 2015, 3-4; Yılcı, Yıldız, ve Kiracı 2013, 1):

- **Karşılaştırma sürecidir**, denetim çeşitli aşamaları olan bir süreçtir ve başından sonuna kadar olan tüm aşamalar belirli bir plan dâhilinde yürütülür.
- **İktisadi faaliyet ve olaylara ilişkin iddialar**, işletmeler tarafından hazırlanmış olan her türlü rapor veya beyan işletme açısından bir iddia niteliği taşımaktadır. Denetim

ilgili taraflara sunulan bu rapor veya beyanların gerçeğe uygunluğunun araştırılmasıdır.

- **Uzman ve bağımsız kişi ve/veya kişilerce yürütülür**, bağımsızlık ve bağımsız davranabilme özelliği denetçi için vazgeçilmez bir özelliktir.
- **Önceden saptanmış ölçütler**, denetçi işletmeler tarafından üretilen rapor ve beyanları önceden saptanmış ölçütler ile karşılaştırarak, doğruluğu ve güvenilirliği hakkında görüş oluşturmaktadır. Ölçütler kanun, anlaşma, işletme yönetimi tarafından saptanmış hedefler ve genel kabul görmüş muhasebe ilkeleridir.
- **Tarafsızca kanıt toplama ve değerlendirme**, denetim sürecinde denetçi, işletme iddialarının doğruluğu ve güvenilirliği ile ilgili görüş oluşturmak maksadı ile kanıt toplar. Kanıt toplama sürecinde denetçi, işletme ve ilgili kurumlardan bağımsız davranmalı, yeterli ve uygun nitelikte, görüşünü destekleyen kanıt toplamalıdır.
- **İlgili taraflara bildirme**, denetçi işletme iddiaları ile ilgili yaptığı denetim sonucunda ulaştığı görüşü yazılı bir rapor halinde ilgililer ile paylaşır. Denetim raporu denetimin son aşamasını oluşturur.

Denetim amacı, “ Denetlenen işletmenin belirli bir döneme ait finansal bilgilerini ve finansal tablolarını önceden saptanmış ve genel kabul görmüş standartlarla karşılaştırarak, bu bilgilerin söz konusu standartlarla uygunluk derecesinin saptanması, işletme yönetimince sunulan finansal tablolardaki bilgilerin güvenilirliğinin artırılması ve usulsüzlük riski faktörlerinin değerlendirilerek finansal tabloların güvenilirliklerini önemli ölçüde etkileyecek temel yolsuzluk, hata ve hilelerin ortaya çıkarılması ve önlenmesi” olarak ifade edilmektedir. (Özbirecikli ve Süslü 2005, 69).

Sanayi devrimi sonrası işletmelerin ortaklık yapılarında yaşanan değişim ve sermaye piyasalarında yaşanan gelişim ile işletmelerin mali durumları ile ilgilenen ilişkili tarafların sayısının artması, denetim faaliyetlerinin gerekliliğini arttırmıştır. Halka arzların arttığı bu süreçte, işletmeler finansal durumu ile ilgili güvenilir bilgi ihtiyacı ön plana çıkmıştır (Kapusuzoğlu 2006, 62; Saygılı 2007, 2).

İktisadi olarak kalkınmış olan her ülkede denetim faaliyetleri bağlamında gerekli yasal düzenlemeler yapılmıştır. Finansal nitelikli olaylar ile ilgili bilgilerin ilgililere sadece gerçeği göstermesi gerekmektedir. Bu da etkin bir denetim mekanizmasının varlığı ile sağlanabilir (Murat Erdoğan 2002, 51-52).

Denetimin sağladığı faydalar şu şekilde sıralanabilir (Murat Erdoğan 2002, 61-62; Kapusuzoğlu 2006, 65-66);

- Üstler, astlarının faaliyetlerini daha iyi değerlendirme imkânı kazanmakta ve astlar ise üstlerinin gelecek ile ilgili planları ve beklentileri ile ilgili daha iyi bilgi sahibi olmaktadır. Bu bağlamda denetim faaliyetlerinin işletme içerisindeki kontrolü ve iletişimi artırdığı söylenebilmektedir.
- Astların, üstler tarafından işletme içinde ve işletmeler arasında uygulanan değerlendirme çalışmaları, işletme içindeki motivasyonu sağlamaktadır.
- Denetim faaliyeti neticesinde hazırlanan raporlar, işletme yönetimine işletmenin mali durumu ile ilgili olumlu ya da olumsuz, tarafsız ve güvenilir bilgi sağlamaktadır. Denetçi tarafında tasdik edilmiş olan mali tablo, yönetimin daha sağlıklı ve doğru karar vermesini sağlayacaktır. Çünkü denetim çalışmaları mali tablolardaki hileleri ve hataları ortaya çıkarmaktadır.
- Denetim faaliyetleri, işletme yönetiminde ya da yönetiminde bulunmayan tüm ortakların ortaklık haklarının korunmasını sağlamaktadır.
- Denetim faaliyetleri işletmenin kredi veren üçüncü şahıslar veya kurumlar ile olan ilişkilerinde yarar sağlamaktadır. Sermaye piyasalarındaki yatırımcılar ile olan ilişkilerde benzer bir etki bulunmaktadır.

### **1.3. Denetim Türleri**

İşletmelerde yürütülen denetim faaliyetleri farklı başlıklar altında sınıflandırılabilir. Genel olarak denetim faaliyetleri yapılaş amacına, nedenine, kapsamına, zamanına ve denetçinin statüsüne göre sınıflandırılmaktadır. Çalışmanın bu kısmında denetim faaliyetlerinin türleri, bu sınıflandırmaya sadık kalınarak incelenmiştir.

#### **1.3.1. Amacına Göre Denetim Türleri**

Yapılış amacına göre denetim mali tablolar denetimi, uygunluk denetimi ve faaliyet denetimi olmak üzere üç başlık altında sınıflandırılmaktadır (Çömlekçe vd. 2004, 6).

##### **1.3.1.1. Mali Tablolar Denetimi**

İşletmelerin iç ve dış çevresinde yaşanan dönüşüm, mali tabloların içerdiği bilginin güvenilirliğinin önemini giderek artırmaktadır. Mali tabloların güvenilirlik derecesinin artırılmasında yararlanılan yapılardan birisi de mali tablolar denetimidir. Mali tablolar denetiminden mali raporların güvenilirliğinin artırılmasında faydalanılmaktadır (Kavut 2001, 2). Mali tablolar denetimi, bir işletmenin belli bir dönem sonunda hazırlanan mali tablolarının önceden saptanmış ölçütlere uygunluk derecesinin araştırılarak, ulaşılan

sonucun raporla bilgi kullanıcılarına sunulması faaliyetidir. Mali tablolar denetimi bağımsız veya bağımlı denetçiler tarafından yapılmaktadır (Aksoy 2005, 118). Mali tablolar denetimi günümüzde en yaygın olarak yürütülen denetim türüdür (Tosun 2021, 101). Mali tablolar denetiminin temel amacı, mali tabloların bir bütün olarak önceden belirlenmiş olan ölçütlere uygunluğunu araştırılmasıdır (Güredin 2014, 16).

### **1.3.1.2. Uygunluk Denetimi**

Uygunluk denetimi, işletme içerisinde veya dışarısından yetkili bir üst makam tarafından saptanmış olan kurallara uyulup uyulmadığının ve uyum derecesinin araştırıldığı denetim faaliyetidir. Uygunluk denetiminin, etkinliğin artırılması için denetimi gerçekleştirecek olan denetçinin konusunda yetkin olması, denetimin yürütüleceği işletmede etkin bir dosyalama sisteminin mevcut olması ve denetim raporlamasının bir sisteme bağlanmış olması gerekmektedir (Güredin 2014, 17).

Uygunluk denetimine konu olan kurallar Uygunluk denetiminde, araştırılan kurallar farklı kişi ya da kurumlarca oluşturulmaktadır. Ayrıca denetim süreci işletme içerisinde yetkilendirilmiş kişi ya da kişilerce de yürütülebileceği gibi Sosyal Güvenlik Kurumu, Hazine ve Maliye Bakanlığı ya da Sayıştay gibi kamu kurumları tarafından da yürütülebilir (Akbiyık 2005, 40) İşletmenin mali işlemlerini konu alan bir denetim uygulamasında, iç denetçi işletme yönetiminin oluşturduğu ve kabul ettiği iç kontrol prosedürlerini dikkate almaktadır (Kepekçi 2004, 3). Uygunluk denetimi sonucunda elde edilen çıktılar tüm çıkar grupları yerine ilgili kişi ya da kurumlara raporlanmaktadır. Bu bağlamda ilgili kurumlar işletme yönetimi ya da bir devlet kurumu olabilir (Çalgan vd. 2008). Uygunluk denetimi genellikle kamu denetçileri veya iç denetçiler tarafından yürütülmektedir. (Cömert vd. 2013, 1:17).

### **1.3.1.3. Faaliyet Denetimi**

İşletmede mali nitelikteki olaylar gibi mali nitelikte olmayan işgücü ve işgücü verimliliği, hammadde verimliliği, makine verimliliği, karlılık, işletmenin faaliyetlerini yürüttüğü sektör, belirlenmiş amaç ve hedeflere ne ölçüde ulaşıldığı gibi konularda denetim faaliyetlerinin kapsamına girmektedir (Akbiyık 2005, 41). Faaliyet Denetimi, işletmede belirlenen amaçlara ulaşıp ulaşılmadığını ve işletmenin verimli çalışıp çalışmadığı tespit etmek amacı ile işletme politikalarının ve bu politikaların sonuçlarının değerlendirilmesini kapsayan denetim faaliyetidir (Toroslu 2012, 68).

Faaliyet denetimi verimlilik ve etkinlik kavramları üzerine kurulmuş bir yapıdır. İşletme faaliyetlerinin gözden geçirilerek belirlenmesi faaliyet denetiminin amaçlarından birisidir. Etkinlik ve verimlilikteki artışların tespiti önemli bir unsurdur. Denetçinin geçmiş ve cari dönem raporlarını inceleyerek, sektör ortalamalarını analiz ederek, işletme içindeki ve dışındaki faaliyetleri gözlemleyerek, işlemleri değerlendirerek, mesleki yargılarına ve tecrübesine dayanarak meydana gelmesi muhtemel gelişmeler için öneriler yapması gerekmektedir. Bu bağlam da gelecek gelişmeler için fırsatların belirlenmesi faaliyet denetiminin bir diğer amacıdır. Ayrıca gelecekteki etkinlikler ve gelişmeler için öneriler oluşturulması faaliyet denetiminin amaçları arasındadır (Kardeş Selimoğlu 1999, 197).

### **1.3.2. Yapılış Nedenine Göre Denetim Türleri**

Yapılış nedenine göre denetim faaliyetleri, yasal (zorunlu) denetim ve isteğe bağlı denetim olmak iki başlık altında sınıflandırılmaktadır (Karanfiloğlu 1999, 32).

#### **1.3.2.1. Yasal (Zorunlu) Denetim**

Kanuni düzenlemeler ve mevzuat hükümleri doğrultusunda yapılması gereken denetim faaliyetleridir. Olumsuz denetim görüşü sebebi ile işletme yasal yaptırımlar ile karşı karşıya kalmaktadır. Zorunlu denetimin kapsamı nasıl, ne zaman, ne şekilde ve kim tarafından yürütüleceği yasa, yönetmelik ve genelge gibi metinlerce tanımlanmaktadır (Karacan ve Uygun 2012, 38).

#### **1.3.2.2. İsteğe Bağlı Denetim**

Herhangi bir yasal zorunluluğun bulunmadığı, işletme ile ilgili çeşitli çıkar gruplarının isteği üzerine yapılan denetim çalışmalarına isteğe bağlı denetim olarak adlandırılmaktadır (Dalak 2000, 69).

### **1.3.3. Kapsamına Göre Denetim Türleri**

Kapsamına göre denetim faaliyetleri, genel denetim ve özel denetim olmak üzere iki başlık altında sınıflandırılmaktadır (Dalak 2000, 68).

#### **1.3.3.1. Genel Denetim**

Muhasebe bilgi sistemi içerisindeki tüm kayıt ve işlemlerin doğruluğunu araştırmak maksadı ile yürütülen denetim faaliyetleri genel denetim olarak adlandırılmaktadır. Genel denetimde mali nitelikteki işlemlerin muhasebe kayıtlarına doğru bir şekilde aktarılıp

aktarılmadığı ve kayıtların belgelere dayalı olup olmadığı araştırılmaktadır. Bu açıdan değerlendirildiğinde mali tablolar denetimi, genel denetim tanımı kapsamına girmektedir (Çalış 2013, 19).

### **1.3.3.2. Özel Denetim**

İşletme yönetiminin ya da ilgili tarafların karar süreçlerine yardımcı olmak maksadı ile ihtiyaç duyulan konu kapsamında yürütülen denetim faaliyetidir. Satın alma, devir, tasfiye, kredi süreci, vergilendirme, yolsuzluk ve ihmal iddialarının araştırılması, menkul kıymet yatırımı, mahkeme incelemeleri ve devlet kurumları tarafından gerçekleştirilen teftişler özel denetim olarak nitelendirilmektedir (Haftacı 2011, 7)

### **1.3.4. Denetçinin Statüsüne Göre Denetim Türleri**

Denetçinin statüsüne göre denetim faaliyetleri, bağımsız denetim, iç denetim ve kamu denetimi olmak üzere üç başlık altında sınıflandırılmaktadır (Dalak 2000, 68)

#### **1.3.4.1. Bağımsız Denetim**

Bağımsız denetim, mali tabloların önemli bir hata veya hile içerip içermediğini tespit etmek amacı ile bağımsız, yeterli mesleki bilgi ve tecrübeye sahip kişilerce icra edilen denetim faaliyetidir. Bağımsız denetim sürecinde denetçi, mali tabloların veya faaliyetlerin önceden belirlenmiş olan standartlara uygunluğunu değerlendirmek maksadı ile kanıt toplamakta ve bu kanıtlardan yola çıkarak ilgililere rapor sunmaktadır (Ertaş ve Kandemir 2020, 84).

Sermaye Piyasası Kurulunun yayınladığı, Sermaye Piyasasında Bağımsız Denetim Hakkındaki Seri:X, No:22' lu tebliğde bağımsız denetim, *“İşletmelerin kamuya açıklanacak veya Kurulca istenecek yıllık finansal tablo ve diğer finansal bilgilerinin, finansal raporlama standartlarına uygunluğu ve doğruluğu hususunda, makul güvence sağlayacak yeterli ve uygun bağımsız denetim kanıtlarının elde edilmesi amacıyla bağımsız denetim standartlarında öngörülen gerekli tüm bağımsız denetim tekniklerinin uygulanarak, defter, kayıt ve belgeler üzerinden denetlenmesi ve değerlendirilerek rapora bağlanmasını”* olarak tanımlamaktadır (Sermaye Piyasası Kurumu 2011).

Bağımsız denetim, bir işletmenin mali tablolarının ve diğer mali bilgilerinin, finansal raporlama standartlarına uygunluğu ve doğruluğu hususunda, gerekli bağımsız denetim teknikleri uygulanarak denetlenmesi ve değerlendirilerek bir rapora bağlanmasıdır (Akdoğan 2020, 2-3).

Bağımsız denetim işletmelerin finansal tablolarının güvenilir olup olmadığını objektif olarak belirleyen ve bunu rapor eden faaliyettir. Bağımsız denetimin amacı ortaklıklar ve sermaye piyasası kurumlarına ait hesapların vergi kanunları bağlamında incelenmesi değildir. Bağımsız denetime tabi tutulacak, halka açıklanacak yada sermaye piyasası kurumuna gönderilecek olan, finansal tablolar uluslararası muhasebe ilke ve standartlarına uygun olup olmadıkları yönünden incelenmektedir (Murat Erdoğan 2002, 59).

Bağımsız denetimin temel işlevi, finansal tablolarda raporlanan olayların, raporlandıkları dönemdeki standart ve ilkelere uygun olarak raporlandığından emin olmaktır. Bu işlev bağımsız denetçinin mali tabloların doğru durumu yansıtıp yansıtmadığı ile ilgili kanuni yükümlülüğünü ifade etmektedir bağımsız denetçinin tüm veri kaynaklarına erişimi mevcuttur bu yolla denetim görüşü oluşturmak için gerekli bilgiye erişim sağlayabilir (Nasta ve Ladar 2015, 48).

#### **1.3.4.2. Kamu Denetimi**

Kamu denetimi, görev ve yetkilerini yasalardan alan ve kamu adına, kamu ihtiyaçlarını karşılamak maksadı ile denetim faaliyetleri yürüten kişilerce gerçekleştirilen, mali tablolar, uygunluk veya faaliyet denetimlerini ifade etmektedir (B. Ataman 2010, 19). Kamu denetimi, devlet ve özel sektör kurumlarının tüm faaliyetlerinin, kamu adına denetim yetkisine sahip kişi ya da birimlerce denetlenmesini kapsamaktadır. Kamu denetimi, kamu görevlisi olan denetçiler tarafından yürütülmektedir. Ancak bazı durumlarda özel yetkileri bulunan yeminli mali müşavirler de kamu adına denetim faaliyetleri yürütmektedir (Uzay 2004, 12).

#### **1.3.4.3. İç Denetim**

İşletmelerde, varlıkların zarar görmesini, çalınmasını ya da kaybolmasını engellemek, muhasebe bilgilerinin hata ve hilelerin önlenmesi yolu ile doğruluğu ve güvenilirliği artırmak ve işletme faaliyetlerinin etkinliği artırmak için çeşitli politika ve prosedürlerden oluşan kontrol mekanizmalarına ihtiyaç duyulmaktadır. İç denetim terimleri sözlüğünde iç denetim, “Kurumun faaliyetlerini geliştirmek ve bunlara değer katmak için tasarlanan bağımsız, objektif güvence ve danışmanlık hizmeti sağlayan birim, bölüm, danışman ekibi veya diğer uygulamacılardır. İç denetim faaliyeti, kurumun risk yönetimi, kontrol ve yönetim süreçlerinin etkinliğini değerlendirmek ve geliştirmek amacıyla yönelik

sistemli ve disiplinli bir yaklaşım getirerek kurumun amaçlarına ulaşmasına yardımcı olur.” şeklinde tanımlanmaktadır (“İç denetim Terimleri Sözlüğü” 2011, 2)

İç denetimin işletmelerdeki fonksiyonu, iç kontrol sisteminin yeterlilik ve etkinliğini değerlendirme yolu ile yönetime hizmet etmektir. İç denetim, faaliyet sonuçlarını prosedürlerde belirtilen amaçlar ile karşılaştırır, farklılıkları analiz eder ve düzeltici önlemler alınması gerekiyor ise alınması için önerilerde bulunur (Güney ve Sarı 2016, 6).

#### **1.4. Denetçi Kavramı ve Denetçi Türleri**

Bir süreç olarak denetim, amaca ulaşmada çeşitli unsurlara sahiptir. Bu unsurlardan birisi denetim faaliyetini yürüten denetçidir. Denetçi, denetim faaliyetlerini yürüten, yeterli mesleki bilgi ve tecrübeye sahip, bağımsız davranabilen ve yüksek ahlaki nitelikler taşıyan kişilerdir (Güredin 2014, 19). Denetim faaliyetleri yürüten denetçiler, iç denetçi, bağımsız denetçi ve kamu denetçisi olmak üzere üç grupta tanımlanmaktadır.

##### **1.4.1. İç Denetçi**

İşletme varlıklarının yeterince korunup korunmadığını araştırmak, muhasebe bilgi ve belgelerinin güvenilirliğini belirlemek, işletme üst yönetimi tarafından belirlenen politika ve prosedürlere uyulup uyulmadığı gözlemlenmek, işletmede yürütülen faaliyetlerin etkinliği hakkında görüş bildirmek, işletme üst yönetiminin talepleri doğrultusunda özel denetimler yürütmek ve bağımsız denetçilere, bağımsız denetim faaliyetlerinde destek vermek iç denetçilerin başlıca görevleri arasındadır (Bozkurt 2018, 39).

İç denetçiler, işletmelerin sürekli personelleridir ve işletme içerisindeki denetim faaliyetlerini yürütmekle sorumlu olan kişilerdir. İç denetçiler, denetim faaliyetlerini işletme yönetiminin talep ve istekleri doğrultusunda yürütürler. İç denetçilerin başlıca görevleri muhasebe bilgi sisteminde oluşabilecek hata ve/veya hileden kaynaklı düzensizlikleri tespit ederek işletme üst yönetimine raporlamak ve faaliyet denetimlerini gerçekleştirmektir. Son yıllarda meydana gelen gelişmeler iç denetçileri, işletmelerde ortaya çıkabilecek riskleri değerlemede önemli bir pozisyona getirmiştir.(Ü. Ataman, Hacirüstemoğlu, ve Bozkurt 2001, 21).

##### **1.4.2. Bağımsız Denetçi**

Bağımsız Denetçi, Kamu Gözetimi ve Denetim Standartları tarafından yayınlanmış olan bağımsız denetim yönetmeliğinde “Bağımsız denetim yapmak üzere, 1/6/1989 tarihli ve 3568 sayılı Serbest Muhasebeci Mali Müşavirlik ve Yeminli Mali Müşavirlik Kanununa

göre yeminli mali müşavirlik ya da serbest muhasebeci mali müşavirlik ruhsatını almış meslek mensupları arasından kurum tarafından yetkilendirilen kişiler” şeklinde tanımlanmaktadır (Kamu Gözetim Kurumu 2012, 1)

Bağımsız denetçiler, denetimi yaptıkları işletme ile herhangi bir işçi-işveren ilişkisi bulunmayan ve işletmelere denetim ve diğer hizmetleri sunan kişilerdir. Bağımsız denetçilerin temel görevleri, işletmelerin mali tablolarının bağımsız denetimlerini yürütmektir. İşletme yönetimi ile işletme ilgilileri arasında bir köprü görevi görmektedirler (Ü. Ataman, Hacırüstemoğlu, ve Bozkurt 2001, 21).

### **1.4.3. Kamu Denetçisi**

Kamu denetçileri, çeşitli kamu kurumlarının denetim birimlerinde görev yapan, kamu adına denetim yapan ve denetim raporunu bağlı buldukları kamu kurumuna sunan denetçilerdir (Cömert vd. 2013, 1:17).

### **1.5. Genel Kabul Görmüş Denetim Standartları**

Genel kabul görmüş denetim standartları, denetim faaliyetlerinin kalitesini ve verimliliğini belirli bir çerçeveye oturtmak ve denetim mekanizmasının günün şartlarına uygun olarak kendini yenilemesi sağlamak ve denetimin genel çerçevesini ortaya koymak maksadı ile 1947 yılında Amerikan Diplomalı Kamu Muhasebecileri Enstitüsü (AICPA) tarafından hazırlanmıştır. Bu standartlar 1972 yılından sonra Denetim standartları komitesi (Auditing Standards of Boards) tarafından gözden geçirilmiş ve SAS (Statement on Auditing Standards) Denetim Standartları Açıklamaları olarak yayınlanmaya başlanmıştır. Denetim standartları komitesi tarafından 1972 yılında yayınlanan standartlar ilk standartların yorumu niteliğindedir ve çok az değişikliğe uğrayarak günümüze kadar gelmiştir. (Dönmez, Berberoğlu, ve Ersoy 2005, 54).

Genel kabul görmüş denetim standartları, bir denetçide bulunması gereken mesleki nitelikleri, denetim sürecinde uygulanması gerekenleri denetim raporunun yazımı aşamasında uyulması gereken kuralları genel hatları ile belirlemektedir (Kavut 2000, 9). Denetçi, genel kabul görmüş denetim standartlarını kullanarak, denetlenen işletme ile ilgili bir görüş oluşturmak için yeterli ve güvenilir kanıt toplamaktadır. Denetçi ihtiyaç duyduğu bu bilgileri, muhasebe sistemini inceleyerek, sisteminin nasıl işlediğini araştırarak, ilgili belgeleri inceleyerek, işletme varlıklarını yerinde görerek, işletme çalışanlarını gözlemleyerek, belirli rakamları tekrar hesaplayarak ve diğer çeşitli yolları

kullanarak elde etmektedir. Denetçinin farklı yollar ile elde ettiği bu bilgilere denetim kanıtı adı verilmektedir (Güredin 2014, 41).

Denetim amaçlarının gerçekleştirilmesi ve ilgili tarafların beklentilerinin karşılanabilmesi için denetimin genel kabul görmüş denetim standartlarına uygun bir şekilde planlanması, yürütülmesi ve sonuçlarının raporlanması gerekmektedir (Kayrak 2015, 91). Genel kabul görmüş denetim standartları Genel Standartlar, Çalışma Alanı Standartları ve Raporlama Standartları olmak üzere üç başlık altında on standarttan oluşmaktadır.

### **1.5.1. Genel Standartlar**

Genel standartlar denetçinin taşıması gereken özelliklerin neler olduğu ile ilgilidir. Genel standartlar ışığında, denetim faaliyetinin yeterli mesleki bilgi ve tecrübeye sahip denetçi adı verilen kişi veya kişilerce yapılması gerektiği, denetçinin veya denetçilerin denetim sürecinin her aşamasında bağımsız davranması gerektiği ve denetçinin denetim faaliyetinin her aşamasında gerekli mesleki özeni göstermesi gerektiği anlaşılmaktadır(Bozkurt 2018, 42)

#### **1.5.1.1. Mesleki Eğitim ve Deneyim**

Denetçilerin, mesleki yeterliliğe ulaşmış olması için muhasebe ve denetim konularını içeren teknik bilgi ve deneyime sahip olmaları gerekmektedir. Denetim faaliyeti sonunda hazırlanan denetim raporundan faydalanan ilgili taraflar, denetim görüşü veren ve raporu hazırlayan denetçiyi tanımamaktadırlar, ilgili taraflar için önemli olan kaliteli ve güvenilir denetim görüşüne ulaşmaktır. Bu sebeple kalite bir denetim görüşüne ulaşmak için, denetçilerin iyi bir mesleki eğitime ve mesleki tecrübeye sahip olması gerekmektedir (Dönmez, Berberoğlu, ve Ersoy 2005, 54; Başpınar 2005, 55).

Mesleki eğitim ve tecrübe genel standartlar kapsamında zorunlu olarak aranmaktadır. Denetim faaliyetini yürütecek olan kişi ve kişilerin sadece muhasebe ve işletmecilik alanında bilgi sahibi olması yeterli değildir. Denetçilerin, kamusal bir işlevi yerine getirdikleri için aynı zamanda genel bilgi düzeyi bakımından da iyi bir durumda olmaları ve iletişim becerilerinin de iyi olması gerekmektedir. Denetçilerin ayrıca sahip oldukları mesleki bilgiyi güncel tutmaları da gerekmektedir. Denetim, denetim standartlarının güncellenmesi ve değiştirilmesi, bilişim teknolojilerindeki yenilikler, yeni denetim tekniklerinin geliştirilmesi, yönetim anlayışının değişmesi ve yeni yasal düzenlemelerin

yürürlüğe konulması gibi gerekçeler nedeniyle sürekli eğitimin gerekli olduğu bir disiplindir(Güredin 2014, 43-44; Kayrak 2015, 100).

### **1.5.1.2. Bağımsızlık**

Denetçinin, denetim faaliyetlerinin yürütülmesinde ve bulguların değerlendirilip denetim raporunun oluşturulmasında bağımsız bir bakış açısına sahip olması gerekmektedir (Bozkurt 2018, 43). Denetçi tarafından incelenen finansal tabloların güvenilirliği denetçinin bağımsızlığı ile doğru orantıya sahiptir. Denetçi, bağımsızlık standardına önem verdikçe, yürütülen denetim çalışmasının da güvenilirliği artacaktır (Dönmez, Berberoğlu, ve Ersoy 2005, 56; Kayrak 2015, 96; Tuan 2015, 295).

Bir denetçinin bağımsızlığı iki aşamada sağlanabilir. Bunlar, karakter yapısı ve şekil özellikleridir. Bağımsızlık sayılan iki unsur açısından da sağlanmalıdır (Bozkurt 2018, 44).

Denetçinin karakter olarak bağımsız olması, onun mesleğini icra ederken bağımsız davranması gerektiği gerçeğinin farkında olduğu anlamına gelmektedir. Ancak bazı durumlarda denetim faaliyetinin şeklinden dolayı, denetçi tam bağımsızlık sahibi olamamaktadır. Aşağıdaki durumlar şekil açısından denetçinin bağımsızlığını bozan durumlara örnek verilebilir (Bozkurt 2018, 44);

- Denetçinin, denetlenen işletmenin ortağı olması,
- Denetçi ile denetlenen işletmenin ortakları veya yöneticileri arasında akrabalık bağı bulunması,
- Denetçi ile denetlenen işletme arasında iş ilişkisinin bulunması,
- Denetçinin, denetlenen işletme ortakları ile borç ilişkisi içerisinde olması,
- Denetçinin, denetlenen işletmeye aynı zamanda muhasebe hizmeti veriyor olması.

### **1.5.1.3. Gerekli Mesleki Özen**

Genel kabul görmüş denetim standartları, denetçinin mesleki eğitimi ve tecrübe ile bağımsızlığı ile ilgili yargıların yanında denetçinin denetim faaliyetinin yürütülmesi aşamasında gerekli mesleki özeni de göstermesi gerekliliğinin ifade etmektedir. Denetçi, denetim faaliyetinin tümünde mesleki sorumluluklarını yerine getirmelidir. Denetim kanıtlarının toplamasında ve değerlendirilmesinde, çalışma kağıtlarının düzenlenmesinde ve denetim raporunun hazırlanmasında yeterli özeni göstermesi gerekmektedir (Bozkurt 2018, 44). Gerekli mesleki özen ve titizliğin gösterilmesi, denetçinin tüm denetim standartlarına eksiksiz bir biçimde uyması ile sağlanmaktadır (Başpınar 2005, 56).

Denetçinin denetim faaliyeti süresince, gerekli mesleki özeni göstermemesi denetim kalitesini ve güvenilirliğini olumsuz yönde etkileyecektir (Dönmez, Berberoğlu, ve Ersoy 2005, 56).

### **1.5.2. Çalışma Alanı Standartları**

Çalışma alanı standartları, denetim faaliyetinin başladığı andan itibaren en azından nelerin yapılması gerektiği ile ilgili açıklamalar getirmektedir. Çalışma alanı ile ilgili standartlar denetimin planlanması ve gözetim, iç kontrol yapısının incelenmesi ve değerlendirilmesi ve yeterli sayıda ve kalitede kanıt toplama olmak üzere üç adettir (Bozkurt 2018, 44-46)

#### **1.5.2.1. Denetimin Planlanması ve Gözetim**

Denetimin planlanması aşama, denetçinin denetimi yürütmek için genel bir denetim stratejisi geliştirirken göz önünde bulunması gereken tüm konuları içermektedir. Temel olarak denetim planı, denetimin verimli bir şekilde nasıl yürütüleceğine odaklanmaktadır (Messier, Glover, ve Prawitt 2016, 78). Denetim çalışmasında, planlama yapılırken işgücü planlaması, zaman planlaması ve kaynak planlamasının yapılması gerekmektedir (Başpınar 2005, 56).

Denetim faaliyetinin planlamasında temel olarak üç amaç bulunmaktadır;

- Yeterli sayıda ve kalitede kanıt elde edebilmek,
- Denetim maliyetlerini mantıklı bir düzeyde tutabilmek,
- Denetlenen işletme ile ilgili yanlış anlaşılmalara önlemek.

Belirtilen amaçları gerçekleştirmek için denetim planında, denetim faaliyeti süresince uygulanacak denetim politikası ve prosedürleri, denetim programları, denetim ekibi, denetim çalışmasının zaman çizelgesi ve denetim maliyetleri, iç denetçiden ne şekilde yararlanılacağı ve gerekmesi durumunda uzman görüşüne başvurulacak olan kişiler ile ilgili bilgilerin bulunması gerekmektedir (Bozkurt 2018, 45).

Denetçi, etkili bir denetim planlaması yapabilmek için işletmenin organizasyon yapısı, konumu, ürettiği mal ve hizmetler, finansal yapısı, tedarikçileri, müşterileri ve ilişki içerisinde olduğu tüm taraflar ile ilgili gerekli bilgilere ihtiyaç duymaktadır. Denetçi anılan konularda bilgi sahibi olmak suretiyle, genel kabul görmüş denetim standartlarının öngördüğü kalitede bir denetim planı hazırlayabilecektir (Başpınar 2005, 56; Güredin 2014, 45).

Denetim planının hazırlanması için gerekli olan bilgiler önceki dönem çalışma kağıtlarından, işletmenin faaliyet gösterdiği sektör hakkında hazırlanmış her türlü yayından, denetlenen işletmeye yapılan ziyaret ve görüşmelerden sağlanmaktadır. Denetlenen işletmenin iç kontrol sistemi ile ilgili elde edilen bilgiler de denetim planının hazırlanması aşamasında önemli rol oynamaktadır.

#### **1.5.2.2. İç Kontrol Sisteminin İncelenmesi ve Değerlendirilmesi**

Çalışma alanı standartlarından ikincisi, iç kontrol sisteminin incelenmesi ve değerlendirilmesidir. İç kontrol, işletmenin faaliyetlerini geliştirmeye ve değer katmaya yönelik geliştirilen politikalar ve uygulamalardır (Göçen 2010, 110). İç kontrol sisteminin etkinliği, mali nitelikli olaylar hakkındaki bilgi üzerinde düzenleyici bir etkiye sahiptir. Ayrıca mali tablolardaki düzensizlikleri de azaltmaktadır. Bu durum denetim riskinin azalmasını sağlayacaktır. Denetim riskinin azalması neticesinde, denetçinin, denetim görüşü oluşturmak için ihtiyaç duyduğu kanıt miktarı ve uygulaması gereken denetim prosedürlerinin kapsamı ve sayısı azalmaktadır (Dönmez, Berberoğlu, ve Ersoy 2005, 56). İç kontrol sisteminin etkinliği ile ilgili, denetçinin oluşturacağı yargı, denetim çalışmalarını yönlendiren en önemli unsurdur (Bozkurt 2018, 45).

Denetçi, iç kontrol sistemini değerlendirme noktasında ilk olarak kontrol ortamı, muhasebe sistemi ve kontrol yordamları ile ilgili genel bilgileri elde eder. Bu bağlamda denetçi işletme çalışanlarını gözlemleyerek, soruşturma yaparak, anket uygulayarak, belgeleri ve kayıtları inceleyerek gerekli bilgileri toplar. Elde edilen bilgilerin analizi yolu ile denetimin ileriki aşamaları ile ilgili kararlar alınmaktadır. Denetçi işletmenin iç kontrol sistemini, denetimin amacına uygun olarak ele alıp değerlendirmelidir. Denetlenen işletmenin büyüklüğü, faaliyet gösterdiği sektör, organizasyon yapısı ve işletme içerisinde kullanılan bilgi sistemi gibi faktörleri göz önüne almalıdır (Tuan 2009, 8-9).

Denetçi tarafından, iç kontrol sisteminin incelenmesinde iki amaç bulunmaktadır (Başpınar 2005, 56);

1. İç kontrol sisteminin etkinliğinin saptaması, böylece denetçi muhasebe sistemi içerisine alınmış olan olayların eksiksiz bir şekilde sisteme aktarıldığı ve düzenlenmiş olan finansal tabloların da bu olayları eksiksiz bir şekilde yansıttığı ile ilgili görüş oluşturulma imkanına sahip olacaktır,

2. Denetim görüşü oluşturmak için uygulanacak olan denetim prosedürlerinin kapsamının ve ayrıntı derecesinin belirlemesi.

### **1.5.2.3. Yeterli Sayı ve Kalitede Kanıt Toplama**

Çalışma alanı standartlarının üçüncüsü, denetçinin finansal tablolarla ilgili görüş oluşturabilmesi için yeterli sayıda ve kalitede denetim kanıtı toplaması gerektiğini belirtmektedir (Bozkurt 2018, 46). Denetim sürecinde toplanacak olan kanıt miktarı denetçinin mesleki yargısına bırakılmıştır (Başpınar 2005, 57).

Başarılı bir denetim çalışması, denetim kalitesi ile doğrudan ilişkilidir ve yeterli, uygun ve objektif denetim kanıtlarına dayanılarak oluşturulmuş denetim görüşünü içeren bir denetimi ifade eder (Kayrak 2015, 98).

### **1.5.3. Raporlama Standartları**

Raporlama standartları, denetçinin denetim çalışması sonucunda hazırladığı denetim raporunda bulunması gereken özellikleri belirtmektedir. Denetim raporunda, denetçinin işletmenin finansal tablolarının genel kabul görmüş muhasebe ilkelerine uygun olarak hazırlanıp hazırlanmadığı, denetimin gerçekleştiği dönemde uygulanan muhasebe politikalarının geçmiş dönemlerde de aynen uygulanıp uygulanmadığını hakkında görüş bildirmelidir. Denetçi raporunda aksi belirtilmediği sürece mali tabloların sahip olduğu açıklayıcı bilgilerin yeterli olduğu kabul edilmektedir. Denetçinin mali tablolar hakkında bir görüş bildirmesi yada görüşe ulaşamamış ise bir görüşe ulaşamadığını bildirmesi gerekmektedir (Bozkurt 2018, 43).

#### **1.5.3.1. Genel Kabul Görmüş Muhasebe İlkelerine Uygunluk**

Raporlama standartlarının ilki, denetçinin işletmenin finansal tablolarının genel kabul görmüş muhasebe ilkelerine göre hazırlanmış olduğu ile ilgili görüş bildirmesini belirtmektedir. İşletmenin mali tablolarının sağlıklı olup olmadığının belirlenmesinde genel kabul görmüş muhasebe ilkeleri çok önemli bir konumdadır (Bozkurt 2018, 46). İşletmenin mali tablolarını hazırlarken mutlaka genel kabul görmüş muhasebe ilkelerine uygun davranması gerekmektedir. Genel kabul görmüş muhasebe ilkelerine uygunluk mali tabloların içerdiği temel iddiadır. Denetçinin de bu iddiaya uygunluğu araştırması ve görüşünü de bu ölçüte göre oluşturması gerekmektedir (Melih Erdoğan vd. 2012, 2:13).

### **1.5.3.2. Devamlılık**

Denetçi denetim raporunda, işletmelerin uyguladıkları muhasebe politikalarında farklı yıllar arasında karşılaştırma yapmayı önleyici değişiklikler bulunması durumunda, bu durumu belirtmeleri gerekmektedir (Bozkurt 2018, 47). Devamlılık ilkesi finansal tabloların karşılaştırılabilir olmasına olanak sağlamaktadır ve finansal tabloların karşılaştırılabilir olması için, hazırlanırken uygulanan muhasebe politikalarının değişmemiş olması gerekmektedir. Ancak işletmelerin uyguladıkları muhasebe politikalarını değiştirmeleri doğal bir durumdur. İşletme yönetiminin muhasebe politikalarında değişiklik yapması durumunda, değişikliğin ilgili dönemin finansal tablolarına yapacağı etki finansal tablo dipnotlarında ayrıntılı şekilde açıklanmalıdır(Güredin 2014, 51).

### **1.5.3.3. Açıklayıcı Bilgiler**

Raporlama standartlarının üçüncüsü, denetçinin, finansal tablo kullanıcılarına finansal tablolar aracılığı ile ulaşmaları gereken tüm bilgilerin işletme tarafından sağlanıp sağlanmadığını belirlemesi gerekliliği belirtilmektedir (Başpınar 2005, 57-58). Denetçi finansal tablolar ile birlikte tablolarda yer alan açıklayıcı bilgi niteliği taşıyan dipnotları da incelemelidir. Denetçi finansal tablolardaki dipnotların yeterli olmadığına kanaat getirirse bu durumu denetim raporunda belirtmelidir. Denetçinin finansal tablolarda bu konu hakkında görüş bildirmemesi durumunda finansal tablolardaki dipnotların yeterli olduğu kabul edilir.

### **1.5.3.4. Denetçi Görüşünün Beyanı**

Yürütülen denetim çalışmasının sonucu denetim raporunda açıklamaktadır. Raporlama standartlarının dördüncüsü, finansal tabloların bir bütün olarak ele alınarak bir denetim görüşü oluşturulmasını ve denetçinin bu görüşünü denetim raporunu aracılığı ile bildirmesinin, ancak bir denetim görüşü oluşturamaması durumunda denetçinin bu durumu nedenleri ile açıklamasını öngörmektedir. Bu standart hem kamu kesimi hem de özel kesim denetimlerinde uyulması gereken bir standarttır (Güredin 2014, 52).

Genel kabul görmüş denetim standartları denetçiye dört farklı denetim görüşü sunmaktadır. Bunlar, olumlu görüş, olumsuz görüş, şartlı görüş ve görüş bildirmekten kaçınmadır (Bozkurt 2018, 47).

## 1.6. Uluslararası Denetim Standartları

Standart, belirli bir ya da birden fazla kritere uygun nitelik veya karşılaştırma ölçüsü olarak tanımlanmaktadır (Erol ve Aslan 2017, 58). Günümüzde pek çok farklı konuda, standart kavramı karşımıza çıkmaktadır. Denetim uygulamalarında da standartlar söz konusudur (Türker vd. 2003, 11). Denetim standartları denetçiler tarafından yürütülen denetimin kalitesinin sağlanması ile ilgili genel ilke ve kurallardır. Uluslararası denetim standartları, Uluslararası Muhasebeciler Federasyonu bünyesinde oluşturulan Uluslararası Denetim ve Güvence Standartları kurulu tarafından yayınlanmaktadır. Uluslararası denetim standartları 15 Aralık 2009 tarihinden sonra başlayan dönemlerde kullanılmak üzere yayınlamıştır (Kavut, Taş, ve Şavlı 2009, 75,76). Uluslararası Denetim ve Güvence Standartları kurulu tarafından yayınlanan standartlar, Türkiye’ de ulusal standartların oluşturulmasında örnek alınmıştır.

Türkiye’ de bağımsız denetim uygulamaları yabancı sermayeli işletmelerin bilgi ihtiyacının karşılanması maksadı ile 1970’ li yıllarda başlamıştır. Denetim uygulamaları bağlamında yasal altyapının oluşturulması ise 1987 yılında Sermaye Piyasası Bağımsız Denetim Yönetmeliği ile başlamıştır. 3568 sayılı yasa ile Türkiye’ de meslek mensuplarının çalışmalarında bağlı kalacakları denetim standartlarının oluşturulması hususunda TÜRMOB görevlendirilmiştir. Bu bağlamda 20 Şubat 2003 tarihinde Türkiye Denetim Standartları kurulu kurulmuştur. Kurulun amacı 3568 sayılı yasaya tabi olan meslek mensuplarının denetim faaliyetlerini disiplinli bir şekilde yürütebilmeleri için Türkiye (Ulusal) Denetim Standartları’ nın saptanması, belirlenmesi ve yayınlanması ile bu standartların güncelliğinin korunmasıdır. 2011 yılında yürürlüğe giren yeni Türk Ticaret Kanunu’nu takiben çıkarılan Bağımsız Denetim Yönetmeliği ve Kamu Gözetim Kurumu düzenlemeleri temel alınarak denetim faaliyetleri yürütülmektedir. Denetim faaliyetlerinde düzenleme yetkisi Kamu Gözetim Kurumuna verilerek uygulamada bütünlük sağlanması hedeflenmiştir (Ertaş ve Kandemir 2020, 84-85; Türker 2006, 93-94).

6102 sayılı Türk Ticaret Kanununun 397 maddesi uyarınca denetime tabi olan işletmelerin mali tablolarının denetiminde uyulacak olan standartlar, 660 nolu kanun hükmünde kararname ile oluşturulan Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumu tarafından hazırlanmakta ve yayınlanmaktadır. Ticaret kanununun aynı maddesi uyarınca, Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumu uluslararası standartlar ile birebir uyumlu standartlar hazırlamak ile görevlidir (*Türk Ticaret Kanunu* 2011, c. 6102,

md. 397; Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumunun Teşkilat ve Görevleri Hakkında Kanun Hükmünde Kararname 2011, c. 660, md. 9).

**Tablo 1.1. Uluslararası Bağımsız Denetim Standartları**

<b>BDS 200-299 Genel İlke ve Sorumluluklar</b>	
BDS 200	Bağımsız Denetçinin Genel Amaçları ve Bağımsız Denetimin Bağımsız Denetim Standartlarına Uygun Olarak Yürütülmesi
BDS 210	Bağımsız Denetim Sözleşmesinin Şartları Üzerinde Anlaşmaya Varılması
BDS 220	Finansal Tabloların Bağımsız Denetiminde Kalite Kontrol
BDS 230	Bağımsız Denetimin Belgelendirilmesi
BDS 240	Finansal Tabloların Bağımsız Denetiminde Bağımsız Denetçinin Hileye İlişkin Sorumlulukları
BDS 250	Finansal Tabloların Bağımsız Denetiminde Mevzuatın Dikkate Alınması
BDS 260	Üst Yönetimden Sorumlu Olanlarla Kurulacak İletişim
BDS 265	İç Kontrol Eksikliklerinin Üst Yönetimden Sorumlu Olanlara ve Yönetime Bildirilmesi
<b>BDS 300-499 Risk Değerlendirmesi ve Değerlendirilmiş Risklere Verilecek Karşılıklar</b>	
BDS 300	Finansal Tabloların Bağımsız Denetiminin Planlanması
BDS 315	“Önemli Yanlılık” Risklerinin Belirlenmesi ve Değerlendirilmesi
BDS 320	Bağımsız Denetimin Planlanması ve Yürütülmesinde Önemlilik
BDS 330	Bağımsız Denetçinin Risk Olarak Değerlendirilmiş Hususlara Karşı Yapacağı İşler
BDS 402	Hizmet Kuruluşu Kullanan Bir İşletmenin Bağımsız Denetiminde Dikkate Alınacak Hususlar
BDS 450	Bağımsız Denetimin Yürütülmesi Sırasında Belirlenen Yanlılıkların Değerlendirilmesi
<b>BDS 500-599 Bağımsız Denetim Kanıtları</b>	
BDS 500	Bağımsız Denetim Kanıtları
BDS 501	Bağımsız Denetim Kanıtları – Belirli Kalemler İçin Dikkate Alınması Gereken Özel Hususlar
BDS 505	Dış Teyitler
BDS 510	İlk Bağımsız Denetimler – Açılış Bakiyeleri
BDS 520	Analitik Prosedürler
BDS 530	Bağımsız Denetimde Örnekleme
BDS 540	Muhasebe Tahminlerinin ve İlgili Açıklamaların Bağımsız Denetimi
BDS 550	İlişkili Taraflar
BDS 560	Bilanço Tarihinden Sonraki Olaylar
BDS 570	İşletmenin Sürekliliği
BDS 580	Yazılı Beyanlar
<b>BDS 600-699 Başkalarının Çalışmalarının Kullanılması</b>	
BDS 600	Özel Hususlar–Topluluk Finansal Tablolarının Bağımsız Denetimi (Topluluğa Bağlı Birim Denetçilerinin Çalışmaları Dâhil)
BDS 610	İç Denetçi Çalışmalarının Kullanılması
BDS 620	Uzman Çalışmalarının Kullanılması
<b>BDS 700-799 Bağımsız Denetim Sonuçları ve Raporlama</b>	
BDS 700	Finansal Tablolara İlişkin Görüş Oluşturma ve Raporlama
BDS 701	Kilit Denetim Konularının Bağımsız Denetçi Raporunda Bildirilmesi
BDS 705	Bağımsız Denetçi Raporunda Olumlu Görüş Dışında Bir Görüş Verilmesi
BDS 706	Bağımsız Denetçi Raporunda Yer Alan Dikkat Çekilen Hususlar ve Diğer Hususlar Paragrafları
BDS 710	Karşılaştırmalı Bilgiler-Önceki Dönemlere Ait Karşılık Gelen Bilgiler ve Karşılaştırmalı Finansal Tablolar
BDS 720	Bağımsız Denetçinin Diğer Bilgilere İlişkin Sorumlulukları
<b>BDS 800-899 Özel Alanlar</b>	
BDS 800	Özel Hususlar – Özel Amaçlı Çerçevelelere Göre Hazırlanan Finansal Tabloların Bağımsız Denetimi
BDS 805	Özel Hususlar – Tek Bir Finansal Tablonun Bağımsız Denetimi ile Finansal Tablolardaki Belirli Unsurların, Hesapların veya Kalemlerin Bağımsız Denetimi
BDS 810	Özet Finansal Tablolara İlişkin Rapor Vermek Üzere Yapılan Denetimler

### **1.7. Bağımsız Denetim Standardı 570 – İşletmenin Sürekliliği Standardı**

Son yıllarda yaşanan küresel krizler ve işletme başarısızlıklarının artması, işletmenin sürekliliği varsayımının gerek işletme yönetimi gerekse denetçi tarafından dikkatlice değerlendirilmesi gerekliliğini ortaya koymuştur. İşletmenin sürekliliği kavramı muhasebenin temel kavramlarından birisidir ve işletmenin faaliyetlerini öngörülebilir bir gelecekte devam ettirebileceği varsayımına dayanmaktadır. İşletme ile ilgili tüm çıkar grupları işletmenin mali durumu ile ilgili kararlarını işletmenin sürekliliği varsayımı altında vermektedir.

İşletmelerin kuruldukları andan itibaren sınırsız bir ömre sahip oldukları kabul edilmektedir. Ancak işletmelerin varlığını sürdürebilmesi faaliyetlerinin devamlılığına, rasyonel yatırımların doğru zamanda yapılmasına, rekabet edebilirlik düzeyindeki başarı yeteneğine ve müşteri istek ve ihtiyaçlarını en iyi şekilde karşılayabilme potansiyeline bağlıdır. Bu anlamda özellikle işletme sahip, ortak ve yöneticilerinin işletme performansını ölçerken geleneksel finansal ölçütlerin yanı sıra işletmenin sürekliliği hakkında bilgi veren birtakım ölçütleri de dikkate alarak hem finansal performansı hem de süreklilik performansını ölçmeleri daha doğru kararların alınmasını mümkün kılacaktır. İşletmenin sürekliliği üzerinde tehlike oluşturabilecek risk faktörlerinin önceden belirlenmesi işletmenin ani iflas riskleri ile karşılaşmasının da önüne geçmiş olacaktır. Günümüzün rekabet ortamında son derece önemli bir unsur olarak öne çıkan işletmenin sürekliliği olgusu işletmeye borç verenler, tedarikçiler, yatırımcılar ve bağımsız denetçiler açısından dikkate alınmaktadır. Özellikle bağımsız denetim açısından işletmenin süreklilik yeteneğinin olup olmadığının denetiminin bağımsız denetçilerin sorumluluğunda olması denetim görüşünün oluşturulmasında önemlidir.

İşletmelerin sahip olduğu varlıklar ve kaynaklar farklı yöntemlere göre değerlendirilmektedir. Bu kuralların uygulanması için ise işletme ömrünün sınırlı olmaması gerekmektedir. Bu sebeple muhasebeleştirme faaliyetleri yürütülürken, işletmenin sürekli olduğu varsayılmaktadır. İşletmenin sürekli olması muhasebe politikalarının devamlılığı açısından çok önemlidir. Aksi takdirde karşılaştırmalı mali tablo çıkarma vb. amaçlar ortadan kalkacaktır (Kaval vd. 2015, 353).

İşletmenin sürekliliği standardının kapsamı BDS-570 İşletmenin Sürekliliği Standardında *“Bu Bağımsız Denetim Standardı (BDS), işletmenin sürekliliğine ve işletmenin sürekliliğinin denetçi raporuna etkilerine ilişkin denetçinin finansal tabloların denetimindeki sorumluluklarını düzenler.”* şeklinde tanımlanmaktadır. (Kamu Gözetim Kurumu 2020, par. 1).

Süreklilik esası ve işletmenin sürekliliğini devam ettirme yeteneğinin değerlendirilmesine ilişkin işletme yönetiminin ve bağımsız denetçinin sorumlulukları, BDS 570-İşletmenin Sürekliliği Standardı kapsamında düzenlenmektedir. Standart kapsamında, mali tablolarda işletmenin sürekliliğini devam ettirme yeteneği hususunda önemli bir belirsizliğe yönelik yeterli açıklamaların yapılması ya da yapılmaması durumunda belirsizliklerin etkisine göre nasıl denetim görüşü vermesinin uygun olacağı açıklanmaktadır. Denetçi, denetim raporunda işletmenin sürekliliği ile ilgili yaptığı

açıklamaları “Dikkat Çeken Hususlar” başlığı altında sunar. İşletmenin sürekliliğine ilişkin belirsizlik vurgusu yapılan olay veya şartın ne olduğu ve bu tür durumların hangi tür denetim görüşleri ile açıklandığı bilgi kullanıcılarının kararlarında etkili olmaktadır (Yılmaz ve Yaşar 2018, 57).

### **1.7.1. İşletmenin Sürekliliği Esası**

Mali tabloların hazırlanmasında kullanılan temel varsayımlardan bir tanesi işletmenin sürekliliği ilkesidir (Bilir 2015, 12). İşletmenin sürekliliği esası, işletmenin kuruluş sözleşmesinde aksine bir kayıt bulunmadıkça, işletmenin sonsuz bir ömre sahip olduğunu ifade etmektedir. Süreklilik esasının göz ardı edilmesi durumunda gelir ve giderlerin karşılaştırılarak karın belirlenmesi anlayışı, işletme varlıklarının tasfiye değeri ile değerleyip, edinme değeri ile karşılaştırarak karı belirleme yaklaşımına yerini bırakır (Sevilengül 2016, 19-20).

BDS-570 İşletmenin Sürekliliği Standardı işletmenin sürekliliği esasını “*İşletmenin sürekliliği esası uyarınca, finansal tablolar, işletmenin sınırsız bir ömre sahip olduğu ve öngörülebilir gelecekte faaliyetlerini sürdüreceği varsayımıyla hazırlanır. Yönetim, işletmeyi tasfiye etme ya da ticari faaliyeti sona erdirmeye niyeti ya da mecburiyeti bulunmadığı sürece işletmenin sürekliliği esasını kullanmaktan sorumludur.*” şeklinde tanımlanmaktadır. (Kamu Gözetim Kurumu 2020, par. 2). BDS-570 İşletmenin Sürekliliği Standardında yer alan tanıma göre mali tablolar işletmenin sınırsız bir ömre sahip olduğu ve öngörülebilir bir gelecekte faaliyetlerini sürdüreceği varsayımına göre hazırlanmaktadır.

“*Yönetimin, raporlama döneminden sonra, işletmeyi tasfiye etme veya işletmenin ticari faaliyetine son verme niyetinin olduğuna ya da bunları yapmak dışında gerçekçi bir alternatifinin bulunmadığına karar vermesi durumunda işletme, finansal tablolarını işletmenin sürekliliği esasına göre hazırlamaz.*” (Kamu Gözetim Kurumu 2018b, par. 14). İşletme yönetimi, işletmeyi tasfiye etme ya da işletmenin faaliyetlerine son verme niyetine sahip değil ise ya da bu şartlar bir zorunluluktan kaynaklanmıyor ise, işletmenin mali durumunu raporlarken işletmenin sürekliliği esasına göre raporlama yapmak ile sorumlu tutulmuştur.

TMS 10 - Raporlama Döneminden Sonraki Olaylar standardına göre, İşletme yönetimi normal iş akışı içerisinde süreklilik esasına bağlı kalmak ile sorumludur. Ancak tasfiye

ya da işletme faaliyetlerine son verme niyeti taşıması durumunda mali tabloların hazırlanması sürecinde süreklilik esası dikkate alınmamaktadır.

*“Faaliyet sonuçlarında ve finansal durumda raporlama döneminden sonra meydana gelen bozulmalar, işletmenin sürekliliği varsayımının geçerliliğini koruyup korumadığının gözden geçirilmesine ihtiyaç olduğunu gösterebilir. İşletmenin sürekliliği varsayımının geçerliliğini yitirmesi durumunda, ortaya çıkan etki oldukça yaygın olduğundan, bu Standart, muhasebenin temel varsayımlarına göre finansal tablolara alınan tutarlarda bir düzeltme yapmak yerine muhasebe esasında temel bir değişikliğe gitmeyi zorunlu kılar.”* (Kamu Gözetim Kurumu 2018b, par. 15).

İşletmenin faaliyet sonuçlarında ve mali durumunda meydana gelen bozulmalar, işletmenin sürekliliği varsayımının geçerliliğini koruyup korumadığının gözden geçirilmesine ihtiyaç duyulduğunu gösterebilmektedir. İşletmenin sürekliliği varsayımının geçerliliğini yitirmesi durumunda ortaya çıkan etki oldukça yaygın olduğundan, TMS 10 – Raporlama Döneminden Sonraki Olaylar standardı, muhasebenin temel varsayımlarına göre mali tablolara alınan tutarlarda bir düzeltme yapmak yerine muhasebe esasında temel bir değişiklik yapmayı zorunlu kılmaktadır.

### **1.7.2. İşletmenin Sürekliliği ile İlgili İşletme Yönetimi ve Bağımsız Denetçinin Sorumlulukları**

Bağımsız denetim raporlarının şeffaf ve hesap verilebilir şekilde hazırlanması raporun kalitesini artıran önemli faktörlerden birisidir. Denetçinin yürüttüğü denetim çalışmasının içeriğinin ve niteliğinin anlaşılması işletmenin mali tablolarının doğruluğu ve güvenilirliğinin sağlanması ise bilgi kullanıcılarının temel beklentisidir (Akdoğan 2020, 2).

#### **1.7.2.1. İşletme Yönetiminin Sorumluluğu**

İşletmenin sürekliliği varsayımının uygun olup olmadığının değerlendirilmesi sürecinde yönetim, bilanço tarihinden itibaren en az bir yıl içerisinde gerçekleşmesi muhtemel olayları yani geleceğe ilişkin tüm bilgileri dikkate alarak karar vermelidir. (Güngör, Ayluçtarhan, ve Sari 2022, 613). İşletmenin sürekliliği ile ilgili değerlendirme, işletme yönetimi tarafından her durumda yapılmak zorundadır (Kaval vd. 2015, 355).

*“Yönetim finansal tabloları hazırlarken, işletmenin sürekliliğini devam ettirebilme gücünü değerlendirir. Yönetimin işletmeyi tasfiye etme veya işletmenin ticari faaliyetine son verme niyeti bulunmadığı ya da bunları yapmak dışında gerçekçi bir alternatifi*

*olduğu sürece işletme, finansal tablolarını işletmenin sürekliliği esasına göre hazırlar. Yönetimin, değerlendirme yaparken, işletmenin sürekli olarak faaliyetlerini devam ettirebilmesini ciddi şekilde şüpheli hale getirebilecek olay veya koşullarla ilgili önemli belirsizliklerin farkına varması durumunda, bu belirsizlikler işletme tarafından açıklanır. Finansal tablolarını süreklilik esasına göre hazırlamaması durumunda işletme, bu durum ile birlikte finansal tablolarını hangi esasa göre hazırladığını ve neden işletmenin sürekliliği esasının geçerli olmadığını açıklar.”* (Kamu Gözetim Kurumu 2018a, par. 25). Bağımsız denetim standardı, işletme yönetiminin mali tabloları hazırlarken işletmenin sürekliliğini devam ettirme gücünü değerlendirmesi gerekliliği belirtilmektedir. İşletme yönetimi, işletmeyi tasfiye etme ya da işletmenin ticari faaliyetlerini durdurma niyeti bulunmadığı durumlarda veya bunlar dışında gerçekçi bir alternatifi olduğu sürece işletmenin mali tablolarını işletmenin sürekliliği esasına göre hazırlamak ile yükümlüdür. İşletme yönetimi süreklilik ile ilgili değerlendirme yaparken işletmenin faaliyetlerinin devam ettirilmesi noktasında önemli belirsizlikler oluşturacak olay veya koşullar tespit ederse bu belirsizlikleri açıklamak ile sorumludur. Ayrıca mali tabloların süreklilik esasına göre hazırlanmaması durumunda, mali tabloların hangi esasa göre hazırlandığı ve hangi nedenle işletmenin sürekliliği esasına göre hazırlanmadığını açıklamalıdır.

*“Yönetim, işletmenin sürekliliği varsayımının geçerli olup olmadığını değerlendirirken, bununla sınırlı olmamakla birlikte, raporlama döneminin sona erdiği tarihten itibaren en az on iki aylık dönemi kapsayacak şekilde gelecek hakkında elde edilebilen bütün bilgileri dikkate alır. Dikkate alınacak hususların kapsamı, her durumun şartlarına bağlıdır. Kârlı faaliyetler yürüttüğü ve finansal kaynaklara kolay ulaştığı bir geçmişinin olması durumunda işletme, ayrıntılı analizler yapmadan süreklilik esasının geçerli olduğu sonucuna ulaşabilir. Diğer durumlarda, işletmenin sürekliliği esasının geçerli olduğuna karar vermeden önce yönetimin, mevcut ve beklenen kârlılık, borç ödeme planları ve yenilenecek finansmanlara ilişkin potansiyel kaynaklarla ilgili pek çok faktörü dikkate alması gerekebilir.”* (TMS 1 - Finansal Tabloların Sunuluşu, 2018, par. 26).

İşletme yönetimi tarafından işletmenin sürekliliği varsayımının geçerli olup olmadığı değerlendirilirken, raporlama döneminin sona erdiği tarih itibarıyla on iki aylık dönemi içine alacak şekilde elde edilebilen tüm bilgiler dikkate alınmalıdır. Değerlendirme sürecinde dikkate alınacak olan hususların kapsamı, durumun şartlarına bağlıdır ve farklılık gösterebilmektedir. İşletmenin süreklilik varsayımı ile ilgili yürüteceği faaliyetlerin kapsamı mevcut durumu incelenerek belirlenebilir. Faaliyet karlılığı yüksek

olan ve finansal kaynaklara kolaylıkla erişebilen bir işletmenin süreklilik esası hakkında değerlendirilmesi için detay analizlere gerek duyulmayabilir. Ancak işletme yönetimi tarafından süreklilik esası hakkında bir karara varılmadan önce mevcut ve beklenen karlılık, borç ödeme planları ve yenilebilecek finansmanlara ilişkin potansiyel kaynaklar ile ilgili pek çok unsurun dikkate alınması gereken durumlar da oluşabilir. İşletmenin sürekliliği hakkında yapılacak olan değerlendirmede finansal tabloların sunuluşu standardı işletmenin faaliyet karlılığına finansal kaynakların erişilebilirliğine, işletmenin mevcut ve gelecek dönemlerde karlılığına ve borç ödeme planlarına atıf yapmaktadır.

“Yönetimin işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yaptığı değerlendirme, olay veya şartların, doğası gereği belirsiz olan gelecekteki sonuçları hakkında belirli bir tarihte muhakemede (yargıda) bulunulmasını içerir. Aşağıdaki unsurlar yapılan bu muhakemeye ilgilidir:

- Bir olay veya şart ya da bunların sonuçları ne kadar ileri bir tarihte meydana geliyorsa, söz konusu olay veya şartların sonucuna ilişkin belirsizlik derecesi de o kadar artar. Bu sebeple, yönetimin işletmenin sürekliliğiyle ilgili açık bir değerlendirme yapmasını zorunlu kılan finansal raporlama çerçevelerinin birçoğu, hangi dönem için yönetimin mevcut tüm bilgileri dikkate almasının zorunlu olduğunu belirler.

- İşletmenin büyüklüğü ve karmaşıklığı, faaliyetlerinin niteliği ve durumu, dış faktörlerden etkilenme derecesi, olay veya şartların sonucuna ilişkin yapılacak muhakemeyi etkiler.

- Geleceğe ilişkin herhangi bir muhakeme, o muhakemenin yapıldığı andaki mevcut olan bilgilere dayanır. Bilanço tarihinden sonraki olaylar, yapıldığı tarihte makul olan muhakemelerle tutarlı olmayan sonuçlar ortaya çıkarabilir.” (Kamu Gözetim Kurumu 2020, par. 5).

İşletmenin sürekliliğini devam ettirme yeteneği hakkında işletme yönetimi tarafından yürütülen değerlendirme olay ve şartların belirsiz bir gelecekteki sonuçları hakkında yargı üretilmesini içermektedir. Bu nedenle işletme yönetiminin değerlendirme sürecinde birtakım unsurlara dikkat etmesi gerekmektedir. İşletmenin sürekliliğini devam ettirme yeteneğini kaybetmesine yol açacak bir olay, şart ve bunların sonuçlarının ortaya çıkışı ne kadar ileri bir tarihte gerçekleşirse, değerlendirme sürecinin taşıyacağı belirsizlikte o derecede artacaktır. Bu nedenle işletme yönetiminin değerlendirme sürecinde dikkate alması dönem ya da dönemler standartları tarafından sınırlandırılmıştır. Değerlendirmenin

yapılacağı dönem ile birlikte işletmenin yapısı ve faaliyet gösterdiği sektör gibi unsurlarında değerlendirme sürecine önemli etkisi bulunmaktadır.

### **1.7.2.2. Bağımsız Denetçinin Sorumluluğu**

Mali tabloların yayımlanması sürecinde bu tablolarda sunulan bilgilerin doğru ve güvenilir olmasının yanı sıra işletmelerin sürekliliğinin de güvence altına alınması ya da süreklilik üzerinde olumsuz etkiye sahip olan faktörlerin dikkate alınarak açıklanması bilgi kullanıcıları açısından önemlidir. Bu çerçevede işletmelerin finansal sonuçlarının iyi olması kadar sürekliliğinin de teminat altına alınmış olması arzu edilen bir durumdur. Sermaye piyasalarından fon ihtiyacını karşılayan işletmelerin kamuyu aydınlatma yükümlülükleri gereği bağımsız denetim yaptırma zorunluluğu kapsamında bağımsız denetçilerin bağımsız görüş oluşturma sürecinde işletmenin raporlama döneminin sona erdiği tarihten itibaren en az bir yıllık süre için sürekliliğe dair kanıt toplaması gerekmektedir. (Vargün ve Yıldız 2021, 158).

Muhasebe standartları uyarınca, işletmenin sürekliliği esasını yönetim uygulamak zorundayken, BDS-570 uyarınca bağımsız denetçinin de yönetimin sürekliliğe ilişkin değerlendirmesinin doğruluğunu ve tarafsızlığını incelemesi gerekmektedir. BDS-570 standardı bağımsız denetçilere, denetimlerinde işletmenin sürekliliğini devam ettirme yeteneğiyle ilgili olumsuz olay ve şartların olması halinde bunları değerlendirmeleri ve paydaşlara net, şeffaf ve güvenilir bilgilerin sağlanması konularında çeşitli sorumluluklar getirmiştir. Aynı zamanda, bağımsız denetçi görüş bildirmek için de bu değerlendirmeyi gerçekleştirmekle yükümlüdür (Güngör, Ayluçtarhan, ve Sari 2022, 611)

Denetçinin sorumlulukları şunlardır;

- Finansal tabloların hazırlanmasında yönetimin işletmenin sürekliliği esasını kullanmasının uygunluğu hakkında yeterli ve uygun denetim kanıtı elde etmek ve yönetimin bu esas kullanmasının uygunluğuna ilişkin bir sonuca varmak,
- Elde edilen denetim kanıtlarına dayanarak işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin önemli bir belirsizliğin mevcut olup olmadığı konusunda bir sonuca varmak.

Bu sorumluluklar, finansal tabloların hazırlanmasında kullanılan finansal raporlama çerçevesi, işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yönetimin özel bir değerlendirme yapmasını gerektirecek açık bir hüküm içermese dahi, mevcuttur. (Kamu Gözetim Kurumu 2020, par. 6).

İşletmenin sürekliliği esasının sınanması hususunda, bağımsız denetçi işletme yönetiminin süreklilik esasını kullanımının uygunluğu hakkında yeterli denetim kanıtı elde etmek ve işletme yönetiminin süreklilik esasını kullanımının uygunluğu hakkında bir yargı üretmek ve elde edilen denetim kanıtlarına dayanarak işletmenin sürekliliğine dair önemli bir belirsizliğin bulunup bulunmadığına dair bir sonuca ulaşmak ile sorumludur. *“Ancak, BDS 2002’de açıklandığı gibi denetçinin önemli yanlışlıkları tespit etme becerisi üzerindeki yapısal kısıtlamaların muhtemel etkileri, işletmenin sürekliliğini devam ettirme kabiliyetini sona erdirebilecek gelecekteki olay veya şartlar söz konusu olduğunda daha büyüktür. Denetçi gelecekteki bu tip olay veya şartları önceden tahmin edemez. Dolayısıyla denetçi raporunda işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin önemli bir belirsizliğe atıfta bulunulmaması, işletmenin sürekliliğini devam ettirme kabiliyeti açısından bir güvence teşkil etmez.”* (BDS 570 - İşletmenin Sürekliliği, 2017, par. 7).

İşletmenin sürekliliğini devam ettirme yeteneğini sona erdirebilecek gelecekteki olay ve şartların mevcut olması, bağımsız denetçinin mali tabloların içerdiği önemli düzensizlikleri tespit etme becerisi üzerindeki yapısal kısıtlamaların muhtemel etkilerini artırır.

### **1.7.3. İşletmenin Sürekliliği Varsayımı Üzerinde Ciddi Şüpheler Doğuracak Olay ve Durumlar**

İşletmenin sürekliliğinin sağlanıp sağlanmadığına yönelik olarak denetçiler tarafından elde edilen kanıtların birlikte değerlendirilmesi kadar bu kanıtların hem finansal hem de faaliyetlere ilişkin bilgileri de içerecek yeterlilikte olması denetim sürecinin kaliteli bir şekilde gerçekleştirilmesinin ön koşulu olarak karşımıza çıkmaktadır. Bu açıdan işletmenin sürekliliği üzerinde ciddi şüpheler uyandıran nakit akışındaki uyumsuzluklar, faaliyet zararının sürekliliği, borç ödeme ve likidite yeterlilik oranlarındaki olumsuzluk, faiz ve kur riski nedeniyle pozisyon açığının karşılanamama ihtimali, karşılanması işletmeyi iflas riski ile karşı karşıya getirebilecek taahhütlerin varlığı, ana tedarikçinin iflası, müşteri kaybında artış yaşanması, faaliyet alanını sınırlandırıcı veya daraltıcı yasal düzenlemeler getirilmesi ve bunun gibi birçok finansal durum ve faaliyetlere ilişkin veriler denetçiler tarafından işletmenin sürekliliğinin değerlendirilmesinde önemli risk göstergeleri olarak kabul edilmektedir (Vargün ve Yıldız 2021, 158)

BDS 570 İşletmenin Sürekliliği standardı hem işletme yönetimini hem de bağımsız denetçiyi süreklilik esasının sınanması hususunda sorumlu tutmaktadır. İşletmenin sürekliliğini devam ettirme becerisi hususunda önemli şüphelerin oluşmasına sebep olacak pek çok faktör bulunmaktadır. Bu bağlamda, BDS 570 İşletmenin Sürekliliği Standardı, Risk Değerleme Prosedürleri ve İlgili Faaliyetler başlığı altında işletmenin sürekliliğini devam ettirme yeteneğine ilişkin ciddi şüphe oluşturabilecek olay ve şartlar hususunda atıflar yapmaktadır. Standart içerisinde işaret edilen olay ve şartlar işletmenin sürekliliğini devam ettirme yeteneği ile ilgili ciddi şüphelerin doğmasına neden olacak olan tüm durumlar yer verilmemektedir. Ayrıca burada sayılan olay ve şartların bir veya birkaçının mevcudiyeti işletmenin sürekliliği ile ilgili bir belirsizliği var olduğu anlamına da gelmemektedir(Kamu Gözetim Kurumu 2020, par. A3):

Mali Durum ile ilgili:

- Net yükümlülük veya kısa vadeli net yükümlülük pozisyonu,
- Sabit vadeli borçların vadelerinde geri ödeneceğine veya yenileneceğine dair gerçekçi olmayan beklentiler veya uzun vadeli varlıkları finanse etmek için kısa vadeli borçlara gereğinden fazla bağlı olunması,
- Kredi verenler tarafından sağlanan finansal desteğin geri çekileceğine dair belirtiler,
- Tarihi veya ileriye yönelik finansal tablolarında gösterilen, işletme faaliyetleriyle ilgili negatif nakit akışları,
- Temel finansal oranlardaki olumsuzluk,
- Önemli tutarlardaki faaliyet zararı veya nakit akışı yaratan varlıkların değerindeki önemli azalışlar,
- Temettü ödemelerindeki gecikme veya kesilmeler,
- Alacaklılara vade tarihinde ödeme yapılamaması,
- Kredi sözleşmelerinin şartlarına uyulamaması,
- Tedarikçilerle yapılan işlemlerde vadeli ödemedi, peşin ödemeye geçilmesi,
- Yeni ürün geliştirilmesi veya diğer önemli yatırımlar için finansman temin edilememesi.

İşletme Faaliyetleri ile İlgili:

- Yönetimin işletmeyi tasfiye etme veya faaliyetleri durdurma niyeti.
- Kilit yönetim kadrosunun boş olarak yerlerinin doldurulmaması.
- Önemli bir pazarın, kilit müşteri veya müşterilerin, imtiyaz, lisans veya ana tedarikçi veya tedarikçilerin kaybedilmesi.

- İşgücüyle ilgili yaşanan sıkıntılar.
- Önemli hammadde darboğazı.
- Oldukça başarılı bir rakibin ortaya çıkması.

Diğer Konular ile İlgili:

- Finansal kuruluşların borç ödeme gücü veya likidite yükümlülükleri gibi sermaye yeterlilik yükümlülüklerine veya diğer yasal yükümlülüklerine aykırılık.
- İşletme aleyhine açılmış, sonuçlanmamış ve işletmenin aleyhine sonuçlanması durumunda işletmenin karşılamayacağı taleplere sebep olabilecek davalar.
- İşletmeyi olumsuz yönde etkilemesi beklenen mevzuat değişiklikleri veya hükümet politikalarındaki değişiklikler.
- Oluştığında sigorta kapsamında bulunmayan veya eksik sigortalanmış felaketler.

#### **1.7.4. İşletmenin Sürekliliğine Dair İşletme Yönetiminin Yaptığı Değerlendirmenin Gözden Geçirilmesi**

İşletme yönetiminin, işletmenin sürekliliğini devam ettirme yeteneğine dair yaptığı değerlendirme bağımsız denetçi tarafından incelenmesi gerekmektedir. “Denetçi, yönetimin işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yaptığı değerlendirmeyi inceler.” (Kamu Gözetim Kurumu 2020, par. 12).

“Yönetimin işletmenin sürekliliğini devamı ettirme kabiliyetine ilişkin yaptığı değerlendirme, yönetimin kullandığı süreklilik esasıyla ilgili denetçinin yapacağı değerlendirmenin odak noktasını oluşturmaktadır.” (Kamu Gözetim Kurumu 2020, A8).

“Yönetimin analiz yapmamasının eksikliğini gidermek denetçinin sorumluluğunda değildir. Ancak bazı durumlarda yönetimin, değerlendirmesini destekleyen ayrıntılı bir analiz yapmaması, yönetimin işletmenin sürekliliği esasını kullanmasının uygun olup olmadığı hakkında denetçinin bir sonuca varmasına engel teşkil etmeyebilir. Örneğin, geçmişte kârlı faaliyetlerde bulunan ve finansal kaynaklara kolay erişen işletmelerde yönetim, değerlendirmesini ayrıntılı bir analiz olmadan yapabilir. Bu durumda, yönetimin finansal tabloların hazırlanmasında işletmenin sürekliliği esasını kullanmasının içinde bulunulan şartlar altında uygun olup olmadığına ilişkin denetçinin bir sonuca varabilmesi için diğer denetim prosedürlerinin yeterli olması hâlinde, yönetimin değerlendirmesinin uygunluğu hakkında denetçinin yapacağı inceleme, ayrıntılı prosedürler uygulanmadan yapılabilir.”(Kamu Gözetim Kurumu 2020, A9).

İşletme yönetiminin, işletmenin sürekliliğini devam ettirme yeteneği üzerinde yaptığı değerlendirme, bağımsız denetçinin yönetimin kullandığı süreklilik esasını ile ilgili değerlendirmenin kapsamını oluşturmaktadır. İşletme yönetiminin süreklilik yeteneği hakkında yürüttüğü analizlerin eksik olması durumunda bu eksiklikleri gidermek bağımsız denetçinin sorumluluğunda değildir. Bazı durumlarda TMS 1 – Finansal Tabloların Sunuluşu standardında belirtildiği gibi, yönetimin detaylı bir analiz yapmamış olması, denetçinin işletme yönetiminin süreklilik esasını kullanımının uygunluğu hakkında bir sonuca varmasını engellemeyebilmektedir.

“Denetçi, yönetimin işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yaptığı değerlendirmeyi incelerken, geçerli finansal raporlama çerçevesi uyarınca, yönetimin değerlendirmeyi yapmak için kullandığı dönem ile aynı dönemi veya mevzuatta daha uzun bir dönem öngörülmesi durumunda bu süreyi dikkate alır. İşletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yönetimin yaptığı değerlendirmenin, BDS 5604’ta tanımlanan finansal tabloların tarihinden itibaren on iki aydan daha kısa bir süreyi kapsamaması durumunda denetçi, yönetimden yaptığı değerlendirmeyi finansal tablo tarihinden itibaren en az on iki aylık dönemi kapsayacak şekilde genişletmesini talep eder.”(Kamu Gözetim Kurumu 2020, par. 13).

Bağımsız denetçi tarafında işletme yönetiminin, işletmenin sürekliliğinin devam ettirme yeteneğine ilişkin yapılan değerlendirme incelenirken yönetimin değerlendirmeyi yapmak için kullandığı dönem ve aynı zamanda mevzuatta daha uzun bir dönem öngörüldüyse bu süre dikkate alınır. Bağımsız denetçi tarafından, işletme yönetiminin yaptığı değerlendirmenin bilanço tarihinden itibaren on iki aylık dönemi kapsamadığı tespit edilirse, yönetimden değerlendirmeyi bilanço tarihinden itibaren en az on iki aylık dönem için yapmasını talep eder.

### **1.7.5 İşletmenin Sürekliliğine Dair İşletme Yönetiminin Yaptığı Değerlendirme Sonrası Dönem**

“Denetçi, yönetimin değerlendirdiği dönemden sonrası için, işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin ciddi şüphe oluşturabilecek olay veya şartlarla ilgili sahip olduğu bilgiler hakkında yönetimi sorgular” (Kamu Gözetim Kurumu 2020, 15).

“11 inci paragrafta zorunlu kılındığı üzere denetçi, yönetimin değerlendirme yaptığı dönemden sonra meydana gelecek ve finansal tabloların hazırlanmasında işletmenin sürekliliği esasının kullanılmasının uygunluğunu şüpheye düşürecek planlanmış veya

planlanmamış olay veya şartların meydana gelme ihtimali konusunda dikkatli davranır. Bir olay veya şartın sonucuyla ilgili belirsizlik derecesi, olay veya şart ne kadar ileri bir tarihte ilgiliyse o kadar artacağı için, bu tür olay veya şartlar gözden geçirilirken, işletmenin sürekliliğine dair sorun oluşturacak göstergelerin önemli hâle gelmesi durumunda, denetçinin ilâve eylemler gerçekleştirmesi gerekir. Bu tür olay veya şartların tespit edilmesi hâlinde denetçi, yönetimden olay veya şartın işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin yapılan değerlendirme üzerindeki muhtemel önemini değerlendirmesini talep edebilir. Bu durumda, 16 ncı paragrafta yer alan prosedürler uygulanır.” (Kamu Gözetim Kurumu 2020, par. A14).

İşletme yönetiminin süreklilik değerlendirmesi yaptığı dönem sonrasında mali tabloların hazırlanmasında işletmenin sürekliliği esasını kullanmasının uygunluğunu şüpheyeye düşürecek planlanmış veya planlanmamış olaylar meydana gelebilir. Bağımsız denetçi, bu olay veya şartlara dikkat etmelidir. Bağımsız denetçi, yönetimin değerlendirdiği dönem sonrasında işletmenin sürekliliğini devam ettirme yeteneğine ilişkin ciddi şüphey oluşturacak olay ve şartlar hakkında işletme yönetiminden bilgi alır.

“Yönetimin değerlendirme yaptığı dönemden (bu dönem, 13 üncü paragrafta belirtildiği gibi finansal tabloların tarihinden itibaren en az 12 ay olacaktır) sonra işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin ciddi şüphey oluşturabilecek olay veya şartları belirlemek için denetçinin, yönetimin sorgulanması dışında başka herhangi bir denetim prosedürünü uygulama sorumluluğu yoktur.” (Kamu Gözetim Kurumu 2020, par. A15).

İşletmenin süreklilik kabiliyetini etki etme ihtimali olan olay ve şart ne kadar ileri bir tarih ile ilgili ise, olay veya şartın sonucu ile ilgili belirsizlik o derecede artacaktır. Bu tür olay veya şartlar gözden geçirilirken işletmenin sürekliliğine dair şüphey ortaya çıkmasına neden olacak göstergelerin ortaya çıkması durumunda, bağımsız denetçinin ek eylemler gerçekleştirmesi gerekmektedir. Bu tür olay veya şartların tespit durumunda bağımsız denetçi, işletme yönetiminden sürekliliği devam ettirme yeteneği üzerinde muhtemel etkilerini değerlendirmesi talep edebilir. Bağımsız denetçinin, işletme yönetiminin değerlendirme yaptığı dönemden sonra ortaya çıkabilecek olan olay ve şartlar hakkında yönetimden bilgi talep etmek dışında herhangi bir denetim prosedürü uygulaması gerekli değildir.

## İKİNCİ BÖLÜM

Bu bölümde makine öğrenmesi kavramı, öğrenme stratejileri, makine öğrenmesi süreci ve makine öğrenmesi algoritmaları hakkında teorik bilgi verilmektedir.

### 2.1. Makine Öğrenmesi

İnsanlık uzun yıllardır, insan benzeri düşünme yeteneğine sahip makineler yaratmayı amaçlamaktadır. Günümüzde makineler sahip oldukları teknoloji ve işlem kapasitesi ile insan aklının belirli işlevlerini yerine getirebilmektedir. Ancak makinelerin insanın çizdiği sınırları geçme yeteneği bulunmamaktadır (Gopal 2018). Bilgisayarların, insanın bilişsel kabiliyetlerini gerektiren karmaşık görevleri yerine getirip getiremeyeceği sorusu, bilgisayar biliminde, yapay zekâ altında geniş bir disiplin olarak sınıflandırılmaktadır. Yapay zekânın literatürde net bir tanımı bulunmamaktadır. Genel bir ifade ile yapay zekâ öğrenme, algılama, akıl yürütme ve problem çözme gibi insan zihni ile ilişkili bilişsel işleri taklit eden bilgisayarları niteleyen bir terim olarak tanımlanmaktadır. Erken dönem yapay zekâ araştırmacıları, bir yapay zekâ sorununun çözümünde bilgisayar programlama sürecinin uygulanabileceği varsaymıştır. Bu varsayım ile bilgisayarlara karmaşık görevleri yerine getirmeyi öğretmek için belirli koşullar altında belirli kararlar vermesini sağlayan ve çok sayıda koşul ifadesi barındıran programlardan yararlanılması mümkün olmaktadır. Yapay zekâ problemlerinin çözümünde alternatif yaklaşım, bilgisayarın problemin çözümünde yeteneklerini deneyim yolu geliştirebileceği algoritmalar tasarlamaktır. Önerilen yaklaşım tecrübenin algoritmaya eğitim veri seti yolu ile aktarılmasını temel almaktadır. Kullanılan algoritmaların tasarımı, biyolojiden esinlenen öğrenme makinelerinden, olasılığa dayalı istatistiksel öğrenme yöntemlerine kadar farklı stratejilere dayanmaktadır. Yapay zekâda makine öğrenmesi adı altında bir alt alan olarak öne çıkan alan, 1980'li yıllardan günümüze pek çok araştırmaya konu olmuştur. Makine öğrenmesinin doğası sembolik yaklaşımların sahip olduğu dezavantajlardan mustarip olmasını engellemektedir. Sembolik yaklaşımların aksine, veriye dayalı makine öğrenmesi algoritmaları yapay zekâ problemlerinin çözümünde, programlama olmadan matematiksel modeller üretmek maksadı ile veriye odaklanmaktadır (Jiang 2021).

Makine öğrenmesi, algoritmik olarak veri hakkında tahminlerde bulunma ya da verinin yapısı üzerinde gözlemler yürütme faaliyetleri olarak tanımlanabilir. Kavram bilgisayarın mevcut veriden bilgi çıkarımını ifade etmekte ve lineer cebir, optimizasyon, yapay zekâ, bayesyen teori, hesaplamalı karmaşıklık teorisi, kontrol teorisi, teorik bilgisayar bilimi,

bilgi teorisi, felsefe, psikoloji ve nörobiyoloji, olasılık ve istatistik gibi pek çok farklı alandan etkilenmektedir (Jung 2022; Mitchell 1997). Makine öğrenmesi, bilgisayarlara açıkça programlamadan öğrenme yeteneği veren bir çalışma alanıdır (Awad ve Khanna 2015; J. Bell 2015; McClendon ve Meghanathan 2015).

Makine öğrenmesi kavramının ortaya çıkışı Arthur Samuel tarafından 1950'li yıllarda IBM' de, ilk bir programı depolayabilen ve çalıştırabilen bilgisayarda, bir dama oyununu geliştirmesine kadar geriye gitmektedir (McCarthy ve Feigenbaum 1990). Makine Öğrenmesi kavramının adlandırılmasının ilk olarak kim tarafından yapıldığı ile ilgili literatürde açık bir bilgi bulunmamaktadır. Terimin ilk olarak Samuel (1959) tarafından kullanıldığına dair literatürde görüşler mevcuttur (El Naqa ve Murphy 2015, 5-6). Ancak kavram Arthur Samuel tarafından açık bir şekilde tanımlanmamaktadır. Samuel (1959), çalışmasında, kavramı niteleyen *“Bir bilgisayarın, programı yazan kişinin oynayabileceğinden daha iyi bir dama oyunu oynamayı öğrenecek şekilde programlanabileceğini doğrulamak için yeterince çalışma yapılmıştır.”* ve *“Bilgisayarları tecrübeden öğrenecek şekilde programlamak, sonunda bu ayrıntılı programlama çalışmasının çoğunu ortadan kaldırmalıdır.”* ifadelerine yer vermektedir. Mitchell (1997), tarafından makine öğrenmesi, *“Eğer bir bilgisayar programının T görevlerindeki, P ile ölçülen performansı, deneyim E ile artıyorsa, programın bazı T görevleri hususunda deneyim E' den öğrendiği söylenmektedir.”* şeklinde tanımlanmaktadır. Tanımdan yola çıkarak bilgisayar programının bir dizi görevi yerine getirirken edinilen deneyimin performansı artırdığı söylenebilir. Ayrıca tanım programa deneyim aktarmak için kullanılan veri setinin niteliğinin önemini de işaret edilmektedir. Makine öğrenmesi, veri setindeki örüntüleri otomatik olarak algılayabilen ve daha sonra elde edilen örüntüleri gelecek verileri tahmin etmek ya da farklı karar verme işlemleri gerçekleştirmek için kullanabilen bir dizi yöntemdir (Murphy 2012, 1).

Makine öğrenmesi metotları, uygulama örnek uzayı içerisindeki veri noktalarını ya da gözlemleri okumaktadır. Her bir gözlem iki özellikte, öznitelikler (feature) ve etiket (label) ile karakterize edilmektedir. Öznitelikler kolaylıkla ölçülebilen ve hesaplanabilen özelliklerdir. Bunun aksine etiketlerin kolaylıkla ölçülebilmesi mümkün olmayabilir ve çoğu zaman ölçülebilmesi için bir insan yargısına ihtiyaç duyulmaktadır. Makine öğrenmesi özniteliklere dayanarak gözlemin etiketini tahmin etmeyi amaçlamaktadır ve değişkenler arasındaki ilişkileri öğrenmek, tahminler yapmak ve değişen durumlarda karar almak için büyük veri kümelerini kullanmaktadır (Hull 2020; Jung 2022). Makine

öğrenmesi, tahmin yapmak için deneyimden faydalanan hesaplamalı yöntemler olarak tanımlanmaktadır. Burada deneyim, tipik olarak toplanan ve analiz için hazır hale getirilen elektronik formattaki geçmiş verileri ifade etmektedir. Verinin kalitesi ve boyutu her durumda yapılacak olan tahminlerin doğruluğu için çok önemlidir. Standart bir gözetimli öğrenme sürecinde yeni gözlemler için doğru tahminler yapmak için etiketlenmiş örneklerden oluşan bir veri setinden faydalanmaktadır. Bu durumda problem, bir fonksiyon kümesinin alt kümesi olan bir hipotez setinden bir fonksiyon seçilmesi olarak ifade edilir. Ardından seçilen fonksiyon, her bir yeni örnek için etiket üretmek maksadı ile kullanılır (Mohri, Rostamizadeh, ve Talwalkar 2018).

Makine öğrenmesi modelleri, üzerinde çeşitli çıktı fonksiyonları tanımlanabilecek olan örüntüler/gözlemler/ desenler kümeleri ile çalışmaktadır ve bu örüntüleri bir “ $S$ ” kümesi ile tanımlanmaktadır. Eğitim tecrübesi, “ $N$ ” bir örüntü şeklinde mevcuttur (Gopal 2018, 36);

$$S^{(i)} \in S; i = 1, 2, \dots, N.$$

Makine öğrenmesi modellerinde bir örüntü sabit sayıda “ $n$ ” özellik/öznitelik  $x_j; j = 1, 2, \dots, n$  ile tanımlanmaktadır ve her öznitelik örüntü için gerçek sayısal değerlere sahiptir.  $x_j$  probleme ait olan bir vektör uzayıdır ve elemanı olan her bir öznitelik  $V_{x_j} \in \mathcal{R}$  kümesine ait değerler içermektedir. Bir veri deseni  $S^{(i)}$ ,  $x_j^{(i)} \in V_{x_j}$  sağlayan,  $\{x_1^{(i)}, \dots, x_n^{(i)}\}$  şeklinde bir öznitelik değer setine sahiptir. Her bir örüntü,  $n$  boyutlu durum uzayında ya da hipotez uzayında  $n$  adet sayısal öznitelik ile görselleştirilmektedir.

$$\mathbf{X} = [x_1 x_2 \dots x_n]^T \in \mathfrak{R}^n$$

Denklemden gösterilen  $\mathbf{X}$  seti, tüm  $n$  örüntüleri için sonsuz bir vektör  $x^{(i)}$  setidir.  $\mathbf{X}$ , hipotez uzayında  $\mathbf{X} \subset \mathfrak{R}^n$  bir bölgeyi temsil etmektedir.

Literatürde yer alan tanımlar, bilgisayarların veriden öğrenme yolu ile belirli görevleri yerine getirebileceği fikrini paylaşmaktadır. Kavramın nitelediği öğrenme davranışı, veri içerisindeki temel örüntülerin bulunması ve bunların matematiksel bir fonksiyon şeklinde tanımlanması sürecini ifade etmektedir. Makine öğrenmesinin temel prensipleri şunlardır (Ghatak 2017, 77; Jiang 2021, 11-15);

Occam's razor, felsefe ve bilimde genel bir problem çözüme ilkesidir. İlke "en basit çözüm büyük olasılıkla doğru olandır" prensibini ifade etmektedir. Makine öğrenmesi bağlamında Occam's razor, iki model arasında yapılacak olan seçimde basitliğin temel alınmasını anlamına gelmektedir. Bu bağlamda, eğer eğitimde iki farklı modelin benzer performans gösterdiği gözlemlenirse, basit olan modelin karmaşık olana tercih edilmesi gerekmektedir. Ayrıca minimum açıklama uzunluğu (minimum description length) ilkesi makine öğrenmesi metodlarının verideki düzenlilikleri bulmayı amaçladıklarını ve verideki düzenlilikleri açıklamak için en iyi modelin (hipotezin) veriyi en fazla sıkıştıran model olduğunu ifade eder,

No Free Lunch teoremi, makine öğrenmesi bağlamında teorem, hiçbir makine öğrenmesi metodunun olası tüm makine öğrenmesi problemleri için evrensel olarak diğer metotlardan üstün olmadığını belirtmektedir. Herhangi iki makine öğrenmesi algoritması verildiğinde, bu iki algoritmayı olası tüm problemleri çözmek için kullanırsak iki algoritmanın performansı aynı olacaktır. Teorem makine öğrenmesi algoritmasının yalnızca eğitim verisinden yararlı bir örüntü öğrenemeyeceğini ifade etmektedir. Bir makine öğrenmesi metodu, bazı problemler için iyi performans gösteriyorsa, metodun eğitim verilerinin ötesinde problemin altında yatanlara ilişkin diğer bilgileri açık bir şekilde ya da üstü kapalı olarak kullanmış olması gerekmektedir,

Smooth World Kanunu, gerçek dünyanın kusursuzluğunun, makine öğrenmesi modellerinin başarılı olmasının temel nedeni olduğunu açıklamaktadır. Enerji ve güç gibi gerçekte var olan katı kısıtlamalar sebebi ile gerçek dünyadaki herhangi bir süreç doğası gereği kusursuzdur. Ayrıca sezgilerimiz ve algılarımız "Smooth World" kanunu üzerine inşa edilmiştir. Bu nedenle gerçek dünyadaki herhangi bir sorunun çözümünde makine öğrenmesinin kullanılırsa, yasa her zaman uygulanabilir ve makine öğrenmesi problemlerinin çoğunu basitleştirir. Kanun, makine öğrenmesi için basit bir strateji önermektedir. Herhangi bir bilinmeyen için gözlem alındığında, eğitim setinde yer alan bilinen tüm örnekler ile bilinmeyen bir gözlem için tahmin yapılabilir,

Curse of Dimensionality, makine öğrenmesinde yüksek boyutlu veri uzaylarında öğrenme ikilemini ifade etmektedir. Curse of Dimensionality, makine öğrenmesi çalışmalarındaki en büyük zorluklardan birisidir. Makine öğrenmesi probleminin boyutu büyüdükçe veri uzayının boyutu da katlanarak artmaktadır. Bu durum, tipik olarak herhangi bir makine öğrenmesi probleminin etkinliğini sağlamak için katlanarak artan miktarda eğitim verisi ve bilgi işlem kaynağına ihtiyaç duyulmasına neden olmaktadır. Makine öğrenmesi

problemlerini çoğu çok boyutludur. Ancak insanın üç boyutlu evren algısı, yüksek boyutlu veri uzaylarında başarısız olmaktadır. Ayrıca yüksek boyutlu hipotez uzaylarında, veri eşit bir dağılım göstermektedir. Makine öğrenmesi algoritmaları düşük boyutlu veri setleri ile iyi performans ortaya koyarken, çok boyutlu veri uzaylarında modellerin genelleştirme performansı düşmektedir. Curse of Dimensionality modelin genelleştirme performansının kontrolünü zorlaştırmaktadır. Düzgün olmayan bir veri dağılımı, verilerin tüm boyutlarının bağımsız olmadığını, ancak birçok boyutun gereksiz olduğunu ve bir şekilde yüksek oranda ilişkili olduğunu göstermektedir. Bir başka ifade ile, veri hakkında herhangi bir bilgi kaybı yaşamadan veri uzayından birçok boyut atılabilmektedir. Bu fikir makine öğrenmesi modellerinde yaygın olarak kullanılan boyut indirgemenin (dimensionality reduction) temel mantığını oluşturmaktadır.

Boyutsallık laneti bir veri setindeki gözlemlerin sayısı ile özniteliklerin sayısı arasındaki oranı, bu veri setindeki gözlemlenebilen olgularla ilişkilendirir. Veri setindeki özniteliklerin sayısı arttıkça, veri hacmi girdi uzayının çeşitli bölgelerinde hiçbir nesne olmaksızın daha seyrek hale gelir. Buna ek olarak, gözlemler arasındaki mesafe aynı değere yakınsar. Bu durumda ise gözlemlerin arasındaki mesafeye dayalı tekniklerin performansını azaltmaktadır. Boyutsallık lanetinin üstesinden gelmenin iki yolu bulunmaktadır. Bunlardan ilki veri setindeki gözlem sayısının artırılmasıdır. İkincisi ise öznitelikleri sayısının azaltılmasıdır. Çoğu durumda veri setindeki gözlemlerin sayısının artırılması mümkün değildir. Bu nedenle tek alternatif öznitelik sayısının azaltılmasıdır (Moreira, Carvalho, ve Horváth 2018).

## 2.2. Öğrenme Stratejileri

Veri setindeki faydalı örüntüleri otomatik olarak tanımlayabilen ve bu örüntüleri gelecek durumları tahmin etmek maksadı kullanabilen makine öğrenmesi algoritmaları temel olarak “Gözetimli Öğrenme” ve “Gözetimsiz Öğrenme” olarak iki kategoride sınıflandırılmaktadır (Abbott 2014, 5; James vd. 2013, 1). Ancak pekiştirmeli öğrenme (Reinforcement Learning) olarak adlandırılan ve diğer makine öğrenmesi türlerinden daha az yaygın olarak kullanılan bir diğer üçüncü tür makine öğrenmesi yöntemi daha bulunmaktadır (Murphy 2012). Pekiştirmeli öğrenme gözetimli öğrenme ile gözetimsiz öğrenme arasında konumlanmaktadır. Pekiştirmeli öğrenmede model eğitim sürecinde, cevabın yanlış olduğu durumda geri bildirim almaktadır. Ancak modele doğru cevabın

nasıl bulunacağı ile ilgili bir sinyal verilmemektedir. Model doğru cevabı bulana kadar farklı olasılıkları keşfetmekte ve denemektedir (Marsland 2015).

Gözetimli öğrenmede eğitim verisi girdi-çıkı vektörlerinden oluşmaktadır. Eğitim verisindeki her bir girdi için algoritmanın gözetim sinyali olarak kullanabileceği çıktı tanımlanmış olmalıdır. Gözetimli öğrenme metotları, makine öğrenmesi literatüründe derinlemesine tartışılmış bir alandır ve yeterli sayıda verinin olması durumunda genellikle iyi bir performans garanti ederler. Ancak, girdi-çıkı çiftlerini oluşturmak ya da veri setini etiketlemede insan yargısına ihtiyaç duyulması sebebi ile maliyetli olabilmektedirler. Gözetimsiz öğrenme metotları olası tüm girdilerin kullanıldığı senaryoda benzer girdileri kümelemeyi amaçlamaktadır. Gözetimsiz öğrenmede veri etiketleri mevcut değildir. Bu nedenle çıkı etiketlerinin mevcut olmadığı durumda hangi girdilerin benzer olduğunun tespiti aşılması gereken temel zorluktur ve gözetim desteği bulunmaması sebebi ile çok daha zordur. Ancak eğitim verilerinin toplanması görece daha az maliyetlidir (Jiang 2021).

### 2.2.1. Gözetimli Öğrenme

Gözetimli öğrenme, genel olarak bir ya da birden fazla girdiye dayanarak bir çıkıyı tahmin etmek amacıyla bir model oluşturmayı ifade etmektedir (James vd. 2013, 1). Gözetimli öğrenme, bilinen hedef değerlerin kullanılması ile her bir yeni gözlem için uygun bir değer belirlemeyi öğrenebilen, tahmine dayalı bir model geliştirme sürecidir. Gözetimli öğrenme sürecinde model, neyi ve nasıl öğreneceği konusunda gerekli talimatlar ile donatılmaktadır. Gözetimli öğrenme, etiketlenmiş veri setlerini referans almaktadır. Süreç, hedef sınıfı, öznitelikler açısından açıklamaktır (Van der Aalst 2016, 93).

Gözetimli öğrenmede temel amaç, aşağıdaki denklemde gösterildiği şekilde, etiketlenmiş olan bir girdi-çıkı seti verildiğinde, girdilerden çıkıtlara doğru bir eşleşmeyi öğrenmektir (Murphy 2012).

$$D = \{(x_i, y_i)\}_{i=1}^N$$

Formülde,  $D$  eğitim seti olarak adlandırılmaktadır,  $N$  ise eğitim gözlemlerinin sayısıdır. Temel olarak, her bir eğitim girdisi  $x_i$  'yi, temsil eden  $D$  boyutlu bir sayı vektörüdür. Bu değerler öznitelik (feature) olarak adlandırılmaktadır.  $D$  eğitim setindeki  $x_i$  yapısal formattaki bir değer ya da bir zaman serisi olabileceği gibi yapısal olmayan formatta e-mail ya da metin de olabilmektedir. Çıkı ya da yanıt değişkeni olan  $y_i$  prensip olarak

herhangi bir şey olabilmektedir. Ancak çoğu makine öğrenmesi yöntemi  $y_i$ 'nin bazı sonlu kümelerden kategorik veya nominal bir değişken olduğunu varsaymaktadır. Yanıt değişkeni  $y_i$  kategorik olduğunda, sınıflandırma problemi ya da örüntü tanıma problemi olarak adlandırılmaktadır. Yanıt değişkeni  $y_i$  gerçek değere sahip olduğunda ise regresyon problemi olarak adlandırılmaktadır. Sınıflandırma problemlerinde temel amaç her sınıfın örneklerinden yararlanarak, eğitime dayalı olarak yeni gelecek olan girdiler için çıktı değerleri tahmin etmektir. Regresyon problemlerinde temel amaç, yeni girdiler için çıktı değerleri tahmin etmek amacı ile fonksiyon oluşturmaktır (Gopal 2018).

Hedef vektörünün, iki elemanı eşit olup olmamaları dışında belli bir yapıya sahip değil ise bu sınıflandırma problemi olarak tanımlanmaktadır. Sınıflandırma problemleri, kategorik bir yanıt değişkenine ihtiyaç duymaktadır. Sınıflandırma problemlerinde amaç, veri setindeki gözlemleri özniteliklere dayanarak sınıflandırmaktır. Hedef vektörünün elemanları “*sınıf*” olarak tanımlanmaktadır. Ayrıca sınıflandırma problemlerinin çözümünde kullanılan algoritmalara sınıflandırıcı adı verilmektedir.

Regresyon problemleri sayısal ya da gerçek değişkenlere sahip yanıt değişkenlerine ihtiyaç duymaktadır. Regresyon problemlerinde amaç, verilere en az hata ile uyum bir fonksiyon tahmin etmektir. En sık kullanılan regresyon tekniği doğrusal regresyondur. Doğrusal model  $\hat{y} = f(x_1, x_2, x_3, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i$  bağımlı değişken  $y$  ile bağımsız değişkenler  $x_1, x_2, x_3, x_4, x_5, \dots, x_n$  veriyken, veri seti üzerinden öğrenilir. Veri setindeki her bir gözlem için  $|y - \hat{y}|$  hatası hesaplanmaktadır. Doğrusal regresyonda yaygın olarak en çok kullanılan yöntem hata karelerin toplamını minimize etmektir (Abbott 2014, 270; Van der Aalst 2016, 93).

Gözetimli öğrenmede, makine, girdi vektörleri ya da durumların verilen değerlerinden ve her duruma verilen yanıtta ya da çıktılardan oluşan bir eğitim setinden yararlanarak öğrenmektedir (Gopal 2018). Gözetimli öğrenme, etiket ve öznitelik içeren veriler ile çalışmaktadır. Etiketler, model ile üretilmesi amaçlanan hedef vektörün sahip olduğu değerlerdir. Öznitelikler ise tahminde kullanılan değişkenlerdir (Hull 2020). Makine öğrenmesi, veriden örüntüleri çıkaran otomatik bir süreç olarak tanımlanabilir. Gözetimli öğrenme teknikleri öznitelikler ile etiketler arasındaki ilişkiyi, geçmiş örnekler üzerinden öğrenmektedir. Model, sonrasında yeni örnekler için tahmin yapmak maksadı ile kullanılmaktadır (Kelleher, Mac Namee, ve D'Arcy 2020). Algoritma, bir dizi etiketli örneği eğitim veri seti olarak almak sureti ile, tüm bilinmeyen gözlemler için tahminlerde bulunmaktadır (Mohri, Rostamizadeh, ve Talwalkar 2018).

### 2.2.2. Gözetimsiz Öğrenme

Gözetimli öğrenmenin aksine, gözetimsiz öğrenme teknikleri veri setindeki gözlemlerin etiketlerine ihtiyaç duymamaktadır (Jung 2022). Gözetimsiz öğrenmenin temel amacı belirli bir değerin tahmin edilmesi değildir. Uygulama veri içerisindeki örüntülerin ortaya çıkarılmasını amaçlamaktadır. Böylece verinin temsil ettiği ortam daha iyi anlaşılabilir (Hull 2020).

Gözetimsiz öğrenme sürecinde girdiler bulunmakla birlikte, gözetimli öğrenmeden farklı olarak çıktı bulunmamaktadır. Gözetimsiz öğrenmeden, veri yapısı ve gözlemler arasındaki ilişkileri öğrenmek amacı ile faydalanılabilir (James vd. 2013, 1). Gözetimsiz öğrenme, girdi veriden, önceden bilgi sahibi olmadan ya da herhangi bir yönlendirme olmaksızın, anlamlı örüntüler elde etmeyi amaçlayan veri analiz tekniğidir. Gözetimli öğrenmeden ve pekiştirmeli öğrenmeden farklı olarak her bir girdi ile ilişkili istenen çıktılar veya çevresel bir geri bildirim bulunmamaktadır. Gözetimsiz öğrenmede yalnızca,  $D = \{x_i\}_{i=1}^N$ , girdileri verilir. Formülde açıkça görüleceği gibi gözetimsiz öğrenmede çıktı ya da yanıt değişkenleri mevcut değildir. Gözetimsiz öğrenmede temel amaç değerli örüntülerin ortaya çıkarılmasıdır.

Gözetimsiz öğrenmenin alt alanlarından birisi kümelemedir. Kümeleme çalışmaları veriyi anlamlı alt kümeler halinde düzenlemeyi amaçlamaktadır. Kümeleme analizi, bilgiyi yapılandırmak ve veriler arasında anlamlı ilişkiler üretmek için faydalı bir analizdir. Kümeleme, genellikle çok büyük veri kümelerini analiz etmek için kullanılmaktadır (Mohri, Rostamizadeh, ve Talwalkar 2018). Gözetimsiz öğrenme makine öğrenmesi alanında kümeleme (clustering), yoğunluk tahmininin (density estimation) ve örüntü tanıma (pattern recognition) amacı ile yaygın olarak kullanılmaktadır (Sanghoon Lee vd. 2020; Raschka 2015).

### 2.2.3. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme bir bilgisayarın ya da bir ajanın bilinmeyen bir ortamda uzun bir etkileşim sürecinde, seçilebilecek en iyi eylemleri nasıl seçeceğinin nasıl öğretilbileceği ile ilgili makine öğrenmesi alanıdır. Gözetimli öğrenmeden farklı olarak, pekiştirmeli öğrenme sürecinde ajan her adımda en iyi eylemin ne olduğu ile ilgili bir destek almamaktadır. Ancak seçimleri ile ilgili geri dönüş sinyalleri almaktadır. Pekiştirmeli öğrenmede temel amaç, uzun vadede ödülü maksimize etmek için genellikle politika olarak adlandırılan her koşulda hangi eylemin seçilmesi gerektiğinin öğrenilmesidir.

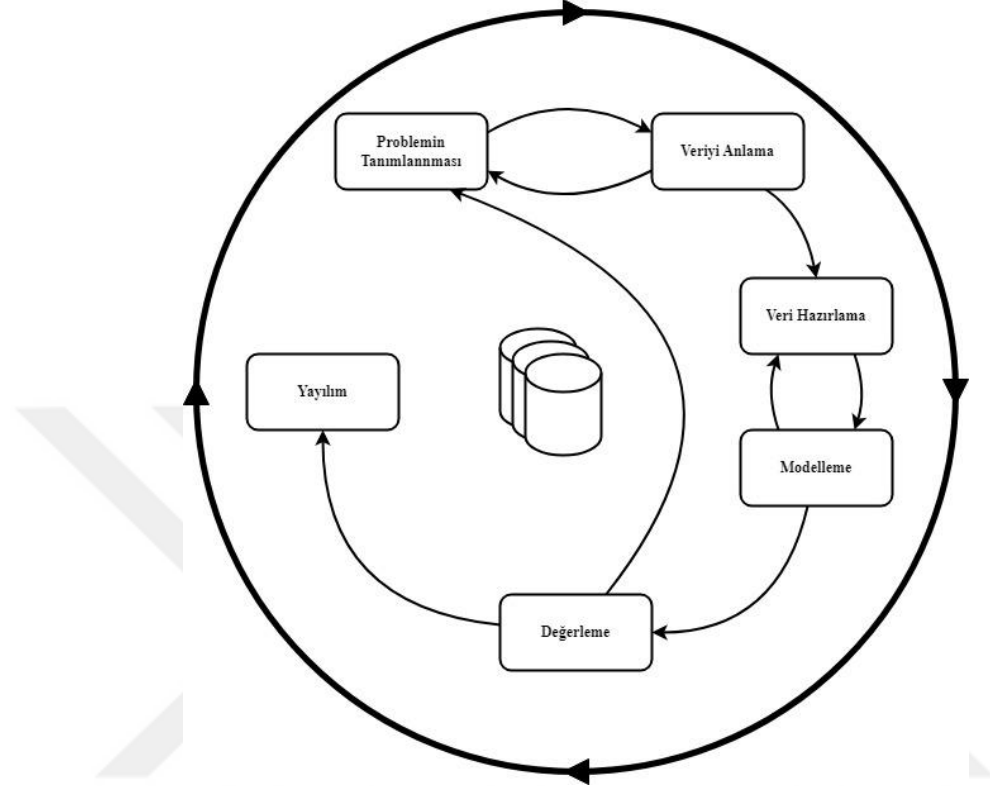
Pekiştirmeli öğrenme genel bir öğrenme çerçevesini tanımlamaktadır. Öğrenme süreci oldukça zor kabul edilmektedir. Çünkü ajanın yalnızca zayıf ödül sinyallerine dayalı olarak büyük arama alanlarını nasıl keşfedebileceğini öğrenmesi gerekmektedir (Jiang 2021).

Pekiştirmeli öğrenmede, gözetimli öğrenmeden farklı olarak etiketlenmiş veri setleri yer almamaktadır. Pekiştirmeli öğrenme, öğretmen ile öğrenme yerine eleştirmen ile öğrenme olarak tanımlanmaktadır. Öğrenme sürecinde öğretmen, öğrenciye eylemin nasıl yapılması gerektiğini öğretmektedir. Ancak eleştirmen sadece yanlış eylemin seçildiği hususunda bir geri bildirim vermektedir. Pekiştirmeli öğrenmede ajan anlık durumlar ve eylemler için değer üretmeyi öğrenmektedir. Süreçte yer alan ödüllendirme sistemi ajanın ödülü maksimize etmek maksadı ile yerel eylemler gerçekleştirmesine olanak tanımaktadır. Öğrenme sürecinde ajan çevre ile etkileşim içerisinde ve bir eylem gerçekleştirildiğinde geri bildirim almaktadır. Bir eylemin ardından tatmin edici bir durum gelmesi eylemi üretme eğiliminin güçlendirmesine ile sonuçlanacaktır. Bu durum pekiştirmeli öğrenmenin altında yatan temel mantıktır. Pekiştirmeli öğrenmede, ajan çevresi ile etkileşim yolu ile bir ödül ya da ceza geri bildirimini alarak performansını artırmaktadır. Pekiştirme sinyali eylemin uygunluğunun bir göstergesidir. Öğrenme gerçekleşirken makineye hangi eylemin yapılmasının gerekli olduğuna dair bir sinyal verilmemektedir. Pekiştirmeli öğrenme sürecinde ajan ve çevre dışında, politika, ödül fonksiyonu, değer fonksiyonu ve karar ufku olmak üzere dört unsur bulunmaktadır. Politika, ajanın belirli bir zaman aralığındaki eylemini tanımlamaktadır. Ödül fonksiyonundan ajanın bir eylemi için anlık olarak bir ödül tanımlamak maksadı ile yararlanılmaktadır. Değer fonksiyonu karar vermek için sonlu ya da sonsuz bir karar ufku tanımlamaktadır. Karar ufkunun sonlu olması sürecin bir şekilde sona ermesini ve gerçekten hiçbir şeyin önemli olmadığı bir süreyi tanımlamaktadır (Gopal 2018; Mohri, Rostamizadeh, ve Talwalkar 2018).

### **2.3. Makine Öğrenmesi Süreci**

Veriden bilgi üretme süreci farklı yetenekleri ve bilgileri gerektiren yaratıcı bir süreçtir. Bu bağlamda, iş dünyasındaki problemlerin anlaşılması ve bunların bir veri madenciliği ya da bir başka ifade ile makine öğrenmesi görevine dönüştürebilmek maksadı ile temel bir sürece ihtiyaç duyulmaktadır. “Cross Industry Standard Process for Data Mining” CRISP-DM sektörlerden ve uygulanacak teknolojiye bağımsız olarak veri madenciliği

ve makine öğrenmesi uygulamalarında ifade edilen problemin aşılmasında yararlanılabilecek bir süreç modeli tanımlamaktadır (Larose ve Larose 2015; Wirth ve Hipp 2000).



Şekil 2.1. CRISP-DM Sürecinin Aşamaları (Wirth ve Hipp 2000, 5)

CRIPS-DM süreci altı aşamadan oluşmaktadır (Wirth ve Hipp 2000, 5-7);

1. Problemin Tanımlanması: Sürecin ilk aşamasında, makine öğrenmesi projesinin hedeflerini ve gereksinimlerini anlama ve takip eden süreçte, üretilen bu bilgiyi bir makine öğrenmesi problemi tanımına ve ön proje planına dönüştürmeye odaklanılmaktadır,

2. Veriyi Anlama: Sürecin ikinci aşamasında, ilk verinin toplanması ile başlamakta ve veriye aşinalık kazanmak, veri kalitesi ile ilgili sorunları tanımlamak, veriyi keşfetmek ve verideki gizli bilgileri açığa çıkarmak amacıyla hipotezler oluşturma faaliyetleri ile devam etmektedir. Veriyi anlama aşaması ile problemin tanımlanması aşaması arasında yakın bir bağlantı bulunmaktadır. Problemin ve sürecin formülasyonu için azda olsa verinin anlaşılması gerekmektedir. Veriyi tanımak ve yapısını anlamak maksadı ile keşfedici veri analizi kullanılmalıdır. Verinin kalitesi değerlendirilmeli ve gerekli görüldüğü takdirde işlenilebilir örüntüler içerebilecek alt kümeler seçilmelidir,

3. Veri Hazırlama: Ham veriden nihai veri setini oluşturmak için gerekli olan tüm faaliyetleri kapsamaktadır. Bu aşamasında yürütülen faaliyetleri belirli bir sırası bulunmamaktadır. Çünkü süreç içerisinde bir faaliyetin birden çok kez yürütülmesi muhtemeldir. Veri hazırlama aşamasında yürütülen görevler arasında verinin temizlenmesi, öznitelik seçimi, veri dönüştürme gibi görevler yer almaktadır. Bu aşamada ham verinin modelleme için hazır hale gelmesi için gerekli tüm faaliyetler yürütülmektedir,

4. Modelleme: Çeşitli makine öğrenmesi yöntemleri seçilir, eğitilir ve parametreleri optimize edilir. Genellikle bir problemin çözümü için birden fazla metot mevcuttur. Bazı metotlar belirli bir veri türünü gerektirmektedir ve genellikle bir problem alanına birden fazla benzer metot uygulanmaktadır. Veri hazırlama ve modelleme aşamaları arasında yakın bir bağlantı bulunmaktadır. Modelleme aşamasında verinin sahip olduğu problemler ile ilgili fikir üretilir. Teknikler arasından uygun olanının seçilmesi ve optimize edilmesi gerekmektedir. Belirli metotların veri gereksinimlerinin veya varsayımlarının sağlanması için veri hazırlama aşamasını dönülmesi ve modellemenin yinelenmesi mümkündür,

5. Değerleme: Sürecin bu aşamasında, modelleme aşamasında eğitilen bir ya da birden fazla model hedeflere ulaşıp ulaşılmadığında emin olmak amacı ile kapsamlı bir şekilde değerlendirilmektedir,

6. Yayılım: Modelin oluşturulması ile süreç tamamlanmaktadır. Modelin ihtiyaca uygun olarak düzenlenmesi ve sunulması gerekmektedir. Bu nedenle eğitilecek olan modellerden gerçekten faydalanabilmek için hangi eylemlerin gerçekleştirilmesi gerektiğinin önceden anlaşılması oldukça önemlidir. Nihai modelin ortaya koyulmuş olması sürecin tamamlandığı anlamına gelmemektedir. Ayrıca bu modelin uygulanması gerekmektedir. Temel konuşlandırma bir rapor üretimi olabilir.

#### **2.4. Veri, Veri Seti ve Veri Ön İşleme Kavramları**

Geleneksel olarak karar problemlerinin çözümünde, problemle ilgili değişkenler tanımlanır, değişkenler arasındaki ilişki geçmişteki bilgilere uyum gösteren ve gözlemlenen verileri kapsayan matematiksel denklemler cinsinden ifade edilir. Kararlar elde edilen bu denklemlerin analitik ve sayısal çözümleri yolu ile verilir. Deneyimden öğrenmek makine öğrenmesi için temel bir kavramdır. Makine öğrenmesi, bilgisayarların

eylemlerini, bu eylemlerin deneyimle doğruluk açısından gelişecek şekilde güncellemesini sağlamak ile ilgilidir (Gopal 2018).

Başarılı bir makine öğrenmesi modelinin oluşturma süreci üç aşamadan oluşmaktadır. İlk aşamada, öğrenme işlevinin gerçekleştirilmesi için gerekli olan verinin toplanması gerekmektedir. Bu aşamada toplanacak olan ve eğitim sürecinde kullanılacak olan veriler, nihai modelin gelecekte tahmin yürüteceği aynı koşullar altında toplanmalıdır. Bu yolla toplanan veriler genellikle alan içi veriler olarak adlandırılmaktadır. Pek çok makine öğrenmesi algoritması, eğitim sürecinden önce insan unsurların verileri etiketlemesini içeren bir etkileşime ihtiyaç duymaktadır. Ayrıca makine öğrenmesi modelinin nihai performansı da büyük ölçüde mevcut eğitim verilerinin miktarı tarafından belirlenmektedir ve daha çok alan içi veriye ulaşmak modelin performansını artırmanın en etkili yollarından birisidir. İkinci aşamada, öznitelikleri ham veriden çıkarmak maksadı ile genellikle alana özgü prosedürler uygulanmaktadır. Bu aşamada özniteliklerin kesif olması ancak aynı zamanda ham verideki en önemli özellikleri koruması oldukça önemlidir. Öznitelik çıkarım sürecinin, verinin doğasına ve alan bilgisine göre manuel olarak tasarlanması gerekmektedir ve bu süreç alandan alana değişiklik göstermektedir. Son aşamada ise özniteliklerden matematiksel modeller oluşturmak maksadı ile bir makine öğrenmesi algoritması seçilmektedir (Jiang 2021).

Makine öğrenmesi sürecinde büyük veri setleri, büyük veri kavramını karşılamamaktadır ve modelleme büyük veri ile yürütülmemektedir. Eğitimin gerçekleştirileceği veri setlerinin modelleme aşaması öncesinde ön işleme tabi tutulması gerekmektedir. Çok yönlü veri araştırması ve büyük boyutlu veri setleri ile modelleme, sürecin zaman maliyetlerini artırmakta ve analizi yürütülemez hale getirebilmektedir. Bir veri setinin boyutsallığı, veri setindeki özniteliklerin sayısı ile kavramlaşmaktadır. Özniteliklerin sayısındaki önemli derecedeki artış, eş zamanlı olarak modelleme sürecini zorlaştırmakta ve bu durum boyutsallık laneti olarak adlandırılmaktadır. Gerekenden fazla, alakasız özniteliğe modelde yer verilmesi makine öğrenmesi algoritmasının performansını da düşürebilmektedir. Bu nedenle öznitelik sayısının optimize edilmesi gerekmektedir. Öznitelik sayısı haricinde veri kalitesini doğrudan etkileyen sorunların önüne geçmek maksadı ile modelleme aşaması öncesinde veri ön işleme tabi tutulmalıdır. Eksik gözlemler ile ilintili sorunların önüne geçmek maksadı ile araştırmacılar pek çok yöntem isnat etmektedir (Singh 2022).

Makine öğrenmesi sürecinde kullanılacak olan tüm veri setleri, ham haldeyken dağınık bir yapıya sahip olma potansiyeline sahiptir. Veri setlerinin eksik gözlemler ve hatalı değerler içeriyor olması muhtemeldir. Yapısal sorunlara sahip olmadığı düşünülen veri setlerine de şüphe ile yaklaşılması gerekmektedir. Veri kalitesinin sağlanmasının tek gerçek yolu doğru veri setinin doğru tasarlanmış olması ve analiz öncesinde veri ön işleme sürecine sokulmuş olmasıdır.

Yapay zekâ uygulamalarında genellikle veriler hata ve gürültü içermekte ve modellemeyi güçleştiren bu durumlar ile birlikte ölçeklenmesi, dönüştürülmesi veya farklı ve muhtemelen homojen olmayan bilgi kaynaklarından toplanması gerekmektedir. Veri üzerinden stokastik ve deterministik olmak üzere iki tip hata tanımlamak mümkündür. Deterministik hataların bazı durumlarda kolaylıkla düzeltilmesi mümkündür. Aykırı değerler ve örtük değerler belirlenebilir ve sonrasında düzenlenebilir ya da veri setinden silinebilirler. Filtreleme yolu ile örtük değerler, aykırı değerler ya da gürültü azaltılabilir (Runkler 2020).

İlk ham veriler her durumunda, makine öğrenmesi süreci için en uygun veriler değildir. Farklı makine öğrenmesi problemlerinin çözümü için daha yararlı verileri üretmek maksadı ile farklı işlemlere ihtiyaç duyulabilmektedir. Veri ön işleme ya da hazırlanma aşamasında iki temel görev bulunmaktadır (Kantardzic 2011).

- Verinin bilgisayar tabanlı araçlarda kullanılmak üzere standart bir forma düzenlenmesi,
- Model performansını maksimize edecek veri setinin hazırlanması.

#### **2.4.1. Veri ve Veri Seti Kavramları**

Herhangi bir makine öğrenmesi yönteminin ya da probleminin en önemli bileşeni veridir. Veri, bilgi taşıyıcılarının atomik birimleri olan bireysel veri noktaları kümesi olarak kabul edilmektedir. Veri noktaları farklı kavramı temsil edebilmektedir. Veri noktaları kavramı çok soyut ve esnek bir tanımlamadır ve temelde farklı uygulama etki alanlarında ortaya çıkan çok farklı nesne türlerini temsil edebilir. Birçok makine öğrenmesi yöntemi bir dizi referans ya da bir başka ifade ile eğitim veri noktasının ortalamasını alarak belirli ilgi düzeyi belirlemek için tahminler oluşturmaktadır. Oluşturulan tahminler ortalamayı hesaplamak maksadı ile kullanılan artan sayıda veri noktası için doğru hale gelmektedir. Bir veri kümesinin anahtar parametresi, içerdiği veri noktalarının " $m$ " sayısı olarak tanımlanmaktadır. Bir veri kümesindeki veri noktalarının sayısı aynı zamanda örneklem

büyüküğü olarak da adlandırılmaktadır. Makine öğrenmesi çalışmalarda çoğu veri noktası için her bir mikro düzeydeki özeliğine ulaşmak mümkün olmamaktadır. Her bir veri noktasının iki farklı özellik grubu bulunmaktadır. Birinci özellik grubu öznelik olarak adlandırılmaktayken, ikinci özellik grubu etiket olarak adlandırılmaktadır. Uygulama alanına bağılı olarak etiketler hedef ya da çıktı deęişkeni olarak tanımlanabilmektedir. Ayrıca veri noktalarının öznelikler bazı durumlarda girdi deęişkeni olarak da adlandırılmaktadır (Jung 2022, 19-21).

Bir veri seti, bir ortamdan ya da bir süreçten elde edilen bir dizi ölçümdür. En temel tanım ile veri seti bir nesne koleksiyonudur ve her nesne için aynı ölçülerini içermektedir. Veri setlerine, farklı biçimlerde veriler gelmektedir. Bu bağlamda bir taksonominin geliştirilip geliştirilemeyeceğı tam olarak belli deęildir. Bunun sebebi, bir durumda önemli olarak nitelenen verinin bir başka durumda önemsiz olarak nitelenebilme olasılığıdır. Veri ilgili alandaki varlıkların, bir deęişkenin deęerini bir varlığın belirli bir özelliğı ile ilişkilendiren bazı ölçüm prosedürleri aracılığıyla sembolik temsile eşleyerek toplanır. Nesnelere arasındaki ilişkiler, deęişkenler arasındaki sayısal ilişkilerle temsil edilir. Bu nicel gösterimler, veri setinde saklanır ve makine öğrenmesi faaliyetinin konusunu oluşturmaktadır (Hand, Mannila, ve Smyth 2001).

Makine öğrenmesi sürecinde temel olarak eğitim, test ve doğrulama olmak üzere üç farklı veri seti kullanılmaktadır. Modelleme aşamasında birden fazla farklı model tahmin edilmekte ve bu modeller arasından seçim yapılmaktadır. Ardından seçilen modelin doğruluğı test edilmektedir. Bu süreçte eğitim, doğrulama ve test olmak üzere üç farklı veri setinden yararlanılmaktadır. Eğitim veri seti modelin eğitiminin gerçekleştirildiğı ya da modelin parametrelerinin tahmini için faydalanılan veri setidir. Doğrulama veri seti model seçimi amacı ile eğitilmiş modellerin riskini tahmin etmek maksadı ile kullanılmaktadır. Model seçim sürecinde tespit edilen en iyi modelin riskini belirlemek için ise test veri setinden faydalanılmaktadır (Hull 2020; Knox 2018, 128-29). Veri setleri temel olarak tek bir olasılık dağılımından üretilmektedir. Veriler üretilirken veri setindeki gözlemlerin birbirlerinden bağımsız olduğı ve aynı şekilde dağıldığı varsayılmaktadır. Makine öğrenmesi sürecinde kullanılacak olan eğitim ve test setlerinin de mutlaka aynı olasılık dağılımından gelmesi gerekmektedir. Eğitim veri seti algoritmanın eğitilmesi ve eğitim hatasının azaltılması maksadı ile kullanılmaktadır. Eğitim aşaması sonrasında, algoritmanın daha önce karşılaşmadığı test setindeki gözlemler ile test hatası hesaplanmaktadır. Test hatası her zaman eğitim hatasında fazla ya da ona eşit

hesaplanmaktadır (Ghatak 2017, 61). Makine öğrenmesi modeli eğitiminde, modelin eğitim veri seti ile iyi uyum göstermesi ancak yeni verileri ile iyi bir genelleme performansı ortaya koyamaması tehlikesi bulunmaktadır. Modelin gelecekte uygulanacak olan verileri temsil gücünün yüksek olması çok önemlidir. Bu nedenle, modelin örneklem dışı test edilmesi gerekmektedir. Bir başka ifade ile model parametrelerinin belirlenmesi amacı ile kullanılan veriden farklı bir veri ile test edilmelidir (Hull 2020).

Bir makine öğrenmesi problemine çözüm aramanın en az iki temel nedeni bulunmaktadır. İlk olarak öğrenme süreci ile ortaya koyulacak olan model sınıflanmamış gözlemler ile ilgili yapılacak olan tahmininde yararlanılacaktır. Bu durum tahmine dayalı öğrenme ya da tahmine dayalı modelleme olarak adlandırılmaktadır. İkinci olarak ise veri üretim süreci hakkında insan anlayışını geliştirmek maksadı ile gözlenen öznitelik vektörleri ile bunlara karşılık gelen sınıf etiketleri tespit etmektir. Bu durum ise açıklayıcı öğrenme ya da açıklayıcı modelleme olarak adlandırılmaktadır (Knox 2018, 13).

Veri noktaları ya da gözlemlerin seçimi gibi gözlemlerin hangi özelliklerinin öznitelik olarak kullanılacağı da makine öğrenmesi sürecinin tasarımı ile ilgilidir. Öznitelikler genel olarak, bir gözleme ait olan kolaylıkla hesaplanabilen ya da gözlemlenebilen özelliklerdir. Ancak bu tanımlamada, bir gözleme ait özelliğin kolay hesaplanabilirliği ile ilgili evrensel bir kriter bulunmamaktadır. Bu nedenle kolay hesaplama olgusu, gözlem özellikleri bağlamında öznel kalmaktadır. Makine öğrenmesi sürecinde kaç adet öznitelik kullanılacağı ile ilgili kesin bir kural bulunmamaktadır. Ancak genel bir kural olarak veri setindeki gözlem sayısının, öznitelik sayısından fazla olması gerektiği kabul edilmektedir. Uygulamada yaygın olarak daha fazla özniteliğin kullanılmasının başarıyı artırdığına yönelik bir kabul bulunmaktadır. Bunun sebebi daha çok özniteliğin daha fazla bilgi aktarımı sağlama potansiyeline sahip olmasıdır. Ancak aşırı sayıda ve gereksiz öznitelige modelde yer verilmesi modelin genelleme performansına zarar verebilmektedir. İstatistiksel olarak modele dahil edilen ek bir öznitelik, veri setine gürültü taşımaktadır. Veri setindeki gözlemlerin özniteliklerinin yanı sıra, gözlem ile ilgili daha yüksek bir olgu ya da ilgi düzeyine temsil eden özellikleri de bulunmaktadır. Bu özelliklere gözlemin etiketi adı verilmektedir. Bir makine öğrenmesi uygulamasında ortaya çıkan veri noktalarının tüm olası etiket değerleri etiket uzayı olarak adlandırılmaktadır. Genel olarak bir gözlemin etiketinin belirlenmesi özniteliklerin belirlenmesinden daha zor bir süreçtir (Jung 2022).

## 2.4.2. Gürültü

Literatürde gürültü kavramının birden fazla tanımı bulunmaktadır. Basit bir ifade ile verinin gürültü içermesi, verinin beklenen işlevi yerine getirememesi durumu olarak açıklanmaktadır. Gürültü, hatalı ya da bozuk ölçümlerden veya insan hatasından kaynaklanabilir. Veride gürültü tespiti sınıflandırma algoritmalarının uyarlanması ya da veri ön işleme için gürültü filtrelerinin kullanılması ile gerçekleştirilebilir. Gürültü hedef sınıfta ve tahminde kullanılan özniteliklerde var olabilmektedir. Bir gözlemin gürültü olduğundan kesin olarak emin olmak mümkün değildir. Gürültü olarak tanımlanan gözlemler mevcut yapıyı takip etmeyen doğru gözlemlerde olabilmektedir. Bu durum ise analistin alan bilgisinin süreçte oldukça önemli olduğunu göstermektedir (Moreira, Carvalho, ve Horváth 2018).

## 2.4.3. Örtük Değer ve Aykırı Değer Kavramları

Örtük değerler ve aykırı değerler (inliers – outliers), bireysel veri değerlerindeki hatalar olarak tanımlanmaktadır. Örtük değerler ve aykırı değerler, stokastik ve deterministik etkilerden aşırı bireysel ölçüm hatalarında, veri iletimindeki kayıplardan ve elle veri girişlerinde verinin yanlış alanlara kaydedilmesi gibi yazım hatalarından kaynaklanabilir. Diğer deterministik hata türleri, türetilmiş verilerin hesaplanması için yanlış formüllerin kullanılması, yanlış ölçeklendirme veya yanlış ölçümden kaynaklanmaktadır. Hata sistematığı biliniyorsa bu tip hatalar düzeltilebilmektedir. Aykırı değerler bir veri setindeki, anormal değerler ya da gözlemlerdir. Aykırı değerler bir ya da birden fazla öznitelik için değeri diğer gözlemleri aynı özniteliklerinde bulunan değerlerinden çok farklı olan nesnelere olarak da tanımlanmaktadır. Verideki gürültünün aksine, aykırı değerler geçerli değerler olabilir (Moreira, Carvalho, ve Horváth 2018).

Veri setinde, genel davranışa uymayan gözlemlerin bulunması makine öğrenmesi çalışmalarında sık sık karşılaşılan bir sorundur. Aykırı değerler ölçüm hatasında kaynaklanabileceği gibi içsel veri değişkenliğinin bir sonucu da olabilir. Örneğin veri setinde bir işletmenin personel sayısının “-1” olarak tanımlanmış olması değerlerin mutlak olarak yanlış olduğu göstermemektedir. Aykırı değer tespitini maksadı ile kullanılan yöntemler bu değerlerin tespit edilip uygun ise veri setinden çıkarılmasını amaçlamaktadır. Aykırı değerler mekanik hatalar, sistem hataları, insan hatası gibi veri popülasyonundaki doğal sapmalardan ortaya çıkmaktadır. Aykırı değerlerin tespiti ve veriden temizlenmesi noktasında analistin çok dikkatli olması gerekmektedir. Analistin

aykırı deęerleri popülasyon dıřarısında bırakmak hususundaki isteęi, dikkatsiz seçimler ile veriden bilginin eksilmesine de neden olabilmektedir (Kantardzic 2011).

Bir veri setindeki aykırı deęerlerin belirlenmesi maksadı ile yararlanılabilecek iki yol bulunmaktadır. Birincisi, her bir öznitelik için minimum ve maksimum deęerlerin incelenmesi ve bunların uygun gözlemler olup olmadığını tespit için alan bilgisinden faydalanmaktır. İkinci yöntem, özniteliğin medyan, minimum, maksimum, 1. çeyrek ve 3. çeyrek deęerleri arasındaki farkı karşılařtırmaktır. 3. çeyrek ile maksimum deęer arasındaki farkın, medyan ile 3. çeyrek arasındaki farktan belirgin şekilde daha büyük olması, maksimum deęerin alışılmadık olduğunu ve gözlemin muhtemel bir aykırı deęer olduğunu gösterir. Benzer şekilde, 1. çeyrek ile minimum deęer arasındaki farkın, medyan ile 1. çeyrek arasındaki boşluktan belirgin şekilde daha büyükse, bu, minimum deęerin alışılmadık olduğunu ve muhtemelen bir aykırı deęer olduğunu gösterir. Kutu grafiklerinde gösterilen aykırı deęerler de bu karşılařtırmanın yapılmasına yardımcı olur. Histogramlardaki üstel veya çarpık daęılımlar da aykırı deęerlerin varlığının iyi bir göstergesidir (Kelleher, Mac Namee, ve D'Arcy 2020, 65-66).

#### **2.4.4. Eksik Gözlem**

Makine öğrenmesi çalışmalarında veri setinin eksik gözlem içermemesi durumu çok nadir karşılaşılan bir durumdur. Bazı makine öğrenmesi yöntemleri eksik gözlem ile çalışabilme yeteneğine sahip olmasına karşın bazıları veri setinde eksik gözlemin bulunmamasını gerektirmektedir. Veri setinin eksik gözlem içermesi, uygulamada bu gözlemlerin ne yapılacağı sorusunu ortaya çıkarmaktadır. Bu bağlamda ilk olarak eksik gözlemler veri setinden çıkartılması tercih edilebilir. Ancak bu seçim büyük veri setlerinde nispi kayıplara neden olurken, görece daha küçük veri setlerinde büyük bilgi kayıplarına neden olabilmektedir. Bu nedenle eęer gözlemlerin veri setinden çıkarılmaması tercih edilirse, deęer üretmek için faydalanılan yaklaşımlar mevcuttur (Kantardzic 2011);

Birinci yaklaşımda veri analisti, alan uzmanları ile birlikte eksik gözlemleri deneyime dayanarak makul ve olası deęerler ile tamamlamayı tercih edebilir. Bu yöntem az sayıda eksik gözlemin bulunduğu ve nispeten küçük veri kümeleri için oldukça basit ve uygundur. Ancak her durum için makul bir deęer bulunmaması veri setine gürültü eklenmesi ile sonuçlanabilmektedir.

İkinci yaklaşımda, eksik gözlemlerin farklı yöntemlerden yararlanılarak tamamlanma amaçlanmaktadır. Tüm eksik gözlemlerin tek bir genel sabit ile tamamlanması, eksik gözlemlerin öznitelik ortalaması ile tamamlanması veya eksik gözlemlerin sınıf ortalaması ile tamamlanması yaklaşımlarından birisi seçilebilir.

#### 2.4.5. Normalizasyon

Farklı aralıklara sahip özniteliklerin modelleme öncesinde sıklıkla standardize edilmesi gerekmektedir (Runkler 2020). Farklı aralıkları kapsayan sürekli öznitelikler içeren veri setleri model eğitiminde çeşitli sorunların ortaya çıkmasına neden olabilmektedir (Kelleher, Mac Namee, ve D'Arcy 2020). Bazı makine öğrenmesi metotları, genellikle  $n$  boyutlu bir uzayda gözlemler arasındaki mesafe hesaplamasına dayalı olanlar, iyi model sonuçları için normalize verilere ihtiyaç duymaktadır. Normalizasyon işlemi ile ölçülen değer  $-1$  ile  $1$  ya da  $0$  ile  $1$  arasında değerlere ölçeklendirilmektedir. Gözlem değerleri normalize edilmez ise doğal olarak mesafe ölçümü ortalama olarak daha büyük değerlere sahip olan öznitelikleri artıracaktır. Normalizasyon için literatürde farklı yöntemler kullanılmaktadır (Kantardzic 2011);

- Ondalık Ölçeklendirme, ondalık hane taşır ancak orijinal basamak değerinin çoğunu korumaktadır. Veri değerlerini  $-1$  ile  $1$  arasında tutmaktadır. Aşağıdaki denklemde  $i$  durumu için  $x$  değerinin nasıl normalize edileceği gösterilmektedir.

$$x'(i) = x(i) / 10^k$$

- Min – Max Normalizasyonu, değerlerin tam bir aralıkta normalize dağılımını elde etmek için kullanılmaktadır.

$$x'(i) = (x(i) - \min[x(i)]) / (\max[x(i)] - \min[x(i)])$$

- Standart Sapma Normalizasyonu, genellikle mesafe ölçüleri iyi uyum göstermektedir. Ancak veriyi orijinal veriden tanınamayacak bir forma dönüştürmektedir. Bir  $x$  özneliği için tüm ortalamalar ve standart sapmalar hesaplanmaktadır. Daha sonra  $i$  gözlemi için öznitelik değeri denklem yardımı ile dönüştürülmektedir.

$$x(i) = (x(i) - \text{ort}(x)) / \text{std}(x)$$

Normalizasyon teknikleri özneliğin değerleri arasındaki görece farkları korurken, sürekli özneliğin değerlerini belirli bir aralığa gelecek şekilde değiştirmek için kullanılmaktadır. Normalizasyonda en basit yaklaşım, özneliğin değerinin belirli bir aralığa doğrusal olarak normalize edilmesini sağlayan aralıklı normalizasyon ya da Min-Max normalizasyonudur (Kelleher, Mac Namee, ve D'Arcy 2020).

#### 2.4.6. Aşırı Uyum ve Yetersiz Uyum Sorunları

Makine öğrenmesinde karşılaşılan en önemli iki sorun aşırı uyum ve yetersiz uyumdur. Makine öğrenmesi yöntemleri, eğitim seti olarak işlev gören etiketli veri noktalarından minimum kayba neden olan bir  $\hat{h} \in H$  hipotezinin öğrenmektedir. Problemi çözen bir hipotez tarafından elde edilen minimum ortalama kayıp, modelin eğitim hatası olarak adlandırılmaktadır. Yüksek boyutlu bir hipotez uzayına sahip olduğunda, tipik olarak eğitim setine mükemmel şekilde uyan bir hipotez elde etmek çok olasıdır. Böylece çok düşük bir eğitim hatası ile karşı karşıya kalınabilir. Ancak eğitim seti dışındaki veri noktalarında kayıp büyüebilmektedir. Modelin öğrendiği bir hipotez tarafından elde edilen minimum eğitim bazı durumlarda oldukça yanıltıcı olabilmektedir. Çok fazla sayıda öznitelik ile öğrenen modeller, küçük bir eğitim hatasına sahip olabilmektedir. Ancak bu modelin öğrendiği hipotezin eğitim seti dışında büyük kayıplara neden olması aşırı uyum (overfitting) olarak tanımlanmaktadır. Aşırı uyumun aksine, yetersiz uyum (underfitting) modelin verilere iyi uyum göstermediği durumlarda gözlemlenmektedir. Bir modelin aşırı uyum durumuna sahip olup olmadığını eğitim ve doğrulama hataları karşılaştırılarak tespit edilebilmektedir. Doğrulama hatası ayrıca tüm bir hipotez uzayı veya model için de bir kalite ölçüsü olarak da kullanılabilir. Bir makine öğrenmesi modeli tarafından öğrenilmiş hipotez, eğitim setindeki gözlemler ile daha iyi performans ortaya koyma eğilimindedir. Bu bağlamda öğrenilmiş olan hipotezi doğrulamanın altında yatan temel mantık, eğitim setinde kullanılmayan gözlemler ile model kaybının tekrar hesaplanmasıdır. Aşırı uyum ve yetersiz uyum sorunları model kapasitesi ile ilgilidir. Kapasitesi düşük olan modeller yetersiz uyum sorunu gösterirken, yüksek kapasiteli modeller aşırı uyum sorunu gösterebilmektedir (Jung 2022; Ghatak 2017).

#### 2.4.7. Öznitelik Seçimi

Makine öğrenmesi uygulamalarında, bazen çok büyük boyutlu veri setlerinden faydalanılmaktadır. Büyük ve çok fazla değişkene sahip olan yüksek boyutlu problemler, boyutsallık lanetini meydana getirmektedir. Bu sorunun çözümü için en tipik yaklaşım, öznitelik seçimidir (Van der Aalst 2016). Makine öğrenmesi uygulamalarında en gelişmiş algoritmalar dahi çok sayıda ve zayıf ilişkili özneliğin üstesinden gelme kabiliyetine sahip değildir. Ayrıca yüksek boyutlu veri setleri hesaplama maliyetlerini artırmaktadır. Yapılan araştırmalar, veri setindeki öznitelik sayısının azaltılmasının hesaplama süresini kısalttığını ortaya koymaktadır. Bununla birlikte gereksiz öznitelikler içeren veri setleri

modellerde aşırı uyum probleminin ortaya çıkmasına da neden olmaktadır. Bu nedenle modelleme öncesinde performans kaybına uğramadan gereksiz özniteliklerin veri seti içerisinde çıkarılması gerekmektedir. Makine öğrenmesi uygulamalarında en az maliyetle maksimum genelleme performansının elde edilmesi amaçlanmaktadır. Bu bağlamda öznitelik seçimi ile aşağıdaki sonuçlara ulaşılması hedeflenmektedir (Kantardzic 2011; Moreira, Carvalho, ve Horváth 2018);

- Daha az veri ile algoritmanın daha hızlı öğrenmesini sağlamak,
- Yüksek genelleme performansı sağlamak için yüksek model doğruluk düzeyi elde etmek,
- Modelin anlaşılabilirliği ve kullanımını kolaylaştırmak için sürecin sonuçlarını basitleştirmek,
- Daha az öznitelik ile eğitim gerçekleştirmek yolu ile sonraki veri toplama aşamalarında kolaylık sağlamak.
- Veri setindeki gürültünün azalmasına olanak tanımak,
- Modellerin basitleşmesine ve daha kolay yorumlanabilmesine imkân vermek,
- Veri seti üzerinden hazırlanan görsellerin anlaşılabilirliğinin artmasına ve büyük veri setlerinin görselleştirilebilmesine olanak tanımak.

Makine öğrenmesi algoritmaları etiketli ve etiketsiz veri setleri üzerinden öğrenmektedir. Bir makine öğrenmesi algoritmasının performansı özniteliklerin doğru seçimine bağlıdır. Makine öğrenmesi problemi ile en alakalı özniteliklerin seçimi için algoritmaya en uygun yöntemin seçilmesi gerekmektedir. Öznitelik seçimi uygulamanın sonucuna önemli katkı sağladığı için bu sürece önem verilmelidir. Genel olarak herhangi bir uygulamada özniteliklerin bir kısmı doğrudan hedef sınıfla ilgili olmaktadır ve bazı öznitelikler dönüştürme teknikleri uygulanarak ilgili kümelere dönüştürülebilmektedir. Bunların dışında kalan öznitelikler problem ile ilgisizdir (Singh 2022).

Öznitelik sayısının azaltılması için iki alternatif yaklaşım mevcuttur. Bunlardan ilki öznitelik birleştirmedir. Öznitelik birleştirmede, bir öznitelik grubu, gruptaki özniteliklerin kombinasyonu ile değiştirilmektedir. Bu süreçte gözlemleri en iyi ayırt eden öznitelikler seçilmektedir. Öznitelik birleştirme metotları orijinal veri setini yeni ancak daha düşük boyutlu bir alana yansıtır ve bu süreçte veri setinde herhangi bir bilgi kaybı oluşmaz. Öznitelik birleştirme maksadı ile temel bileşen analizi (principal component analysis), bağımsız bileşen analizi (independent component analysis), çok boyutlu ölçekleme (multi dimensional scaling) gibi tekniklerden faydalanılmaktadır.

İkinci yaklaşım öznitelik seçimidir. Öznitelik seçim sürecinde veri setindeki özniteliklerin sayısının bilgi kaybına neden olmadan azaltılmak amaçlanmaktadır. Öznitelik seçimi, öğrenme sürecini daha az sayıda işlem gerektirmesine olanak tanıdığı için hızlandırmaktadır (Moreira, Carvalho, ve Horváth 2018).

Öznitelik seçimi, makine öğrenmesi uygulamasının yürütüldüğü alan bilgisine ve uygulamanın hedeflerine bağlı olarak, analist tarafından ilk veri setindeki özniteliklerin bir alt kümesinin seçilmesi anlamına gelmektedir. Bu süreç analist tarafından manuel olarak da yürütülebileceği gibi analist çeşitli tekniklerden de faydalanabilmektedir. Genel olarak öznitelik seçim sürecinde faydalanan teknikler filtreleme, sarmalama ve gömülü olmak üzere üç sınıfta tartışılmaktadır. Bu üç yaklaşım temel olarak, öğrenme algoritmasının seçim sürecine nasıl dahil olduğuna göre farklılaşmaktadır. Filtreleme yöntemleri herhangi bir öğrenici algoritma dahili olmadan bir veri ön işleme faaliyet şeklinde yürütülmektedir. Süreç içerisinde özniteliklerin bir alt kümesini seçmek için yararlanılan bir arama elemanından yararlanılmaktadır. Sarmalama yöntemlerinde bir öğrenici algoritma kullanılmaktadır ve özniteliklerin alt kümesi bu öğrenici algoritmanın genelleme performansına dayanarak belirlenmektedir. Bu yaklaşımın temel dezavantajı, öznitelik kombinasyonlarının ardışık olarak denenmesi sebebi ile hesaplama karmaşıklığının artmasıdır. Gömülü yöntemler, öznitelik seçimi ve öğrenme algoritmasını tek bir optimizasyon formülasyonunda birleştirmektedir. Veri setindeki gözlemlerin sayısı ile veri setinin boyutları çok fazla olduğunda hesaplama verimliliği elde etmek ve öğrenme metodolojisinin sübjektifliğinden sakınmak maksadı ile genellikle filtreleme yaklaşımları tercih edilmektedir. (Kantardzic 2011).

Öznitelik seçimi maksadı ile ister bir filtreleme yöntemi isterse de bir sarmalama yöntemi seçilmiş olsun, en iyi öznitelik alt kümesinin nasıl belirleneceğini tanımlayan bir arama kriterine mutlaka sahip olunması gerekmektedir. En basit arama stratejisi, tüm olası öznitelik kümelerinin değerlendirildiği ve en iyi alt kümenin seçildiği ayrıntılı aramadır. Öznitelik sayısı eğer çok fazla ise bu süreç oldukça uzun sürebilmektedir. Öznitelik seçimi için basit ve verimli olan iki diğer arama stratejisi ileriye doğru seçim ve geriye doğru seçimdir. İleriye doğru seçimde, seçim süreci boş bir öznitelik kümesi ile başlamaktadır. Ardından mevcut her bir öznitelik başına bir model eğitilir ve test edilir. En iyi tahmin performansına sahip öznitelik seçilir. Daha sonra, öznitelik alt kümesini değerlendirmek için kullanılan ölçümle ilgili olarak iki özniteliğin en iyi alt kümesini veren kalan özniteliklerden en iyi ek özniteliği seçerek ikinci bir öznitelik eklenir. Bu ek

özniteliklerden hiçbiri öznitelik alt kümesini iyileştirmese hiçbiri eklenmez ve seçim durur. İkinci bir öznitelik eklenirse, süreç, alt küme iyileştirmeyi durdurana kadar kalan öznitelikler için birer birer eklenerek devam eder. Geriye doğru seçim ters yönde çalışmaktadır. Süreç tüm öznitelikleri içeren bir küme ile başlar. Ardından, alt kümenin tahmin performansı azalmaya başlayana veya öznitelik sayısı bire eşitlenene kadar öznitelikler birer birer kümeden çıkartılır. Bir özelliğin çıkartılması her değerlendirildiğinde, en az zarara neden olan özellik çıkartılmaktadır (Moreira, Carvalho, ve Horváth 2018).

#### **2.4.8. Çapraz Doğrulama**

Makine öğrenmesi modelinin başarısı, hipotez uzayı karmaşıklığı ve örneklem karmaşıklığına dayanmaktadır. Ancak iki unsur birbirinden bağımsızdır. Makine öğrenmesinin temel amacı, karmaşıklık açısından en basit ve verilerdeki ampirik hata açısından en iyi fonksiyonu tespit etmektir. Bu seçimin, modelin genelleme performansı açısından en iyi sonucu vermesi beklenmektedir ve ortaya koyulan modelin yeni veriler üzerinde nasıl bir sonuç ortaya koyacağı belirsizdir. Eğitim verilerinin ortaya koyduğu performans, eğitimde yer almayan verilerin performansının bir göstergesi olmaya uygun değildir. Ayrıca test verileri de hiçbir durumda model eğitmek için kullanılamamaktadır. Makine öğrenmesinde bazı eğitim şemaları iki aşama içermektedir. Birinci aşamada veri içerisindeki temel yapı ya da bir başka ifade ile örüntüler ortaya çıkarılmaktadır. İkinci aşamada ise parametreler optimize edilmektedir. Bu iki aşama için farklı veri setlerinin kullanılması gerekmektedir. Bu gibi durumlarda, eğitim, doğrulama ve test olmak üzere üç farklı veri seti dikkate alınmaktadır. Optimizasyon veya seçim aşamasında mükemmel performans elde etmek için doğrulama seti eğitim setinden farklı olmalıdır ve gerçek değerlerin güvenilir bir tahminini elde etmek için test verileri her ikisinden de farklı olmalıdır (Gopal 2018, 56-57).

Büyük veri kümeleri, modelin eğitimi için bir eğitim veri seti, modeli doğrulamak için bir doğrulama veri seti ve modeli değerlendirmek için bir test veri setine bölünebilir. Ancak büyük bir veri setine sahip olunmadığı durumlarda, çapraz doğrulama özellikle yararlıdır. Çapraz doğrulama, model seçimine ve modeldeki hatayı tahmin etmeye olanak tanımaktadır. Çapraz doğrulama kullanılarak model seçimi, belirli bir veri seti ile eğitilmiş bir dizi model arasından bir model ve/veya belirli bir modelin hiperparametrelerini seçerek yapılmaktadır. Temel ayırma yaklaşımının, eğitim seti ve

doğrulama seti olarak, birçok varyasyonu ve uzantısı bulunmaktadır. K-katmanlı çapraz doğrulamada bu uzantılardan en popüler olanıdır. Çapraz doğrulama işlemi ilk veri setinin  $k$  adet alt veri setine ayrılması ile başlar. Bu çapraz doğrulamanın temel prensibidir. Ayrılan her bir alt veri seti  $k-1$  kombinasyonu eğitim için kullanılır, dışarıda kalan alt veri seti ile test yürütülür. Sonuç olarak işlem  $k$  defa tekrarlanır ve tüm tekrarlar üzerinden eğitim ve doğrulama hatalarının ortalaması alınır. Burada süreçteki temel problem  $k$ 'nın kaç olacağıdır. Düşük bir  $k$  değeri modelin yanlılığı artırırken, yüksek  $k$  değeri ise varyansı artırmaktadır. Bu nokta genel kural  $k$  değerinin 10 kabul edilmesidir. Çapraz doğrulama ile elde edilen ortalama hata tek bir bölme ile elde edilen hatadan daha iyi bir tahmin edicidir. Çapraz doğrulama, model hatasını genelleme hatasına mümkün olduğu kadar yakınsamasını sağlamaktadır. Ayrıca her yinelemede modeldeki varyansın gözlemlenmesine olanak tanımaktadır. Farklı modeller ve farklı hiperparametreler ile çapraz doğrulama yapıldıktan sonra hata ve varyans açısından en iyi performansı gösteren model seçilmektedir (Ghatak 2017, 72; Jung 2022, 120).

Çapraz doğrulama işleminde 10 katman kullanılması hatanın tahmin için standart bir yaklaşımdır. Çeşitli öğrenme teknikleri ve çok sayıda veri kümesi üzerinde yürütülen kapsamlı testler en iyi sayısı uygulamada 10 olduğunu ortaya koymuştur. Ayrıca bunu destekleyen teorik kanıtlarda mevcuttur. Kesin bir tahmin araması durumunda standart süreç, 10 katmanlı çapraz doğrulama prosedürünün on kez tekrarlanması ve ortalamasının alınmasıdır (Gopal 2018, 57-58).

## **2.5. Makine Öğrenmesi Algoritmaları**

### **2.5.1. Lojistik Regresyon**

Boyutsallık lanetinin önüne geçmenin bir yolu, veri yapısı ya da veri dağılımı ile ilgili varsayımlarda (sınıflama problemleri için,  $P(Y/X)$ , kümeleme problemleri için  $P(X)$ ) bulunmaktadır. Tümevarımsal yargı olarak bilinen bu varsayımlar genel olarak sabit sayıda parametreye sahip istatistiksel bir model olan parametrik bir model şeklinde somutlaştırılır.

Doğrusal regresyon, regresyon modelleri arasında en yaygın olarak kullanılan modeldir. Doğrusal regresyon bağımlı değişkenin, girdilerin ya da bağımsız değişkenlerin doğrusal bir fonksiyonu olduğunu varsaymaktadır (Murphy 2012, 19). Doğrusal regresyon, bağımlı değişkenin sürekli olduğu bir regresyon modelidir. Lojistik regresyon ya da logit regresyon ya da logit model bağımlı değişkenin (hedef sınıfın) kategorik olduğu bir

regresyon modelidir. Lojistik regresyon, sınıflama problemlerinin çözümünde yaygın olarak kullanılmaktadır (Ahmed ve Pathan 2019).

$$y(x) = w^T x + \epsilon = \sum_{j=1}^D w_j x_j + \epsilon$$

Yukarıda gösterildiği gibi, genellikle  $\epsilon$ 'nin doğrusal regresyon modellerinde Gauss yani normal dağılıma sahip olduğu varsayılmaktadır. Aşağıda normal dağılım ve doğrusal regresyon arasındaki bağlantı daha açık bir şekilde gösterilmektedir;

$$p(y|x, \theta) = \mathcal{N}(y|x(x), \sigma^2(x))$$

Doğrusal regresyon modeli üzerinde iki düzenleme yapmak sureti ile, sınıflandırma problemleri için uygun hale getirilmektedir. Birinci düzenleme,  $y$  için normal dağılımın, hedef sınıf ikili olduğunda daha uygun olan Bernoulli dağılımı ile değiştirilmesi,  $y \in \{0,1\}$ ;

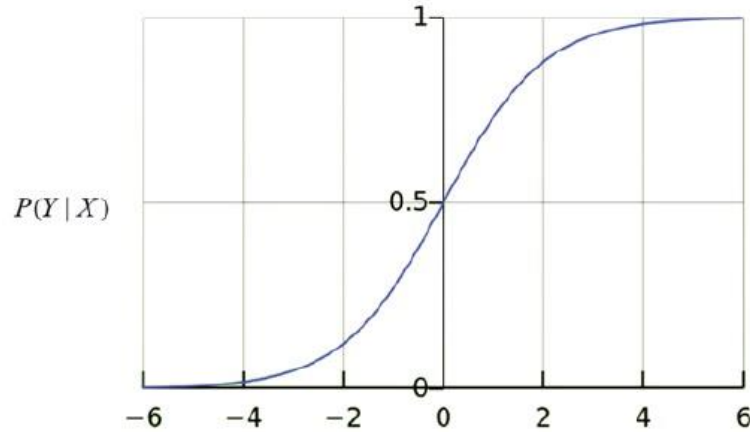
$$p(y|x, w) = \text{Ber}(y|\mu(x))$$

$$\mu(x) = \mathbb{E}[y|x] = p(y = 1|x)$$

İkinci düzenleme ise, doğrusal regresyonda olduğu gibi girdilerin doğrusal bir kombinasyonunun hesaplanmasını içerir ancak  $0 < \mu(x) < 1$  koşulunu sağlayan bir fonksiyondan tanımlanır.

$$\mu(x) = \text{sigm}(w^T x)$$

Yukarıdaki denklemde  $\text{sigm}(w^T x)$  sigmoid fonksiyonu tanımlamaktadır. Sigmoid fonksiyon aynı zamanda lojistik fonksiyon veya logit fonksiyon olarak da adlandırılmaktadır.



Şekil 2.2. Sigmoid Fonksiyon (Awad ve Khanna 2015, 23)

Sigmoid, S şeklinde anlamı taşımaktadır. Lojistik fonksiyon aşağıdaki gibidir;

$$Sigm(\eta) \triangleq \frac{1}{1 + \exp(-\eta)} = \frac{e^\eta}{e^\eta + 1}$$

Lojistik fonksiyon,  $-\infty$  - ve  $+\infty$  arasında bir dizi değeri ( $w_0 + w_1x_i$ ) alabilir ve fonksiyonun çıktısı ( $p(y|x, w)$ ) sadece 0 ile 1 arasındaki değerler ile sınırlıdır.

$$0 < sigm(\eta) < 1$$

Gerekli olan iki düzenleme sonrasında aşağıdaki bağıntı elde edilir;

$$p(y|x, w) = Ber(y|sigm(w^T x))$$

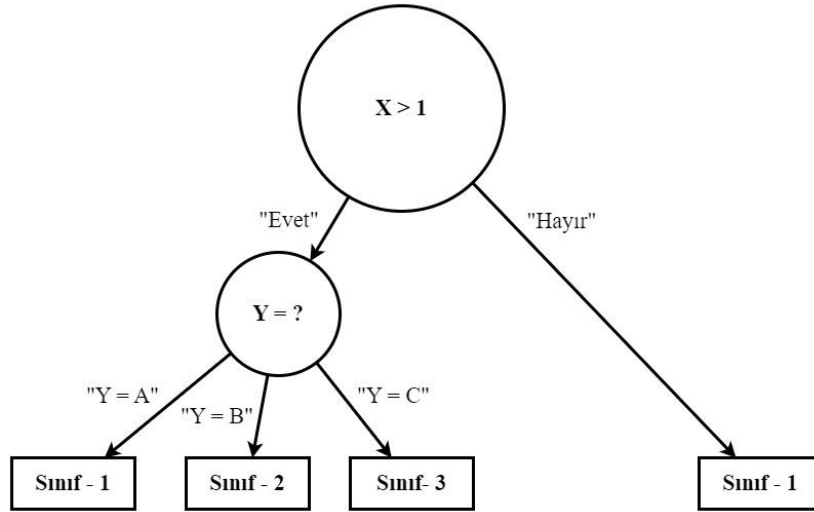
Yukarıdaki denklem “*lojistik regresyon*” olarak adlandırılmaktadır. Bir regresyon modeli olmamasına ve sınıflama modeli olmasına rağmen doğrusal regresyona benzerliği sebebi ile bu şekilde adlandırılmıştır (Murphy 2012, 21).

Lojistik regresyon sonuçları kolaylık yorumlanabilmektedir ve model hiper parametreye sahip değildir. Bu özellikler lojistik regresyonun avantajları olarak sıralanabilmektedir. Lojistik regresyon ikili sınıflama görevleri ile sınırlıdır. Model öznelik sayısından fazla gözlem gerektirmektedir ve aralarında korelasyon olan özneliklere ve aykırı gözlemlere karşı duyarlıdır. Bu hususlar da lojistik regresyonun dezavantajları olarak sıralanmaktadır (Moreira, Carvalho, ve Horváth 2018).

### 2.5.2. Karar Ağaçları

Karar ağacı algoritmaları, ağaç yapısı şeklinde bir model oluşturmaktadır. Model, bir öznelik üzerinden verilebilecek olan kararı gösteren karar düğümleri ile akış şemasına benzer bir dizi mantıksal karar içermektedir. Model içerisindeki düğümler, seçimi gösteren dallara ayrılmakta ve karar ağacı, karar seçimleri kombinasyonunu izlemenin sonucunu gösteren yaprak düğümler (terminal düğüm) ile sonlanmaktadır. Sınıflandırılacak olan veri, öznelik değerlerine göre karar ağacındaki çeşitli kararlardan geçireceği kök düğümden başlar. Karar ağaçlarının hiyerarşik ayrıştırma özelliği, sınıflamada mevcut özneliklerin daha iyi kullanılmasına olanak tanır ve hesaplama verimliliği sağlar. Karar ağacı algoritmaları bazı istatistiksel yöntemlerin aksine tek modlu veya çok modlu veriler ile aynı şekilde eğitim gerçekleştirebilir. Deterministik alanlarda, hedef sınıf ve öznelikler tam olarak belirlenebilmektedir. Ancak bunun aksine eksik problemlerde tam olarak belirlenememektedir. Karar ağacı algoritmalarından her iki durumda da aynı kolaylıkla faydalanılabilir. Karar ağaçları basit ve kolay anlaşılabilir bir şekilde sınıflandırma yapar (Murthy 1998, 346).

Karar ağacı algoritmaları tahmine dayalı modelleyicilerdir. Halen kullanılmakta olan AID (Automatic Interaction Detection) algoritması 1963 yılında geliştirilmiştir. Daha sonra algoritmaya, değişken seçimini iyileştirmek için Ki-kare testi eklenmiş ve CHAIN (Chi square Automatic Interaction Detection) algoritması geliştirilmiştir. Algoritma 1980 yılından günümüze halen kullanılmaktadır. 1970'li yıllarda, (1986 yılında tanıtılmıştır) ID3 algoritması Ross Quinlan tarafından geliştirmiştir. ID3 algoritması ayırma kriteri olarak bilgi kazanımını kullanmaktadır (Quinlan 1986). Quinlan (1993), ID3 algoritmasının, ayırma kriteri olarak bilgi kazanımı yerine, Normalleştirilmiş Bilgi Kazanımı – Kazanım Oranını kullanan geliştirilmiş C4.5 algoritması tanıtmıştır. Sonrasında C4.5 algoritması da Quinlan tarafından geliştirilerek C5.0 algoritması olarak tanıtılmıştır. Algoritmanın karar ağacı oluşturma hızı oldukça artmış ve doğruluk derecesi korunurken daha az sayıda ayırma ile ağaç oluşturulmasına imkân tanınmıştır. Bir diğer karar ağacı algoritması ID3 algoritması ile aynı dönem geliştirilen CART (Classification and Regression Tree) algoritmasıdır (Breiman 1998). CART algoritması Gini indeksini birinci bölme kriteri olarak kullanmaktadır (Abbott 2014).



**Şekil 2.3.** X ve Y Özniteliği ile Eğitilen Basit Bir Karar Ağacı (Kantardzic 2011, 172)

Karar ağacı algoritmalarının sahip olduğu en büyük avantaj, özellikle çok fazla sayıda öznitelik içeren veri setleri ile eğitim yapıldığında, mutlaka tüm öznitelikleri göz önüne almamaları ve öznitelik dereceleri hakkında ek bilgi sağlamalarıdır. Algoritmaların en büyük dezavantajı ise ya ayrıık öznitelikler ile ya da sürekli öznitelikler ile çalışıyor olmalarıdır. Karar ağacı algoritmaları, sınıf sınırlayıcıları yalnızca koordinat eksenine

paralel olabilir, bu sebeple karmaşık doğrusal olmayan sınıf sınırlarının modellenmesi mümkün değildir (Runkler 2020, 113).

Bir karar ağacı, kök düğümden başlayan, yaprak düğümler ile sona eren ve dallarla birbirine bağlanan karar düğümleri topluluğudur. Veri setinde yer alan öznitelikler her bir karar düğümünde test edilir ve alt küme oluşturan bir bölme işlemi ile sonuçlanır. Bir karar ağacı algoritması kabaca aşağıdaki gibi tanımlanmaktadır (Ghatak 2017, 145).

1. Süreç boş bir karar ağacı ile başlar,
2. Bir bölme işlemi gerçekleştirmek için en aza hatayı veren öznitelik seçilir,
3. Her bir bölme işlemi için aşağıda sayılan durdurma koşulları dikkate alınmaktadır,
  - Eğer gözlem sayısı belirtilen minimumdan az ise, bir tahmin yap ve dur,
  - Eğer tüm gözlem noktaları spesifik bir etikete ait ise, dur,
  - Eğer bölünecek bir öznitelik bulunmuyorsa dur,
  - Eğer tanımlanmış olan maksimum ağaç derinliğine ulaşıldı ise dur.
4. Eğer hiçbir durdurma kriterine ulaşılmadı ise, ikinci adıma dön.

CART algoritması, yalnızca ikili bölmelere sahip ağaçlar oluşturmaktadır. Bu durum bölme kriterini basitleştirmektedir ve ikiden fazla bölme durumları için bir ceza belirlenmesi gerekmemektedir. CART algoritması bölme kriteri olarak İtalyan iktisatçı ve istatistikçi Corrado Gini' nin adını taşıyan Gini indeksini kullanmaktadır. Simetrik maliyetleri içerebilecek şekilde genişletilebilmesi ve bilgi kazancından daha hızlı hesaplanabilirliği, Gini indeksinin tercih edilme nedenleridir. Gini indeksi her bir düğümdeki özniteliği seçmek maksadı ile kullanılır. CART algoritması ayrıca, çok sınıflı problemler için kullanılabilen ikiye bölme kriterini de desteklemektedir. Her düğümde, sınıflar ayrık ve karşılıklı olarak kapsamlı sınıflar içeren iki süper sınıfa ayrılmaktadır. İki sınıflı bir problem için ayırma kriteri, kriteri optimize eden özniteliği ve iki süper sınıfı bulmak için kullanılır. Bu yaklaşım, daha dengeli sonuçlar verse de Gini indeksinden daha yavaş hesaplanmaktadır. CART algoritması budama tekniğine, minimum maliyetli karmaşıklık budaması adı verilmektedir. Bu yaklaşım, bir ağacın yeniden ikame hatasındaki yanlılığın yaprak düğüm sayısı ile doğrusal olarak arttığını varsaymaktadır. Bir alt ağaca atanan maliyet, yeniden ikame hatası ve yaprak sayısı ile " $\alpha$ " karmaşıklık parametresinin çarpımının toplamına eşittir. Her " $\alpha$ " değeri için ağacın maliyetini en aza indiren bir en küçük ağaç olduğunu gösterebilmektedir. CART algoritması, bölmede kullanılan öznitelik için bir bilinmeyen değer mevcutsa, bölme kriterini cezalandırmamaktadır. Yalnızca değeri bilinen örnekler kullanılmaktadır. CART

algoritması, orijinal bölme yerine kullanılacak vekil bölmeler tespit eder ve sınıflandırma sırasında, bilinen bir öznitelik değerine dayanarak ilk vekil bölme kullanılır. Vekil bölmeler, orijinal bölme kriterine göre seçilmezler. Çünkü her düğümdeki alt ağaç, seçilen orijinal bölmeyle dayalı olarak oluşturulur. Bu nedenle vekil bölmeler, orijinal bölme ile tahmine dayalı ilişkinin bir ölçüsünü en üst düzeye çıkarmak için seçilir. Bu prosedür, seçilen öznitelikle yüksek oranda ilişkili nitelikler varsa iyi çalışmaktadır (Gopal 2018, 418; Kantardzic 2011, 189-91).

Gini indeksi, bir “ $F$ ” veri seti için şu formül ile hesaplanmaktadır.

$$Gini(F) = 1 - \sum_{i=0}^{c-1} p_i^2$$

Formülde;

- $c$  ön tanımlı olan sınıfları,
- $C_i$ ,  $i = 1, \dots, C - 1$  e kadar olan sınıfları,
- $s_i$ ,  $C_i$  sınıfına ait olan gözlem sayısını,
- $p_i = s_i/S$ ,  $C_i$  sınıfının set içerisindeki göreceli sıklığını göstermektedir.

Gini indeksi her bir öznitelik için ikili bölmeyle dikkate almaktadır. İlk olarak, sıralanmış bitişik değerlerin her bir çifti arasındaki orta nokta olası bir bölme noktası olarak ele alınır. Bu yaygın olarak kullanılan basit bir politikadır. İlgili öznitelik için minimum gini indeksini veren nokta bölme noktası olarak kabul edilir.  $C_i$  nin bir bölme noktası,  $F_1$ ,  $C_i \leq$  bölme noktasını sağlayan demetlerin sayısıdır.  $F_2$  ise  $C_i >$  bölme noktasını sağlayan demetlerin sayısıdır. Minimum gini indeksine sahip olan öznitelik bölme için seçilir. İlk bölme sonrasında  $F_1$  ve  $F_2$ ’ den birer öznitelik seçilir ve benzer şekilde yaprak düğümlere ulaşana kadar bölme devam eder (Gopal 2018, 417-18).

Bu metrik  $F$  veri setinin bölme saflığını göstermektedir. İki sınıfa sahip olunan durumlarda sınıf tahmini için gini indeksi, “0,0,5” arasında yer almaktadır.  $S$ ’ teki tüm veriler aynı sınıfa ait ise metrik sifıra eşittir. Bu durumda  $F$ ’ in saf olduğu anlamına gelmektedir. Metrik 0,5’ e eşit ise veriler iki sınıf arasında eşit olarak dağıtılır. Bu, bir sınıfı destekleyen bir bölünme olduğu anlamına gelmektedir. Bölünmenin 70/30 şeklinde gerçekleşmesi için metriğin 0,42 değerine eşit olması gerekmektedir. Bir özniteliğin  $k$  adet alt kümeyle ( $F_i$ ) bölünmesinin kalitesi daha sonra, ortaya çıkan alt kümelerin gini indekslerinin ağırlıklı toplamı olarak hesaplanmaktadır (Kantardzic 2011, 189-90);

$$Gini_{bölme} = \sum_{i=0}^{k-1} n_i / Gini(F_i)$$

Formülde;

- $n_i$ , bölme sonrasında alt küme  $F_i$ ' de kalan gözlemlerin sayısını,
- $n$ , düğümde verilmiş olan toplam gözlem sayısı göstermektedir.

CART algoritmasının aykırı değerlere ve gürültü içeren verilere karşı sağlam oluşu temel avantajları olarak sayılmaktadır. Genellikle bölme algoritması, tek bir düğüm veya düğümlerdeki aykırı değerleri izole eder. Ayrıca, CART' in önemli bir pratik özelliği, sınıflandırma veya regresyon ağaçlarının yapısının, bağımsız değişkenlerin monoton dönüşümlerine göre değişmez olmasıdır. Bu, veri setinde yer alan herhangi bir değişkenin logaritması veya karekök değeri ile değiştirilebilmesine imkân tanımaktadır. Ancak ağacın yapısı değişmemektedir. CART' in dezavantajlarından biri, sistemin kararsız karar ağaçlarına sahip olabilmesidir. Birkaç gözlemin ortadan kaldırılması gibi öğrenme örneklerinde yapılan önemsiz değişiklikler, bir karar ağacında radikal değişikliklere yol açabilmektedir (Kantardzic 2011, 191).

### 2.5.3. Destek Vektör Makineleri

Destek vektör makineleri (DVM), doğrusal ikili sınıflandırıcılar oluşturma amacı taşıyan bir makine öğrenmesi tekniğidir. DVM amaca uygun olarak veri setini iki sınıfa ayıran bir hiperdüzlem oluşturmaktadır (Maheshwari 2020). Destek vektör makinelerinin altında sağlam bir teori yatmaktadır. Temelleri istatistik, optimizasyon ve makine öğrenmesi dayanmaktadır. DVM maksimum marjine sahip sınıflandırıcılar oluşturabilmek maksadı ile 1960'lı yıllardan günümüze tartışılan pek çok kavramı bir araya getirmektedir. Basit bir teoriye sahip olmasına karşın, model genelleme yetenekleri ve başarılı global optimum çözümler sunmaları, DVM' nin en popüler gözetimli öğrenme algoritmaları arasına girmelerine olanak tanımıştır. Destek vektör makinelerinden dış bükey optimizasyon problemlerini analitik olarak çözüme kavuşturması sebebi ile sınıflandırma problemlerinin çözümünde yaygın olarak yararlanılmaktadır. DVM, genetik algoritmalar ve çok katlı algılayıcılardan farklı olarak her durumda aynı optimal hiperdüzlem parametresini döndürmektedir. DVM, yerel minimum ya da boyutsallık lanetinden etkilenmemektedir. Makine öğrenmesi modelinin eğitim setindeki tüm gözlemlere sıfır hata ile uyum göstermeye çalışmasından kaynaklanan aşırı uyum durumunun yapısal riski en aza indirgeyen DVM modelleri önleyebilmektedir. DVM,

öznitelik setini sınırlayarak, destek vektörlerin sayısını seçerek model karmaşıklığını otomatik olarak belirlemektedir (Awad ve Khanna 2015).

Destek Vektör Makineleri bir kernel tekniğidir. Diğer makine öğrenmesi tekniklerinde olduğu gibi optimumların sezgisel olarak değil analitik olarak bulunduğu dış bükey bir optimizasyon tekniği olarak veriyi daha yüksek boyutlu bir uzaya eşlemek amacı ile kernel hilesini kullanmaktadır. Genellikle gerçek hayattaki veriler doğrusal olarak ayrılamamaktadır. Doğrusal olmayan bir ayırma sınırını öğrenmeye çalışmak, ayırma yüzeyi az ikinci dereceden olacağı için optimizasyon aşamasında hesaplama gereksinimi artıracaktır. DVM veri setindeki verileri, önceden tanımlanmış kernel işlevlerini kullanarak, doğrusal bir ayırıcının farklı sınıflar arasında ayırım yapabileceği yeni ancak daha yüksek boyutlu bir alana eşler. Optimizasyon aşaması böylece haritalanan uzayda yalnızca doğrusal bir ayırt edici yüzeyin öğrenilmesini gerektirecektir. Bu nedenle kernel fonksiyonunun seçimi ve parametrelerin seçimi DVM için oldukça önemlidir. DVM eğitim setine dayalı olarak, hatayı veya maliyeti en aza indirmenin ötesinde optimizasyon sorununa ek bir kısıtlama getirmektedir. DVM hiper düzlemin hedeften maksimum uzaklıkta olacak şekilde konumlandırılması gerekmektedir. DVM bir maksimum marjın ayırıcısıdır. DVM yapısal risk minimizasyonunu kullanır. Dualite ve dışbükeylik gereksinimlerini karşılar. Yapısal risk minimizasyonu, sonlu bir eğitim seti ile eğitim gerçekleştirmek için uygun bir model seçen tümevarımsal bir ilkedir. Yapısal risk minimizasyonu alan karmaşıklık hipotezi ve ampirik hata arasında (Vapnik-Chervonenkis boyutları) bir değiş tokuş önermektedir. Kernel uzayında, girdi uzayında yüksek dereceden sınıflayıcı bir hiperyüzey tespit etmek yerine sınıflama probleminin çözümü için doğrusal bir hiperdüzlem tahmini belirlenebilir. Bir kernel, hermiyen ve pozitif yarı kesin bir matris olmak zorundadır ve Mercer teoremini sağlamalıdır. Bu durum kernel ya da gram matrisinin tüm gözlem çiftlerinin pozitif ve yarı kesin olarak değerlendirilmesi anlamına gelmektedir. Yaygın olarak kullanılan bazı kernel fonksiyonları şunlardır (Awad ve Khanna 2015);

Doğrusal kernel fonksiyonu:

$$K(x, u) = X^T \cdot u$$

Polinomsal kernel fonksiyonu:

$$K(x, u) = (ax^T u + c)^q, q > 0$$

Hiperbolik tanjant (sigmoid) kernel fonksiyonu:

$$K(x, u) = \tan h(\beta x^T u + \gamma)$$

Gauss radyal temelli kernel fonksiyonu:

$$K(x, u) = \exp\left(-\frac{\|x-u\|^2}{\sigma^2}\right)$$

Laplace radyal temelli kernel fonksiyonu:

$$K(x, u) = \exp\left(-\frac{\|x-u\|^2}{\sigma}\right)$$

ANOVA radyal temelli kernel fonksiyonu:

$$K(x, u) = \sum_{k=1}^n \exp(-\sigma(x^k - u^k)^2)^d$$

Doğrusal spline kernel fonksiyonu(birinci boyutta):

$$K(x, u) = 1 + x \cdot u \cdot \min(x, u) - \frac{x + u}{2} (\min(x, u))^2 + \frac{1}{3} \min(x, u)^3$$

#### 2.5.4. Topluluk Öğrenmesi (Ensemble Learning)

Topluluk öğrenmesi yaklaşımdan tahmine dayalı modellerin doğruluğunu arttırmada yararlanılmaktadır. Topluluk öğrenmesinin temelinde farklı modellerden elde edilen sonuçları birleştirme amacı bulunmaktadır. Grup ortalaması bireysel tahminlerden daha iyi sonuçlar verecektir. Böylece model hatasını azaltmak hedeflenmektedir (Kantardzic 2011, 235-36).

İyi performansa sahip topluluk öğrenmesi modelinin temel sınıflayıcıların tahmin performansı ve tahmine dayalı çeşitliliği olmak üzere iki gereksinimi bulunmaktadır. Tahmine dayalı çeşitlilik gereksinimi temel sınıflayıcıların bağımsız olması gerektiği anlamına gelmektedir. Tahmin performansı ise temel sınıflandırıcıların çoğunluk sınıf sınıflandırıcılardan daha iyi bir tahmin performansına sahip olması gerekliliğidir. Bir topluluk öğrenmesi modeli oluşturmak maksadı ile yararlanılabilecek üç farklı yaklaşım bulunmaktadır. Bunlar paralel, ardışık ve hiyerarşik yaklaşımlardır. Paralel yaklaşımda, temel sınıflayıcılar, paralel şekilde eğitilir ve sonrasında bir başka bir ajan ile tahmin sonuçları birleştirilir. Ardışık yaklaşımda, bir temel sınıflandırıcının eğitiminde, önceden eğitilmiş olan bir sınıflandırıcının tahminleri kullanılmaktadır. Son yaklaşımda ise birinci ve ikinci yaklaşımlar birleştirilmektedir (Moreira, Carvalho, ve Horváth 2018).

Topluluk öğrenmesi süreci eğitim ve test olmak üzere iki ardışık aşamadan oluşmaktadır. Eğitim aşamasında eğitim verisi kullanılarak birkaç farklı temel sınıflayıcı model oluşturulur. Sonrasında test verilerinin değerlerini tahmin etmek için, eğitilen sınıflayıcıların çıktıları birleştirilir. Topluluk öğrenmesi modeli, birkaç farklı temel sınıflayıcıdan ve bir birleştirme kuralında oluşmaktadır. Topluluk öğrenmesi nispeten

yeni bir çalışma alanıdır ve çoklu sınıflayıcıların kombinasyonu, sınıflandırıcı füzyonu, uzmanların karışımı konsensus toplama gibi farklı eş anlamlı ifadeler ile nitelenmektedir. Bir topluluk öğrenmesi modeli oluşturulurken, temel sınıflayıcıların nasıl eğitileceği ve temel sınıflayıcılardan elde edilen sonuçların nasıl birleştirileceği yanıtlanması gereken iki temel sorundur. Aşağıda bir topluluk öğrenmesi modeli oluşturmada bu sorunların üstesinden gelmek maksadı yararlanılan yaklaşımlar sıralanmaktadır (Kantardzic 2011, 236,238);

- Destek vektör makineleri, yapay sinir ağları veya karar ağaçları gibi farklı makine öğrenmesi algoritmaları kullanmak,
- Farklı modeller üretmek maksadı ile aynı algoritmanın farklı hiper parametreler ile eğitilmesi. Örneğin, yapay sinir ağının farklı sayıda gizli katman ile eğitimi,
- Bir veri setinde farklı özneliklerin alt kümelerinin kullanılması,
- Aynı sınıflayıcı algoritmanın, veri setinin farklı alt kümeleri ile eğitilmesi.

#### **2.5.4.1. Bootstrap Aggregating (Bagging)**

Makine öğrenmesi uygulamalarında bir hipotez fonksiyonunun çıktısı belirlenmiş olan problem için uygun olmayabilir. Bu durumda problemin çözümü için birden fazla fonksiyonun çıktıları kullanılabilir. Algoritma bu görevi, bir hipotezler topluluğundan öğrenerek ve bunları birleşik bir biçimde kullanarak yerine getirir. Bagging ve Boosting topluluk öğrenmesi metotları içerisinde en yaygın kullanılanlarıdır. Bu metotlar sınıflandırma ve regresyon problemlerinin çözümünde kullanılabilir ve tahmin performansını artırmaktadır (Gopal 2018, 54).

Bagging metodunda, topluluğun her bir üyesi farklı bir eğitim veri setinden oluşturulur. Tahminler tek tip ortalama alma veya sınıf etiketleri üzerinden oylama yoluyla birleştirilmektedir. Her bir veri seti toplam  $N$  gözlemden örnekleme yolu ile oluşturulur ve gözlemler değiştirme yolu ile tek tip olarak rastgele seçilmektedir. Her bir örnek bir yeniden örnekleme (bootstrap) olarak tanımlanmaktadır. Bagging, Bootstrap Aggregating 'den kısaltılarak türetilmiştir. Yeniden örnekleme (bootstrapping),  $N$  öğeyi değiştirme ile tek biçimli olarak rastgele örneklemediğinden, herhangi bir bireysel veri ögesinin seçilmeme olasılığı  $p = (1 - 1/N)^N$  dir. Bu nedenle büyük bir veri seti ile yeniden örneklemenin orijinal setin yaklaşık %63,2' sini içermesi beklenmektedir. Bu durumda orijinal verinin %36,8' i seçilmemiş olacaktır. Pek çok topluluk öğrenmesi modelinde olduğu gibi Bagging metodu da kararsız modeller ile bir başka ifade ile aşırı uyuma yatkın modeller

ile daha iyi sonuçlar ortaya koymaktadır. Bu sınıflama algoritmalarına karar ağaçları ve yapay sinir ağları örnek verilebilir (Sammut ve Webb 2017, 395).

#### **2.5.4.1.1. Rastgele Orman Algoritması (Random Forest)**

Breiman (2001) tarafından tanıtılmış olan rastgele orman algoritması bir bagging metottur. Algoritmada temel sınıflayıcı olarak karar ağaçları kullanılmaktadır. Rastgele orman algoritmasında her karar ağacı orijinal veri setinden alınan rastgele bir örneklem ile eğitilmektedir. Rastgele örnekleme maksadı ile random subspace metodundan yararlanılmaktadır. Karar ağacı algoritmalarının eğitiminde, oluşturulan ağaçlar eğitim sonrasında budanmaktadır ancak rastgele orman algoritmasında budama kuralı bulunmamaktadır. Böylece karar ağaçlarının oluşturuldukları alt örnekleme kısmen uyum sağlamasına olanak tanınmaktadır. Sınıflandırıcıları daha fazla çeşitlendirmek için, ağaçtaki her dalda, hangi özelliğin ayrılacağına ilişkin karar, tam özellik kümesinden  $n$  boyutundan rastgele bir alt küme ile sınırlandırılmaktadır. Rastgele seçilen alt küme, her dallanma noktası için yeniden seçilir,  $n$ ' in, tüm veri setinin  $N$  ile gösterdiği durumda,  $\log_2(N + 1)$  olması beklenmektedir (Kantardzic 2011, 239; Sammut ve Webb 2017, 1054).

Bagging metodunun başarısı, büyük ölçüde, eğitilen tüm temel sınıflandırıcı modellerin yeterince çeşitli olup olmamasına bağlıdır. Çünkü tüm temel sınıflayıcı modeller yüksek oranda ilişkili ise birleşik topluluk modeli benzer sonuçları üretecektir. Bu bağlamda, rastgele orman algoritmasında her biri aynı eğitim veri setinde öğrenen modelleri çeşitlendirmek için aşağıdaki teknikler kullanılmaktadır (Jiang 2021, 208-9);

1. Satır Örnekleme: Her bir karar ağacı modelini oluşturmak için bir yeniden örnekleme oluşturmak maksadı ile değiştirme ile orijinal veri setini örnekleme için bagging metodunun kullanılır.
2. Sütun Örnekleme: Elde edilen her bir yeniden örnekleme için karar ağacı oluşturma maksadı ile kullanılan özniteliklerin yalnızca bir rastgele bir alt kümesini tutmak için ilgili öznitelikteki tüm girdi öğeleri de örneklenir.
3. Yetersiz Bölme: Bir karar ağacı oluşturmak maksadı ile ikinci adımda oluşturulan rastgele alt küme kullanılır.

Rastgele orman algoritmasında öznitelik örnekleme çok önemlidir. Çünkü öznitelik örnekleme ile karar ağacı çeşitliliği önemli ölçüde artırılabilir. Öznitelikler rastgele örnekleme ile çeşitliliği artırmak maksadı ile eğitilen bazı karar ağaçlarında zayıf

özniteliklerden de yararlanılabilmektedir. Sonuç olarak, rastgele orman algoritması, pratik olarak çok güçlü bir topluluk öğrenmesi metodudur. Çünkü saf bir karar ağacı yönteminden önemli ölçüde daha iyi bir performans gösterebilmektedir.

#### 2.5.4.2. Boosting

Boosting, son derece güçlü bir makine öğrenmesi tekniğidir. Sürecin sonunda çok sayıda temel sınıflandırıcı model basitçe birleştirilerek daha doğru bir model ortaya çıkarılabilmektedir. Her temel sınıflandırıcınının zayıf olduğu kabul edilmektedir. (Jiang 2021, 209-10).

Topluluk öğrenmesi modeli içerisinde temel sınıflayıcıları birleştirmek maksadı ile doğrusal bir yöntem kullanılırsa, aşağıdaki formülde gösterilen model elde edilmektedir;

$$F_m(x) = w_1f_1(x) + w_2f_2(x) + \dots + w_mf_m(x)$$

Her bir temel sınıflandırıcı model  $f_m(x) \in \mathcal{M}$  ön tanımlı bir model uzayından  $\mathcal{M}$  eğitilmektedir ve  $w_m(x) \in \mathbb{R}$  modelin topluluk ağırlığıdır. Tüm temel sınıflandırıcı modeller  $\mathcal{M}$  model uzayından seçildiğinde dahi, topluluk öğrenmesi modeli,  $F_m(x)$  'ye ait olmak zorunda değildir. Bunun yerine  $\mathcal{M}$  'deki herhangi bir fonksiyonun olası tüm kombinasyonlarını içeren  $dog(\mathcal{M})$  şeklinde gösterilen genişletilmiş bir model uzayına aittir. Genellikle,  $dog(\mathcal{M})$  uzayı  $\mathcal{M}$  ' e eşit değildir. Ancak  $dog(\mathcal{M}) \supseteq \mathcal{M}$  ilişkisi kolaylıkla doğrulanabilmektedir. Bunun dışında,  $l(f(x), y)$  kayıp fonksiyonu,  $dog(\mathcal{M})$  model uzayında bir fonksiyon olarak ele alındığında, topluluk öğrenmesi problemi aşağıdaki fonksiyonel minimizasyon problemi şeklinde gösterilebilmektedir;

$$F(x) = \underset{f \in dog(\mathcal{M})}{agr} \min \sum_{n=1}^N l(f(x), y_n)$$

Boosting, tüm temel sınıflandırıcı algoritmaların ardışık bir şekilde eğitildiği özel bir tür topluluk öğrenmesi metodudur. Metodun her bir adımında yeni bir temel sınıflandırıcı model  $f_m(x)$  ve topluluk ağırlığı  $w_m$  öğrenmek amaçlanmaktadır. Böylece topluluğa eklendiğinde modelin  $F_{m-1}(x)$  daha geliştirilmesi sağlanacaktır;

$$F_m(x) = F_{m-1}(x) + w_mf_m(x)$$

##### 2.5.4.2.1. Adaboost Algoritması

Adaboost (Adaptive Boosting) algoritması, ek bir topluluk üyesi eklerken, kümülatif öğrenme gelişimine yinelemeli olarak uyum sağlayan ilk boosting metodudur. Algoritma özellikle, yeni bir temel sınıflayıcı eğitimden önce eğitim setindeki mevcudun kümülatif

hatalarını yansıtacak şekilde, eğitim verilerindeki ağırlıklarını değiştirmektedir. Adaboost, teorik olarak performans artışını garanti etmesi sebebi, makine öğrenmesi alanına büyük katkılar sağlamıştır. Performans artışının sağlanması için algoritma yeterli sayıda veriye ve rastgele tahminden daha iyi ve güvenilir tahmin yapabilen bir temel sınıflayıcıya ihtiyaç duymaktadır. Çeşitli aşamalara sahip bu metot tüm öznitelik uzayında doğru bir performans gerektirmemektedir. Model, tasarımı basit tahminler yapmak yerine daha iyi performans ortaya koyabilecek temel sınıflandırıcı algoritmaların tespitine odaklanmaktadır. Algoritma, yanlılığın önüne geçmek amacı ile çok derin karar ağaçları üreten bagging metotlarından önemli ölçüde ayrılmaktadır. Metot sıg karar ağaçları üretmektedir ve genellikle tek bir bölünmeden oluşan bu ağacalar ile iyi performans ortaya koymaktadır. Algoritma eşit ağırlıklı bir veri seti ile başlamaktadır ve sonrasında örnek dağılımı sırayla değiştirilmektedir. Her yineleme sonrasında, yanlış sınıflandırılmış gözlemlerin ağırlıkları artırılmakta ve doğru tahmin edilen gözlemlerin ağırlıkları azaltılmaktadır. Böylece sonraki sınıflandırıcı sorun teşkil eden durumlara daha fazla odaklanabilmektedir. Modelin eğitimi tamamlandıktan sonra, karar ağacı topluluğunun eğitim hatasını azaltmaya olan katkısını gösteren bir ağırlıkla topluluğa dahil edilmektedir (Jansen 2018, 315-16);

Adaboost algoritması, bir temel sınıflandırıcı topluluğu  $h_m(x), m = 1, \dots, M$ , için aşağıdaki gibidir,

1. Gözlemler,  $i = 1, \dots, N$  için örnek ağırlıklarının  $w_m = \frac{1}{N}$  ilklendir,
  2. Her bir temel sınıflandırıcı  $h_m(x), m = 1, \dots, M$ , için aşağıdaki adımlar uygula,
    - $w_i$  ile ağırlıklandırılmış eğitim verileri ile  $h_m(x)$  eğit,
    - Temel sınıflandırıcının eğitim verisindeki ağırlıklandırılmış hatasını  $\varepsilon_m$ ' i hesapla,
    - Aşağıdaki formülden gösterildiği şekilde, temel sınıflandırıcının, kendi hatasının bir fonksiyonu olarak topluluk ağırlığını  $\alpha_m$ 'i hesapla,
- $$\alpha_m = \log\left(\frac{1 - \varepsilon_m}{\varepsilon_m}\right)$$
- $w_i * \exp(\alpha_m)$ ' e göre yanlış sınıflandırılmış olan gözlemlerin ağırlıklarını güncelle.
3. Aşağıda gösterilen formüle göre, topluluk üyelerinin ağırlıklı toplamı pozitif olduğunda pozitif sınıfı tahmin et. Aksi durumda negatif sınıfı tahmin et.

$$H(x) = \text{sing}\left(\sum_{m=1}^M \underbrace{\alpha_m h_m(x)}_{\text{ağırlıklandırılmış temel sınıflayıcı}}\right)$$

### 2.5.4.2.2. Gradient Boosting

Bir topluluk öğrenmesi metodu olarak boosting, hipotez uzayındaki hatayı minimize edecek olan çözümü ardışık bir süreç içerisinde tespit etmektedir. Boosting sürecinde kritik bir konu, modele yeni bir temel sınıflandırıcı eklenmesi durumunda, modelin performansını artırmayı garanti edecek olan yeni temel sınıflandırıcının nasıl seçileceğidir. Kayıp fonksiyonu  $l(f(x), y)$  'i fonksiyon uzayının bir üyesi kabul edilmesi durumunda, gradyan  $\frac{\partial l(f(x), y)}{\partial f}$ , fonksiyon  $f$  teki kaybın en hızlı artış yönünü gösteren yeni fonksiyonu temsil etmektedir. Gradient Boosting metodu mevcut topluluktaki gradyanın yönü boyunca yeni bir model tahmin etmeyi amaçlamaktadır (Jansen 2018; Jiang 2021);

$$-\nabla l(F_{m-1}(x)) \triangleq -\left. \frac{\partial l(f(x), y)}{\partial f} \right|_{f=F_{m-1}}$$

Ancak, negatif gradyanın  $-\nabla l(F_{m-1}(x))$ , yeni bir temel model olarak kullanılması, model uzayına ait olmama ihtimali sebebi ile mümkün değildir. Bu nedenle, Gradient Boosting metodu, model uzayında belirtilen gradyana en çok benzeyen fonksiyonu bulmayı amaçlamaktadır. İlk olarak tüm eğitim seti kullanılarak,  $f(\cdot)$  ve  $g(\cdot)$  gibi herhangi iki fonksiyon arasında bir iç çarpım tanımlanılır:

$$\langle f, g \rangle \triangleq \frac{1}{N} \sum_{i=1}^N f(x_i) g(x_i)$$

Bu aşamadan sonra Gradient Boosting metodunu gerçekleştirmenin bir yolu, negatif gradyanlı iç çarpımı maksimize etmek için her adımda model uzayında yeni bir model aramaktır:

$$f_m = \arg \max_{f \in \mathcal{M}} \langle f, -\nabla l(F_{m-1}(x)) \rangle$$

Buna alternatif olarak  $f(\cdot)$  ve  $g(\cdot)$  fonksiyonları arasında eğitim seti kullanılarak aşağıdaki gösterilen metrik de tanımlanabilmektedir;

$$\|f - g\|^2 = \frac{1}{N} \sum_{i=1}^N (f(x_i) - g(x_i))^2$$

Bu metriği kullanarak, model uzayında negatif gradyandan uzaklığını minimize edecek bir temel model tespiti edilebilmektedir:

$$f_m = \arg \min_{f \in \mathcal{M}} \|f + \nabla l(F_{m-1}(x))\|^2$$

$$= \text{agr} \min_{f \in \mathcal{M}} \sum_{n=1}^N (f(x_n) + \nabla l(F_{m-1}(x_n)))^2$$

Son olarak açıklanan yöntemleri kullanarak yeni sınıflandırıcı  $f_m$  belirlendikten sonra, optimal topluluk ağırlığı aşağıdaki formül yardımı ile belirlenmektedir:

$$w_m = \text{agr} \min_w \sum_{n=1}^N l(F_{m-1}(x_n) + w f_m(x_n), y_n)$$

#### 2.5.4.2.3. Extreme Gradient Boosting (XGBoost)

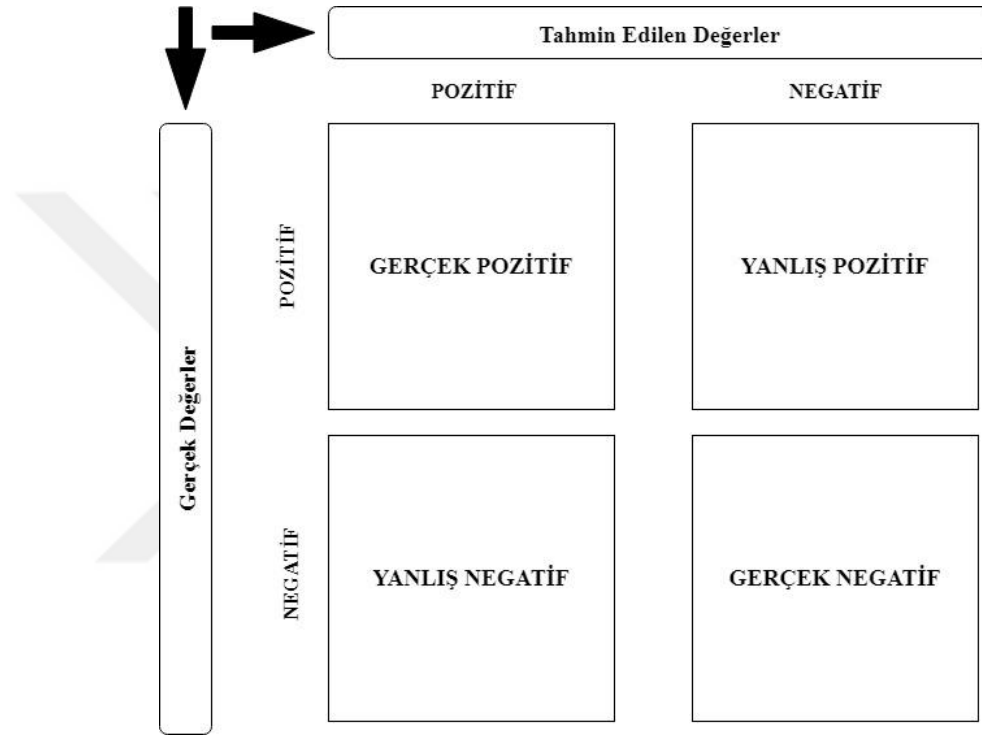
“Extreme Gradient Boosting” yaygın olarak kullanılan kısaltması ile XGBoost, popüler ve “tree boosting” yönelimli, ağırlıklanabilir bir topluluk öğrenmesi sistemidir. Gradient Boosting metodunun en yaygın kullanılan varyantlarından birisidir. Yöntem, Chen ve Guestrin (2016) tarafından geliştirilmiştir. Yöntem, ağırlıklanabilirliği artırmak maksadı ile Gradient Boosting’ e çeşitli optimizasyonlar ve iyileştirmeler eklemektedir. Metodun sahip olduğu en büyük iyileştirme, kayıp fonksiyonuna daha basit ve daha üretken topluluklar oluşturmayı amaçlayan bir düzeltme bileşeni eklemesidir. XGBoost paralel hesaplama yapılabilmesine olanak sağlamaktadır. Bu da modellemenin hızlı ve verimli yürütülmesine olanak tanımaktadır. Ayrıca farklı makine öğrenmesi problemlerinin çözümü için uygulanabilir, çok yönlü bir yapıdır. Metot öznelik önemi belirleyebilmektedir. Verilerin ölçeklendirilmesi ve normalizasyonuna ihtiyaç duymamaktadır. Sayılan üstünlüklerinin aksine XGBoost yalnızca sayısal veriler ile çalışmaktadır ve hiperparametrelerini doğru bir şekilde optimize edilmemesi durumunda aşırı uyum göstermeye yatkındır (Ghatak 2017; Rokach 2019).

#### 2.6. Model Değerleme

Model değerlendirme aşamasında, eğitilen model ya da modellerin amaca uygunluğunun değerlendirilmesi amaçlanmaktadır. Modelleme aşaması bir ya da birden fazla model ile sonuçlanmaktadır. Modeller kalite ve etkinlik açısından konuşlandırma ya da nihai uygulama öncesinde mutlaka değerlendirilmelidir. Ayrıca, modelin ilk aşamada belirlenen amaca uygunluk gösterip göstermediği de belirlenmelidir.

Karmaşıklık matrisi, sınıflama modellerinin değerlendirilmesi maksadı kullanılmaktadır. Karmaşıklık matrisinden sınıflama modellerini değerlendirme maksadı ile gerçek pozitif

oranı, doğruluk ve kesinlik gibi pek çok metrik üretilmektedir. Karmaşıklık matrisi  $m \times m$  boyutlu ( $m$  modeldeki sınıf sayısı) bir matristir. Matrisin satırlarında soldan sağa doğru, tahmin edilen değerler yani doğru tahmin edilen ve yanlış tahmin edilen örnek sayısı yer almaktadır. Matrisin sütunlarında ise yukarıdan aşağıya, gerçek değerler yer almaktadır. Satır ve sütunların kesişimleri gerçek pozitif (true positive), gerçek negatif (true negative), yanlış pozitif (false positive) ve yanlış negatif (false negative) tahmin edilmiş değerler göstermektedir (Singh 2022).



**Şekil 2.4.** Karmaşıklık Matrisi

Karmaşıklık matrisinde sütunlar tahmin edilen değerleri, satırlar ise gerçek değerleri temsil etmektedir. Matriste yer alan unsurlar önce satır sonra sütun olarak okunmaktadır. Bu nedenle gerçek değerleri belirtmek için gerçek ve yanlış terimleri, tahmin değerlerini belirtmek için ise pozitif ve negatif terimleri kullanılmaktadır. Karmaşıklık matrisi ile değerlendirme metrikleri üretmek için aşağıdaki değerler kullanılmaktadır.

1. Yanlış negatif: Modelin tahmini ettiği sınıf negatiftir. Ancak gerçek sınıf pozitiftir. Tahmin edilen sınıf hatalıdır.
2. Yanlış pozitif: Modelin tahmin ettiği sınıf pozitiftir. Ancak gerçek sınıf negatiftir. Tahmin edilen sınıf hatalıdır.
3. Gerçek negatif: Modelin tahmini gerçek değer ile eşleşmektedir.
4. Gerçek pozitif: Modelin tahmini gerçek değer ile eşleşmektedir.

Karmaşıklık matrisi kullanılarak elde edilebilecek metrikler (Singh 2022, 55-67);

1. Doğruluk (Accuracy): Doğruluk, model tarafından doğru tahmin edilen gözlem sayısının, toplam gözlem sayısına oranıdır.

$$\text{Doğruluk} = \frac{GP + GN}{GP + YP + GN + YN}$$

2. Gerçek Pozitif Oranı / Duyarlılık (Sensitivity): Gerçek pozitif oranı, doğru tahmin edilen pozitif gözlemlerin, toplam gerçek pozitif gözlemlere oranıdır. Oran duyarlılık olarak adlandırılmaktadır. Gerçek pozitif oranı, tüm gerçek pozitif örneklerin kaçta kaçının doğru tahmin edildiğini göstermektedir. Duyarlılık pozitif etikete sahip örneklerin kaçta kaçının model tarafından tahmin edildiğine yönelik bilgi sağlamaktadır. Bu oran modelin başarısı ile doğru orantıya sahiptir.

$$\text{Duyarlılık} = \frac{GP}{GP + YN}$$

3. Yanlış Negatif Oranı / Tip-II Hata: Yanlış negatif oranı, yanlış tahmin edilen pozitif gözlemlerin toplam gerçek pozitif gözlemlere oranıdır. Oran modelin iyiliği ile ters orantılıdır. Oran azaldıkça model daha iyi hale gelmektedir.

$$\text{Yanlış Negatif Oranı} = \frac{YN}{GP + YN}$$

4. Gerçek Negatif Oranı / Seçicilik (Specificity): Gerçek negatif oranı, doğru tahmin edilen negatif gözlemlerin toplam gerçek negatif gözlemlere oranıdır. Oran modelin iyiliği ile doğru orantılıdır. Oran artıkça model daha iyi hale gelmektedir. Bu nedenle doğruluk metriğine bir alternatif olarak kullanılabilir.

$$\text{Gerçek Negatif Oranı} = \frac{GN}{YP + GN}$$

5. Yanlış Pozitif Oranı / Tip-I Hata: Yanlış pozitif oranı yanlış tahmin edilen negatif etiketlerin toplam gerçek negatif etiketlere oranıdır. Oran modelin iyiliği ile ters orantılıdır. Oran azaldıkça modelin iyiliği artmaktadır. Bu nedenle doğruluk metriğine bir alternatif olarak kullanılabilir.

$$\text{Yanlış Pozitif Oranı} = \frac{YP}{YP + GN}$$

6. Kesinlik (Precision): Kesinlik, pozitif etiketli gözlemlerin kaçta kaçının doğru şekilde tahmin edildiğini göstermektedir. Oran, model pozitif bir tahmin yaptığından bu tahminin ne sıklıkla doğru çıktığını göstermektedir. Kesinlik oranı pozitif etikete sahip olduğu

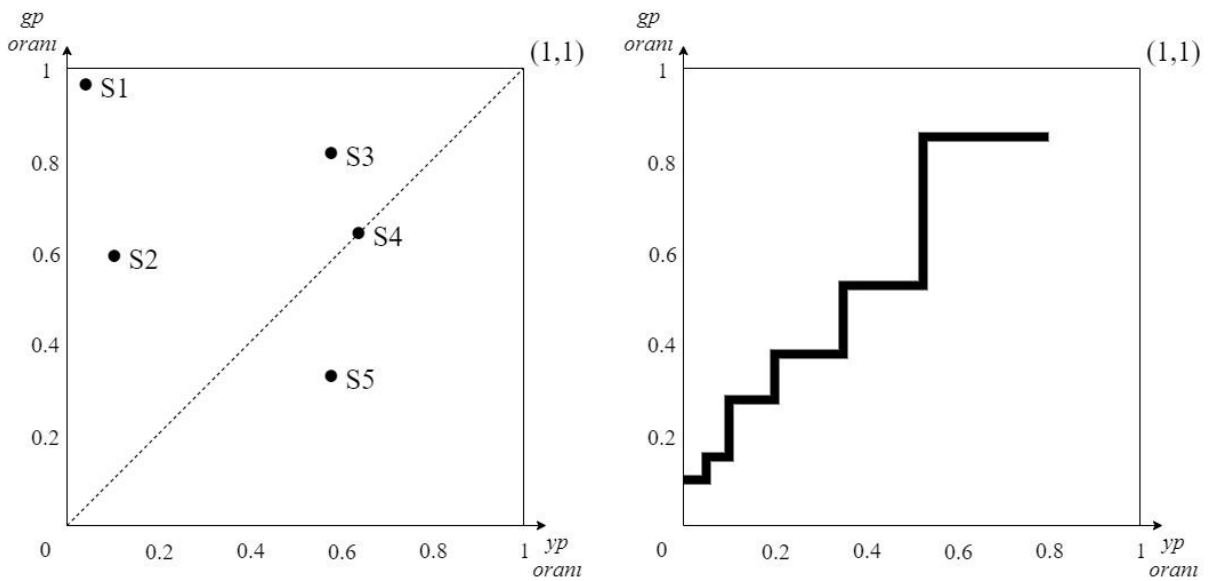
tahmin edilen bir gözlemin gerçekte pozitif etikete sahip olduğuna ne kadar emin olabileceğimizi göstermektedir.

$$Kesinlik = \frac{GP}{GP + YP}$$

Hem kesinlik oranı hem de duyarlılık oranının yüksek değerler almasının modelin iyiliğini göstermektedir.

## 7. ROC (Receiver Operating Characteristic) Eğrisi ve ROC Eğrisinin Altında Kalan Alan - AUC (Area Under the Curve)

ROC (Receiver Operating Characteristic) eğrisi model tahmin metrikleri kullanılarak hesaplanan ve model değerlendirme maksadı ile yaygın bir şekilde kullanılan performans ölçüsüdür. Modele ait GPO ve GNO metriklerinin hesaplanan değerleri, model için belirlenen ayırma kriterine bağlıdır. Örneğin 0.5' lik bir ayırma değerine bağlı olarak eğitilen bir modelin performans metrikleri, ayırma kriteri 0.75 değerine yükseltildiğinde değişecektir. Benzer şekilde ayırma kriterinin azalması da metriklerin değişmesine neden olacaktır. Başka bir ifade ile her olası eşik değeri için farklı GPO ve GNO belirlenecektir. Model performans metriklerindeki bu değişimi ortaya koymak ROC eğrisinin temel amacıdır. ROC eğrisi dikey ekseninde GPO' nun ve yatay ekseninde YPO' nun (1-GNO) yer aldığı bir grafikdir. ROC eğrisi, her bir uygun ayırma kriteri ya da eşik için hesaplanan bir nokta belirlenerek ve bunlar birleştirilerek çizilmektedir (Kelleher, Mac Namee, ve D'Arcy 2020, 558-60).



Şekil 2.5. ROC Grafiği ve ROC Eğrisi Örneği (Gopal 2018, 66,67)

Karmaşıklık matrisinde yer hesaplanan gerçek pozitif, gerçek negatif, yanlış pozitif ve yanlış negatifler farklı sınıflandırma modellerine göre farklı maliyet ve faydalara sahiptir. Yanlış negatif bir tahmin, bir maliyet iken, gerçek pozitif bir tahmin model açısından bir fayda olarak tanımlanmaktadır. Modelin fayda ve maliyeti, makine öğrenmesi probleminin yapısına göre değişmektedir. ROC grafiği sayılan maliyet ve faydalar arasındaki dengeyi ortaya koymaya yarar. Şekilde gösterilen ROC grafiğinde farklı sınıflayıcılara ait gerçek pozitif ve yanlış pozitif oranları yer almaktadır. S1 sınıflayıcısı (0,1) eksenine en yakın olan sınıflayıcıdır ve neredeyse mükemmel sınıflamayı sahip olan modeldir. Sağ üst nokta (1,1) koşulsuz pozitif sınıflamayı temsil etmektedir. (0,0) noktasın ise yanlış bir sınıflamanın yer almadığı ancak gerçek pozitif bir sınıflamanın da bulunmadığı noktadır. ROC grafiklerinde sol üst köşeye yakın noktalar yani gerçek pozitif tahminlerin sayısının fazla ancak yanlış pozitif tahminlerin sayısının az olduğu durumlar tercih edilmektedir. S2 benzeri sınıflayıcılar “korunumlu” olarak adlandırılmaktadır. Bu sınıflayıcılar güçlü kanıtlar üzerinde pozitif tahminler oluşturmaktadır. Bu sınıflayıcılar düşük yanlış pozitif tahmin oluşturmaktadır. Ancak gerçek pozitif tahminlerinin sayısı da düşük kalmaktadır. Sınıflayıcı sağ üst köşeye yaklaştığında “liberal” olarak adlandırılmaktadır. Zayıf kanıtlar ile pozitif tahmin üretme becerisine sahiptirler. Ancak yüksek pozitif tahmin oranına sahiptirler. Köşegen üzerindeki bir sınıflayıcı rastgele tahmin etme stratejisini temsil etmektedir. Altta yer alan üçgende bulunan bir sınıflayıcı rastgele tahminden daha kötü bir performans ortaya koymaktadır. ROC grafiğinde yer alan sınıflayıcılar bir test setinde uygulanan ve bunun tahminleri ile oluşturulan karmaşıklık matrisi sonuçlarını göstermektedir. Farklı eşik değerleri kullanarak elde edilecek olan noktalar ile ROC eğrisi oluşturulmaktadır.

## ÜÇÜNCÜ BÖLÜM

Bu bölümde işletmenin sürekliliğinin tahmini hakkında geniş bir literatür taraması sunulmaktadır. Süreklilik tahmini literatürü istatistiksel metotlardan faydalanılan çalışmalar ve makine öğrenmesi metotlarından faydalanılan çalışmalar olarak iki başlık altında organize edilmiştir.

### 3.1. Süreklilik Tahmini

İşletmenin sürekliliğinin tahmini iktisadi olarak büyük bir önem taşımaktadır. Büyüklükleri ve faaliyet gösterdikleri sektörden bağımsız olarak günümüzde pek çok işletme için en önemli tehdit süreklilik yeteneğini kaybetmek ve iflas etmektedir (Charitou, Neophytou, ve Charalambous 2004). İşletmenin sürekliliği, kamuyu, işletmenin faaliyet gösterdiği sektörde yer alan diğer işletmeleri, yatırımcıları ve politika yapıcıları etkilemektedir. İşletmenin süreklilik kabiliyetini kaybetmesinin yüksek sosyal ve iktisadi maliyetler bulunmaktadır (L. Zhang vd. 2013). İşletmelerin süreklilik yeteneklerini kaybetmeleri piyasa oynaklığını da artırıcı bir etkiye de sahiptir. Ayrıca bu durum, ekonomik durgunluk ortaya çıkmasına ve yatırımcılar, tedarikçiler ve işletme çalışanları gibi pek çok farklı paydaşın önemli ekonomik maliyetler ile karşı karşıya kalmasına neden olabilmektedir (Alam, Gao, ve Jones 2021). Denetçi açısından, gelecekteki potansiyel iflas adayı işletmenin sağlıklı sınıflandırılması denetçinin başarısızlığı olarak tanımlanmaktadır (Taylor ve Glezen 1991). Bu sebeplerle işletmenin sürekliliğinin tahminini yaklaşık bir asırdır oldukça popüler bir konudur ve daha sofistike ve güvenilir süreklilik tahmin modelleri geliştirmeye yönelik çalışmalar yaygınlaşmıştır. (Ahn ve Kim 2009; Alam, Gao, ve Jones 2021; L. Zhang vd. 2013).

Süreklilik tahmini, durum ve tarihsel veriler hakkındaki uzman bilgilerini birleştiren geniş bir alandır. Süreklilik tahmini çalışmalarında tipik olarak, işletmeler geçmiş gözlemlerin kullanıldığı matematiksel bir modeli ortaya koymak için çok sayıda gösterge ile ölçülmektedir (E. I. Altman ve Hotchkiss 2006). İşletmelerin sürekliliğini kaybetmesine neden olan faktörler çeşitlilik göstermektedir. Araştırmacılar bu fenomeni faiz oranları, resesyon nedeni ile azalan kar ve artan borç yüküne dayandırmaktadır. Bunlar dışında hükümet politikaları ve sektör dinamikleri de işletme başarısızlığına katkı sağlayabilmektedir (Charitou, Neophytou, ve Charalambous 2004). Süreklilik tahmini birçok işletme paydaşı tarafından başarısızlığı tahmin etmek ve kaçınmak maksadı ile kullanılmaktadır. Yüksek doğruluk, uzun tahmin ufku ve düşük maliyet, süreklilik

tahmini modellerinden beklenen özelliklerdir. İşletme paydaşları potansiyel iflaslardan kaçınmayı ve işletmenin süreklilik yeteneğini yanlış yorumlamamayı amaçlamaktadır (E. I. Altman vd. 2020). Tip-2 hata, sürekliliğini kaybetmiş olan bir işletmenin sağlıklı olarak sınıflandırılması anlamına gelmektedir. Tip-1 hata ise sağlıklı bir işletmenin sürekliliğini kaybetmiş bir işletme olarak sınıflandırılması anlamına gelmektedir.

İşletmenin sürekliliğinin tahmini, yaklaşık bir asırdır araştırmacıların dikkatini çeken bir konudur. Günümüzde de halen süreklilik tahmini çalışmalarında, birleşik bir teori bulunmamaktadır. Bu eksiklik, literatürdeki çalışmaların çoğunda, modelin doğruluğunu artırmaya odaklanılmasına ve modelin yorumlanmasına yeterince dikkat gösterilmemesine neden olmaktadır. Ancak, modelin sunacağı çıktı işletmenin ve denetçinin kararlarında kullanılacağı için modelin yorumlanılabilirliğinin önemi açıktır. Literatürde tahmin modellerini gerekliliği ile ilgili genel bir fikir birliği bulunmasına karşın modellerde işletmenin başarısız olduğu durumun tanımlama ile ilgili fikir birliği bulunmamaktadır (Alfaro vd. 2008; Appiah, Chizema, ve Arthur 2015).

Literatürdeki çalışmalar incelendiğinde araştırmacıların, çalışmalarında farklı tanımlamalar yaptığı görülmektedir. Bu bağlamda mali sıkıntı, iflas, işletme başarısızlığı ve süreklilik tahmini problemin tanımlanması maksadı ile kullanılmıştır. Çalışmalarda araştırmacılar yasal tanımlamalar ile birlikte literatüre dayanan tanımlamalar da kullanmıştır. Süreklilik tahmini literatüründe birleşik bir teorinin mevcut bulunmaması araştırmacıların işletmenin durumunu tanımlayan öznel göstergeler üretmesine neden olmaktadır. Ancak bu göstergelerin başarılı modeller ile nasıl birleştirileceği açık değildir (Zięba, Tomczak, ve Tomczak 2016).

Araştırmacılar işletmelerin piyasa değerlerinin nihai başarısızlıklarından önce önemli ölçüde azaldığını göstermektedir (Charalambous, Charitou, ve Kaourou 2000). İşletmenin sürekliliğinin tahmini, gelecekteki kurumsal başarısızlık sebebi ile sıkıntı yaşayabilecek olan işletmelerin sağlıklı olanlardan ayırmayı amaçlamaktadır (Constand ve Yazdipour 2011). Bir başka ifade ile işletmenin, denetçinin ve diğer ilgililerin işletmenin sıkıntı anını tahmin edebilmesi amacı ile bir model inşa edilmektedir. Modelin inşası için işletme ile ilgili bilgilerin mevcut bulunması ve erişilebilir olması gerekmektedir. Bu bilgiler temel olarak mali oranlardır ancak gerekli olması halinde ek bilgiler de dikkate alınabilmektedir (Alfaro vd. 2008). İşletmenin sürekliliğinin tahmininde mali oranların öznitelik olarak kullanılması oldukça yaygındır ve çok uzun

bir geçmişe sahiptir (S. Jones 2017). Beaver, McNichols, ve Rhie (2005), çalışmalarında mali oranların süreklilik tahmininde kullanılmasının güvenli olduğunu belirtmektedir.

Literatürde mali oranların seçiminde kullanılan çeşitli kriterler önerilmektedir. Seçilecek olan mali oranın literatürde yaygın olarak kullanılmış olması, oranın popülaritesi ve önceki çalışmalardaki öngörü başarısı, oranın hesaplanması için gerekli olan bilginin mevcut olması ve araştırmacının kişisel yargısı, mali oran seçiminde araştırmacıların başvurduğu kriterlerdir (Alfaro vd. 2008; Appiah, Chizema, ve Arthur 2015).

### **3.2. İşletmenin Sürekliliğinin Tahmininde İstatistiksel Metotlardan Faydalanılan Çalışmalar**

İşletmenin sürekliliğinin tahmini 1960' lı yıllardan günümüze önemli bir araştırma alanı haline gelmiş ve muhasebe, denetim ve finans literatüründe büyük bir ilgi görmüştür. Literatürde, en başarılı süreklilik tahmini modelinin bulmak amaçlanmış pek çok çalışma yer almaktadır. Araştırmacılar, işletme yönteminin, denetçilerin ve diğer işletme paydaşlarının karar vermesine yardımcı olma maksadı ile bir dizi istatistiksel yöntemi önermişlerdir (Murugan Anandarajan, Lee, ve Anandarajan 2001). Süreklilik riskini tahmin etmek ve işletmelerin mali durumlarına göre doğru bir sınıflandırma yapmak maksadı ile pek çok araştırmacı farklı varsayımlara ve belirli bir hesaplama karmaşıklığına sahip çeşitli teknikler kullanmıştır. İşletmenin sürekliliğinin tahmini literatürü kullanılan yöntemler açısından, istatistiksel yöntemler ve makine öğrenimi yöntemleri olarak kategorize edilmektedir (Yeh, Chi, ve Lin 2014b).

Beaver (1966), E. I. Altman (1968), Meyer ve Pifer (1970), Deakin (1972), Ohlson (1980) ve Zmijewski (1984) süreklilik tahmini modellerinin geliştirilmesine katkıda bulunan bazı önemli araştırmacılarıdır. Beaver (1966), tarafından yürütmüş olan çalışma tek değişkenli çalışmalar arasındaki en yaygın kabul görenidir. Çalışmada işletmelerin faaliyet gösterdiği sektör ve varlık büyüklüğü temel alınarak toplam 38 sektörden 79 başarısız ve 79 başarılı işletme seçilmiştir. Veri setini oluşturan işletmelerin ait olduğu sektörler arasında altı işletme ile en çok temsilcinin yer aldığı sektör elektronik ekipman üreticileridir. İşletmelerin varlık büyüklükleri 0,6 milyon ile 45 milyon dolar arasında değişmektedir. Çalışmada 30 mali oranın iflas etmiş işletmeler ile etmemiş işletmeleri sınıflandırmada öngörücü kabiliyetleri test edilmektedir. Yürütülen analiz sonucunda, net gelir / toplam borçlar oranının iflastan bir yıl önce en iyi tahmin performansını ortaya

koyduğu tespit edilmiştir. Bu oranı sırası ile net gelir / satışlar, net gelir / özsermaye, nakit akışları / toplam borçlar ve nakit akışları / toplam varlıklar oranları izlemektedir.

Beaver (1966) çalışmasında, işletmenin sürekliliğinin tahmininde birden fazla mali oran kullanımının daha başarılı sonuçlar ortaya koyma ihtimaline işaret etmektedir. Altman (1968) tarafından yürütülen çalışma literatürde birden fazla mali orandan faydalanılan ilk çalışmadır (Bellovary, Giacominio, ve Akers 2007). Çalışmada işletmenin iflasını tahmin etmek maksadı ile beş değişkenin yer aldığı bir model çoklu diskriminant analizinden faydalanılarak geliştirilmiştir. Araştırma işletme başarısızlığının tespitinde literatüründe ilk olarak istatistiksel bir yöntem kullanılan çalışmadır. Çalışmada ayrıca oran analizinin etkinliğini işletmenin sürekliliği bağlamında test etmek de amaçlanmaktadır. Bu bağlamda 1946 – 1965 yılları arasında iflas etmiş olan 66 işletmenin verilerinden yararlanılmıştır. Veri setinde iflas etmiş olan işletmeler varlık büyüklükleri dikkate alınarak aynı sayıda işletme ile eşleştirilmiştir. Z skoru olarak adlandırılan model, çalışma sermayesi/toplam varlıklar, dağıtılmamış karlar/toplam varlıklar, vergi ve faiz öncesi kar/toplam varlıklar, özsermayenin piyasa değeri/toplam borçların defter değeri ve son olarak satışlar/toplam varlıklar olmak üzere beş mali oran içermektedir. Çalışma sonucu genel işletmeler üç grup ayrılmaktadır. Birinci gruptaki işletmeler z skoru 2.99’ dan büyük olan risksiz işletmelerdir. İkinci grupta 1.81 ile 2.99 arasında z skoruna sahip olan, bu grup çalışmada gri alan olarak tanımlanmaktadır, riskli ya da risksiz olduğu belirli olmayan işletmeler yer almaktadır. Çalışmada ayrıca bu grup için cehalet bölgesi tanımı da kullanılmaktadır. Bu ifade işletmeler yerine analiste atıf yapmaktadır. Riskli işletmeler ise 1.81 değerinin altında kalmaktadır.

Altman (1968) tarafından yürütülen çalışma sonrasında işletmelerin başarısızlığının tahminini amaçlayan modellerin sayısı ve önerilen modellerin karmaşıklığı giderek artmıştır. Meyer ve Pifer (1970), iflas etmiş olan bankalar üzerinde yürüttükleri çalışmalarında, doğrusal olasılık modelinden faydalanmaktadırlar. Çalışmanın amacı geçmişte yaşanan başarısızlıkların yapısını ortaya koymaktan çok gelecekte yaşanacak olan başarısızlıkları tahmin etmektir. Bu bağlamda çalışmada 1948 ile 1965 arasında iflas etmiş ve aynı periyotta ödeme gücünü koruyan bankalara yer verilmiştir. Belirtilen dönemde iflas etmiş olan 55 bankadan, çalışmada açıklanan iki kriter dikkate alınarak 39 adedi seçilmiştir. Çalışma sonucunda iki önemli bulguya dikkat çekilmektedir. Bunlardan ilki, mali oranların işletmelerin görece gücünü değerlendirebileceğidir. İkincisi ise,

bankalar üzerinde yürütülen tahmin çalışmalarında mevcut mali durum göstergelerin fazlasının gerek olduğudur.

İşletmenin sürekliliğinin tahmini, başarısızlığı ortaya çıkaran sebepler ne olursa olsun, kayıpları azaltmak ve hatalı kaynak tahsislerinin önüne geçmek maksadı ile önemli bir konudur. Bu maksatla P. A. Meyer ve Pifer (1970) çalışmalarında, Amerika Birleşik devletlerindeki bankalarda iflas tahmini maksadı ile işletmelerin bilanço ve gelir tablolarından elde ettikleri toplam 32 değişken ile regresyon modeli test etmektedir. Çalışmanın ampirik sonuçları bağımsız değişkenlerin sayısının arttığı ölçüde, modelin tip-I ve tip-II hatasının azalma eğiliminde olduğunu göstermektedir. Ayrıca araştırmanın sonuçları süreklilik tahmini bağlanımda regresyon modellerinin tahmin yeteneklerini ortaya koymaktadır.

İşletme sürekliliğini tahmin etmek maksadı ile alternatif bir model Deakin (1972), tarafından ortaya koyulmuştur. Çalışmada Beaver (1966) tarafından ortaya koyulan ampirik sonuçların daha büyük bir tahmin yeteneğine sahip olduğuna işaret edilmektedir. Ayrıca Altman (1968), tarafından kullanılan metodolojinin daha sezgisel olduğu ifade edilmektedir. Çalışmada ilk olarak Beaver (1966) tarafından kullanılan mali oranların aynıları kullanılarak, aynı metodoloji bir farkla tekrarlanmaktadır. Sonrasında Beaver (1966) tarafından kullanılan mali oranların doğrusal kombinasyonları, işletmenin sürekliliğini kaybetmesinden beş yıl öncesinden başlamak sureti ile her yıl için sınırlanmaktadır. Beaver (1966) tarafından kullanılan metodolojinin sonuçlarının oldukça yüksek doğruluk derecesine sahip olduğu raporlanmaktadır. Ancak, işletmenin iflasından önceki yılda tespit edilen %20'lik hatanın üstesinden gelmek mümkün olmamaktadır. Bu nedenle diskriminant analizine başvurulmuştur. Çalışma sonuçları diskriminant analiz özelinde istatistiksel tekniklerin başarısızlığı tahmin etmek maksadı ile kullanılabilceğini ortaya koymaktadır. Ayrıca muhasebe verilerinin başarısızlıktan üç yıl öncesine kadar tahmin modellerinde kullanılabilceği de ifade edilmektedir.

Ohlson (1980) çalışmasında işletmelerin kurumsal başarısızlıklarını özellikle iflas olaylarını tahmin etmeye odaklanmaktadır. İflas tahmini maksadı ile bir logit modelden faydalanılan çalışmada yararlanılan veri seti 1970 ile 1976 yılları arasını kapsamaktadır. Veri seti oluşturulurken ABD' de faaliyet gösteren işletmelerin 10-K raporları dikkate alınmıştır. Ohlson (1980) çoklu diskriminant analizini eleştirmemek ile birlikte logit model kullanarak, modeli sezgisel yorumlara uygun hale getirmiştir. Ohlson (1980) geçmiş çalışmaların başarısının, varsayımlar ile uyumsuz olmaları sebebi ile belirsiz

olduğunu işaret etmektedir. Özellikle Altman (1968) tarafından önerilen modelin katı istatistiksel varsayımlarına ve sezgisel gücünün düşük oluşuna işaret etmektedir. Ohlson (1980) geliştirdiği logit modelde işletme büyüklüğü, toplam yükümlülükler / toplam varlıklar, çalışma sermayesi / toplam varlıklar, kısa vadeli yükümlülükler / dönen varlıklar, kısa vadeli yükümlülükleri ile toplam varlıkların arasındaki farkı temsil eden bir değişken (değişken eğer kısa vadeli yükümlülükler toplam varlıklardan büyük ise 1 değerini almaktadır, aksi durumda ise 0 değerini almaktadır), net gelir / toplam varlıklar, faaliyetlerden nakit akışları / toplam yükümlülükler, net gelir ya da zararı temsil eden bir değişken (değişken eğer net gelir geçmiş iki yılda negatif ise 1 değerini almaktadır, aksi durumda ise 0 değerini almaktadır), net gelirdeki değişim oranına yer vermektedir. Ohlson (1980) çalışmasında önerdiği logit modelin Altman (1968) tarafından önerilen çoklu diskriminant analizi modelinden daha yüksek tahmin başarısını sahip olduğunu ortaya koymaktadır.

Zmijewski (1984) çalışmasında, Altman (1968) ve Ohlson (1980) tarafından geliştirilen tahmin modelleri ile oldukça popüler olan bir diğer tahmin modelini sunmaktadır. Çalışmada yer alan probit model 1972 ile 1978 yılları arasında kapsamakta ve Birleşik Devletlerde sermaye piyasalarında işlem görmekte olan işletmeleri içermektedir. Zmijewski (1984) literatürde yer alan çalışmalarda seçime dayalı örneklem yanlılığı ve örneklem seçim yanlılığı olarak adlandırılan iki temel hata tanımlamaktadır. Çalışmada, literatürde yer alan modellerde araştırmacıların örneklem seçiminde rastlantısal olmayan bir yol tercih ettiklerini ve bu tercihin uygun bir yöntem seçilmemesi durumunda yanlış parametreler ortaya çıkarabileceğini belirtmektedir. Bu durumu seçime dayalı örneklem yanlılığı olarak tanımlamaktadır. Örneklem seçimi yanlılığı ise verilerinde eksik bulunan işletmelerin veri seti dışarısında tutulması sonucunda ortaya çıkmaktadır. Bu durum ayrıca doğal olarak seçime dayalı örneklem yanlılığının da ortaya çıkmasına neden olmaktadır. Çalışmada geliştirilen model net gelir / toplam varlıklar, toplam yükümlülükler / toplam varlıklar ve dönen varlıklar / kısa vadeli yükümlülükler olmak üzere üç adet bağımsız değişken içermektedir. Ohlson (1980) tarafından geliştirilen logit model gibi probit model de 0 ile 1 arasında sonuçlar üretmektedir ve 0.5' ten az olması işletmenin mali sağlığına işaret etmektedir.

### **3.3. İşletmenin Sürekliliğinin Tahmininde Makine Öğrenmesi Metotlarından Faydalanılan Çalışmalar**

Süreklilik tahmini üzerine uzun yılları kapsayan kapsamlı bir literatür bulunmaktadır. Literatürde pek çok farklı istatistiksel yöntemin kullanıldığı çalışma mevcuttur. Ancak bu yöntemlerin sahip olduğu istatistiksel varsayımlar ve sınırlı işlem kapasitesi araştırmacıların dikkatini makine öğrenmesi algoritmalarına çevirmesine neden olmuştur (Alam, Gao, ve Jones 2021). Denetimde makine öğrenmesi metotlarının kullanımı yaygınlaştırmak ve işletmenin sürekliliği bağlamında riskli işletmeleri sınıflandırmak maksadı ile pek çok araştırmacı makine öğrenmesi metotlarını benimsemiştir (Liang vd. 2016).

M Anandarajan ve Anandarajan (1999) süreklilik görüşünün tahminini kolaylaştırmak maksadı ile geri yayımlı yapay sinir ağları, karar ağaçları (ID3 algoritması) ve çoklu diskriminant analizini karşılaştırmaktadır. Çalışmada New York Borsası (NYSE) ve Amerikan Menkul Kıymetler Borsasında (AMEX) işlem gören işletmelerin mali raporlarının yer aldığı mali tablolar veri tabanı örnek seçmek maksadı ile kullanılmıştır. Veri setinde yer alan işletmelerin belirlenmesi için mali tablonun SAS.58 ve SAS.59 düzenlemelerinden sonra yayınlanmış olması, işletmelerin 1990-1991 yılları arasında şartlı görüş ya da sorumluluk reddi beyannameşi alamamış olması ve sayılan bu durumların işletmenin sürekliliği ile ilgili olması gerekmektedir. Yapılan inceleme sonucunda 1991 yılında süreklilik görüşü almış olan 45 işletme ile süreklilik görüşü almamış rastgele seçilmiş olan 45 işletme belirlenmiştir. Detaylı bir literatür taraması neticesinde seçilen 14 mali oran öznitelik olarak seçilmiştir. Çalışmanın ampirik sonuçları yapay sinir ağlarının karşılaştırılan diğer modellere göre daha başarılı tahmin performansına sahip olduğunu göstermektedir.

Aktaş, Doğanay, ve Yıldız (2003) çalışmalarında mali sıkıntı tahmininde istatistiksel yöntemler ile yapay sinir ağı algoritmalarını karşılaştırmaktadır. Çalışmada Yıldız (2001) tarafından kullanılan veri seti kullanılmış olup, iki veri seti arasındaki tek farklılık kullanılan öznitelik sayısının farklı olmasıdır. Veri seti 1983-1997 yılları arasında İMKB' de işlem gören sanayi, ticaret ve hizmet işletmelerini kapsamaktadır. Gözlemleri etiketlemek maksadı ile yedi kriter kullanılmıştır. Bunlar; iflas, sermayesinin yarısını kaybetmiş olması, varlıklarının %10'unu kaybetmiş olması, üç yıl üst üste zarar etmiş olması, borç ödeme zorluğu içine düşmüş olması, üretimi durdurmuş olması ve borçlarının aktif tutarını aşmış olmasıdır. Bu kriterler ışığında 53 başarısız işletme

belirlenmiş ve belirlenen işletmeler 53 sağlıklı işletme ile eşleştirilmiştir. Veri setindeki gözlemleri temsil etmesi için 23 adet mali oran, literatürde üzerinde ittifak olan, yaygın olarak kullanılan ve kolay hesaplanabilir olmalarına dikkat edilerek belirlenmiştir. Çalışmada yapay sinir ağları çoklu regresyon analizi, diskriminant analizi ve lojistik regresyon ile karşılaştırılmaktadır. Çalışmanın ampirik sonuçları yapay sinir ağlarının tahmin başarısını göstermektedir.

Lojistik regresyon literatürde, süreklilik tahmini çalışmalarında yaygın olarak kullanılan bir yöntemdir. Charitou, Neophytou, ve Charalambous (2004), lojistik regresyon ve YSA metotlarından faydalanmaktadır. Çalışmada 1988-1997 arasında Birleşik Krallıkta iflas etmiş olan kamu sanayi kuruluşları incelenmektedir. Çalışmada yer alan veri setinde 51 tanesi iflas etmiş olan 102 işletme yer almaktadır. Çalışma sonuçları YSA' nın genel olarak lojistik regresyondan daha yüksek sınıflama performansı ortaya koyduğunu göstermektedir. Ancak lojistik regresyon, YSA' ya göre daha düşük tip-1 hataya sahiptir. Çalışmanın sonuçları hem lojistik regresyonun hem de YSA' nın süreklilik tahmini için uygun alternatifler olduğunu ortaya koymaktadır. Süreklilik tahmini çalışmalarında YSA, yüksek tahmin performansı sebebi ile yaygın olarak kullanılan bir algoritmadır. Ancak YSA modellerinin aşırı uyum sorununa yatkın olması, yorumlanabilirliğinin sınırlı olması araştırmacıların YSA algoritmalarını kullanma hususunda çekingen davranmalarına neden olmaktadır (Ahn ve Kim 2009).

Kılıç ve Seyrek (2012) çalışmalarında İMKB' de işlem gören imalat işletmelerinin mali başarısızlıklarına ilişkin faktörleri belirlemeyi amaçlamaktadır. Çalışmada, yapay sinir ağları, tahmin modeli oluşturmak maksadı ile kullanılmıştır. Araştırmacılar 2005-2010 yılları arasında İMKB' de işlem görmüş olan 137 imalat işletmesinin verisini kullanmıştır. Çalışmada mali başarısızlığı tanımlamak için dört kriter kullanılmaktadır. Bunlar iflas, İMKB gözüaltı pazarında yer almak, aktif tutarının %10'unu kaybetmiş olmak ve iki yıl üst üste zarar etmiş olmaktır. Bu kriterlerden en az birisini sağlayan işletmeler mali sıkıntılı olarak etiketlenmiştir. Sayılan kriterler ışığında 191 gözlem mali sıkıntı olarak 604 işletme ise mali açıdan sağlıklı kabul edilmiştir. Gözlemleri temsil etmesi için 14 adet mali oran öznitelik olarak seçilmiştir. Çalışmanın ampirik sonuçları faaliyet karlılığının işletmenin başarısızlığını en fazla etkileyen oran olduğunu ortaya koymaktadır. Faaliyet karlılığını, kaldıraç oranı, kısa vadeli yabancı kaynaklar / toplam kaynaklar, net kar marjı, faaliyet kar marjı ve faaliyet giderleri/net satışlar takip etmektedir.

Karar ağaçları hem sınıflandırma hem de regresyon problemlerinde kullanılan popüler bir metodolojidir. Karar ağaçları, mevcut veri ile tümevarımsal akıl yürütmeyi kullanarak ağaç şeklinde bir yapı üretmektedir. Karar ağaçları iyi bilinen ve birçok başarılı uygulama geliştirilmiş bir tekniktir (Ravi Kumar ve Ravi 2007). Karar ağaçlarının temel algoritmaları arasında ID3 (Interactive Dichotomiser 3), C5.0, sınıflandırma ve regresyon ağaçları (CART) ve CHAID sayılmaktadır (M.-Y. Chen 2011a).

Chye Koh ve Kee Low (2004) lojistik regresyon, yapay sinir ağları ve karar ağaçlarının bir işletmenin sürekliliğinin tahmininde kullanılmasının sağlayacağı faydayı araştırmakta ve algoritmaları tahmin başarıları itibari ile karşılaştırmaktadır. Çalışmada oluşturulan üç süreklilik tahmini modelinin eğitiminde Koh (1991) tarafından yürütülmüş olan çalışmada kullanılan veri seti kullanılmıştır. Veri seti 165 sürekli ve 165 sürekliliğe sahip olmayan işletmeye ait mali oranları içermektedir. Veri setinde her gözlem için altı mali oran öznitelik olarak tanımlanmıştır. Çalışmanın ampirik sonuçları eğitilen üç modelinde birbirine çok yakın tahmin sonuçlarına sahip olduğunu ortaya koymaktadır. Ancak Karar ağaçları, lojistik regresyon ve yapay sinir ağlarından karşılaştırmalı olarak daha iyi tahmin performansına sahiptir. Çalışmanın sonuçları makine öğrenmesi algoritmalarının işletmenin sürekliliğinin tahmini bağlamında sahip olduğu potansiyel faydaları da göstermektedir. Özellikle yapay sinir ağlarının ve karar ağaçlarının karmaşık, doğrusal olmayan ve etkileşimli ilişkileri analiz etmekte güçlü araçlar olduğunu işaret etmektedir. Gepp, Kumar, ve Bhattacharya (2010) çalışmalarında farklı karar ağacı algoritmalarını işletmenin başarısızlığı alanına uygulamaktadır. Çalışmada araştırmacılar tarafından karşılaştırma yapmaya imkân sağlaması maksadı ile Frydman, Altman, ve Kao (1985) tarafından kullanılan veri setini kullanılmıştır. Veri seti COMPUSTAT veri tabanından elde edilmiştir. ABD’li imalat ve perakende işletmelerini içermekte ve 1971-1981 yılları arasını kapsamaktadır. Veri setinde 58 adedi başarısız olarak etiketlenmiş olan 200 işletme gözlem olarak yer almakta ve gözlemler 20 adet mali oran ile temsil edilmektedir. Araştırmacılar çalışmada RPA karar ağaçları, CART ve C5.0 karar ağacı algoritmaları ile çoklu diskriminant analizi metoduna yer vermiştir. Çalışmanın ampirik sonuçları CART ve RPA karar ağaçlarının çok benzer sonuçlar ürettiğini ve genel olarak en iyi tahmin ediciler olduğunu göstermektedir.

M.-Y. Chen (2011), çalışmasında karar ağaçlarını ve lojistik regresyonu karşılaştırmaktadır. Çalışmanın ana amaçları, karar ağaçlarının ve lojistik regresyonun alanda kullanımını benimsetmek, modellerin doğruluğu artırmak maksadı ile mali ve mali

olmayan verileri kullanmak ve iyi bir tahmin modeli ortaya koymak olarak ifade etmektedir. Çalışmada kullanılan veri, Tayvan Borsası' nın veri tabanı kullanılarak elde toplanmıştır. Veri setinde yer alan gözlemler borsada işlem görmektedir ve Ocak 2000 ile Mayıs 2007 tarihlerini kapsamaktadır. Veri seti toplam 100 işletmeyi içermektedir. Veri setinde yer işletmelerin 50 adedi mali sıkıntılı olarak etiketlenmiştir. Denetim raporlarında herhangi bir mali sıkıntı belirtisi bulunmayan 50 işletme ise sağlıklı olarak etiketlenmiştir. Çalışmada kullanılan öznitelikler literatüre dayanarak seçilmiştir. Veri setinde altı başlık altında toplam 37 adet öznitelik yer almaktadır. Çalışmada modellerin eğitimi işletmelerin mali sağlıklarını kaybetmelerinden 2, 4, 6 ve 8 dönem önceki veriler ile ayrı ayrı yapılmıştır. Çalışmanın ampirik sonuçlarına göre karar ağaçlarının özellikle kısa vadede işletmenin mali sıkıntı riski tahmininde lojistik regresyona göre daha iyi sonuçlar ortaya koyma potansiyelini ortaya koymaktadır. Lojistik regresyon ise karar ağaçlarının aksine mali sıkıntının uzun vadeli tahmininde daha başarılıdır.

Yakut ve Elmas (2013) çalışmalarında işletmelerin mali başarısızlığının tahmininde C5.0 karar ağacı algoritması ve diskriminant analizini karşılaştırmaktadır. Çalışmada 2005-2008 yılları arasında İMKB' de işlem görmekte olan 140 sanayi işletmesinin yer aldığı bir veri seti kullanılmıştır. Veri setinde yer alan işletmeler dört adet kriter ile mali başarılı ya da mali başarısız olarak sınıflandırılmıştır. Bunlar işletmenin iflas etmiş olması, borsada tahtasının kapanmış olması, işletmenin faaliyetinin durdurulmuş olması ve üst üste iki yıl veya daha fazla yıl zarar etmiş olmasıdır. Sayılan kriterlere göre yapılan değerlendirme sonucunda 38 işletme başarısız ve 102 işletme ise başarılı olarak etiketlenmiştir. Veri setinde yer alan gözlemleri temsil etmesi maksadı ile literatüre dayalı olarak 23 adet mali oran seçilmiştir. Çalışmanın ampirik sonuçları karar ağacı algoritmasının, diskriminant analizine göre daha başarılı tahmin sonuçları ortaya koyduğunu göstermektedir.

Destek vektör makineleri, süreklilik tahmini çalışmalarında kullanılan en başarılı algoritmalarından birisidir (Shin, Lee, ve Kim 2005). Algoritma, girdi vektörlerini doğrusal olmayan bir şekilde yüksek boyutlu bir öznitelik uzayına eşleyerek, doğrusal olmayan sınıf sınırlarını uygulamak maksadı ile doğrusal bir model kullanmaktadır (Vapnik 1999). DVM ayrıca aşırı uyum sorununa karşı da oldukça dirençli bir metottur ve sınıflandırma problemlerinde yüksek genelleme performansına sahiptir. DVM' den faydalanılan çalışmalarda farklı sınıflandırma problemlerinin çözümünde farklı kernel çekirdeklerinden faydalanılmaktadır.

Hua vd. (2007) çalışmalarında, tahmin doğruluğunu artırmak maksadı ile DVM ve entegre ikili diskriminant kuralı entegrasyonunu önermektedir. Çalışmada lojistik regresyon yöntemi, veri setindeki öznitelikler arasındaki ilişkileri incelemek ve DVM'nin ampirik risk minimizasyonu amacıyla kullanılmıştır. Önerilen modelin sonuçları, geleneksel DVM ile karşılaştırılmıştır. Çalışmada kullanılan veri seti, Şanghai Menkul Kıymetler Borsasında işlem gören işletmelerin denetlenmiş mali raporlarından elde edilmiştir. Veri seti 1999-2004 yılları arasını kapsamaktadır ve 60 adedi ST etiketi almış olan toplam 120 işletmenin verilerini içermektedir. Çalışmada öznitelik seçimi iki aşamalı olarak yapılmıştır. İlk aşamada tüm işletmeler için hesaplanmış olan 22 mali oran içerisinde t-testi kullanılarak 10 adedi seçilmiştir. İkinci aşamada boyutsallığı azaltmak maksadı ile diskriminant analizinden faydalanılmış ve öznitelik sayısı yediye indirilmiştir. Çalışmanın ampirik sonuçları önerilen DVM modelinin, geleneksel DVM modeline kıyasla daha başarılı sonuçlar ortaya koyduğunu göstermektedir.

Ding, Song, ve Zen (2008) destek vektör makinelerini çoklu diskriminant analizi, lojistik regresyon, geri yayımlı yapay sinir ağları ile karşılaştırılmaktadır. Araştırmacılar çalışmada mali sıkıntı tanımı olarak "special treatment" ST etiketini kullanmaktadır. Çalışmada yer alan veri setinde 56 adedi ST etiketi almış toplam 250 adet gözlem bulunmaktadır. Araştırmacılar öznitelik seçimi için t-testinden ve Kolmogorov-Smirnov testinden yararlanmışlardır. Çalışmanın ampirik sonuçları DVM mali sıkıntı alanında karşılaştırılan algoritmalara göre daha başarılı olduğunu göstermektedir. Ancak araştırmacılar DVM modelinin tahmin doğruluğunun yüksek olmasına rağmen model hatasının da yüksekliğine dikkat çekmektedir. Çalışmada bu durumun mali sıkıntı tanımı olarak ST etiketinin kullanılması ile ilişkili olabileceği belirtilmektedir. Çünkü ST etiketi ile işletmeler sınıflandırılırken sadece işletmenin karlılığı ile ilgili mali oranlara odaklanılmaktadır. Ancak problem modellerde kar dışında oranlarla da açıklanmaktadır. Yeh, Chi, ve Hsu (2010) çalışmalarında kaba küme teorisi ile destek vektör makineleri ve yapay sinir ağlarını entegrasyonunu, işletmenin verimliliğinin öznitelik olarak yer aldığı bir modelde sınamayı amaçlamaktadır. Çalışmada veri zarflama analizi işletme verimliliğini ölçmek maksadı ile kullanılmıştır. Veri seti Tayvan Menkul Kıymetler Borsası ve Tayvan Ekonomi Dergisinin veri tabanlarında alınmıştır ve 2005-2007 döneminde iflas başvurusunda bulunan bilişim ve elektronik imalatçısı işletmeleri içermektedir. Sürekliliğinin kaybetmiş olan 38 işletme 76 süreklilik yeteneğini koruyan işletme ile eşleştirilmiştir. Kaba küme teorisi çalışmada öznitelik seçimi maksadı

kullanılmıştır. Veri ön işleme ve öznitelik seçimi sonrasında veri setinde gözlemleri temsil eden işletmenin verimliliği dahil sekiz adet öznitelik yer almıştır. Çalışmanın ampirik sonuçları destek vektör makinelerinin, yapay sinir ağlarından daha başarılı sonuçlar ortaya koyduğunu göstermektedir.

Kim (2011) Güney Kore’ de konaklama işletmeleri üzerinde yürüttüğü çalışmasında yapay sinir ağlarının ve destek vektör makinelerinin sınıflandırma performanslarını, kendi aralarında ve geleneksel istatistiksel metotlar ile karşılaştırmayı amaçlamaktadır. Çalışmada yapay sinir ağları ve DVM, lojistik regresyon ve çoklu diskriminant analizi ile karşılaştırılmaktadır. İşletmenin temerrüt durumu ile karşılaşması, işletmenin borçlarını ödeyememesi sebebi ile yeniden yapılandırma başvurusunda bulunarak iflas açıklaması, iflas sebebi ile mali müşavirlik servisinden kaydını sildirmesi, işletmenin iflas ettiğini açıklaması ve işletmenin iflas sebebi ile üçüncü bir işletme tarafından devralınması ya da ilgili işletme ile birleşmesi veri setinin oluşturulmasında dikkate alınan kurallardır. Çalışmada kullanılan veri seti Ulusal Bilgi ve Kredi Değerleme şirketinin veri tabanından alınmıştır. Veri seti 1995-2002 yıllarını kapsamakta ve 18 adedi iflas etmiş olan 33 işletmeye ait mali verileri içermektedir. Çalışmada kullanılan öznitelikler kapsamlı bir literatür taraması ile belirlenmiş ve likidite, karlılık, faaliyet, büyüme ve istikrar başlıkları altından kategorize edilmiştir. Çalışmanın ampirik sonuçları yapay sinir ağlarının ve DVM ’nin iflas tahmininde geleneksel istatistik yöntemlerden daha başarılı olduğunu ortaya koymaktadır. Ayrıca çalışma sonucunda yapay sinir ağlarının ve DVM’ nin tahmin doğruluğu arasında anlamlı bir fark olmadığı raporlanmaktadır.

M.-Y. Chen (2012) çalışmasında doğrusal diskriminant analizi, lojistik regresyon, C5.0, CART, özdüzenleyici haritalar (self-organizing map/SOM), öğrenmeli vektör kuantalama (learning vector quantization/LVQ) ve DVM’nin tahmin performansına odaklanmakta ve sıralanan yöntemleri karşılaştırmaktadır. Çalışmada kullanılan veri seti Tayvan Borsasının veri tabanından elde edilmiştir. Veri seti Aralık 2000 ile Ocak 2010 arasındaki on yıllık dönemi kapsamaktadır ve Tayvan Borsasında işlem görmekte olan 200 işletmeyi içermektedir. Çalışmada dengesiz bir veri seti kullanılmıştır. Bu sorunun üstesinden gelmek maksadı ile herhangi bir işlem yapılmamıştır. Araştırmada bu seçimin daha iyi doğrulama sonuçları elde etmek maksadı taşıdığı ifade edilmektedir. (Loh ve Shih 1997; Ohlson 1980). Çalışmada kullanılan öznitelikler alandaki geçmiş çalışmalar incelenilerek ya da bir başka ifade ile literatüre dayalı olarak belirlenmiştir. Bu amaçla çalışmada toplam 50 adet öznitelige yer verilmiştir. Çalışmanın ampirik sonuçları DVM,

karar ağaçlarına ve yapay sinir ağlarına göre işletmenin sürekliliğinin tahmininde daha uygun bir metot olduğunu göstermektedir. Ayrıca çalışmada, genetik algoritmalar ve parçacık sürü optimizasyonu ile DVM, algoritmanın yerel optimuma takılma olasılığını azaltma ve global optimum arama yeteneğini artırmak maksadı ile entegre edilmiştir. Çalışma sonuçları ayrıca, önerilen bu modellerin potansiyel tahmin araçları olarak düşünülebileceğini göstermektedir.

Bae (2012) çalışmasında Kore Kredi Garanti Fonu (KODIT) tarafından derlenmekte olan 1888 imalat işletmesine ait mali tabloyu kullanmaktadır. Çalışmada DVM algoritması (radyal tabanlı fonksiyon), lojistik regresyon, çoklu diskriminant analizi, çok katmanlı algılayıcı, karar ağaçları ve bayes ağları metotları kullanılmaktadır. Analizde kullanılan işletmeler varlık büyüklükleri bir ile yedi milyon dolar arasında değişmektedir. Veri seti 1999-2005 yılları arasında iflas etmiş olan 944 işletmeyi ve 944 iflas etmemiş olan işletmeyi içermektedir. İlk olarak belirlenen 83 öznitelikten 54 adedi seçilmiştir. Bu aşamayı takiben stepwise lojistik regresyon kullanılarak eğitimde kullanılan 11 adet öznitelik seçilmiştir. Çalışmanın ampirik sonuçları DVM'lerinin karşılaştırılan diğer yöntemlerden önemli ölçüde daha iyi performans gösterdiğini ortaya koymaktadır.

Hsieh, Hsiao, ve Yeh (2012) çalışmalarında, daha etkin tahmin sonuçları üretmek için ceza fonksiyonu karakteristiklerinden faydalanan bir DVM modeli önermektedir. Modele parçacık sürü optimizasyonu özellikleri eklemek maksadı ile evrimsel bir arı kolonisi algoritması sunulmuştur. Önerilen ceza fonksiyonunun DVM' ni optimize etmesi için, her bir arıya bir hız ve uçuş yönü belirlenmiştir. Örneklemeye hangi gözlemlerin dahil edileceğinin belirlenmesi maksadı ile iki kriter kullanılmıştır. İlk olarak genel mali yapılarının farklı olması sebebi ile işletmelerin sigorta sektöründe faaliyet gösterip göstermediği incelenmiştir. İkinci olarak, işletmelerin altı yıllık verisinin erişilebilir olup olmadığı incelenmiştir. Önerilen modelin değerlendirilmesi maksadı ile iki farklı veri seti oluşturulmuştur. İlk veri seti 1999-2006 yılları arasını kapsamaktadır. Bu veri setinin 1999-2005 yılları arasında yer alan gözlemleri, 93 işletme başarısız ve 186 işletme başarılı olarak etiketlenmiştir, eğitim seti olarak kullanılmıştır. 2006 yılına ait olan gözlemlerden 10 işletme başarısız ve 1188 işletme başarılı olarak etiketlenmiş ve test seti olarak kullanılmıştır. İkinci veri seti 2000-2007 yılları arasını kapsamaktadır. Bu veri setinin 2000-2006 yılları arasında yer alan gözlemleri, 88 işletme başarısız ve 176 işletme başarılı olarak etiketlenmiştir, eğitim seti olarak kullanılmıştır. 2007 yılına ait olan gözlemler, 8 işletme başarısız ve 1250 işletme başarılı olarak etiketlenmiştir, test seti

olarak kullanılmıştır. Veri setine dahil edilen işletmelere ait mali bilgiler Tayvan Ekonomi Dergisi mali veri arşivinden ve örneklemedeki işletmelerin kurumsal raporlarından toplanmıştır. Çalışmanın ampirik sonuçları ceza fonksiyonu entegrasyonun DVM' nin tahmin başarısını artırdığını göstermektedir.

Huang vd. (2012) çalışmalarında iflas tahmini için yarı gözetimli DVM ile kernel yerel diskriminant analizi entegrasyonunu önermektedir. Yarı gözetimli DVM Belkin vd., (2006) Sindhvani vd., (2005) tarafından önerilen manifold tabanlı düzenleyiciyi kullanmaktadır. Çalışmada kullanılan veri seti Tayvan Menkul Kıymetler Borsasında işlem gören işletmeleri içermektedir ve 2000-2007 dönemini kapsamaktadır. Kottan çıkarılmış, iflas etmiş ve likidite sorunu yaşayan işletmeler mali sıkıntılı olarak etiketlenmiştir. Veri seti 50 adedi mali sıkıntılı olmak üzere toplam 100 işletmeye ait veriyi içermektedir. Araştırmada, beş farklı mali indeks altında organize edilmiş 54 mali oran Tayvan Ekonomi Dergisinin veri tabanından toplanmıştır. Eksik gözlem içeren mali oranlar silinmiş ve sonuç olarak 48 mali oran öznitelik olarak kullanılmıştır. Önerilen DVM modeli J48 karar ağaçları, k-en yakın komşu algoritması, bayes ağları, radyal tabanlı sinir ağları, geleneksel DVM ile karşılaştırılmıştır. Çalışmada J48 karar ağaçları ve DVM ile boosting temelli topluluk öğrenmesi modelleri oluşturulmuştur. Çalışmanın ampirik sonuçları önerilen modelin karşılaştırıldığı temel sınıflandırıcılara ve topluluk öğrenmesi modellerine kıyasla tahmin gücünü ve doğruluğunu ortaya koymaktadır.

Gordini (2014) genetik algoritmalar, lojistik regresyon ve DVM'ni karşılaştırdığı çalışmasında KOBİ'lerde iflas tahmini için güvenilir bir model aramaktadır. Çalışmada Ciampi & Gordini (2013) tarafından yapılmış olan temerrüt tanımı kabul edilmiştir. Buna göre temerrüt, iflas ya da tasfiye gibi bir borç için yasal sürecin başlaması olarak tanımlanmaktadır. Çalışmada kullanılan veri seti CERVED veri tabanından toplanmıştır. İlk veri setinde 3584 adet imalat sektöründe İtalya' da faaliyet gösteren KOBİ yer almaktadır. Verideki hata ve gürültüyü engellemek maksadı ile her bir mali oranın üst %1'lik kısmı ve alt %1'lik kısmı veri setinden çıkarmıştır. Bu işlemde sonra 1500 temerrüde düşmüş işletmeyi içeren 3100 gözlemin yer aldığı bir veri seti kalmıştır. Literatürdeki popülerliğine göre 28 mali oran, öznitelik olarak seçilmiştir. Öznitelik seçimi iki adımda yapılmıştır. İlk adımda varyans büyütme faktörü yöntemi ile çoklu bağlantı analizi yürütülmüştür. Öznitelikler arasındaki korelasyonu minimize etmek maksadı ile üçün üzerinde varyans büyütme faktörüne sahip öznitelikler veri setinde çıkarılmıştır. Bu adım sonrasında veri setinde on sekiz öznitelik kalmıştır. İkinci adımda

boyutsallığı azalmak ve öznitelik seçmek için aşamalı (stepwise) bir yöntem kullanılmış ve nihai olarak sekiz öznitelik ile modelleme yapılmıştır. Çalışmanın ampirik sonuçları genetik algoritmaların ve DVM birbirine çok yakın doğruluk değerlerine sahip olduğu göstermektedir. Ancak genetik algoritmalar karşılaştırmalı olarak daha düşük bir tip-2 hata oranına sahiptir.

Goo, Chi, ve Shen (2016) çalışmalarında kesin ve güvenilir bir süreklilik tahmini modeli oluşturmayı amaçlamaktadır. Bu bağlamda çalışmada üç farklı sınıflandırma algoritmasına yapay sinir ağlarına, CART karar ağaçlarına ve DVM ' ne yer verilmiştir. Araştırmada kullanılan veri seti 2003-2013 yılları arasında Tayvan' da süreklilik görüşü almış ve almamış işletmeleri içermektedir. Toplam 192 işletmeden oluşan örneklem içerisinde 48 işletme süreklilik görüşü almıştır. Veri setindeki işletmeleri temsil etmek maksadı ile literatüre dayanarak (M Anandarajan ve Anandarajan 1999; Behn, Kaplan, ve Krumwiede 2001; Kirkos, Spathis, ve Manolopoulos 2007; Martens vd. 2008; Yeh, Chi, ve Hsu 2010) 22 mali oran seçilmiştir. Çalışmada öznitelik seçimi amacı ile LASSO regresyonu kullanılmıştır. Çalışmanın ampirik sonuçları DVM'nin, CART ve yapay sinir ağlarına göre daha başarılı sonuçlar ortaya koyduğunu göstermektedir. Ancak modellerin tip-1 ve tip-2 hataları ile yapılan karşılaştırmada yapay sinir ağlarının hata oranlarının daha düşük olduğu görülmektedir.

Fallahpour, Lakvan, ve Zadeh (2017) çalışmalarında mali sıkıntı tahmini için açık ve güvenilir bir model önermeyi amaçlamaktadır. Bu bağlamda, çalışmada öznitelik seçim için bir sarmalama metodu olan "sequential floating forward selection", SFFS kullanılmaktadır. DVM ise sınıflama algoritması olarak seçilmiştir. Önerilen yöntem, yapay arı kolonisi (artificial bee colony – ABC), genetik algoritmaları, sıralı ileriye doğru seçim (SFS), temel bileşen analizi, relief algoritması ve bilgi kazanımı metotları ile karşılaştırılmaktadır. Çalışmada dört sarmalama metodu ve üç filtreleme metodu kullanılarak yedi farklı DVM modeli eğitilmiştir. Çalışmada bir işletmenin yanlış sınıflandırılmasının hem tip-1 hem de tip-2 hata için ortaya çıkardığı maliyetin farklı olmadığı belirtilmektedir. Bu nedenle çalışmanın birincil amacının önerilen modelin doğruluğunun değerlendirilmesi olduğu ifade edilmektedir. Örneklem 1996-2014 yılları arasında kapsamaktadır. Seçilen örnekler Tahran Menkul Kıymetler Borsasında işlem görmektedir. Veri seti toplam 180 işletmeden oluşmaktadır ve işletmelerin 90 adedi mali sıkıntılı olarak etiketlenmiştir ve toplam 29 mali oran öznitelik olarak kullanılmıştır. Araştırmacılar İran veri setinin dışında önerilen modeli Pietruszkiewicz (2008) tarafından

kullanılan veri seti ile deęerlendirmektedir. Pietruszkiewicz (2008) tarafından kullanılan veri seti 112 adedi mali sıkıntı olarak etiketlenmiş 240 işletmeden oluşmaktadır. Gözlemleri 30 mali oran temsil etmektedir. Çalışmanın ampirik sonuçları, öznitelik seçimi için SFFS metodunun kullanıldığı DVM modelinin diğer yöntemlerden açıkça daha yüksek doğruluğa sahip olduğunu göstermektedir.

Topluluk öğrenmesi, daha iyi tahmin performansı elde etmek maksadı ile temel sınıflandırma algoritmalarının eğitim çıktılarını birleştirilmesini içermektedir. AdaBoost, gradient boosting ve rastgele orman algoritması gibi topluluk öğrenmesi metotları tahmin güçleri ve matematiksel yapıları nedeniyle yaygın olarak kullanılmaktadır (Alam, Gao, ve Jones 2021). Makine öğrenmesi sürecinde ölçüm hatalarının önüne geçmek maksadı ile özniteliklerin normalleştirilmesi ya da standartlaştırılması gerekmektedir. Topluluk öğrenmesi fenomeni farklı yaklaşımları içerisinde barındırmaktadır.

Tsai ve Wu (2008) çalışmalarında, yapay sinir ağlarından oluşan bir topluluk öğrenmesi modeli ile tek bir yapay sinir ağının tahmin performansını üç farklı veri seti üzerinde yapılan eğitim ile karşılaştırmaktadır. Çalışmada eğitilen modeller hem doğruluk parametresi hem de Tip-1 ve Tip-2 hata düzeyleri itibari ile karşılaştırılmaktadır. Eğitim maksadı ile Avustralya, Almanya ve Japonya veri setleri kullanılmıştır. Eğitilen modeller 3 katmanlı geri yayımlı yapay sinir ağıdır. Çalışmanın ampirik sonuçları temel yapay sinir ağı yapısının, topluluk öğrenmesi modelinden ortalama tahmin performansının daha iyi olduğunu ortaya koymaktadır. Ancak Tip-1 ve Tip-2 hata değerleri üzerinden yapılan analizde temel yapay sinir ağlarının topluluk öğrenmesi modeline kesin bir üstünlük sağlamadığı görülmektedir.

Sun ve Li (2009) yürüttükleri çalışmada mali sıkıntı tahmininde temel sınıflayıcılar ve topluluk öğrenmesi metotlarını karşılaştırmaktadır. Çalışmada temel sınıflayıcıların ardışık olarak yer aldığı bir topluluk yapısı yer almaktadır ve toplulukta çoklu diskriminant analizi, lojistik regresyon, yapay sinir ağı, karar ağaçları, destek vektör makineleri ve vaka temelli çıkarsama kullanılmaktadır. Veri seti oluşturulurken “ST” etiket temel alınmıştır. Buna göre 2000-2005 yılları arasında Shenzhen borsasında ve Şanghai Borsasında “ST” etiketi almış olan 135 işletme aynı sayıda sağlıklı işletme ile eşleştirilmiş ve ilk veri seti oluşturulmuştur. Veri setindeki işletmeleri temsil etmesi için ilk olarak işletmelerin karlılık, faaliyet yeteneği, borç ödeme kabiliyeti ve büyüme kabiliyetini gösteren 35 mali oran seçilmiştir. Sonrasında öznitelik seçimi ile bu oranlar arasından yedi adedi öznitelik olarak belirlenmiştir. Araştırmanın ampirik sonuçları

topluluk öğrenmesi modelinin karşılaştırmalı olarak temel sınıflayıcılara bir kesin üstünlük kuramadığı göstermektedir.

Chandra, Ravi, ve Bose (2009) çalışmalarında, internet şirketlerinin iflas tehlikesi için bir erken uyarı sistemi geliştirmeyi amaçlamaktadır. Bu bağlamda çalışmada Bose (2006) ve Bose ve Pal (2006) tarafından kullanılmış olan veri seti üzerinden yararlanılmıştır. Kullanılan veri seti WRDS' ten alınmıştır ve 120 adedi başarısız olarak etiketlenmiş toplam 240 adet internet şirketini içermektedir. Chandra, Ravi, ve Bose (2009) veri setinde yer alan 240 işletmenin temsili için literatüre dayalı olarak 24 mali oran belirlemiştir. Belirlenen mali oranlardan 15 adedinin literatürde kullanılan popüler oranlar olduğu ifade edilmektedir. Araştırmada bir filtreleme yöntemi olan t-testini kullanarak ilk olarak belirlenen 24 mali orandan 10 adedi seçilmiştir. Çalışma da bir boosting modeli önerilmektedir. Önerilen model çok katmanlı algılayıcılar, rastgele orman algoritması, lojistik regresyon, DVM ve sınıflama ve regresyon ağaçları kullanılmaktadır ve model çoğunluk oylamasına dayanan bir topluluk öğrenmesi modeli ile karşılaştırılmaktadır. Çalışmanın ampirik sonuçları t-testi ile yapılan öznitelik seçiminin modelin tahmin başarısını artırdığını ve boosting modelin aynı veri seti ile yapılmış olan diğer çalışmalara kıyasla daha üstün sonuçlar ortaya koyduğunu göstermektedir.

Nanni ve Lumini (2009) çalışmalarında iflas tahmininde topluluk öğrenmesi metotlarının performansını değerlendirmeyi amaçlanmaktadır. Çalışmada bagging, random subspace, sınıf değiştirme (class switching) ve rotation forest metotlarının iflas tahmininde performansları değerlendirilmektedir. Model eğitimi ve testi için, Tsai ve Wu (2008) tarafından kullanılan veri setini kullanılmıştır. Çalışmanın ampirik sonuçları analiz edilen topluluk öğrenmesi yöntemleri arasından random subspace metodunun daha iyi performans gösterdiğini ortaya koymaktadır.

M.-J. Kim ve Kang (2010) çalışmalarında bagging ve boosting yöntemlerini çok katmanlı algılayıcılar üzerinde uygulamakta ve model sonuçlarını temel yapay sinir ağı modeli ile karşılaştırmaktadır. Çalışmada kullanılan veri seti Kore' de faaliyet gösteren bir bankadan elde edilmiştir. Veri seti 2002-2005 yılları kapsamakta ve yarısı iflas eden toplam 1458 imalat işletmesine ait bilgileri içermektedir. Literatür taraması ve temel istatistiksel testler yardımı ile ilk olarak belirlenen 32 mali oran içerisinde 7 mali oran öznitelik olarak belirlenmiştir. Çalışmanın ampirik sonuçları bagging ve boosting yöntemlerinin yapay

sinir ağlarının tahmin performansını artırdığını ortaya koymaktadır. Özellikle bagging yönteminin kullanıldığı model daha iyi tahmin performansı ortaya koymuştur.

Geng, Bose, ve Chen (2015) yürüttükleri çalışmalarında, mali sıkıntı tahmininde en iyi performansı hangi yöntem vermektedir, halka açık işletmelerde mali sıkıntı ne kadar erken tespit edilebilir ve son olarak mali sıkıntının belirlenmesinde hangi mali göstergeler önem taşımaktadır sorularına cevap aramaktadır. Bu bağlamda araştırmacılar yapay sinir ağları, karar ağaçları, DVM algoritmaları ile çoğunluk oylaması metotlarını sınamaktadır. Çalışmada kullanılan veri seti 2008-2011 yılları arasında Şanghai ve Shenzhen borsalarında işlem görmekte olan 214 işletmeyi kapsamaktadır. Veri setinde 107 adet işletme mali sıkıntılı olarak etiketlenmiştir. İşletmeler ilgili dönemde ST etiketi alma durumuna göre mali sıkıntı olarak etiketlenmiştir. Çalışmada mali sıkıntının ne kadar erken tahmin edilebileceği incelendiği için işletmelerin ST etiketi almasında üç, dört ve beş yıl önceki veri toplanmış ve bu verilerden model eğitiminde faydalanılmıştır. Ancak modellerin testi ST etiketi aldığı yıldaki veriler ile yapılmıştır. Veri setindeki gözlemleri temsil etmesi için 31 adet mali oran seçilmiş ve öznitelik olarak belirlenmiştir.

Bagging ya da boosting gibi topluluk öğrenmesi metotları veri analistlerinin, genellikle benzersiz bir tahmin kuralından oluşan modellerden daha doğru olan modeller üretmesine olanak tanımaktadır. Du Jardin (2016) çalışmasında farklı işletme modellerine uyan bir modeller tasarlamayı amaçlamaktadır. Bu amaca ulaşmak için Kohonen haritalarını kullanarak işletme tipolojileri tahmin edilmiştir. Üretilen tipolojiler işletmelerin mali durumlarını karakterize eden profiller oluşturmak maksadı ile kullanılmıştır. Çalışmada literatürde en sık kullanılan sınıflandırma algoritmalarına yer verilmiştir. Bunlar, CART, diskriminant analizi, lojistik regresyon, çok katmanlı algılayıcılardır. Ayrıca temel sınıflandırıcılar bagging, boosting ve random subspace metotları ile desteklenmiştir. Çalışmada önerilen yaklaşımın performansını test etmek maksadı ile oluşturulan mali profiller kullanılmadan da model oluşturulmuştur. Çalışmada kullanılan veri seti Diane veri tabanından toplanmıştır. Araştırmacı, 11 eğitim örnekleme ve 11 test örnekleme toplamıştır. Eğitim örnekleri 2002-2011 yılları arasını, test örnekleri ise 2003-2012 yılları arasını kapsamaktadır. Çalışmada sadece işletmelerin bilanço ve gelir tablolarından elde edilen mali oranlar kullanılmıştır. Çalışmanın ampirik sonuçları, topluluk öğrenmesi modellerinin rastgele oluşturdukları çeşitlilik sayesinde, profillemede temel sınıflayıcıların yakalamadığı bazı varyasyonları yakalama yeteneğine sahip olduğunu göstermektedir.

Choi, Son, ve Kim (2018) Kore inşaat sektöründe faaliyet gösteren taşeron işletmelerin mali sıkıntı riskini tahmin etmeyi amaçlayan bir topluluk öğrenmesi modeli önermeyi amaçlamaktadır. Yararlanılan topluluk öğrenmesi modeli oy temellidir. Çalışmada oylamaya dayalı bir topluluk öğrenmesi mimarisi kullanılmasının sebebi, birden fazla temel sınıflayıcının üreteceği bilgiden istifade etmektir. Eğitilmiş olan temel sınıflandırıcılar, DVM, yapay sinir ağları, C4.5 karar ağacı algoritması, lojistik regresyon, Bayes ağları, K-en yakın komşu algoritmasıdır. Topluluk öğrenmesi modelinde kullanılacak olan temel sınıflayıcılar eğitim sonrasında “Weighted Count of Errors and Correct Results” WCEC metodu ile seçilmiştir. Literatürde yöntemin topluluk oluşturmak için temel sınıflandırıcı seçiminde oldukça faydalı olduğu belirtilmektedir (Aksela ve Laaksonen 2006; P. Du vd. 2012). Çalışmada 2007-2012 yılları arasında yayınlanmış olan 1651 taşeron işletmenin yıl sonu mali raporlarından faydalanılmıştır. Eksik verisi olan işletmelerin veri setinden çıkarılması sonucunda veri setinde toplam 385 taşeron işletme kalmıştır. Araştırmacılar mali sıkıntı tanımı olarak Çin’ de kullanmakta olan “special treatment” (ST) etiketini kullanmaktadır. Bu bağlamda ST etiketi kullanımında Sun ve Li (2009) ve Li ve Sun (2008) takip edilmektedir. Yapılan değerlendirme sonucunda 2011 yılında 41 ve 2012 yılında ise 49 işletme mali sıkıntılı olarak etiketlenmiştir. Literatür incelenerek 21 mali oran öznelik olarak seçilmiştir. Çalışmada önerilen topluluk öğrenmesi modelinin temel sınıflayıcılara kıyasla daha başarılı sonuçlar ortaya koyduğunu göstermektedir.

De Bock (2017) çalışmasında, “spline-rule” topluluk öğrenmesi algoritmasının işletmenin başarısızlığının tahmini alanında sınımlanmaktadır. Önerilen modelin sınanması ve değerlendirilmesi maksadı ile çalışmada lojistik regresyon, doğrusal diskriminant analizi, kuadratik diskriminant analizi, C4.5 karar ağacı algoritması, çok katmanlı algılayıcılar, destek vektör makineleri, k-en yakın komşu algoritması, bagging, adaboost ve rastgele orman algoritmalarından faydalanılmıştır. Çeşitli sektör ve ülkelerden gözlemler içeren 21 farklı veri seti ile modelleme yapılmıştır. Veri setleri Belçika, Fransa ve İtalya’ da faaliyet gösteren işletmelerin bilgilerini içermektedir. Bir filtreleme metodu olan korelasyona göre öznelik seçim metodundan yararlanılmıştır. Çalışmanın ampirik sonuçları üç aşamalı olarak raporlanmaktadır. Birinci aşamada “spline-rule” topluluk öğrenmesi modeli, çalışmada yer verilen diğer topluluk öğrenmesi modelleri ile karşılaştırılmaktadır. İkinci olarak çalışmada faydalanmış olan düzenleme yöntemleri, ridge regresyonu, lasso regresyonu ve elastic net, bağlamında bir değerlendirme

yapılmaktadır. Üçüncü olarak, “spline-rule” topluluk öğrenmesi modeli, literatürde yaygın olarak faydalanılmış ve çalışmada da yer verilmiş olan temel sınıflayıcılar ile karşılaştırılmaktadır. Çalışmanın ampirik sonuçları “spline-rule” topluluk öğrenmesi modellerinin diğer topluluk öğrenmesi modellerinden ve temel sınıflayıcılardan daha iyi sonuçlara sahip olduğunu göstermektedir.

Wang, Chen, ve Chu (2018) çalışmalarında mali sıkıntılı işletmelerin sınıflandırılmasında nitel ve nicel veri ile desteklenmiş olan random subspace metodunun kullanıldığı bir topluluk öğrenmesi modeli önermektedir. Araştırmacılar önerdikleri modele “incorporating sentiment and textual information into a random subspace method (IST-RS)” olarak adlandırmaktadır. Çalışmada önerilen modeli karşılaştırmak maksadı ile DVM, bagging ve boosting metotları da kullanılmıştır. Veri seti 2011-2015 yılları arasında kapsamaktadır ve toplam 1726 işletmenin verilerini içermektedir. Seçilen işletmeler Shenzhen ve Şanghai Borsalarında işlem görmektedir. ST etiketi mali sıkıntıyı tanımlamak maksadı ile kullanılmıştır ve bu dönemde toplam 129 gözlem ST etiketi almıştır. Nitel ve nicel verilerin bir arada kullanıldığı çalışmada 39 mali oran seçilmiş ve öznitelik olarak veri setinde yer almıştır. Araştırmanın ampirik sonuçları önerilen IST-RS modelinin mali sıkıntının tahmininde oldukça yüksek performans gösterdiğini ortaya koymaktadır.

Barboza, Kimura, ve Altman (2017) çalışmalarında bagging, adaboost, rastgele orman algoritması, DVM (doğrusal ve radyal tabanlı fonksiyonlar), yapay sinir ağları, lojistik regresyon ve çoklu diskriminant analizine yer vermiştir. ABD’li ve Kanadalı işletmelere ait veriler Compustat veritabanından elde edilmiştir. Toplanan veri 1985-2013 yılları arasında kapsamaktadır. İşletmelerin iflasına ilişkin bilgiler NYU’ nun Salomon Center veri tabanından toplanmıştır. Çalışmanın ampirik sonuçları bagging, adaboost ve rastgele orman algoritmasının kullanılan diğer makine öğrenmesi metotlarına göre daha yüksek tahmin başarısına sahip olduğunu göstermektedir. Özellikle rastgele orman algoritması eğitim seti ve test seti ve ayrıca sadece Altman (1968) tarafından kullanılmış olan öznitelikler ile birlikte eğitilmiş tüm modellerden daha iyi sonuçlar ortaya koymuştur.

Terzi (2022) çalışmasında işletmenin sürekliliğinin tahmininde, yapay sinir ağları, karar ağaçları, DVM, rastgele orman algoritması, k-en yakın komşu algoritması, lojistik regresyon ve gaussian naive bayes algoritmasının performansını karşılaştırmaktadır. Çalışmada kullanılan veri seti Borsa İstanbul’ da 2010-2019 yılları arasında kesintisiz işlem gören 136 işletmenin verilerini içermektedir. Analize toplam 1360 gözlem dahil

edilmiştir. Süreklilik açısından riskli işletmelerin etiketlenmesi maksadı ile literatürde kullanılan kriterler esas alınmıştır. Bunlar dönem zararı, negatif özkaynak ya da sermaye kaybı, geçmiş yıl zararı, faaliyet zararı, negatif çalışma sermayesi, işletme faaliyetlerinden negatif nakit akışı, denetçi görüşünde süreklilik ile ilgili görüş bulunması ve z-skorudur. Bu kriterler ışığında 410 adet gözlem süreklilik açısından riskli olarak etiketlenmiştir. Veri setinde bulunan gözlemleri temsil etmesi maksadı ile literatüre dayalı olarak 12 adet öznel belirlenmiştir. Çalışmanın ampirik sonuçları rastgele orman algoritmasının ve karar ağaçlarının, tahmin başarısının ve hata düzeylerinin karşılaştırılan metotlara kıyasla daha iyi olduğunu ortaya koymaktadır.

Garcia (2022) çalışmasında temel olarak dengesiz veri seti ile çalışmanın ortaya çıkardığı zorluklara odaklanmaktadır. Ayrıca iflas tahmini literatüründe kullanılan algoritmalar hakkında anlayışın derinleştirilmesini de amaçlamaktadır. Çalışmada dengesiz veri seti sorununun üstesinden gelmek maksadı ile altı farklı örnekleme metodu test edilmektedir. Bunlar SMOTE (Synthetic Minority Oversampling Technique), ADASYN, Borderline SMOTE, DB SMOTE, Safe-Level SMOTE ve hibrid SMOTE Cluster -Based Undersampling metotlarıdır. Çalışmada veri seti dengelendikten sonra, her bir metodun ürettiği çıktı on bir farklı sınıflandırma algoritmasının eğitimi için kullanılmış ve sonuçlar karşılaştırılmıştır. Kullanılan sınıflandırma algoritmaları lojistik regresyon, boosted lojistik regresyon, doğrusal diskriminant analizi, en küçük kareler diskriminant analizi, naive bayes sınıflayıcıları, k-en yakın komşu algoritması, yapay sinir ağları, DVM-polinomal fonksiyon, XGBoost, rastgele orman algoritması, rastgele orman algoritması topluluğudur. Çalışmada kullanılan veri seti, Ocak 2010 ile Aralık 2018 dönemini kapsamaktadır. Toplam halka açık 1824 işletmeyi içine alan örneklem 41933 çeyrek dönemlik gözlemi içermektedir. Problem tanımı yasal bir düzeyde gerçekleşmiştir. Bir işletmenin bir çeyreklik dönemde iflas etmiş ya da iflas başvurusunda bulunmuş olması (ABD iflas kanununa göre Bölüm-7 veya Bölüm-11 içerisinde değerlendirilmek) işletmenin iflas etmiş olarak etiketlenmesi için kullanılan kuraldır. Örnek içerisinde Finans ve kamu sektöründen işletmeler yer almamaktadır. İlgili dönemde toplam 47 işletme tespit edilmiştir. Çalışmanın ampirik sonuçları hibrid SMOTE Clustering-Based Undersampling metodunun karşılaştırılan yöntemlere göre daha başarılı sonuçlar verdiğini ve sınıflandırmada kullanılan rastgele orman topluluğunun ise genel tüm örnekleme metotları ile diğer modellerden daha başarılı olduğunu ortaya koymaktadır.

Boosting metotların arkasındaki temel fikir güçlü bir model ortaya koymak için çok sayıda temel sınıflayıcının çıktılarını birleştirmektir (Alam, Gao, ve Jones 2021).

Adaboost, eğitim veri setinin farklı alt kümelerini kullanarak temel sınıflayıcıları ardışık olarak eğiten, yeni bir topluluk öğrenmesi algoritmasıdır. Alfaro vd. (2008), çalışmalarında, önceki çalışmalar ile karşılaştırmayı kolaylaştırmak maksadı ile yasal bir bakış açısı ile problemi tanımlamaktadır. Bu bağlamda çalışmalarına yalnızca iflas etmiş ya da geçici kayyum atanmış olan işletmeleri dahil etmişlerdir. Çalışmada kullanılan veri seti Bureau Van Dijk' e ait olan SABİ veri tabanından elde edilmiştir. Çalışmada başarısız olarak sınıflanmış olan işletmeler 2000-2003 döneminden seçilmiştir. İşletmelerin seçiminde başarısız olduğu kabul edilen yılda ve bir önceki yılda eksik verinin bulunmaması dikkate alınmıştır. Başarısızlık tahmini çalışmalarında genel bir yaklaşım olan işletme-yıl olarak gözlem seçme yaklaşımı tercih edilmiştir. Bu yolla veri setinde gözlemlerin sayısının artırılması hedeflenmektedir. Çalışmada 2003 yılı sonu itibari ile aktif durumda bulunan ve son üç yıl içinde sürekli olarak zarar etmemiş olan işletmeler başarılı olarak etiketlenmiştir. Bu kriterler etrafında toplam 1180 işletme seçilmiştir. Çalışmada on beş mali oran ve işletme büyüklüğünün göstergesi olarak işletme varlıklarının doğal logaritması öznitelik olarak kullanılmıştır. Araştırmanın sonuçları Adaboost algoritmasının YSA, DDA ve karar ağaçlarına kıyasla daha iyi performans ortaya koyduğunu göstermektedir.

Heo ve Yang (2014) inşaat işletmelerinin iflas riskinin tahmininde adaboost algoritmasının öngörü yeteneğinin test etmektedir. Adaboost algoritmasının görece tahmin gücünü test etmek ve kıyaslamak maksadı ile çalışmada yapay sinir ağları, DVM ve karar ağacı metotlarına da yer verilmiştir. Ayrıca z-skoru da geleneksel yöntemler ile değerlendirme yapmak maksadı ile kullanılmıştır. Modelleme için ihtiyaç olan veri Kore' de bulunan tüm inşaat işletmelerinin verileri arşivleyen NICE DnB' ten elde edilmiştir. Veri setindeki iflas etmiş olan 1381 işletme 2008-2012 yılları arasındaki dönemden seçilmiştir. 2012 itibari ile faaliyetlerine devam eden 28481 işletme veri setinde yer almıştır. Sürekliliğini kaybetmemiş olan işletmelerin sadece 2011 yılı verileri kullanılmıştır. İflas eden işletmelerin ise iflastan bir yıl önceki verileri kullanılmıştır. Ayrıca veri setinde yer alan işletmeler sermaye büyüklüklerine göre küçük, orta ve büyük olarak sınıflandırılmıştır. Bu yolla algoritmaların farklı büyüklükteki işletme grupları ile olan performanslarının da test edilmesi amaçlanmıştır. Çalışmanın ampirik sonuçları

adaboost algoritmasının işletme büyüklüklerine göre en yüksek tahmin performansını sahip olduğunu ortaya koymaktadır.

Süreklilik tahmininde gradient boosting metodunun kullanılmasının birçok faydası bulunmaktadır. İlk olarak metot yüksek boyutlu ve doğrusal olmayan veri yapıları ile iyi performans ortaya koyabilmektedir. Ayrıca kirli veriler ile de iyi tahmin performansı elde etmek mümkündür. Başka bir ifade ile aykırı gözlemlere ve eksik değerlere karşı duyarlı değildir. Metot öznitelikler arasındaki çoklu eş-doğrusallığın üstesinden de gelebilmektedir (White 2000; S. Jones 2017).

Jones (2017), çalışmasında istatistiksel yöntemlerin sınırlılıklarını aşan, gradient boosting algoritmasını tanıtmayı amaçlamaktadır. Çalışmada ABD’ de iflas etmiş olan kamu işletmelerinin yer aldığı geniş bir veri seti kullanılmaktadır. Capital IQ veri tabanından elde edilen veriler ile oluşturulan veri seti 1987 – 2013 yılları arasını kapsayan 1115 iflas etmiş olan işletmeyi içermektedir. İflas etmemiş olan işletmeler, Capital IQ veri tabanında yer alan tüm aktif olan işletmelerden oluşmaktadır. Çalışmada iflas etmiş olan işletmelerin iflas açıkladıkları yıldan önceki üç yıla ait olan verileri kullanılmıştır. İflas etmemiş sağlıklı işletmelerin ise Capital IQ veri tabanındaki en son mali tabloları kullanılmıştır. Çalışmada gradient boosting metodu ve lojistik regresyon metodu kullanılmaktadır. Çalışmanın ampirik sonuçları iflasın yüksek boyutlu veri ortamlarında daha iyi açıklanabildiğini ve tahmin edilebildiğini ortaya koymaktadır.

Altman vd. (2020), çalışmalarında lojistik regresyon, karar ağaçları, gradient boosting, yapay sinir ağları ve DVM algoritmalarının filtreleme,  $R^2$  metodu, ve sarmalama, stepwise lojistik regresyon, öznitelik seçim metotları kullanılarak süreklilik tahmin yeteneklerinin ölçülmesini amaçlamaktadır. Çalışmada araştırmacılar, 2003-2013 yılları arasını kapsayan ve toplam 59099 işletmeden oluşan örneklemeden yararlanmaktadır. Özel mülk sahipleri, vakıflar, dernekler, finans kurumları ve gayrimenkul şirketleri örneklem içerisinden çıkartılmıştır. Araştırmacıların kullandıkları kriterler ile yapılan eleme sonrasında son veri setinin %3,3’ünü oluşturan 1967 iflas etmiş olan işletme veri seti içerisinde yer almıştır. Araştırmacılar parametre araması yapmayarak DVM, gradient boosting ve karar ağacı algoritmalarının parametrelerini varsayılan değerleri ile kullanmıştır. Çalışmada kullanılan mali öznitelikler literatüre dayanarak seçilmiştir. Mali nitelikli öznitelikler ile birlikte mali nitelikte olmayan özniteliklere de yer verilmiştir. Bunun sebebi öznitelikler arasında yüksek korelasyon kaçınmak ifade edilmektedir. Çalışma sonuçları küçük işletmelerin hem mali nitelikli hem de mali

nitelikli olmayan çok geniş bir öznitelik grubuna dayanan verimli tahmin modellerinin geliştirilebileceğini ortaya koymaktadır. Ayrıca çalışma sonuçları lojistik regresyon yöntemi ile yapay sinir ağları ve DVM kadar iyi performans elde edilebileceğini ortaya koymaktadır. Öznitelik seçim yöntemleri arasında sarmalama yöntemleri filtreleme yöntemlerine göre daha başarılı sonuçlar elde edilmesini sağlamıştır.

X. Du vd. (2020) çalışmalarında makine öğrenmesi sürecinin her aşamasına öznitelik seçiminden model eğitime kadar uygulamayı amaçlamaktadır. Ayrıca çalışmada literatürde ilk olarak kümeleme tabanlı under-sampling metodundan (Clustering Based Under-Sampling – CUS) yararlanılmıştır. Araştırmacılar çalışmalarında, literatürde ilk olarak CUS metodunu ağaç temelli gradient boosting algoritması ve XGBoost algoritması ile entegre edilen bir model önermektedir. Kullanılan veri seti Çin Sermaye Piyasası ve Muhasebe Araştırması Veri Tabanında yararlanılarak derlenmiştir. Veri seti 2017-2018 yılları kapsamaktadır ve dahil edilen işletmeler Şanghai ve Shenzhen borsalarında işlem görmektedir. Çalışmada araştırmacılar problem tanımlama ve etiketleme maksadı ile “Special Treatment” ST etiketini kullanmaktadır. Toplam olarak veri seti içerisinde 7154 gözlem yer almaktadır. ST etiketi almış olan işletmelerin sayısı iki yıl için toplam, 256 adettir. Gözlemler 89 adet mali oran ile temsil edilmektedir. Çalışmanın ampirik sonuçları önerilen modelin karşılaştırıldığı topluluk öğrenmesi algoritmalarında çok daha başarılı sonuçlar ortaya koyduğunu göstermektedir. Ayrıca çalışmada mali sıkıntı alanında topluluk öğrenmesi modellerinin tek bir temel sınıflandırıcı kullanılan modellere kıyasla üstünlüğünü ortaya koymaktadır.

Jabeur vd. (2021) çalışmalarında süreklilik tahmini için karar ağacı temelli gradyan boosting metodu, CatBoost’ u önermektedir. Önerilen modelin performansının karşılaştırılması maksadı ile çalışmada diskriminant analizi, lojistik regresyon, destek vektör makineleri, yapay sinir ağları, rastgele orman algoritması, gradyan boosting, derin nöral ağlar ve XGBoost algoritması kullanılmaktadır. Araştırmacılar, 2017 yılında iflas başvurusunda bulunmuş ve iflas başvurusundan önceki üç yılın mali raporları erişilebilir olan Fransa’ da faaliyet gösteren 133 işletme ile veri setini oluşturmuştur. İhtiyaç duyulan veriler Orbis veri tabanından yararlanılarak derlenmiştir. Literatüre dayanarak belirlenen 18 mali oran, öznitelik olarak kullanılmıştır. Çalışmanın ampirik sonuçları önerilen modelin diğer yaklaşımlara kıyasla tahmin gücünün daha iyi olduğunu göstermektedir. XGBoost, kayıp fonksiyonlarının Taylor açılımını optimize etmek maksadı yeniden düzenlenmiştir. Metot gradient boosting’ e çeşitli optimizasyonlar getirmektedir. Zięba,

Tomczak, ve Tomczak (2016), çalışmalarında, XGBoost algoritmasının süreklilik tahmini alanına uygulanmasının nedenini, işletmeyi karakterize eden tahmin edicilerinin nispeten az sayıda örneğin neden olduğu yüksek varyansla karakterize edilmesi olarak ifade etmektedir. Başka bir ifade ile bu durum, değerlerin dar bir segmentte toplandığı anlamına gelmektedir. Ancak bazı işletmeler, veri setinde yer alan özniteliklerin nispeten yüksek ya da küçük değerleri ile de tanımlanabilmektedir. Araştırmacılar, çalışmalarında bu durumun, özellikle yapay sinir ağları ve lojistik regresyon gibi gradyan temelli yöntemler ile eğitim sorunlarına ve düşük tahmin performansına neden olabileceğini ifade etmektedir. Verilerin normalleştirilmesi de bu problemin üstesinden gelmek için yeterli olmamaktadır. Çalışmada, açıklanan problemin üstesinden gelmek maksadı ile bir topluluk öğrenmesi yöntemi olan XGBoost önerilmektedir. Çalışmada, modellerin eğitilmesi maksadı ile Polonya’ da faaliyet gösteren işletmelere ait veriler kullanılmıştır. “Emerging Markets Information Service” veri tabanının elde edilen veri seti 2007-2013 yılları arasında iflas etmiş olan işletmelerin verilerini içermektedir. İlgili dönemde yaklaşık olarak 70 işletmenin iflas ettiği tespit edilmiştir. İflas etmemiş olan işletmeler 2010-2012 yılları arasından seçilmiştir. Toplam 10700 imalat işletmesine ait gözlemi içeren dengesiz veri setinde gözlemler 64 adet mali oran ile temsil edilmektedir. Çalışmada kullanılan mali oranlar, verinin erişilebilirliği ve literatürde kullanılabilir olup olmamasına göre belirlenmiştir. Çalışmada doğrusal diskriminant analizi, çok katmanlı algılayıcılar, JRip, JRip algoritmasının maliyet duyarlı versiyonu, J48 karar ağacı algoritması, J48 karar ağacı algoritmasının maliyet duyarlı versiyonu, lojistik regresyon, maliyet duyarlı lojistik regresyon, AdaBoost, AdaCost, DVM, Maliyet duyarlı DVM, rastgele orman algoritması, XGBoost ve XGBoost algoritmasının iki farklı versiyonu modellenmiştir. Çalışma sonuçları seçilen algoritmaların arasından işletmelerin iflasını tahmin etmede XGBoost temelli yöntemlerin referans yöntemlere göre daha başarılı sonuçlar ortaya koyduğunu göstermektedir.

Y.-P. Huang ve Yen (2019) çalışmalarında çeşitli gözetimli ve gözetimsiz algoritmalarının performanslarını karşılaştırmaktadır. Çalışmada kullanılmış, gözetimli öğrenme metotları geleneksel DVM, hibrid çevirili ilişkisel bellek, hibrid genetik algoritmalar bulanık kümeleme ve XGBoost algoritmalarıdır. Gözetimsiz öğrenme algoritmaları ise derin inanç ağları ve hibrid derin inanç ağları destek vektör makineleridir. Modelleme maksadı ile kullanılacak olan veri setinde Tayvan Ekonomi Dergisi veri tabanında 2010-2016 yılları arasında mali sıkıntı olarak etiketlenmiş olan 32

iřletmeye yer verilmiřtir. Bu iřletmeler aynı yıllardan 32 adet mali sıkıntılı olmayan iřletme ile eřleřtirilmiřtir. Gzlemleri temsil etmesi maksadı ile 16 mali oran literatre dayalı olarak seilmiřtir. alıřmanın ampirik sonuları karřılařtırılan gzetimli ğrenme metotları arasında XGBoost algoritmasının en bařarılı tahminci olduėunu gstermektedir. Ayrıca hibrit derin inan aėları destek vektr makinelerinin geleneksel DVM ve derin inan aėlarına kıyasla daha doėru tahminler rettiėini ortaya koymaktadır.



## **DÖRDÜNCÜ BÖLÜM**

### **4.1. İşletmenin Sürekliliğinin Tahminine Yönelik Borsa İstanbul AŞ. İmalat Sektöründe Bir Uygulama**

Çalışmanın bu bölümünde, teorik olarak açıklanmış olan makine öğrenmesi algoritmaları Borsa İstanbul AŞ. İmalat sektöründe faaliyet gösteren işletmelerin yer aldığı bir veri setinden faydalanılarak eğitilmiş ve sonuçları detaylı bir şekilde tartışılmıştır.

### **4.2. Araştırmanın Amacı ve Kapsamı**

Bu çalışmanın amacı, işletmenin sürekliliğinin tahmininde makine öğrenmesi metodlarının etkinliğinin tespit edilmesidir. Bu bağlamda çalışmada 2017-2021 yılları arasını kapsayan, Borsa İstanbul A.Ş pay piyasası imalat sanayi ve alt sektörlerinde işlem görmekte olan süreklilik görüşü verilmiş olan 16 işletmeye ait 47 gözlem yılı ve süreklilik görüşü verilmemiş 198 işletmeye ait 958 gözlem yılının yer aldığı bir veri seti kullanılarak lojistik regresyon, sınıflandırma ve regresyon ağacı, destek vektör makineleri (doğrusal, polinomsal ve radyal tabanlı kernel fonksiyonları), rastgele orman algoritması, adaboost algoritması, gradient boosting algoritması ve extreme gradient boosting algoritması on katmanlı çapraz doğrulama prosedürü uygulanılarak eğitilmiştir. Çalışmada kullanılan mali oranlar Datastream veri tabanından elde edilmiştir. Modellerin eğitiminde R programlama dili ve R Studio kullanılmıştır.

### **4.3. Araştırmanın Önemi**

İşletmenin sürekliliği varsayımı, BDS- 570 İşletmenin Sürekliliği Standardında yer alan pek çok bilgi sistemi, yönerge ve prosedürü kapsamaktadır. Varsayımın kapsadığı prosedürlerin ve yönergelerin, işletme yönetimi ve bağımsız denetçinin işletmenin sürekliliği hakkında yürüttüğü inceleme için de gerekli ve yararlı oldukları kabul edilmektedir. İşletmenin sürekliliği varsayımı, işletmenin varlığını sürdürme ve gelecek dönemlerde önemli bir müdahale olmaksızın faaliyetlerine devam edebilme kabiliyetini işaret etmektedir ve mali tablolar süreklilik varsayımı altında hazırlanmaktadır. BDS-570 İşletmenin Sürekliliği Standardı, işletme yönetimi ve bağımsız denetçiye, işletmenin sürekliliği sürdürme kabiliyetini değerlendirme konusunda sorumluluklar yüklemektedir. Bağımsız denetçi, işletmenin sürekliliği devam ettirme kabiliyeti konusunda önemli bir belirsizliğin mevcut olması durumunda, ilgili duruma denetim raporunda yer vermektedir (Zdolšek, Jagrič, ve Kolar 2022). Bağımsız denetim, ekonomide istikrar sağlayıcı bir

unsurdur. Denetçiler, objektif muhakemelerini kullanırken aynı zamanda belirli yönergeler çerçevesinde çalışmalarını yürütmek ile yükümlüdür. Denetçinin denetim sürecinde takip etmesi gereken temel sorumluluklarından biri, işletmenin sürekliliğini devam ettirebilme kabiliyetinin sınanmasıdır. Bir başka ifade ile denetçinin, işletmenin faaliyetlerini sürdürme kabiliyeti konusunda bir risk olup olmadığını belirlemek için denetlenen işletmenin mali durumunu incelemesi gerekmektedir (E. Altman 1982). Bu varsayım altında denetçiler, işletmenin gelecek bir yıllık dönem boyunca sürekliliğini devam ettirme kabiliyetine ilişkin önemli şüpheleri değerlendirmesi gerekmektedir. İşletmenin sürekliliğini değerlendirme sürecinde denetçilerin en önemli aracı, mesleki muhakemeleridir. Denetçinin, işletmenin süreklilik kabiliyeti hakkındaki görüşü, işletmenin ciddi bir mali sıkıntı ile karşı karşıya olduğunun kamuya ilk olarak duyurulması olarak değerlendirilmektedir (Kida 1980). İşletmenin sürekliliğini devam ettirme kabiliyeti hakkında önemli şüphe ifadesi tam olarak açık bir şekilde tanımlanmamıştır. Bu husustaki karar denetçinin muhakemesine bırakılmıştır. Denetçilerin temel görevi, işletmenin faaliyetlerine devam etme veya faaliyetlerini durdurmalarına yol açabilecek mali sıkıntılarla karşılaşma olasılığını değerlendirmek maksadı ile işletmenin mali tablolarından elde edilen mali bilgileri araştırmaktır. Ayrıca, işletme yönetimi genel amaçlı bir mali tablo aracılığıyla işletmeyi tasfiye etme niyetini beyan etmedikçe, işletmeler mali tablolarını işletmenin sürekliliği esasına göre hazırlamaktadır. Bu nedenle, mali tablo kullanıcılarına ve diğer paydaşlara mali tabloların önemli bir yanlışlık içermediği ve süreklilik esasına göre hazırlandığının güvencesini sağlamak denetçinin sorumluluğundadır. Literatürde yer alan araştırmalar, süreklilik görüşü almış olan işletmelerin iflas ile karşılaşma olasılığının daha yüksek olduğunu göstermektedir (Foster ve Shastri 2016).

Çoğu durumda işletmenin sürekliliğini kaybetmesi, işletmenin mali durumuna yansıyan işletme içerisindeki çeşitli faktörlerin neticesidir. İşletmenin sürekliliğini kaybetmesi ve sonuç olarak iflas ile karşı karşıya kalmasında faaliyet gösterdiği sektörün özellikleri, sektördeki belirsizlik, kurumsal yönetimdeki yetersizlik, inovasyon eksikliği, işletme yönetiminin performansındaki eksiklik, yüksek faiz oranları, ekonomik durgunluk ve belirsizlik ve işletmenin yüksek borç stoğu gibi pek çok faktör etkili olmakta ve işletmenin örgütsel hedeflere ulaşmasını engellemektedir (Gordini 2014; du Jardin 2015). Kötü yönetimin eşlik ettiği sürekli mali kayıpların birikimi, işletmenin sürekliliğini koruma kabiliyetini olumsuz etkilemekte ve işletmenin paydaşlarının da yer aldığı bir

senaryoda büyük ekonomik maliyetlere yol açabilmektedir. Bu nedenle, işletmenin süreklilik kabiliyeti hakkında erken uyarı veren sistemlerinin geliştirilmesi önemli bir konu haline gelmekte ve hem akademisyenler hem de işletme paydaşları tarafından işletmenin sürekliliğinin tahminini amaçlayan pek çok model geliştirilmektedir (Alam, Gao, ve Jones 2021; Murugan Anandarajan, Lee, ve Anandarajan 2001).

İşletme büyüklüğü ne olursa olsun süreklilik problemi ile karşı karşıya kalan işletmelerin, problem ortaya çıkmadan önce piyasa değerlerinde büyük kayıplar yaşadığı göstermektedir. Literatürde süreklilik görüşünün piyasalar üzerinde de etkisi olduğunu ortaya koyan çalışmalar bulunmaktadır (Fleak ve Wilson 1994; F. L. Jones 1996; Loudder vd. 1992). Denetçinin denetim raporunda, işletmenin sürekliliğini devam ettirme kabiliyeti hakkında ciddi şüphe oluşturabilecek durumlara işaret etmesi durumu piyasanın işletme hakkındaki değerlendirme mekanizmasında değişikliklere yol açabilecek önemli bir iletişim aracıdır (Blay, Geiger, ve North 2011). Bu durum özellikle iç ve dış paydaşları ciddi anlamda olumsuz etkilemektedir (Chandra, Ravi, ve Bose 2009). Ayrıca işletmenin süreklilik bağlamında karşılaşması muhtemel riskler, denetçilerin işletmenin sürekliliği ile ilgili erken uyarı sinyalleri sağlayan nitelikli denetim görüşü vermesi gerekmesi sebebiyle yasal riskler ile karşı karşıya kalmasına neden olabilmektedir (Murugan Anandarajan, Lee, ve Anandarajan 2001; Lisic vd. 2015). İşletmenin süreklilik görüşü almasından önce piyasalar temel olarak işletmenin net karına odaklanmaktadır. Ancak işletmenin süreklilik görüşü alması sonrasında piyasa katılımcıları odaklarını gelir tablosundan, işletmenin varlıklarına ve yükümlüklerine revize etmektedir. Süreklilik görüşü yatırımcıya işletme hakkında kamuya paylaşılandan daha fazla bilgi aktarmaktadır. Başka bir ifade ile süreklilik görüşü gizli bir bilgi kaynağıdır. Bu argümanlardan yola çıkarak denetim raporunun tüm çıkar grupları için önemli bir bilgi kaynağı olduğu açıktır (Foster ve Shastri 2016).

İşletmenin sürekliliğinin tahmini, işletmenin durumunun (işletmenin sürekliliği açısından başarılı ya da başarısız olarak) nasıl sınıflandırılacağı sorusuna çözüm aramaktadır. Beaver (1966) ve Altman (1968) gibi işletmenin sürekliliğinin tahmininde öncü çalışmaları takip eden ve pek çok farklı istatistiksel yöntemle yer veren araştırma yapılmıştır. Araştırmacılar tahmin modellerinin başarısını artırmak maksadı ile pek çok farklı yöntemden faydalanmıştır (Bae 2012). Yürütülen ilk çalışmalarda, işletmenin sürekliliğini devam ettirme kabiliyeti hakkında denetçi yargısını desteklemek maksadı ile diskriminant analizinden ve takip eden çalışmalarda bağımsız değişkenleri açıklama

gücünü test etmek maksadı ile yaygın olarak lojistik regresyon modellerinden faydalanılmıştır (Menon ve Schwartz 1987; Gaeremynck ve Willekens 2003; T. B. Bell ve Tabor 1991; Behn, Kaplan, ve Krumwiede 2001; Alam, Gao, ve Jones 2021; Martens vd. 2008; Murugan Anandarajan, Lee, ve Anandarajan 2001). Ancak katı varsayımlara sahip istatistiksel metotların yerine daha güçlü modeller geliştirmek maksadı ile makine öğrenmesi metotları ikame edilmiştir (Barboza, Kimura, ve Altman 2017; Heo ve Yang 2014; C.-F. Tsai, Hsu, ve Yen 2014; du Jardin 2015; M.-Y. Chen 2011a).

İşletmelerin mali durumu hakkındaki büyük miktardaki veri, işletmenini sürekliliği hakkında değerlendirme yapan birçok araştırmacının sınıflandırma yaparken makine öğrenmesi modellerine yönelmesine neden olmuştur. İşletmenin sürekliliğinin tahmini amacı ile geliştirilen modellerde zaman içerisinde farklı teknikler kullanılmıştır. Araştırmacılar modellerin sınıflandırma yeteneklerini geliştirmek maksadı ile temel sınıflayıcılardan, topluluk öğrenmesi metotlarına kadar pek çok farklı yaklaşımı test etmektedir. Alanda kullanılan modellerin sayısının artması, aynı zamanda modellerin karmaşıklığının ve hesaplama maliyetlerinin de artmasına neden olmaktadır. Ancak, model karmaşıklığının artması ve yorumlanılabilirliğin azalması daha iyi modeller elde edilmesine olanak tanımaktadır (Sangjae Lee ve Choi 2013; G. Wang, Ma, ve Yang 2014; Barboza, Kimura, ve Altman 2017; Choi, Son, ve Kim 2018; du Jardin 2021).

İşletmeler rekabet gücünü korumak maksadı ile farklı ve yenilikçi yönetim yaklaşımları ile yeni bilişim sistemlerini benimsemekte ve uygulamaktadır. İşletmeler, veri odaklı teknolojilerin geliştirilmesi ve adaptasyonu konusunda oldukça açık bir tutum sergilemektedir. Özellikle yapay zekâ ve ilgili teknolojiler geleceğin muhasebe ve denetim uygulamalarında öneminin artacağı düşünülmektedir (Elliott 1992). Bu durum nedeniyle denetim, veri analitiği ve yapay zeka teknolojilerindeki gelişmelerden etkilenmekte ve bir dönüşüm geçirmektedir (Lombardi, Bloch, ve Vasarhelyi 2014; Wagner 2016). Literatürde yapay zekanın ve makine öğrenmesinin denetim uygulamaları bağlamında daha fazla incelenmesi gerektiği belirtilmektedir (Debreceny 2011). Denetim sürecinde otomasyonun kullanımının artması, süreçte insan kaynaklı hataların azalmasına ve performansın artmasına katkı sağlamaktadır. Veri analitiğinin denetçiler adına faydalı ve anlamlı hale getirilmesi için otomatikleştirilmiş süreçlerin denetim sürecine adapte edilmesi gerekmektedir (T. Sun ve Vasarhelyi 2017). Denetim sürecinde yapay zekâ ve veri analitiğinin kullanımı, denetim kalitesinin artmasına olanak tanıyacak unsurlardan birisidir. Denetim sürecinde, yapay zeka ve veri analitiğinin kullanılması denetlenen

işletmenin süreçlerinin, faaliyetlerinin ve muhasebe bilgi sisteminin risk yönetimi bağlamında daha iyi analiz edilmesine olanak tanıyacaktır (Manita vd. 2020).

Denetim sürecinin en önemli parçası olan denetçiler, yeni teknolojik gelişmeleri, denetim sürecine dahil etme konusunda teşvik edilmekte ve böylece denetim sürecinin etkinliğinin artması beklenmektedir (Dai ve Vasarhelyi 2016; Manyika vd. 2017). Ayrıca denetim sürecinin kendini tekrarlayan bir yapıya sahip olması sebebi ile yapay zekâ teknolojilerinin sürece entegrasyonu, denetim sürecinin otomatikleşmesine büyük katkı sağlayacaktır ve denetim sürecinde kendini tekrarlayan unsurların yönetimini kolaylaştıracaktır. Bu yolla denetçi, denetim sürecinde değer yaratma adına daha fazla imkân elde edebilecek ve denetçinin süreç içerisinde değer yaratma potansiyelinin artması denetimin verimliliğini de artıracaktır. Yakın gelecekte, denetim firmaları yeni denetim metotları ile geleneksel denetim uygulamalarını terk edecektir (Luo, Meng, ve Cai 2018; Raphael 2015). Denetim sürecinde, denetçiler denetim verimliliğini arttırmak için yeni araçlara ihtiyaç duymaktadır. Denetim doğası gereği veri analitiği ve makine öğrenmesi metotlarının kullanımı için araştırmacı ve uygulamacılara motivasyon sağlamaktadır. Denetim süreci risk, tekrarlayan yapısal ve yapısal olmayan karar, eksik bilgi ve belirsizliğin değerlendirilmesi süreçlerini kapsamaktadır (Baldwin, Brown, ve Trinkle 2006). Denetim sözleşmesinin yapılmasını takiben denetim süreci risk değerlendirme ve denetimin kapsamını ve hedeflerini kapsayan bir denetim planının oluşturulması ile devam etmektedir. Bunu takiben denetçi, denetim kanıtı toplar ve analiz eder ve ayrıca yönetim tarafından sağlanan bilgilerin güvenilirliğinin yanı sıra iç kontrol yapısına ilişkin görüş oluşturur. Denetim süreci denetçinin denetlenen işletme ile ilgili görüş oluşturması ile tamamlanır.

İşletmenin sürekliliğinin tahmini için başarılı modeller geliştirilmesi işletme yatırımcıları, alacaklıları, çalışanları, tedarikçileri gibi işletme paydaşları ve diğer sektör katılımcıları için çok önemli bir bilgilendirici rol oynamaktadır (Alfaro vd. 2008). Veri analitiği ve makine öğrenmesi metotlarının denetim süreci içerisinde farklı amaçlar ile kullanımını örnekleyen pek çok çalışma mevcuttur (Amani ve Fadlalla 2017). Denetim sürecinde, denetlenen işletmenin süreklilik kabiliyeti denetçinin değerlendirmesi gereken konulardan birisidir. İşletme paydaşlarının kararları üzerinde de büyük bir etkiye sahip olması sebebi ile süreklilik riski literatürde güncel bir konudur. Bu kapsamda denetçinin işletmenin süreklilik kabiliyeti hakkındaki görüşünü desteklemek amacı ile farklı yöntemler ile modeller geliştirilmiştir (Cecchini vd. 2010; Chye Koh ve Kee Low 2004a;

Fallahpour, Lakvan, ve Zadeh 2017a; Fernández-Gámez vd. 2020; Gepp, Kumar, ve Bhattacharya 2010a; S. Y. Kim 2011a; S. Y. Kim ve Upneja 2014; Li ve Sun 2011; Liang vd. 2020a; Martens vd. 2008; Olson, Delen, ve Meng 2012a; Sánchez-Serrano vd. 2020; J. Sun vd. 2020; Veganzones ve Séverin 2018; Volkov, Benoit, ve Van den Poel 2017; Desheng(Dash) Wu, Liang, ve Yang 2008a; Xu ve Wang 2009a; Yeh, Chi, ve Lin 2014a; Lukason ve Camacho-Miñano 2019; Jayasekera 2018; Alam, Gao, ve Jones 2021; S. Jones ve Hensher 2008).

İşletme yatırımcılarının bakış açısından işletmenin süreklilik riskini dolayısıyla işletmenin süreklilik kabiliyetini yitirmesi nedeni ile kaynaklanabilecek olan tüm riskleri en aza indirmek çok önemlidir. Yatırımcılar, sermaye kazançlarını korumak ve temettü ödemesi elde etmek maksadı ile işletmenin süreklilik yeteneğini değerlendirmeyi amaçlamaktadır. Ayrıca işletmenin süreklilik kabiliyetini koruması, operasyonlarının devamlılığı ve mali durumunu anlama ihtiyacı içerisindeki diğer paydaşlar için de oldukça önemlidir (Zdolšek, Jagrič, ve Kolar 2022). İşletme yatırımcıları denetimden geçmiş olan mali tabloları birer değerlendirme ve erken uyarı aracı olarak görmektedir (Francis 2004; Smolarski, Wilner, ve Yang 2011; Agustia, Muhammad, ve Permatasari 2020). İşletmenin sürekliliğini devam ettirme kabiliyetine yönelik önemli bir şüphenin varlığını tespitin en kolay yolu denetim raporunun incelenmesidir. Bu bağlamda denetçi, işletmenin geleceği hakkında içeriden bilgi sağlayan önemli bir kaynaktır (Barnes 2004). İşletme sürekliliğinin tahmini, işletme paydaşlarının rasyonel ekonomik kararlar almasına yardımcı olması açısından büyük önem taşımaktadır. İşletmenin mevcut durumu piyasa katılımcılarının yanı sıra politika yapıcılar ve küresel ekonomi için de büyük önem arz etmektedir. Bu nedenle, işletmenin sürekliliğini koruma kabiliyetini yitirmesinden kaynaklanan iktisadi ve sosyal maliyetler, araştırmacıların dikkatini çekmekte ve araştırmacıların işletmenin mali durumunu yansıtan göstergeler kullanarak gelecek ile ilgili değerlendirmeler yapmaya yönelmektedir (Jabeur vd. 2021). Mali oranlar işletmenin sürekliliğinin tahmininin amaçlandığı çalışmalarda uzun süredir kullanılmaktadır (S. Jones 2017; Brabazon ve O'Neill 2004). İşletmenin sürekliliğinin tahmini maksadı ile başarılı modeller geliştirilmesi için işletmenin mevcut durumu hakkında gerekli bilgilere erişimin bulunması gereklidir. Mali oranlar ihtiyaç duyulan bilgiyi sağlamaktadır (Alfaro vd. 2008). Mali oranların hesaplandığı bilgileri içeren, mali tablolar ve dipnotları, işletmenin süreklilik riski ile ilgili önemli sinyaller içermektedir. Denetim raporunda ise denetçinin mesleki muhakemesine göre, işletmenin gelecekte

gelirleri elde edebilme ya da varlıklarının kullanımına devam etmesi ve yükümlülüklerini yerine getirebilmesi üzerinde önemli şüphe oluşturan olaylar hakkında bilgi verilmektedir. Mali tablo kullanıcıları işletmenin düşük riske sahip olması durumunda, işletmenin varlığının sürdürebileceği varsayımı ile hareket etmektedir. İşletmenin süreklilik kabiliyetinin denetim raporunda açıkça sorgulanmaması durumu, işletmenin gelecekteki gelirleri sonucunda ortaya çıkacak net varlıklara işaret etmektedir. Ancak süreklilik varsayımının zarar görmesi, gelecekteki gelirlerin olasılığını ortadan kaldırmakta ya da önemli ölçüde azaltmaktadır (Blay, Geiger, ve North 2011). Bu bağlamda işletmelerin mali raporlama sistemleri paydaşları tarafından, durumu hakkında bilgi üretmek için kullanılan önemli göstergeler olarak kabul edilmektedir (D. A. Williams 2016). Ayrıca işletmenin sürekliliği hakkındaki denetçi görüşü mali oranlardan üretilmiştir. Denetim raporunun işletmenin sürekliliği hakkında aktardığı bilgi, işletmenin yıllık mali oranlarının bir parçasıdır (Blay, Geiger, ve North 2011). Mali raporlama, işletme yönetimi tarafından tüm paydaşlara mali bilgileri sunmak maksadı ile kullanılan bir iletişim dilidir. Mali raporlama sistemi içerisinde işletme nicel ve nitel nitelikli bilgi sunumu gerçekleştirmektedir. Nitel nitelikli bilgi paylaşımı işletme yönetiminin varsayımları ve tahminleri ile bunları yansıtmak için uyguladığı muhasebe politikalarını içermektedir. Nicel nitelikteki bilgi sunumu ise mali raporlamanın geçerli bir düzenleyici çerçeve esasında temelini oluşturmaktadır. Nicel ya da mali nitelikteki açıklamaların temel amacı bilgi kullanıcılarına yüksek kalitede bilgi sağlamaktır (Ball 2006). İşletmeler tarafından paylaşılan tüm mali tablolar, paydaşların karar süreçlerinde temel farklılıklar oluşmasına yol açmaktadır (Barth, Landsman, ve Lang 2008). Bu durum mali tabloların ve aktardığı bilginin önemine işaret etmektedir. Literatürde yer alan araştırmalar, işletmelerin mali yapısı ile süreklilik yeteneği arasında önemli bir ilişki bulunduğunu ortaya koymaktadır (Mutchler 1985; Mutchler, Hopwood, ve McKeown 1997; Behn, Kaplan, ve Krumwiede 2001; Parker, Peters, ve Turetsky 2005). Ayrıca mali oranlar halka açık ve nesnel nitelik taşımaktadır. Bu özellikleri onların kolaylıkla elde edilebilir ve güvenilir olmalarını sağlamaktadır (Balcaen ve Ooghe 2006). Yıllık mali tablolar vasıtası ile aktarılan bilgilerin mali oran şeklinde süreklilik tahmininde en iyi sonuçları verdiği yönünde de literatürde genel bir kabul bulunmaktadır (Adnan Aziz ve Dar 2006).

Denetçinin, işletme hakkındaki karar süreci doğası gereği subjektiftir ve denetçinin bu süreçte mesleki yargıda bulunması gerekmektedir. Literatürde, bağımsız denetçinin

işletmenin sürekliliği hakkındaki kararının belirleyicileri hakkında yürütülmüş olan çalışmalar mevcuttur. Denetçiler işletmenin sürekliliğini devam ettirme kabiliyeti hakkında belirli temel göstergeleri ölçü almaktadır (Mutchler 1984). Denetçi karar sürecinde işletmenin faaliyetlerinden nakit akışları, cari oranı, işletmenin net değeri, uzun vadeli borçları, toplam borçları ve net karı gibi temel mali oranları göz önüne almaktadır. Sayılan mali oranların denetçinin yargısı üzerinde güçlü bir etkiye sahip olduğu literatürde ifade edilmektedir (Mutchler 1985). Ayrıca denetçinin işletmenin sürekliliği hakkındaki karar sürecinin incelendiği çalışmalarda farklı mali oranlardan da süreçte yararlandığı görülmektedir. (Reynolds ve Francis 2000; DeFond, Raghunandan, ve Subramanyam 2002; Krishnan ve Wang 2015).

Literatürde, işletmenin sürekliliğinin tahminin en zor ve belirsizliğin en fazla olduğu denetim görevlerinden birisi olduğu değerlendirilmektedir (Carcello ve Neal 2000). Değerlendirme sadece denetçinin mesleki muhakemesini içermesi sebebi ile değil, aynı zamanda denetçi görüşünün işletmenin mevcudiyeti için bir felaket olma olasılığı sebebi ile oldukça zor bir süreçtir. Bununla birlikte son yıllarda ortaya çıkan denetim başarısızlıkları da işletmenin sürekliliğinin değerlendirilmesinin çok zor bir süreç olduğunu göstermektedir (Martens vd. 2008). Süreklilik tahmin modelleri, işletmenin süreklilik kabiliyeti hakkında öngörülebilir bulunmada, mali tablolardan daha etkili enstrümanlardır (E. Altman ve McGough 1974). Benzer olarak, pek çok araştırmacı, geliştirilen modellerin kullanılabilirliğini inceleyerek modellerin üstünlüklerini ortaya koymuş ve geliştirilen modeller işletmenin sürekliliğini değerlendirme sürecine katkı sağlayabilecek birer araç olarak değerlendirmiştir (Mutchler 1985; Levitan ve Knoblett 1985; Hian C. Koh ve Killough 1990).

İşletmenin sürekliliği varsayımı muhasebenin temel kavramlarından birisidir. Varsayım öngörülebilir bir gelecekte işletmenin faaliyetlerini sürdürebileceği kabulünü ifade etmektedir. İşletmenin süreklilik kabiliyetini hakkında önemli bir belirsizliğin ya da riski mevcut olması, işletmenin iç ya da dış çevresindeki tüm paydaşlarını etkilemektedir ve önemli ekonomik kayıplara yol açabilmektedir. Bu bağlamda süreklilik kabiliyeti hakkındaki değerlendirmelerin işletme paydaşlarının karar süreçlerindeki etkisinin çok büyük olduğu açıktır. İşletmenin mevcut durumu hakkında bilgi içeren mali tablolar ve mali tablolardan üretilen çeşitli mali oranlar işletmenin süreklilik kabiliyetinin değerlendirilmesinde faydalanabilecek olan nesnel araçlardır. 1960' lı yıllardan günümüze pek çok araştırmacı süreklilik tahmini ya da süreklilik riskinin tespiti maksadı

ile çeşitli modeller geliştirmiştir. 1990' lı yıllardan itibaren katı istatistiksel varsayımları sebebi ile geleneksel istatistiksel metotların yerini yapay zekâ ve makine öğrenmesi metotları almıştır. Karmaşık bir görev olan, mesleki muhakemeye ve deneyime dayalı işletmenin süreklilik kabiliyetinin değerlendirilmesi sürecinde faydalanılacak gelişmiş ve tahmin başarısı yüksek olan modellerin geliştirilmesi işletme paydaşlarının ve denetçinin önemli kayıplardan kaçınmasına ve sürecin etkinliğinin artmasına olanak tanıyacaktır.

#### **4.4. Araştırmanın Kısıtları**

Çalışmanın sahip olduğu kısıtlar şunlardır;

- Çalışmanın birinci kısıtı kapsadığı dönem ile ilgilidir. Çalışma Borsa İstanbul A.Ş' de 2017 - 2021 yılları arasında işlem görmekte olan işletmeleri kapsamaktadır. BDS-570 İşletmenin Sürekliliği Standardı 2017 yılı itibari ile yürütülecek olan denetim çalışmalarında uygulanmaya başlanmıştır. Çalışmada kullanılan veri seti 2022 yılı itibari ile derlenmiştir. Sayılan iki sebep çalışmanın kapsadığı dönemin tanımlanmasına temel oluşturmaktadır,
- Çalışmanın ikinci kısıtı veri setindeki işletmelerin faaliyet gösterdiği sektör ile ilgilidir. Çalışmada kullanılan veri setinde sadece Borsa İstanbul A.Ş imalat ve alt sektörlerinde işlem gören işletmelere yer verilmiştir. Çalışmada kullanılacak olan veri setinde yer alan mali oranlar arasında homojenliği sağlamak ve yürütülecek olan analizlerde daha doğru sonuçlar elde etmek maksadı ile veri seti imalat işletmeleri ile sınırlandırılmıştır,
- Çalışmada yer alan üçüncü kısıt işletmelerin faaliyet gösterdiği ülke ile ilgilidir. Çalışmada sadece Türkiye' de faaliyet gösteren işletmelere yer verilmiştir,
- Çalışmada yer alan dördüncü kısıt kullanılan öznitelikler ile ilgilidir. Denetçinin işletmenin sürekliliğine ilişkin değerlendirmesi birçok faktörün eş zamanlı incelenmesini gerektirmektedir. Ancak bu çalışmada süreçle ilgili sadece yapılandırılmış verilerden ve mali oranlardan faydalanılmıştır. Mali oranların literatürde süreklilik tahmini maksadı ile yoğun olarak kullanıldığı görülmektedir. Kullanılan mali oranları işletmelerin yıllık mali raporlarında elde edilmiştir.

#### **4.5. Araştırmanın Yaygın Etkisi**

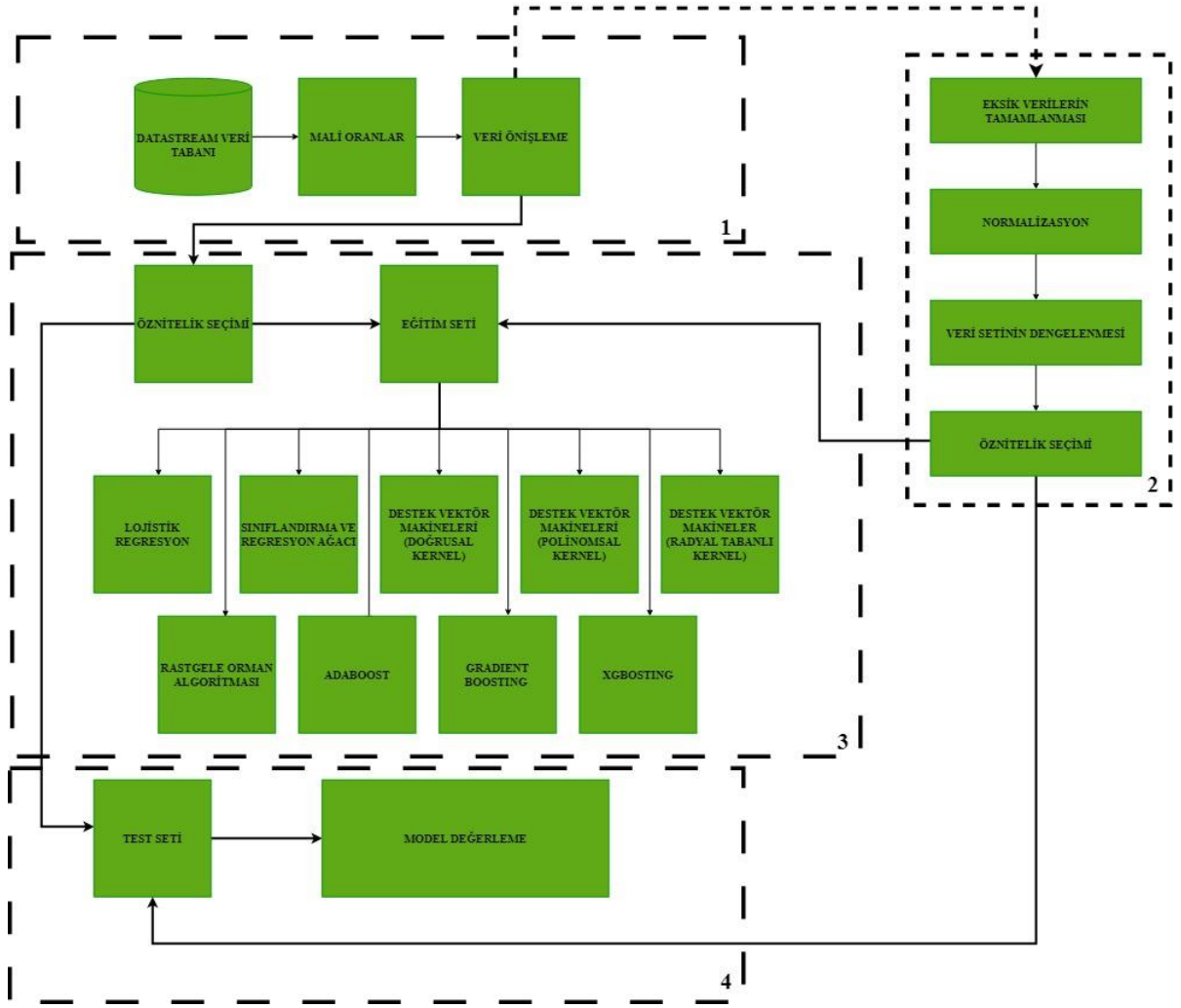
Makine öğrenme yöntemleri 1990' lı yıllardan günümüze işletmenin sürekliliğinin tahmini literatüründe yoğun bir şekilde kullanılmaktadır. Literatürde farklı ülkelerde farklı veri setleri ile yürütülmüş olan çalışmalar göstermektedir ki makine öğrenmesi

metotları işletmenin sürekliliğinin tahmininde başarılı sonuçlar vermektedir. Yürütülen çalışmada işletme yönetimi, denetçi ve işletme paydaşları tarafından işletmenin sürekliliğinin tahmininde literatürde yaygın olarak kullanılan makine öğrenmesi metotlarının etkinliği test edilmektedir. Bu bağlamda çalışma özellikle Türkiye’deki makine öğrenmesi ve denetim literatürüne katkı sağlamaktadır. Çalışmanın sayılan dışında sahip olduğu yaygın etkileri şunlardır;

- Geliştirilecek olan bir modelin bağımsız denetim programlarında, denetçinin karar sürecini destekleyecek şekilde kullanılması planlanmaktadır. Ayrıca önerilen modelin işletmelerde günümüzde yaygın olarak kullanılan Kurumsal Kaynak Planlama Yazılımlarında ve üst düzey karar destek yazılımlarında, üst düzey yöneticilerin karar süreçlerinde kullanılabileceği öngörülmektedir. Önerilen model ilişkili taraflar tarafından işletmenin süreklilik kabiliyeti hakkındaki karar sürecinde kullanılabilecektir.
- Çalışmanın kapsamının genişletilmesi yolu ile makine öğrenmesi metotları ve doğal dil işleme algoritmalarının kullanıldığı gerçek zamanlı denetim karar destek yazılım projesi için bir temel oluşturacağı düşünülmektedir. Denetim karar destek yazılımı, gerçek zamanlı olarak, XBRL taksonomileri kullanılan mali tabloların analizi yolu ile denetim faaliyetini sürekli kılacaktır. Böylece işletme yönetimi ile ilişkili tarafların bilgi ihtiyacı kesintisiz olarak sağlanacaktır.

#### **4.6. Araştırmanın Yöntemi**

İşletmenin sürekliliğinin tahmininde makine öğrenmesi metotlarının etkinliğinin ölçülmesi amaçlanan bu çalışmada standart bir makine öğrenmesi süreci uygulanmıştır. Süreç, şekil 4.1’de detaylı bir şekilde gösterilmektedir. Uygulanılan süreç veri elde etme (1), veri ön işleme (2), modelleme (3) ve değerlendirme (4) olmak üzere dört aşamadan oluşmaktadır. Sürecin birinci aşaması ham verinin Datastream veri tabanından elde edilmesi ile başlamaktadır. Ham verinin elde edilmesi sonrasında Ms Excel programında veri setinde yer alacak olan mali oranlar hesaplanmış ve R Studio ortamına aktarılmıştır. Veri ön işleme, modelleme ve değerlendirme aşamaları bütün olarak R Studio ortamında yürütülmüş olup uygulama eksizsiz bir şekilde ve detaylı olarak bölümde açıklanmaktadır.



Şekil 4.1. Çalışmada Kullanılan Makine Öğrenmesi Süreci

Verinin R studio ortamına aktarılmasını takiben ilk olarak çalışılacak verinin yapısını daha iyi tanımak maksadı ile veriye ait olan temel göstergeler incelenmiş olup, bu surette veri ön işleme aşamasında ihtiyaç olan işlemlerin tespiti amaçlanmıştır. Veri ön işleme aşamasında verinin modellemeye hazırlanması maksadı ile eksik verilerin tamamlanması, normalizasyon, veri setinin dengelenmesi ve öznetelik seçimi işlemleri yürütülmüştür. Bu aşamayı modelleme aşaması takip etmektedir. Modelleme aşamasında lojistik regresyon, sınıflama ve regresyon ağacı, destek vektör makineleri (doğrusal, polinomsal ve radyal tabanlı kernel fonksiyonları), rastgele orman algoritması, adaboost, gradient boosting, ve XGBoost algoritmaları eğitilmiş olup eğitilen modeller değerlendirme aşamasında literatürde yaygın olarak kullanılan kriterler kullanılarak değerlendirilmiş, sonuçları raporlanmış ve tartışılmıştır.

#### 4.7. Çalışmada Kullanılan Veri Seti

BDS-570 İşletmenin Sürekliliği Standardı, 01.01.2017 tarihinde ve sonrasında başlayacak hesap dönemlerinin denetiminde uygulamaktadır (Kamu Gözetim Kurumu 2020, 2). Çalışmada standardın uygulamaya geçtiği tarih, verinin tanımlanması için referans olarak kabul edilmiştir. Çalışmada kullanılan veri seti 03.08.2022 tarihinde derlenmiştir. Bu tarihte 2022 yılı mali tabloları halka açıklanmış değildir. Bu kısıtlar altında işletmenin sürekliliğinin tahmininin amaçlandığı bu çalışmada veri setinin 2017 ile 2021 yılları arasındaki dönemi kapsamına karar verilmiştir. Sonuç olarak veri seti 2017, 2018, 2019, 2020 ve 2021 yıllarında halka açıklanmış olan mali tablolardan elde edilen verileri içermektedir. Veri seti Borsa İstanbul A.Ş. pay piyasası imalat sanayi ve alt sektörlerinde işlem görmekte olan işletmeleri kapsamaktadır. Finans kuruluşlarının mali yapılarının farklı olması sebebi ile çalışmada kullanılacak olan veri seti sadece imalat işletmeleri ile sınırlı tutulmuştur. Borsa İstanbul A.Ş. pay piyasası imalat ve alt sektörlerinde Mayıs 2023 tarihi itibarı ile 209 işletme işlem görmektedir. Ancak veri setinin derlendiği tarihte 201 işletme yer almaktadır.

Gözlem yıllarının etiketlenmesi maksadı ile imalat sektöründe işlem gören işletmelerin Kamuyu Aydınlatma Platformu' nun web sitesinde halka açıklanan denetim raporlarının incelenmiştir. Buna göre bağımsız denetçinin, denetim raporunda işletmenin sürekliliğini devam ettirme kabiliyeti hakkında önemli bir belirsizliğe atıf yaptığı toplam 16 işletmeye ait toplam 47 gözlem yılı işletmenin sürekliliği görüşü olarak etiketlenmiştir. Bu işletmelere karşılık 198 işletmeye ait olan ve işletmenin sürekliliği hakkında herhangi bir olumsuzluğa atıfta bulunulmayan 958 gözlem yılı ise süreklilik kabiliyeti açısından sağlıklı olarak kabul edilmiştir. Çalışmada süreklilik görüşü almış olan gözlem yılları referans sınıf olarak kabul edilmiş ve buna göre ilgili sınıf "1" ile etiketlenmiştir. Süreklilik görüşü almamış olan gözlem yılları "0" olarak etiketlenmiştir. Çalışmada yer alan veri setinde bir işletmeye ait 2017-2021 yılları arasındaki dönemde birden fazla yıla yer verilmiş ve gözlem yılı olarak adlandırılmıştır. İşletmenin sürekliliğinin tahminini amaçlayan çalışmalarda, süreklilik görüşü almış olan işletmelerin sayısının azlığı sebebiyle veri setinin boyutunu artırmak maksadı ile bir işletmeye ait birden fazla yıla veri setinde yer verilmesi yaygın bir yaklaşımdır. Çalışmada da literatürde yer alan bu yaklaşım benimsenmiştir.(Alfaro vd. 2008).

**Tablo 4.1** Veri Setinde Yer Alan Gözlem Yıllarının Sayısı

	2017	2018	2019	2020	2021	Toplam
Sürekli Görüşü Almış İşletmeler	9	7	11	10	6	43
Sürekli Görüşü Almamış Olan İşletmeler	151	159	157	158	163	788
Toplam	160	166	168	168	169	831

Veri setinde yer alan mali oranların hesaplanması sonrasında, eksik verilerin tamamlanması maksadı herhangi bir işlem yapılmadan önce özniteliklerin büyük bir bölümü mevcut olmayan gözlem yılları derlemeden çıkarılmıştır. Bu suretle süreklilik görüşü almış olan 4 adet gözlem yılı ve süreklilik görüşü almamış olan 170 gözlem yılı veri seti dışında bırakılmıştır. Bu işlem dahilinde veri setinden çıkarılan gözlem yılları %100 ile %61 arasında eksik gözlem içermektedir. Tablo 4.1. veri ön işleme aşamasında kullanılan nihai veri setindeki gözlem yıllarının sayılarını içermektedir. Veri setinde, 2017 yılından toplam 160 adet gözlem yılı yer almaktadır. Bunların 6 adedi süreklilik görüşü almış olup, 151 adedi süreklilik görüşü almamıştır. Veri setinde, 2018 yılından toplam 166 adet gözlem yılı yer almaktadır. Bunların 7 adedi süreklilik görüşü almış olup, 159 adedi süreklilik görüşü almamıştır. Veri setinde, 2019 yılından toplam 168 adet gözlem yılı yer almaktadır. Bunların 11 adedi süreklilik görüşü almış olup, 157 adedi süreklilik görüşü almamıştır. Veri setinde, 2020 yılından toplam 168 adet gözlem yılı yer almaktadır. Bunların 10 adedi süreklilik görüşü almış olup, 158 adedi süreklilik görüşü almamıştır. Veri setinde, 2021 yılından toplam 169 adet gözlem yılı yer almaktadır. Bunların 6 adedi süreklilik görüşü almış olup, 163 adedi süreklilik görüşü almamıştır. Veri setinde belirtilen yıllara ait olarak süreklilik görüşü almış olan 43 adet gözlem yılı mevcut iken, süreklilik görüşü almamış olan toplam 788 adet gözlem yılı bulunmaktadır.

#### **4.8. Veri Setinde Yer Alan Mali Oranlar**

İşletmenin sürekliliğinin tahmininde kullanılacak olan özniteliklerin neler olacağı ve sayısının ne olacağı çok önemli bir sorudur (E. I. Altman vd. 2020). Bellovary, Giacomino, ve Akers (2007) süreklilik tahmini literatüründe 752 farklı öznitelik tespit etmiştir. Bu bağlamda özniteliklerin seçimi çok büyük bir önem arz etmektedir (du Jardin 2009). Literatürde, veri setinde yer alacak olan özniteliklerin nasıl seçileceği hakkında bir görüş birliği bulunmamaktadır. BDS- 570 işletmenin sürekliliği standardında işletmenin mali durumu ile ilgili, işletme faaliyetleri ile ilgili ve diğer durumlar ile ilgili olmak üzere

üç ana başlıkta atıf yapılan, işletmenin sürekliliğini devam ettirme kabiliyetine ilişkin ciddi şüphe oluşturabilecek olaylar veya şartlar hakkında referanslar verilmektedir (Kamu Gözetim Kurumu 2020, 13-15). Çalışmada ilgili referanslar göz önünde bulundurularak, detaylı ve kapsamlı literatür taraması yürütülmüş olup ilk aşamada 75 adet mali oran belirlenmiştir. Detaylı bir literatür taraması yapılmak sureti ile özniteliklerin seçimi literatürde sıklıkla kullanılmış ve anlamlı sonuçlar vermiş olan mali oranlar, öznitelik olarak belirlenmiştir (M.-Y. Chen 2011b). Veri setinde yer verilen öznitelikler tablo 4.2.'de yer almaktadır (Alfaro vd. 2008; Zięba, Tomczak, ve Tomczak 2016; Zhou, Lai, ve Yen 2014; 2012; Yeh, Chi, ve Lin 2014b; Alam, Gao, ve Jones 2021; Ahn ve Kim 2009; Z. Zhang vd. 2022; Xiao vd. 2012; Chandra, Ravi, ve Bose 2009; Cheng, Chan, ve Sheu 2019; Yang, You, ve Ji 2011; Murugan Anandarajan, Lee, ve Anandarajan 2001; W.-S. Chen ve Du 2009; Zhao vd. 2022; L. Wang ve Wu 2017; L. Zhang vd. 2013; Zeng vd. 2020; Xu ve Wang 2009b; Yeh, Chi, ve Hsu 2010; Zhou, Lu, ve Fujita 2015; Desheng Wu, Ma, ve Olson 2022; Barboza, Kimura, ve Altman 2017; E. I. Altman vd. 2020; Zhou, Si, ve Fujita 2017; G. Wang vd. 2020; Desheng(Dash) Wu, Liang, ve Yang 2008b; Bunyaminu ve Issah 2012; Y.-P. Huang ve Yen 2019; Olson, Delen, ve Meng 2012b; Zhou, Lai, ve Yen 2014; Li, Andina, ve Sun 2013; Mselmi, Lahiani, ve Hamza 2017; Shin, Lee, ve Kim 2005; J. Sun vd. 2017; G. Wang, Chen, ve Chu 2018; X. Du vd. 2020; Liang vd. 2020b; Pham Vo Ninh, Do Thanh, ve Vo Hong 2018). Tablo 4.2.'de yer alan öznitelikler genetik algoritmalar ile öznitelik seçimi süreci öncesinin temsil etmektedir. Öznitelik seçimi sonrasında, veri setinden çıkarılan öznitelikler ve modelleme aşamasına aktarılan öznitelikler ilgili bölümde paylaşılmıştır.

**Tablo 4.2.** Veri setinde Yer Alan Öznitelikler

Kodu	Özniteliğin Adı	Hesaplamada Kullanılan Formül
O1	İşletme Büyüklüğü	Log("WC02999" - Total Assets)
O2	ROA – Varlık Karlılığı	"WC08326" - Return On Assets
O3	ROE - Özsermaye Karlılığı	"WC08301" - Return On Equity Total %
O4	Nakit / Toplam Varlıklar	("WC02003" - Cash) / ("WC02999" - Total Assets)
O5	Nakit / Toplam Borçlar	("WC02003" - Cash) / ("WC03255" - Total Debt)
O6	Nakit ve Nakit Benzerleri / Toplam Dönen Varlıklar	("WC02005" - Cash & Equivalents Generic) / ("WC02201" - Current Assets Total)
O7	Nakit Oran	("WC02005" - Cash & Equivalents Generic) / ("WC03101" - Current Liabilities Total)
O8	Likidite Oranı	"WC08101" - Quick Ratio
O9	Cari Oran	"WC08106A" - Current Ratio
O10	Çalışma Sermayesi / Toplam Satışlar	("WC03151" - Working Capital) / ("WC01001" - Net Sales Or Revenues)
O11	Çalışma Sermayesi / Toplam Varlıklar	("WC03151" - Working Capital) / ("WC02999" - Total Assets)
O12	Toplam Alacaklar / Toplam Yükümlülükler	("WC02051" - Receivables (Net)) / ("WC03351" - Total Liabilities)
O13	Toplam Dönen Varlıklar / Toplam Varlıklar	("WC02201" - Current Assets Total) / ("WC02999" - Total Assets)

**Tablo 4.2. (Devam) Veri setinde Yer Alan Öznitelikler**

O14	Toplam Duran Varlıklar / Toplam Varlıklar	$(\text{"WC02999"} - \text{Total Assets} - \text{"WC02201"} - \text{Current Assets Total}) / (\text{"WC02999"} - \text{Total Assets})$
		$\text{Fixed Assets} = (\text{"WC02999"} - \text{Total Assets} - \text{"WC02201"} - \text{Current Assets Total})$
O15	Toplam Varlıklar / Toplam Yükümlülükler	$(\text{"WC02999"} - \text{Total Assets}) / (\text{"WC03351"} - \text{Total Liabilities})$
O16	Toplam Varlıklar / Toplam Öz Kaynaklar	$(\text{"WC02999"} - \text{Total Assets}) / (\text{"WC03501"} - \text{Common Equity})$
O17	Kısa Vadeli Borçlar & Uzun Vadeli Borçların Bir Yıl İçinde Kısa Vadeli Olacak Kısımları / Toplam Yükümlülükler	$(\text{"WC03051"} - \text{Short Term Debt} \& \text{Current Portion Of Long Term Debt}) / (\text{"WC03351"} - \text{Total Liabilities})$
O18	Uzun Vadeli Borçlar / Toplam Özkaynak	$(\text{"WC03251"} - \text{Long Term Debt}) / (\text{"WC03501"} - \text{Common Equity})$
O19	Toplam Borçlar / Toplam Varlıklar	$(\text{"WC03255"} - \text{Total Debt}) / (\text{"WC02999"} - \text{Total Assets})$
O20	Toplam Borçlar / Toplam Özkaynak	$(\text{"WC03255"} - \text{Total Debt}) / (\text{"WC03501"} - \text{Common Equity})$
O21	Toplam Kısa Vadeli Yükümlülükler / Toplam Varlıklar	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O22	Toplam Kısa Vadeli Yükümlülükler / Toplam Yükümlülükler	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O23	Toplam Kısa Vadeli Yükümlülükler / Toplam Özkaynak	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03501"} - \text{Common Equity})$
O24	Toplam Kısa Vadeli Yükümlülükler / Toplam Satışlar	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O25	Toplam Yükümlülükler / Toplam Varlıklar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC02999"} - \text{Total Assets})$
O26	Toplam Yükümlülükler / Ortalama Toplam Borçlar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC03255"} - \text{Total Debt})$
O27	Toplam Yükümlülükler / Toplam Özkaynaklar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC03501"} - \text{Common Equity})$
O28	Dağıtılmamış Karlar / Toplam Varlıklar	$(\text{"WC03495"} - \text{Retained Earnings}) / (\text{"WC02999"} - \text{Total Assets})$
O29	Toplam Özkaynak / Toplam Borçlar	$(\text{"WC03501"} - \text{Common Equity}) / (\text{"WC03255"} - \text{Total Debt})$
O30	Alacakların Devir Hızı (Toplam Satışlar / Ortalama Alacaklar)	$(\text{"WC01001"} - \text{Net Sales Or Revenues}) / ((\text{"WC02051"} - \text{Receivables (Net)}_{(\text{Dönem Başı})} + \text{"WC02051"} - \text{Receivables (Net)}_{(\text{Dönem Sonu})}) / 2)$
O31	Stokların Devir Hızı	$\text{WC08136} - \text{Inventory Turnover}$
O32	Çalışma Sermayesi Devir Hızı (Toplam Satışlar / Ortalama Çalışma Sermayesi)	$(\text{"WC01001"} - \text{Net Sales Or Revenues}) / ((\text{"WC03151"} - \text{Working Capital}_{(\text{Dönem Başı})} + \text{"WC03151"} - \text{Working Capital}_{(\text{Dönem Sonu})}) / 2)$
O33	Dönen Varlıkları Devir Hızı	$(\text{"WC01001"} - \text{Net Sales Or Revenues}) / ((\text{"WC02201"} - \text{Current Assets Total}_{(\text{Dönem Başı})} + \text{"WC02201"} - \text{Current Assets Total}_{(\text{Dönem Sonu})}) / 2)$
O34	Varlıkların Devir Hızı	$\text{"WC08401"} - \text{Total Asset Turnover}$
O35	Borçların Devir Hızı (Satışların Maliyeti / Ortalama Ticari Borçlar)	$(\text{"WC01051"} - \text{Cost Of Goods Sold (Excl Depreciation)}) / ((\text{"WC03040"} - \text{Accounts Payable}_{(\text{Dönem Başı})} + \text{"WC03040"} - \text{Accounts Payable}_{(\text{Dönem Sonu})}) / 2)$
O36	Brüt Kar / Kısa Vadeli Yükümlülükler	$(\text{"WC01100"} - \text{Gross Income}) / (\text{"WC03101"} - \text{Current Liabilities Total})$
O37	Faaliyet Harcamaları / Toplam Yükümlülükler	$(\text{"WC01249"} - \text{Operating Expenses Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O38	VFÖK Marjı	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O39	Vergi ve Faiz Öncesi Kar / Toplam Varlıklar	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC02999"} - \text{Total Assets})$
O40	Faiz Karşılama Oranı	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC01251"} - \text{Interest Expense On Debt})$
O41	Net Kar Marjı	$\text{"WC08366"} - \text{Net Margin}$
O42	Net Kar / Ortalama Dönen Varlıklar	$(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}) / ((\text{"WC02201"} - \text{Current Assets Total}_{(\text{Dönem Başı})} + \text{"WC02201"} - \text{Current Assets Total}_{(\text{Dönem Sonu})}) / 2)$
O43	Net Kar / Ortalama Duran Varlıklar	$(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}) / ((\text{Fixed Assets}_{(\text{Dönem Başı})} + \text{Fixed Assets}_{(\text{Dönem Sonu})}) / 2)$
		$\text{Fixed Assets} = (\text{WC02999} - \text{Total Assets} - \text{WC02201} - \text{Current Assets Total})$
O44	Net Kar / Toplam Borçlar	$(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}) / (\text{"WC03255"} - \text{Total Debt})$
O45	Hisse Başına Kar	$\text{"WC05201"} - \text{Earnings Per Share}$
O46	Hisse Başına Faaliyet Karı	$\text{"WC05509"} - \text{Operating Profit Per Share}$
O47	Hisse Başına Net Varlık	$((\text{"WC02999"} - \text{Total Assets}) - (\text{"WC03351"} - \text{Total Liabilities})) / (\text{"WC05191"} - \text{Common Shares Used To Calculate EPS})$
O48	Hisse Başına Nakit Akışı	$\text{"WC05502"} - \text{Cash Flow Per Share Fiscal}$
O49	Hisse Başına Net Faaliyet Nakit Akışı	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC05191"} - \text{Common Shares Used To Calculate EPS})$
O50	Nakit Akışı / Toplam Varlıklar	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC02999"} - \text{Total Assets})$
O51	Nakit Akışı / Toplam Borçları	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC03255"} - \text{Total Debt})$
O52	Nakit Akışı / Toplam Kısa Vadeli Yükümlülükler	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC03101"} - \text{Current Liabilities Total})$

**Tablo 4.2. (Devam) Veri setinde Yer Alan Öznitelikler**

O53	Nakit Akışı / Toplam Yükümlülükler	(“WC04201” - Funds From Operations) / (“WC03351” - Total Liabilities)
O54	Nakit Akışı / Toplam Özkaynak	(“WC04201” - Funds From Operations) / (“WC03501” - Common Equity)
O55	Nakit Akışı / Toplam Satışlar	“WC08311” - Cash Flow/Sales
O56	Net Faaliyet Nakit Akışı / Toplam Varlıklar	(“WC04860” - Net Cash Flow Operating Activities) / (“WC02999” - Total Assets)
O57	Net Faaliyet Nakit Akışı / Toplam Borçlar	(“WC04860” - Net Cash Flow Operating Activities) / (“WC03255” - Total Debt)
O58	Net Faaliyet Nakit Akışı / Toplam Kısa Vadeli Yükümlülükler	(“WC04860” - Net Cash Flow Operating Activities) / (“WC03101” - Current Liabilities Total)
O59	Net Faaliyet Nakit Akışı / Toplam Yükümlülükler	(“WC04860” - Net Cash Flow Operating Activities) / (“WC03351” - Total Liabilities)
O60	Net Faaliyet Nakit Akışı / Faiz Harcamaları	(“WC04860” - Net Cash Flow Operating Activities) / (“WC01251” - Interest Expense On Debt)
O61	Net Faaliyet Nakit Akışı / Toplam Satışlar	(“WC04860” - Net Cash Flow Operating Activities) / (“WC01001” - Net Sales Or Revenues)
O62	Çalışma Sermayesindeki Yıllık Büyüme	(“WC03151” - Working Capital <sub>(t)</sub> ) - (“WC03151” - Working Capital <sub>(t-1)</sub> ) / (“WC03151” - Working Capital <sub>(t-1)</sub> )
O63	Varlıkların Büyüme Hızı	(“WC02999” - Total Assets <sub>(t)</sub> ) - (“WC02999” - Total Assets <sub>(t-1)</sub> ) / (“WC02999” - Total Assets <sub>(t-1)</sub> )
O64	Toplam Borçlardaki Yıllık Büyüme	(“WC03255” - Total Debt <sub>(t)</sub> ) - (“WC03255” - Total Debt <sub>(t-1)</sub> ) / (“WC03255” - Total Debt <sub>(t-1)</sub> )
O65	Dağıtılmamış Karlardaki Yıllık Büyüme	(“WC03495” - Retained Earnings <sub>(t)</sub> ) - (“WC03495” - Retained Earnings <sub>(t-1)</sub> ) / (“WC03495” - Retained Earnings <sub>(t-1)</sub> )
O66	Satışlardaki yıllık Büyüme	(“WC01001” - Net Sales Or Revenues <sub>(t)</sub> ) - (“WC01001” - Net Sales Or Revenues <sub>(t-1)</sub> ) / (“WC01001” - Net Sales Or Revenues <sub>(t-1)</sub> )
O67	Faaliyet Nakit Akışlarındaki Yıllık Büyüme	(“WC04860” - Net Cash Flow Operating Activities <sub>(t)</sub> ) - (“WC04860” - Net Cash Flow Operating Activities <sub>(t-1)</sub> ) / (“WC04860” - Net Cash Flow Operating Activities <sub>(t-1)</sub> )
O68	Sermaye Harcamalarındaki Yıllık Değişim	(“WC04601” - Capital Expenditures (Additions To Fixed Assets) <sub>(t)</sub> ) - (“WC04601” - Capital Expenditures (Additions To Fixed Assets) <sub>(t-1)</sub> ) / (“WC04601” - Capital Expenditures (Additions To Fixed Assets) <sub>(t-1)</sub> )
O69	Sermayedeki Yıllık Değişim	(“WC03501” - Common Equity <sub>(t)</sub> ) - (“WC03501” - Common Equity <sub>(t-1)</sub> ) / (“WC03501” - Common Equity <sub>(t-1)</sub> )
O70	Net Kardaki Yıllık Değişim	(“WC01651” - Net Income Before Preferred Dividends <sub>(t)</sub> ) - (“WC01651” - Net Income Before Preferred Dividends <sub>(t-1)</sub> ) / (“WC01651” - Net Income Before Preferred Dividends <sub>(t-1)</sub> )
O71	Özkaynakların Piyasa Değeri	“WC08001” - Market Capitalization
O72	Özkaynakların Piyasa Değeri / Toplam Borçlar	(“WC08001” - Market Capitalization) / (“WC03255” - Total Debt)
O73	Sermaye Harcamaları / Toplam Varlıklar	(“WC04601” - Capital Expenditures (Additions To Fixed Assets)) / (“WC02999” - Total Assets)
O74	Fiyat / Defter Oranı	“PTBV” - Price To Book Value
O75	Z Skor	(0,012 * (“WC03151” - Working Capital / “WC02999” - Total Assets)) + (0,014 * (“WC03495” - Retained Earnings / “WC02999” - Total Assets)) + (0,033 * (“WC18191” - Earnings Before Interest And Taxes (EBIT) / “WC02999” - Total Assets)) + (0,006 * (“WC08001” - Market Capitalization / “WC03255” Total Debt)) + (0,999 * (“WC01001” - Net Sales Or Revenues / “WC02999” - Total Assets))

Çalışmada kullanılan tüm veri Datastream veri tabanından elde edilmiştir. Tablo 4.2.’ de özniteliğin adı ve hesaplamada kullanılan formül yer almaktadır. Datastream veri tabanında her bir mali oran ya da mali nitelikteki veri özgün bir kod içermektedir. Örneğin işletmelerin toplam varlıkları “WC02999” kodu ve “Total Assets” adı ile veri tabanında yer almaktadır. Tablo 4.2.’ de özniteliklerin hesaplanmasında kullanılan formüller, Datastream kodu ve adı ile birlikte sunulmaktadır.

#### 4.9. R Studio Sürümü ve Kullanılan Kütüphaneler

Çalışmada, Datastream veri tabanından elde edilen veriler kullanılarak, mali oranların ya da öz niteliklerin hesaplaması Microsoft Excel programı kullanılarak yapılmıştır. Yapılan hesaplama sonrasında yürütülen tüm işlemler R programlama dili kullanılarak ve R Studio ortamında gerçekleştirilmiştir. R programlama dili yayımlandığı 2000 yılından günümüze pek çok araştırmacının ve veri bilimcisinin katkısı ile farklı amaçlarla hazırlanmış toplam 19834 paketi kullanıcılara sunmaktadır (“R Mevcut Paketler” t.y.). Çalışmada kullanılan R programlama dilinin sürümü, R Studio’ nun sürümü ve faydalanılan paketlere ilişkin atıflar aşağıdaki gibidir;

- R programlama dili sürümü, “4.1.3”,
- R Studio sürümü, “RStudio 2023.03.0+386 "Cherry Blossom" Release (3c53477afb13ab959aeb5b34df1f10c237b256c3, 2023-03-09) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) RStudio/2023.03.0+386 Chrome/108.0.5359.179 Electron/22.0.3 Safari/537.36”,
- “adabag” (Alfaro, Gámez, ve García 2013),
- “Amelia” (Honaker, King, ve Blackwell 2011),
- “caret” (Kuhn 2008),
- “rattle” (G. Williams 2011),
- “naniar” (Tierney ve Cook 2023),
- “gbm” (Greenwell vd. 2022),
- “kernlab” (Karatzoglou vd. 2004),
- “mice” (Buuren ve Groothuis-Oudshoorn 2011),
- “ranger” (Wright ve Ziegler 2017),
- “readxl” (Wickham ve Bryan 2023),
- “rpart” (Therneau ve Atkinson 2022),
- “unbalanced” (Dal Pozzolo, Caelen, ve Bontempi 2015),
- “Tidyverse” (Wickham vd. 2019),
- “xgboost” (T. Chen vd. 2022),
- “e1071” (D. Meyer vd. 2023),
- “dplyr” (Wickham vd. 2023),
- “plyr” (Wickham 2011),
- “pROC” (Robin vd. 2011).

#### **4.10. Veri Ön İşleme Süreci**

Makine öğrenmesi sürecinde yüksek genelleme performansı sahip modellerin eğitiminde, faydalanılan veri setinin kalitesi çok önemli bir rol oynamaktadır (Hagendorff 2021). Makine öğrenmesi uygulamalarında, verinin modellemeye hazır olarak elde edilmesi mümkün değildir. Verinin modellemeye hazır bir forma dönüştürmesi maksadı ile farklı işlemlerin yürütülmesi gerekmektedir. Temel olarak ham veri, ön işleme aşamasında eksik verilerin tamamlanması, öznitelik seçimi gibi işlemlerden geçirilmektedir. Uygulanacak olan işlemlere bazı durumlarda verinin yapısı gereği ihtiyaç duyulurken, algoritma tercihi de ham veri üzerinde çeşitli işlemler yürütülmesini gerekli kılabilir. Çalışmada bu bağlamda veri ön işleme aşamasında eksik verilerin tamamlanması, normalizasyon, veri setinin dengelenmesi ve öznitelik seçimi olmak üzere dört işlem yürütülmüştür. Veri ön işleme aşamasında yer alan tüm faaliyetler R Studio ortamında yürütülmüştür. Veri Setinde yer alan eksik veriler ilgili özneliğin ortalamasından faydalanılarak tamamlanmıştır. Veri setinin normalizasyonu maksadı ile Min-Max normalizasyon metodundan yararlanılmıştır. Veri setinin dengelenmesi maksadı ile SMOTE metodundan faydalanılmıştır. Öznitelik seçimi maksadı ile genetik algoritmalar kullanılmıştır.

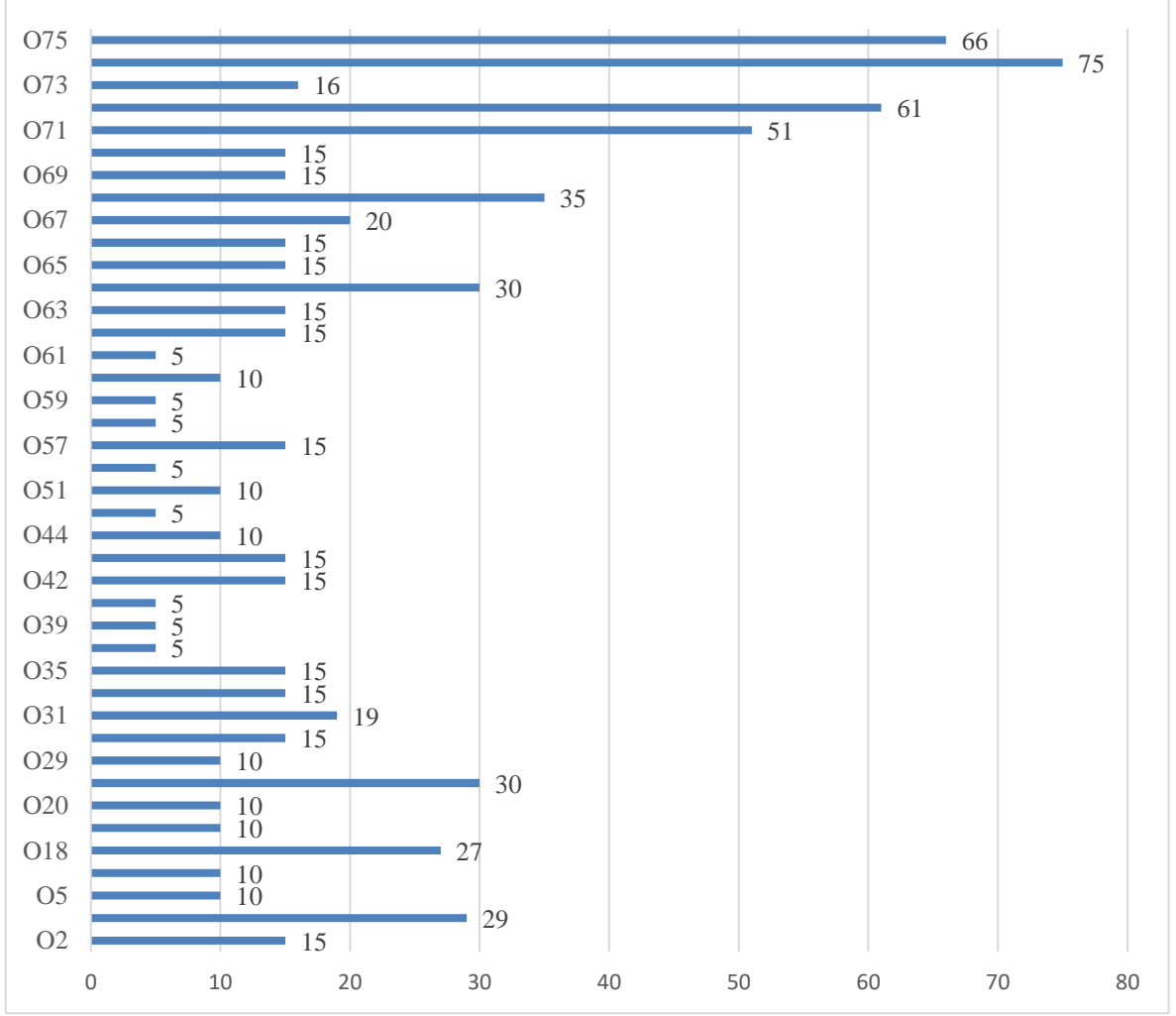
##### **4.10.1. Eksik Verilerin Tamamlanması**

Uygulamada veri setlerinde eksik değerlerin bulunması sıklıkla karşılaşılan bir durumdur. Bu nedenle, eksik verilerin model üzerindeki olumsuz etkilerini ortadan kaldırmak maksadı ile model eğitimi öncesinde eksik verilerin tamamlanması gerekmektedir (W. Zhang, He, ve Zhang 2019; Zhou ve Lai 2017). Literatürde gradient boosting gibi eksik veriye karşı duyarlılığı az olan metotlar mevcut olmasına karşın eksik veri modelin genelleme performansını olumsuz etkileyen bir unsurdur (Alam, Gao, ve Jones 2021). Çalışmada kullanılan veri setinin R Studio ortamına aktarımı sonrasında ilk olarak verinin genel yapısı incelenmiştir. Takip eden adımda her bir özneliğin içerdiği eksik veri noktaları tespit edilmiştir. Ham veri seti ilk olarak incelendiğinde toplam 62.325 veri noktası içerinden 774 gözlemin eksek veri içerdiği tespit edilmiştir.

**Tablo 4.3.** Veri Setinde Yer Alan Özniteliklerin İçerdiği Eksik Gözlem Sayıları

KODU	EKSİK	KODU	EKSİK	KODU	EKSİK	KODU	EKSİK	KODU	EKSİK
O1	0	O16	0	O31	19	O46	0	O61	5
O2	15	O17	10	O32	15	O47	0	O62	15
O3	29	O18	27	O33	0	O48	0	O63	15
O4	0	O19	10	O34	0	O49	5	O64	30
O5	10	O20	10	O35	15	O50	0	O65	15
O6	0	O21	0	O36	0	O51	10	O66	15
O7	0	O22	0	O37	0	O52	0	O67	20
O8	0	O23	0	O38	5	O53	0	O68	35
O9	0	O24	0	O39	5	O54	0	O69	15
O10	0	O25	0	O40	5	O55	0	O70	15
O11	0	O26	30	O41	0	O56	5	O71	51
O12	0	O27	0	O42	15	O57	15	O72	61
O13	0	O28	0	O43	15	O58	5	O73	16
O14	0	O29	10	O44	10	O59	5	O74	75
O15	0	O30	15	O45	0	O60	10	O75	66

Tablo 4.3.' de özniteliklerin içerdiği eksik veri noktalarının toplam sayısı gösterilmektedir. Eksik verilerin ortaya çıkaracağı problemin önüne geçilmesi maksadı ile pek çok ampirik çalışmada kayıp veri noktalarının silinmesi tercih edilmektedir. Ancak bu yaklaşım gözlem kaybına bir başka ifade ile bilgi kaybına neden olmaktadır (Alam, Gao, ve Jones 2021). Gözlem kaybının önüne geçmek maksadı ile eksik veri içeren veri noktaları özniteliklerin ortalamaları ile tamamlanmıştır (Z. Zhang vd. 2022; W. Zhang, He, ve Zhang 2019).



**Şekil 4.2.** Veri Setinde Yer Alan Eksik Gözlem İçeren Öznitelikler

Şekil 4.2. veri setinde eksik gözlem içeren öznitelikleri ve içerdikleri eksik gözlem adetlerini göstermektedir.

#### 4.10.2. Normalizasyon

Normalizasyon makine öğrenmesinde doğru sonuçların elde edilmesi için çok önemli bir rol oynayan veri ön işleme sürecidir. İşlem verinin sahip olduğu orijinal dağılımı bozmadan, verileri tipik olarak 0 ile 1 arasına ölçeklendirmeyi amaçlamaktadır. Süreç veri setinde var olan eşitsizliklerin önüne geçmeyi hedeflemektedir. Veri setinde yer alan özniteliklerin büyüklükleri karşılaştırmalı olarak büyük farklılıklar içeriyor olabilir. Bu durum modellerin tahmin başarısını olumsuz etkileyebilmektedir (Z. Zhang vd. 2022) Normalizasyon işlemi verilerin birbirine yakınsama oranını artırmaktadır (Hastie, Tibshirani, ve Friedman 2009). Normalizasyon işlemi gözlemler arasındaki eşitsizlikleri azaltma, model performansını geliştirme, mesafeye dayalı algoritmaların eğitimini

kolaylaştırma, model katsayılarının yorumlanabilirliğini artırma, farklı birimlere sahip özniteliklerin karşılaştırılabilirliğini artırma, hesaplama karmaşıklığını azaltma, veri görselleştirmeyi kolaylaştırma ve aykırı değerleri etkisini azaltma noktasında fayda sağlamaktadır. Çalışmada, Min-Max normalizasyon metodu ile normalleştirilme işlemi gerçekleştirilmiştir. Metot veriyi 0 ile 1 aralığında olacak şekilde ölçeklendirmektedir. Normalizasyon işlemi ölçeklenmiş veriler ile ilk veri seti arasındaki ilişkiyi bozmamaktadır (Sree ve Bindu 2018; W. Zhang, He, ve Zhang 2019).

#### **4.10.3. Veri Setinin Dengelenmesi**

Süreklilik tahmini literatüründe dengesiz veri seti sorunu yaygın olarak karşılaşılan bir sorundur (Alam, Gao, ve Jones 2021; Bae 2012). İşletmenin sürekliliği tahmini çalışmalarında, süreklilik görüşü almış olan işletmelerin sayısının, süreklilik görüşü almamış olan işletmelere göre az olması, verimlilik problemi ortaya çıkarmaktadır (Martin 2000). Sınıflandırma algoritmaları, veriye dayanan tahminlerinin genel doğrululuğunu optimize etmeyi amaçlamaktadır. Bu nedenle sınıfların dağılımlarını dikkate almamaktadırlar. Dengesiz veri setleri, modellerin azınlık sınıfı göz ardı etmesine ve dolayısıyla yetersiz tahmin modellerinin geliştirilmesine yol açmaktadır (G. Wang vd. 2020; Veganzones ve Séverin 2018).

Çalışmada veri setinin dengelenmesi maksadı ile SMOTE metodundan faydalanılmıştır. SMOTE (Synthetic Minority Over-SAMpling Technique), dengesiz veri setlerinin üstesinden gelmek maksadı ile yaygın olarak kullanılan bir oversampling metodudur (J. Sun vd. 2020; Fernandez vd. 2018). Metot, Chawla vd. (2002) tarafından geliştirilmiştir. Metot bir arada bulunan azınlık sınıf örneğini enterpolasyon yapmak suretiyle yapay örnekler üretmektedir. Başka bir ifade ile metot gözlemleri aynı sınıftaki en yakın komşularını dikkate alarak, veri setini dengelemektedir (Veganzones ve Séverin 2018). Dengeleme sonrasında veri setinde süreklilik riski taşıyan 817 ve süreklilik riski taşımayan 774 gözlem yer almaktadır.

#### **4.10.4 Öznitelik Seçimi**

Bir veri ön işleme aşaması olan öznitelik seçimi, makine öğrenmesi sürecinin en önemli adımlarından birisidir. Öznitelik seçimi, gereksiz ya da ilgisiz öznitelikleri veri setinden çıkarmayı amaçlamaktadır. Çalışmada öznitelik seçimi maksadı ile genetik algoritmalarından faydalanılmıştır. Genetik algoritmalar literatürde yaygın olarak

kullanılmaktadır. (Verikas vd. 2010; Ahn ve Kim 2009; Liang, Tsai, ve Wu 2015; K. Kim, Lee, ve Ahn 2018).

Öz nitelik seçim süreci neticesinde Nakit / Toplam Borçlar, Nakit Oran, Cari Oran, Kısa Vadeli Borçlar & Uzun Vadeli Borçların Bir Yıl İçinde Kısa Vadeli Olacak Kısımları / Toplam Yükümlülükler, Alacakların Devir Hızı, Dönen Varlıkları Devir Hızı, Varlıkların Devir Hızı, Brüt Kar / Kısa Vadeli Yükümlülükler, Net Kar Marjı, Net Kar / Ortalama Dönen Varlıklar, Nakit Akışı / Toplam Varlıklar, Nakit Akışı / Toplam Yükümlülükler, Varlıkların Büyüme Hızı, Faaliyet Nakit Akışlarındaki Yıllık Büyüme, Sermayedeki Yıllık Değişim, Özkaynakların Piyasa Değeri, Sermaye Harcamaları / Toplam Varlıklar, Fiyat / Defter Oranı veri seti dışarısında bırakılmıştır.

**Tablo 4.4. Öz nitelik Seçim İşlemi Sonrası Veri Setinde Yer Alan Öz nitelikler**

Kodu	Öz niteliğin Adı	Hesaplama Kullanan Formül
O1	İşletme Büyüklüğü	$\text{Log}(\text{"WC02999"} - \text{Total Assets})$
O2	ROA – Varlık Karlılığı	$\text{"WC08326"} - \text{Return On Assets}$
O3	ROE - Özsermaye Karlılığı	$\text{"WC08301"} - \text{Return On Equity Total \%}$
O4	Nakit / Toplam Varlıklar	$(\text{"WC02003"} - \text{Cash}) / (\text{"WC02999"} - \text{Total Assets})$
O6	Nakit ve Nakit Benzerleri / Toplam Dönen Varlıklar	$(\text{"WC02005"} - \text{Cash \& Equivalents Generic}) / (\text{"WC02201"} - \text{Current Assets Total})$
O8	Likidite Oranı	$\text{"WC08101"} - \text{Quick Ratio}$
O10	Çalışma Sermayesi / Toplam Satışlar	$(\text{"WC03151"} - \text{Working Capital}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O11	Çalışma Sermayesi / Toplam Varlıklar	$(\text{"WC03151"} - \text{Working Capital}) / (\text{"WC02999"} - \text{Total Assets})$
O12	Toplam Alacaklar / Toplam Yükümlülükler	$(\text{"WC02051"} - \text{Receivables (Net)}) / (\text{"WC03351"} - \text{Total Liabilities})$
O13	Toplam Dönen Varlıklar / Toplam Varlıklar	$(\text{"WC02201"} - \text{Current Assets Total}) / (\text{"WC02999"} - \text{Total Assets})$
O14	Toplam Duran Varlıklar / Toplam Varlıklar	$(\text{"WC02999"} - \text{Total Assets} - \text{"WC02201"} - \text{Current Assets Total}) / (\text{"WC02999"} - \text{Total Assets})$
		$\text{Fixed Assets} = (\text{"WC02999"} - \text{Total Assets} - \text{"WC02201"} - \text{Current Assets Total})$
O15	Toplam Varlıklar / Toplam Yükümlülükler	$(\text{"WC02999"} - \text{Total Assets}) / (\text{"WC03351"} - \text{Total Liabilities})$
O16	Toplam Varlıklar / Toplam Öz Kaynaklar	$(\text{"WC02999"} - \text{Total Assets}) / (\text{"WC03501"} - \text{Common Equity})$
O18	Uzun Vadeli Borçlar / Toplam Özkaynak	$(\text{"WC03251"} - \text{Long Term Debt}) / (\text{"WC03501"} - \text{Common Equity})$
O19	Toplam Borçlar / Toplam Varlıklar	$(\text{"WC02201"} - \text{Total Debt}) / (\text{"WC02999"} - \text{Total Assets})$
O20	Toplam Borçlar / Toplam Özkaynak	$(\text{"WC03255"} - \text{Total Debt}) / (\text{"WC03501"} - \text{Common Equity})$
O21	Toplam Kısa Vadeli Yükümlülükler / Toplam Varlıklar	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O22	Toplam Kısa Vadeli Yükümlülükler / Toplam Yükümlülükler	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O23	Toplam Kısa Vadeli Yükümlülükler / Toplam Özkaynak	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC03501"} - \text{Common Equity})$
O24	Toplam Kısa Vadeli Yükümlülükler / Toplam Satışlar	$(\text{"WC03101"} - \text{Current Liabilities Total}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O25	Toplam Yükümlülükler / Toplam Varlıklar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC02999"} - \text{Total Assets})$
O26	Toplam Yükümlülükler / Ortalama Toplam Borçlar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC03255"} - \text{Total Debt})$
O27	Toplam Yükümlülükler / Toplam Özkaynaklar	$(\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC03501"} - \text{Common Equity})$
O28	Dağıtılmamış Karlar / Toplam Varlıklar	$(\text{"WC03495"} - \text{Retained Earnings}) / (\text{"WC02999"} - \text{Total Assets})$
O29	Toplam Özkaynak / Toplam Borçlar	$(\text{"WC03501"} - \text{Common Equity}) / (\text{"WC03255"} - \text{Total Debt})$
O31	Stokların Devir Hızı	$\text{WC08136} - \text{Inventory Turnover}$
O32	Çalışma Sermayesi Devir Hızı (Toplam Satışlar / Ortalama Çalışma Sermayesi)	$(\text{"WC01001"} - \text{Net Sales Or Revenues}) / ((\text{"WC03151"} - \text{Working Capital}_{(\text{Dönem Başı})} + \text{"WC03151"} - \text{Working Capital}_{(\text{Dönem Sonu})}) / 2)$
O35	Borçların Devir Hızı (Satışların Maliyeti / Ortalama Ticari Borçlar	$(\text{"WC01051"} - \text{Cost Of Goods Sold (Excl Depreciation)}) / ((\text{"WC03040"} - \text{Accounts Payable}_{(\text{Dönem Başı})} + \text{"WC03040"} - \text{Accounts Payable}_{(\text{Dönem Sonu})}) / 2)$
O37	Faaliyet Harcamaları / Toplam Yükümlülükler	$(\text{"WC01249"} - \text{Operating Expenses Total}) / (\text{"WC03351"} - \text{Total Liabilities})$
O38	VFÖK Marjı	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O39	Vergi ve Faiz Öncesi Kar / Toplam Varlıklar	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC02999"} - \text{Total Assets})$

**Tablo 4.4. (Devam) Öznitelik Seçim İşlemi Sonrası Veri Setinde Yer Alan Öznitelikler**

O40	Faiz Karşılama Oranı	$(\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)}) / (\text{"WC01251"} - \text{Interest Expense On Debt})$
O43	Net Kar / Ortalama Duran Varlıklar	$(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}) / ((\text{Fixed Assets (Dönem Başı)} + \text{Fixed Assets (Dönem Sonu)}) / 2)$
O44	Net Kar / Toplam Borçlar	$\text{Fixed Assets} = (\text{WC02999} - \text{Total Assets} - \text{WC02201} - \text{Current Assets Total})$ $(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}) / (\text{"WC03255"} - \text{Total Debt})$
O45	Hisse Başına Kar	$\text{"WC05201"} - \text{Earnings Per Share}$
O46	Hisse Başına Faaliyet Karı	$\text{"WC05509"} - \text{Operating Profit Per Share}$
O47	Hisse Başına Net Varlık	$(\text{"WC02999"} - \text{Total Assets}) - (\text{"WC03351"} - \text{Total Liabilities}) / (\text{"WC05191"} - \text{Common Shares Used To Calculate EPS})$
O48	Hisse Başına Nakit Akışı	$\text{"WC05502"} - \text{Cash Flow Per Share Fiscal}$
O49	Hisse Başına Net Faaliyet Nakit Akışı	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC05191"} - \text{Common Shares Used To Calculate EPS})$
O51	Nakit Akışı / Toplam Borçları	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC03255"} - \text{Total Debt})$
O52	Nakit Akışı / Toplam Kısa Vadeli Yükümlülükler	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC03101"} - \text{Current Liabilities Total})$
O54	Nakit Akışı / Toplam Özkaynak	$(\text{"WC04201"} - \text{Funds From Operations}) / (\text{"WC03501"} - \text{Common Equity})$
O55	Nakit Akışı / Toplam Satışlar	$\text{"WC08311"} - \text{Cash Flow/Sales}$
O56	Net Faaliyet Nakit Akışı / Toplam Varlıklar	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC02999"} - \text{Total Assets})$
O57	Net Faaliyet Nakit Akışı / Toplam Borçlar	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC03255"} - \text{Total Debt})$
O58	Net Faaliyet Nakit Akışı / Toplam Kısa Vadeli Yükümlülükler	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC03101"} - \text{Current Liabilities Total})$
O59	Net Faaliyet Nakit Akışı / Toplam Yükümlülükler	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC03351"} - \text{Total Liabilities})$
O60	Net Faaliyet Nakit Akışı / Faiz Harcamaları	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC01251"} - \text{Interest Expense On Debt})$
O61	Net Faaliyet Nakit Akışı / Toplam Satışlar	$(\text{"WC04860"} - \text{Net Cash Flow Operating Activities}) / (\text{"WC01001"} - \text{Net Sales Or Revenues})$
O62	Çalışma Sermayesindeki Yıllık Büyüme	$(\text{"WC03151"} - \text{Working Capital}_{(t)}) - (\text{"WC03151"} - \text{Working Capital}_{(t-1)}) / (\text{"WC03151"} - \text{Working Capital}_{(t-1)})$
O64	Toplam Borçlardaki Yıllık Büyüme	$(\text{"WC03255"} - \text{Total Debt}_{(t)}) - (\text{"WC03255"} - \text{Total Debt}_{(t-1)}) / (\text{"WC03255"} - \text{Total Debt}_{(t-1)})$
O65	Dağıtılmamış Karlardaki Yıllık Büyüme	$(\text{"WC03495"} - \text{Retained Earnings}_{(t)}) - (\text{"WC03495"} - \text{Retained Earnings}_{(t-1)}) / (\text{"WC03495"} - \text{Retained Earnings}_{(t-1)})$
O66	Satışlardaki yıllık Büyüme	$(\text{"WC01001"} - \text{Net Sales Or Revenues}_{(t)}) - (\text{"WC01001"} - \text{Net Sales Or Revenues}_{(t-1)}) / (\text{"WC01001"} - \text{Net Sales Or Revenues}_{(t-1)})$
O68	Sermaye Harcamalarındaki Yıllık Değişim	$(\text{"WC04601"} - \text{Capital Expenditures (Additions To Fixed Assets)}_{(t)}) - (\text{"WC04601"} - \text{Capital Expenditures (Additions To Fixed Assets)}_{(t-1)}) / (\text{"WC04601"} - \text{Capital Expenditures (Additions To Fixed Assets)}_{(t-1)})$
O70	Net Kardaki Yıllık Değişim	$(\text{"WC01651"} - \text{Net Income Before Preferred Dividends}_{(t)}) - (\text{"WC01651"} - \text{Net Income Before Preferred Dividends}_{(t-1)}) / (\text{"WC01651"} - \text{Net Income Before Preferred Dividends}_{(t-1)})$
O72	Özkaynakların Piyasa Değeri / Toplam Borçlar	$(\text{"WC08001"} - \text{Market Capitalization}) / (\text{"WC03255"} - \text{Total Debt})$
O75	Z Skor	$(0,012 * (\text{"WC03151"} - \text{Working Capital} / \text{"WC02999"} - \text{Total Assets})) + (0,014 * (\text{"WC03495"} - \text{Retained Earnings} / \text{"WC02999"} - \text{Total Assets})) + (0,033 * (\text{"WC18191"} - \text{Earnings Before Interest And Taxes (EBIT)} / \text{"WC02999"} - \text{Total Assets})) + (0,006 * (\text{"WC08001"} - \text{Market Capitalization} / \text{"WC03255"} - \text{Total Debt})) + (0,999 * (\text{"WC01001"} - \text{Net Sales Or Revenues} / \text{"WC02999"} - \text{Total Assets}))$

Tablo 4.4.' de öznitelik seçim süreci sonrasında veri setinde yer alan mali oranlar yer almaktadır. Modelleme aşamasında kullanılan veri setinde toplam 57 adet öznitelik yer almıştır.

## 4.11. Araştırmanın Bulguları

### 4.11.1. Özet İstatistikler

Tablo 4.5.' de veri ön işleme süreci öncesi veri setine ait özet istatistikler yer almaktadır. Buna göre tabloda her bir öz niteliğe ait minimum, 1.çeyrek, medyan, ortalama, 3.çeyrek ve maksimum değerleri gösterilmektedir.

**Tablo 4.5.** Öz niteliklere ait Özet İstatistikler

	Minimum	1. Çeyrek	Medyan	Ortalama	3. Çeyrek	Maksimum
O1	3.356	5.218	5.712	5.744	6.195	8.101
O2	-64.190	3.498	8.375	9.823	14.790	137.590
O3	-1572.320	1.955	12.885	9.822	27.872	1235.580
O4	0.000000	0.006818	0.030080	0.069425	0.098177	0.573386
O5	0.000	0.026	0.122	24.014	0.430	16075.750
O6	0.00000	0.03178	0.10180	0.15499	0.23732	0.93659
O7	0.00000	0.03671	0.13784	0.37524	0.38150	18.36564
O8	0.030	0.560	0.840	1.172	1.295	22.130
O9	0.030	1.055	1.360	1.838	2.085	25.280
O10	-17.03541	0.02567	0.17065	0.19645	0.39356	33.12458
O11	-4.93436	0.02331	0.14911	0.14627	0.31423	0.82743
O12	0.007215	0.242073	0.451378	0.556205	0.712355	6.211675
O13	0.02222	0.42581	0.58447	0.56732	0.71307	0.99682
O14	0.00318	0.28693	0.41553	0.43268	0.57419	0.97778
O15	0.1764	1.3888	1.7234	2.4596	2.8076	47.3371
O16	-172.826	1.587	2.448	3.116	3.587	112.745
O17	0.0000	0.1357	0.2762	0.2824	0.4108	0.8659
O18	-44.06480	0.02234	0.20385	0.45246	0.56072	30.77772
O19	0.0000	0.1052	0.2783	0.2879	0.4174	2.8934
O20	-102.2142	0.1674	0.6739	1.1360	1.4049	74.6958
O21	0.07626	0.63149	0.76659	0.74947	0.88559	2.03128
O22	0.07626	0.63149	0.76659	0.74947	0.88559	2.03128
O23	-127.5247	0.4189	0.9694	1.5458	1.8389	111.5564
O24	0.0402	0.3095	0.4820	0.6385	0.6904	23.7492
O25	0.02113	0.35617	0.58023	0.56858	0.72007	5.66884
O26	0.000	1.612	2.139	10.779	3.508	1954.400
O27	-173.8262	0.5392	1.3872	2.0799	2.5402	111.7450
O28	-20.508370	-0.000998	0.147441	0.066281	0.305491	0.867798
O29	-67.4	0.6	1.3	488.5	4.1	325163.7
O30	0.2223	2.9733	4.1772	5.4470	5.8097	121.5943
O31	0.260	2.790	4.200	7.231	6.588	450.480
O32	-3231.040	1.300	3.315	2.238	8.333	1134.273
O33	0.01758	1.13573	1.52764	1.63386	1.89946	6.42559
O34	0.0100	0.5900	0.8000	0.8851	1.0700	4.1600

**Tablo 4.5. (Devam) Özniteliklere ait Özet İstatistikler**

<b>O35</b>	0.2092	3.0928	4.6717	6.7695	7.4771	120.4895
<b>O36</b>	-1.6269	0.2660	0.4409	0.5988	0.7392	12.6144
<b>O37</b>	0.06363	0.95587	1.40759	1.84364	2.22186	16.82756
<b>O38</b>	-3.30829	0.04288	0.10504	0.11051	0.17905	3.23090
<b>O39</b>	-0.75551	0.03429	0.08739	0.09559	0.14793	0.91592
<b>O40</b>	-149.280	1.076	3.163	47.400	10.772	7836.529
<b>O41</b>	-436.020	0.790	5.800	4.444	12.520	321.830
<b>O42</b>	-6.78826	0.01330	0.10050	0.08922	0.22185	1.75118
<b>O43</b>	-3.79177	0.01577	0.13490	0.25001	0.33111	7.18284
<b>O44</b>	-965.50	0.01	0.16	182.62	0.61	134442.33
<b>O45</b>	-9.9570	0.0245	0.2470	1.5689	0.9980	191.5950
<b>O46</b>	-8.844	0.005	0.318	1.495	1.154	113.037
<b>O47</b>	-3.923	1.101	2.257	7.609	6.613	355.046
<b>O48</b>	-11.0300	0.0785	0.4850	2.0585	1.5050	152.2700
<b>O49</b>	-17.1644	-0.1063	0.1605	1.3350	1.0078	100.1159
<b>O50</b>	-4.54582	0.02618	0.08081	0.08249	0.14658	0.84775
<b>O51</b>	-915.33	0.07	0.29	213.31	0.82	159467.00
<b>O52</b>	-13.59722	0.06464	0.21702	0.34358	0.43095	15.70046
<b>O53</b>	-3.57320	0.05242	0.15233	0.26961	0.31172	12.93803
<b>O54</b>	-10.3445	0.0600	0.2089	0.2681	0.3769	21.1218
<b>O55</b>	-474.360	3.460	9.870	8.872	17.235	492.440
<b>O56</b>	-4.65947	-0.02392	0.03838	0.03716	0.10927	0.81965
<b>O57</b>	-1357.83	-0.07	0.13	66.51	0.51	49523.33
<b>O58</b>	-5.77801	-0.06156	0.09397	0.20064	0.32072	14.21341
<b>O59</b>	-5.30059	-0.04258	0.07023	0.15641	0.22126	11.71261
<b>O60</b>	-1487.429	-0.708	1.474	18.876	6.856	6427.235
<b>O61</b>	-34.59136	-0.02931	0.04962	-0.01717	0.12814	0.97052
<b>O62</b>	-577.5424	-0.2939	0.1634	-0.2811	0.6982	163.5534
<b>O63</b>	-0.9724	0.1049	0.2483	0.3708	0.4660	13.0870
<b>O64</b>	-1.000	-0.018	0.188	15.987	0.512	9024.000
<b>O65</b>	-76.18182	-0.08067	0.17151	0.92675	0.54801	240.87097
<b>O66</b>	-0.9070	0.1174	0.2707	0.4163	0.4896	19.2226
<b>O67</b>	-285.5000	-1.2620	-0.2665	1.1397	0.9519	480.6000
<b>O68</b>	-1.0000	-0.2879	0.2648	1.7952	1.3149	69.6283
<b>O69</b>	-72.3013	0.0379	0.2058	0.4622	0.5078	108.8154
<b>O70</b>	-46.0484	-0.7080	0.1887	0.8037	1.2441	84.9924
<b>O71</b>	2369	146302	506700	2893384	1895463	98560000
<b>O72</b>	0.0	1.3	3.1	1294.6	11.1	629078.3
<b>O73</b>	0.00000	0.01512	0.03483	0.05329	0.06200	1.07955
<b>O74</b>	-90.130	1.087	1.730	3.262	2.942	342.700
<b>O75</b>	0.000	0.636	0.854	8.675	1.234	3775.808

#### 4.11.2. Eğitilen Modellerin Sonuçları

Makine öğrenmesi modellerinin performansının, problemin belirli özelliklerine ve veri setinin yapısına bağlı olması, bir yöntemin en iyi yöntem olarak öne çıkmasına engel

olmaktadır (Duéñez-Guzmán ve Vose 2013). Çalışmada bu nedenle birden fazla makine öğrenmesi algoritması eğitilmiş ve literatürde yaygın olarak kullanılan metriklere göre değerlendirilerek raporlanmıştır (M.-Y. Chen 2011b). Lojistik regresyon, destek vektör makineler (doğrusal, polinomsal ve radyal tabanlı kernel fonksiyonlar), sınıflandırma ve regresyon ağacı, rastgele orman algoritması, adaboost algoritması, gradient boosting algoritması ve extreme gradient boosting algoritmasının ampirik sonuçları sırası ile detaylı olarak paylaşılmış ve tartışılmıştır. Modelleri değerlemek maksadı ile literatürde yaygın olarak kullanılan ve karmaşıklık matrisi üzerinden hesaplanan doğruluk, duyarlılık, seçicilik, kesinlik metrikleri ile tip-I ve tip-II hataları ve AUC skorları kullanılmıştır (Alfaro vd. 2008; E. I. Altman vd. 2020; Murugan Anandarajan, Lee, ve Anandarajan 2001; Barboza, Kimura, ve Altman 2017; Bunyaminu ve Issah 2012; Chandra, Ravi, ve Bose 2009; W.-S. Chen ve Du 2009; Cheng, Chan, ve Sheu 2019; Cho, Hong, ve Ha 2010; G. Wang, Chen, ve Chu 2018; Cleofas-Sánchez vd. 2016; De Bock 2017; Fallahpour, Lakvan, ve Zadeh 2017a; Fedorova, Gilenko, ve Dovzhenko 2013; Yeh, Chi, ve Lin 2014a; Verikas vd. 2010; Zhou ve Lai 2017; Garcia 2022). Test veri seti ile her bir model ile yapılan tahminlere ait sonuçlara ilişkin karmaşıklık matrisleri ve hesaplanan doğruluk, duyarlılık, seçicilik, kesinlik, tip-I ve tip-II hataları ve AUC skorları paylaşılmıştır.

Veri ön işleme aşaması sonrasında elde edilen veri seti pek çok farklı araştırmada kullanılan 80/20 oranına uygun olarak rastgele ikiye ayrılarak eğitim ve test veri setleri oluşturulmuştur (Harris 2015; He, Zhang, ve Zhang 2018). Böylece veri setinin 0.8'i eğitim verisi ve 0.2' si test verisi olarak ayrılmıştır. Eğitilen modellerin tahmin başarısı güvence etmek maksadı ile literatürde yaygın olarak kullanılan ve etkinliği kanıtlanmış olan 10 katmanlı tek tekrarlı çapraz doğrulama prosedürü uygulanmıştır (Alfaro vd. 2008; Chandra, Ravi, ve Bose 2009; G. Wang, Chen, ve Chu 2018; Cleofas-Sánchez vd. 2016; De Bock 2017; Fallahpour, Lakvan, ve Zadeh 2017a).

#### **4.11.2.1. Lojistik Regresyon Modeli**

Lojistik regresyon, ikili yanıt değişkeni ile açıklayıcı değişkenler arasındaki ilişkiyi araştırmak maksadı ile kullanılan bir makine öğrenmesi metodudur. Lojistik regresyondan, literatürde farklı çalışmalarda pek çok araştırmacı tarafından faydalanılmıştır. (M.-Y. Chen 2011a; Hua vd. 2007; G. Zhang vd. 1999; Chandra, Ravi, ve Bose 2009; Chye Koh ve Kee Low 2004a).

**Tablo 4.6.** Lojistik Regresyon Modeline Ait Karmaşıklık Matrisi

TAHMİN	REFERANS		
	147 (GP)	8 (YP)	155
	16 (YN)	146 (GN)	162
	163	154	

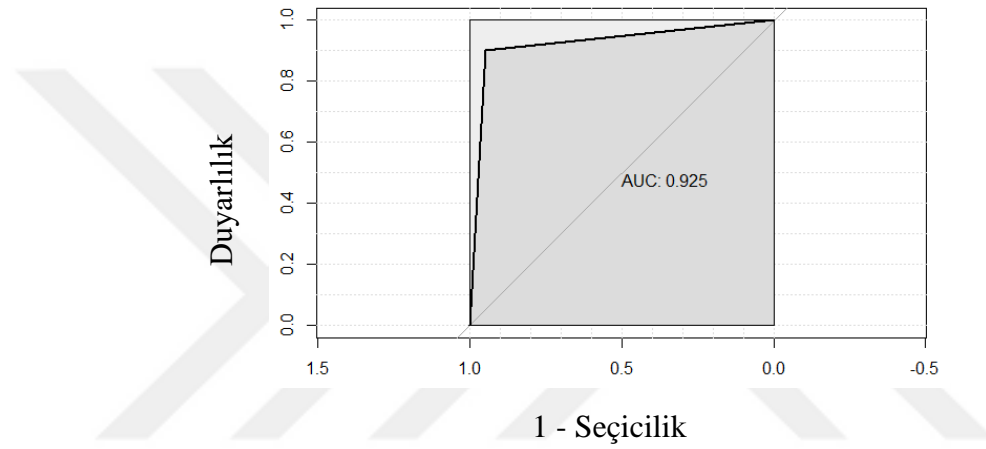
Lojistik regresyon modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.6.' da gösterilmektedir. Lojistik regresyon modeli test veri setinde yer alan toplam 163 riskli gözlemden 147 adedini doğru ve 16 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 146 adedini doğru ve 8 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 155 adedini işletmenin sürekliliği açısından riskli ve 162 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.7.** Lojistik Regresyon Modeline ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9243	0.9018	0.9480	0.0519	0.0981	0.9483

Tablo 4.7.' da lojistik regresyon modeline ait model değerlendirme metrikleri gösterilmektedir. Lojistik regresyon modelinin doğruluk metriği 0.9243 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, lojistik regresyon modeli test veri seti ile yapılan tahminlerin %92.43' ünü doğru tahmin etmiştir. Modelin duyarlılık metriği 0.9018 ve seçicilik metriği 0.9480 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %90,18' ini doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %94.80' ini doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatkınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası 0.0519' dur. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası

0.0981' dir. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığına ölçüsüdür. Modelin kesinlik metriği 0.9483' tür. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %94.83'ünün süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.3.** Lojistik Regresyon Modele ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.3.' te gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayırt edilebilirliğinin ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %92.5 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığını bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

Lojistik regresyon işletmenin sürekliliğinin tahmini literatüründe istatistiksel dönemde pek çok çalışmada, süreklilik riskinin ölçülmesi maksadı ile temel yöntem ve makine öğrenmesi metotlarının istatistiksel tekniklerin yerini aldığı 2000'li yıllardan itibaren ise önerilen modeli karşılaştırmak maksadı ile yaygın olarak kullanılmıştır. Xiao vd. (2012)' nin 276 gözlemin yer aldığı veri seti ile yürüttükleri çalışmalarında yer alan lojistik

regresyon modeli işletmenin riskli kabul edilmesinden t-2 ve t-3 dönem öncesi ait veri setleri ile eğitilmiştir. T-2 dönem de lojistik regresyon modeli doğruluk metriği 0.8346 ve t-3 dönem de lojistik regresyon modeli doğruluk metriği 0.6400 olarak hesaplanmıştır. Çalışma sonuçları işletmenin riskli kabul edildikleri döneme yaklaştıkça tahmin başarısının arttığını göstermektedir. L. Zhang vd. (2013)' nin 92 işletmenin yer aldığı ve sadece mali oranların kullanıldığı veri seti ile yürüttükleri çalışmada lojistik regresyon modeli doğruluk metriği 0.9048 olarak hesaplanmıştır. Zhou, Lai, ve Yen (2014)' in 7728 gözlemin yer aldığı veri seti ile yürüttükleri çalışmalarında eğittikleri lojistik regresyon modelinin doğruluk metriği 0.7399, duyarlılığı 0.7828 ve seçiciliği 0.6971 olarak tespit edilmiştir. C.-F. Tsai (2014) Avustralya, Almanya, Japonya, İflas ve UCcompetition veri setleri ile beş farklı lojistik regresyon modeli eğitmiştir. Avustralya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8714, tip-I hatası 0.1245 ve tip-II hatası 0.1330 olarak hesaplanmıştır. Almanya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.7722, tip-I hatası 0.0981 ve tip-II hatası 0.5137 olarak hesaplanmıştır. Japonya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8574, tip-I hatası 0.1171 ve tip-II hatası 0.1632 olarak hesaplanmıştır. İflas veri ile eğitilmiş olan modelin doğruluk metriği 0.8291, tip-I hatası 0.1386 ve tip-II hatası 0.2078 olarak hesaplanmıştır. UCcompetition veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8224, tip-I hatası 0.2293 ve tip-II hatası 0.1255 olarak hesaplanmıştır. Veganzones ve Séverin (2018) 1500 gözlemin yer aldığı, mali oranlardan oluşan veri seti ile eğittikleri lojistik regresyon modelinin duyarlılık metriği 0.500, seçicilik metriği 0.965 ve AUC skoru 0.771 olarak tespit edilmiştir. W. Zhang, He, ve Zhang (2019) çalışmalarında beş farklı veri seti kullanmıştır. Avustralya veri seti ile eğitilen lojistik regresyon modeli doğruluk metriği 0.8616 ve AUC skoru 0.9255, Almanya veri seti ile eğitilen lojistik regresyon modeli doğruluk metriği 0.7317 ve AUC skoru 0.7684, Japonya veri seti ile eğitilen lojistik regresyon modeli doğruluk metriği 0.8507 ve AUC skoru 0.9263, PPDai veri seti ile eğitilen lojistik regresyon model doğruluk metriği 0.8713 ve AUC skoru 0.6506 ve GMSC veri seti ile eğitilen lojistik regresyon modeli doğruluk metriği 0.9313 ve AUC metriği 0.6892 olarak hesaplanmıştır. Z. Zhang vd. (2022)'nin çalışmalarında 2007 ile 2020 yılları arasında 2260 gözlemi kapsayan veri seti ile eğitilen lojistik regresyon modelinin doğruluk metriği 0.699, duyarlılık metriği 0.684, kesinlik metriği 0.705 ve AUC skoru 0.754' tür. Çalışmada yer alan veri setinde sadece mali oranlara yer verilmiştir. Metrikleri paylaşılan modeller, yer aldıkları çalışmalarda bir temel sınıflayıcı olarak, önerilen modelleri için karşılaştırma

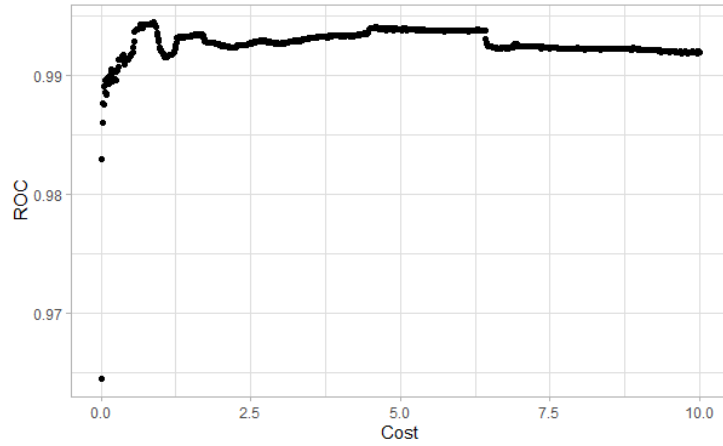
aracı olarak kullanılmıştır. Literatürde yer alan farklı çalışmalarda eğitilen lojistik regresyon modellerinin hesaplanan değerlendirme kriterleri ile yürütülen çalışmada yer alan lojistik regresyon modelinin değerlendirme metrikleri karşılaştırıldığında, çalışmada eğitilen lojistik regresyon modelinin daha iyi bir tahmin performansına sahip olduğu görülmektedir.

#### **4.11.2.2. Destek Vektör Makinesi Modelleri**

Destek vektör makineleri ikili sınıflandırma problemleri için yaygın olarak kullanılan başarılı bir makine öğrenmesi metodudur (Ding, Song, ve Zen 2008; J. Sun vd. 2017; 2021; J. Sun ve Li 2012). Metot veri setinde yer alan gözlemleri, bazı doğrusal olmayan girdi vektörlerini yüksek boyutlu öznitelik uzayına taşımak için doğrusal yaklaşım kullanmaktadır. Yeni bir gözlem uzayında oluşturulan doğrusal model, orijinal gözlem uzayın gerçekte doğrusal olmayan bir karar sınırını temsil etmektedir. Parametre araması ve optimizasyonu algoritmaların eğitiminde çok önemlidir. Ancak hem parametre optimizasyonu hem de en doğru kernel fonksiyonunun seçimine yönelik genel bir yaklaşım literatürde bulunmamaktadır. Kernel fonksiyonları arasında radyal tabanlı fonksiyonun kullanılmasına dair genel bir kabul bulunmaktadır (Bae 2012). Kernel fonksiyonları arasında karşılaştırma yapan araştırmacılar polinomsal kernel fonksiyonunun eğitiminin radyal tabanlı kernel fonksiyonuna göre daha maliyetli olduğunu ve daha kötü sonuçlar verdiğini belirtmektedir (Ding, Song, ve Zen 2008; Z. Huang vd. 2004). Çalışmada doğrusal, polinomsal ve radyal tabanlı olmak üzere üç farklı kernel fonksiyonundan faydalanılmıştır.

##### **4.11.2.2.1. Doğrusal Destek Vektör Makinesi Modeli**

Doğrusal kernel fonksiyonu kullanılan destek vektör makinelerinin sahip olduğu tek parametre C parametresidir. Parametre, modelin yanlış sınıflamadan ne derecede kaçınması gerektiğini tanımlamaktadır. Parametre 0.001 ile 10 aralığında her bir sınamada 0.01 artırılarak ve ROC skoru ölçüsü kullanılarak sınanmıştır ve en iyi ROC skorunu veren değer seçilmiştir. Buna göre C parametresinin 0.871 olduğu durumda en iyi ROC skoru 0.9944 elde edilmiştir.



**Şekil 4.4.** Doğrusal Destek Vektör Makineleri Parametre Optimizasyonu (Cost= C)

Şekil 4.4.' de parametre optimizasyonu sürecinde sınanan değerler ve hesaplanan ROC skorlarının yer aldığı grafik gösterilmektedir.

**Tablo 4.8.** Doğrusal Destek Vektör Makineleri Modeline Ait Karmaşıklık Matrisi

		REFERANS		
TAHMİN	163 (GP)	8 (YP)	171	
	0 (YN)	146 (GN)	146	
	163	154		

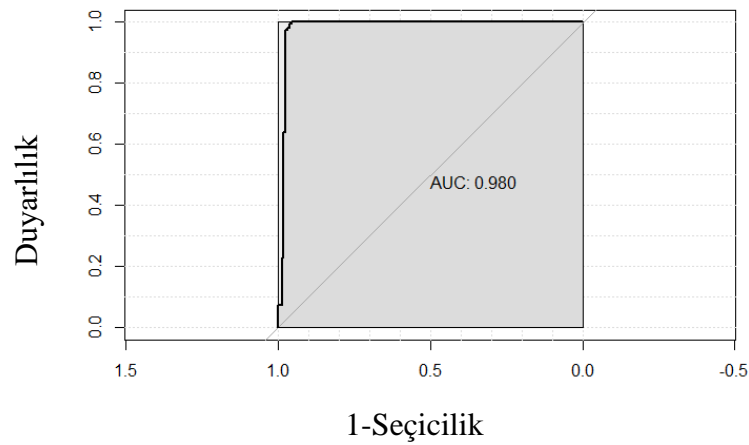
Doğrusal destek vektör makinesi modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.8.' de gösterilmektedir. Model test veri setinde yer alan toplam 163 riskli gözlemin tamamını doğru sınıflandırmıştır. Model riskli olmayan 154 gözlemin 146 adedini doğru ve 8 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 171 adedini işletmenin sürekliliği açısından riskli ve 146 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.9.** Doğrusal Destek Vektör Makinesi Modeline ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9748	1	0.9480	0.0519	0	0.9532

Tablo 4.9.' de doğrusal destek vektör makinesi modeline ait model değerlendirme metrikleri gösterilmektedir. Doğrusal destek vektör makineleri modelinin doğruluk metriği 0.9748 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir.

Hesaplanan doğruluk metriğine göre, doğrusal destek vektör makineleri modeli test veri seti ile yapılan tahminlerin %97.48' ini doğru tahmin etmiştir. Modelin duyarlılık metriği 1 ve seçicilik metriği 0.9480 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden tamamını doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %94.80'ini doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatınlığa sahip olduğunu göstermektedir. Model özellikle süreklilik riski taşıyan gözlemleri doğru sınıflandırmada yüksek kabiliyete sahiptir. Modelin tip-I hatası 0.0519' dur. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Model tip-II hataya sahip değildir. Karmaşıklık matrisi yolu ile hesaplanan kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığını ölçüsüdür. Modelin kesinlik metriği 0.9532' dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %95.32'sinin gerçekten süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.

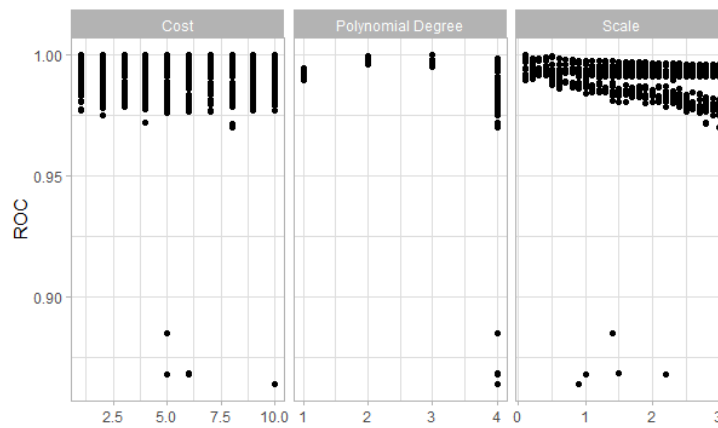


Şekil 4.5. Doğrusal Destek Vektör Makinelere ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.5.' de gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemleri ile süreklilik riski taşımayan gözlemlerin ayırt edilebilirliğinin ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %98 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahiptir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

#### 4.11.2.2.2. Polinomsal Destek Vektör Makinesi Modeli

Polinomsal kernel fonksiyonu kullanılan destek vektör makineleri C, scale ve degree olmak üzere üç farklı parametreye sahiptir. C parametresi modelin yanlış sınıflandırmadan ne derece kaçınması gerektiğini tanımlamaktadır. Parametre 1 ile 10 aralığında her bir sınamada 1 artırılarak sınanmıştır. Degree parametresi fonksiyonun derecesini tanımlamaktadır. Parametre 1 ile 4 aralığında her sınamada 1 artırılarak sınanmıştır. Scale modeldeki normalleştirme parametresidir. Parametre 0.1 ile 3 aralığında her sınamada 0.1 artırılarak sınanmıştır. Parametre optimizasyonu ROC skoru ölçüsü kullanılarak yürütülmüş ve en iyi ROC skorunu veren parametre değerleri seçilmiştir. Buna göre C parametresinin 3, degree parametresinin 3 ve scale parametresinin 0.1 olduğu koşulda en iyi ROC skoru 0.9998 elde edilmiştir.



Şekil 4.6. Polinomsal Destek Vektör Makineleri Parametre Optimizasyonu (Cost = C  
Polynomial Degree=degree, Scale=scale)

Şekil 4.6.’ da parametre optimizasyonu sürecinde sınanan değerler ve hesaplanan ROC skorlarının yer aldığı grafik gösterilmektedir.

**Tablo 4.10.** Polinomsal Destek Vektör Makinelerine Ait Karmaşıklık Matrisi

TAHMİN	REFERANS		
	162 (GP)	3 (YP)	165
1 (YN)	151 (GN)	152	
163	154		

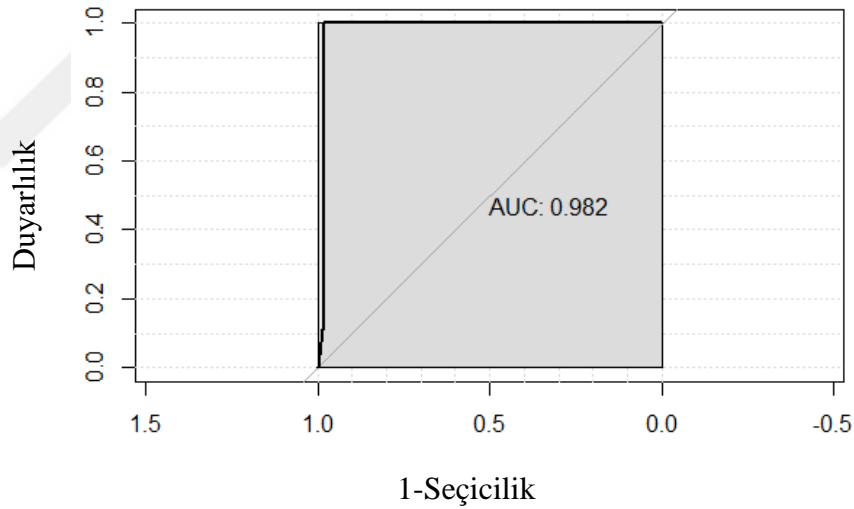
Polinomsal destek vektör makineleri modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.10.’ da gösterilmektedir. Polinomsal destek vektör makineleri modeli test veri setinde yer alan toplam 163 riskli gözlemden 162 adedini doğru ve 1 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 151 adedini doğru ve 3 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 165 adedini işletmenin sürekliliği açısından riskli ve 152 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.11.** Polinomsal Destek Vektör Makinesi Modeline ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9873	0.9938	0.9805	0.0194	0.0061	0.9818

Tablo 4.11.’ de polinomsal destek vektör makineleri modeline ait model değerlendirme metrikleri gösterilmektedir. Polinomsal destek vektör makineleri modelinin doğruluk metriği 0.9873 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, polinomsal destek vektör makineleri modeli test veri seti ile yapılan tahminlerin %98.73’ ünü doğru tahmin etmiştir. Modelin duyarlılık metriği 0.9938 ve seçicilik metriği 0.9805 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %99.38’ ini doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan

gözlemlerden %98.05'ini doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatkınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası 0.0194' tür. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası 0.0061' dir. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığını ölçüsüdür. Modelin kesinlik metriği 0.9818' dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %98.18'inin süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.7.** Polinomsal Destek Vektör Makinesi Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.7.' de gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayırt edilebilirliğinin ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %98.2 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC

skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

#### 4.11.2.2.3 Radyal Destek Vektör Makinesi Modeli

Radyal tabanlı kernel fonksiyonu kullanılan destek vektör makineleri sigma ve C olmak üzere iki parametreye sahiptir. C parametresi modelin yanlış sınıflandırmadan ne derece kaçınması gerektiğini tanımlamaktadır. Parametre 1 ile 200 aralığında her bir sınamada 1 artırılarak sınanmıştır. Sigma parametresi kernel genişliğini tanımlamaktadır. Kernel genişliği, kernel fonksiyonunun yayılmasını kontrol etmektedir. Parametre 0.1 ile 5 aralığında her sınamada 0.1 artırılarak sınanmıştır. Parametre optimizasyonu ROC skoru ölçüsü kullanılarak yürütülmüş ve en iyi ROC skorunu veren değerler seçilmiştir. Buna göre C parametresinin 2 ve sigma parametresinin 0.1 olduğu koşulda en iyi ROC skoru 0.9999 olarak elde edilmiştir.

**Tablo 4.12.** Radyal Destek Vektör Makinesi Modeline Ait Karmaşıklık Matrisi

		REFERANS		
TAHMİN	161 (GP)	0 (YP)	161	
	2 (YN)	154 (GN)	156	
	163	154		

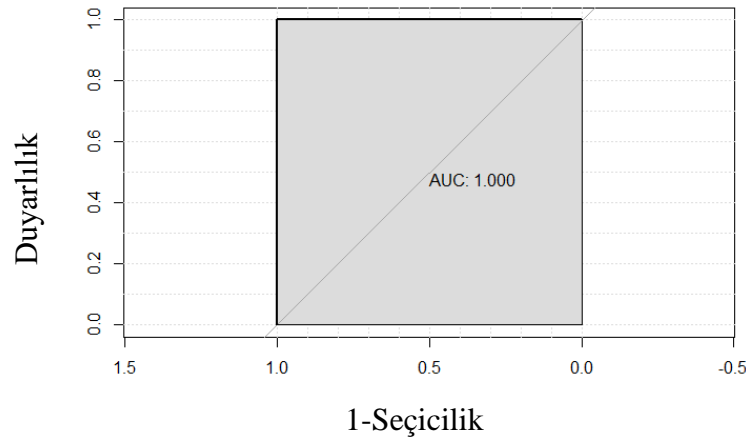
Radyal destek vektör makinesi modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.12.' de gösterilmektedir. Radyal destek vektör makinesi modeli test veri setinde yer alan toplam 163 riskli gözlemden 161 adedini doğru ve 2 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 154 adedini doğru sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 161 adedini işletmenin sürekliliği açısından riskli ve 156 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.13.** Radyal Destek Vektör Makinesi Modeline Ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9936	0.9877	1	0	0.0122	1

Tablo 4.13.' de radyal destek vektör makinesi modeline ait model değerlendirme metrikleri gösterilmektedir. Radyal destek vektör makinesi modelinin doğruluk metriği 0.9936 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, radyal destek vektör makinesi modeli test veri seti ile yapılan tahminlerin %99.36' sını doğru tahmin etmiştir. Modelin duyarlılık metriği

0.9877 ve seçicilik metriği 1 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %98.77' sini doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden tamamını doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası bulunmamaktadır. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası 0.0122' dir. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığının ölçüsüdür. Modelin kesinlik metriği 1' dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin tamamının süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.8.** Radyal Destek Vektör Makinesi Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.8.' de gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayırt edilebilirliğinin ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %100 ihtimalle doğru

şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

Destek vektör makinelerinden, süreklilik tahmini literatüründe, özellikle mali nitelikte veriler ile genelleme performanslarının yüksek olması sebebi ile hem çalışmalarda önerilen modelin eğitiminde hem de karşılaştırma maksadı ile yaygın olarak faydalanılmaktadır. Yeh, Chi, ve Hsu (2010) 114 gözlemin yer aldığı bir veri seti ile yürütülmüş olan çalışmalarında gaussian radyal tabanlı kernel çekirdeği kullanan iki destek vektör makinesi modeli eğitmiştir. Modellerden ilki, çalışmada ek bilgi çıkarımı maksadı ile kullanılan veri zarflama analizinin çıktısını içermemektedir. Modelin doğruluk metriği 0.8333, tip-I hatası 0.1579 ve tip-II hatası 0.1842 olarak tespit edilmiştir. İkinci destek vektör makinesi modeli veri zarflama analizinin çıktısını içermektedir. Modelin doğruluk metriği 0.8684, tip-I hatası 0.1447 ve tip-II hatası 0.1053 olarak hesaplanmıştır. Yang, You, ve Ji (2011)' nin çalışmalarında 120 gözlemin yer aldığı veri seti ile eğittikleri radyal tabanlı destek vektör makinesi modelinin doğruluk metriği 0.8667, duyarlılık metriği 0.8226, seçicilik metriği 0.9138 ve kesinlik metriği 0.9170 olarak hesaplanmıştır. Xiao vd. (2012)' nin 276 gözlemin yer aldığı veri seti ile yürüttükleri çalışmalarında yer alan destek vektör makinesi modeli işletmenin riskli kabul edilmesinden t-2 ve t-3 dönem öncesi ait veri setleri ile eğitilmiştir. T-2 dönem de destek vektör makinesi modeli doğruluk metriği 0.8543 ve t-3 dönem de destek vektör makinesi modeli doğruluk metriği 0.6440 olarak hesaplanmıştır. Çalışma sonuçları işletmenin riskli kabul edildikleri döneme yaklaştıkça tahmin başarısının arttığını göstermektedir. L. Zhang vd. (2013)' nin 92 işletmenin yer aldığı ve sadece mali oranların kullanıldığı veri seti ile yürüttükleri çalışmada eğitilen radyal tabanlı destek vektör makineleri modelinin doğruluk metriği 0.9048 olarak hesaplanmıştır. G. Wang, Ma, ve Yang (2014) çalışmalarında Pietruszkiewicz (2008) ve Shmueli, Patel, ve Bruce (2010) tarafından derlenmiş olan iki farklı veri setinden faydalanarak destek vektör makinesi modeli eğitimi gerçekleştirilmiştir. Sırasıyla birinci modelin doğruluk metriği 0.7221, tip-I hatası 0.2501 ve tip-II hatası 0.3019, ikinci modelin doğruluk metriği 0.7999, tip-I hatası 0.1819 ve tip-II hatası 0.2155 olarak tespit edilmiştir. Zhou, Lai, ve Yen (2014)' nin 6509 gözlemin yer aldığı çalışmalarında eğitilen genetik algoritmalar ile öznelik seçimi uygulanan ve radyal tabanlı kernel fonksiyonu kullanılan destek vektör makineleri modelinin doğruluk

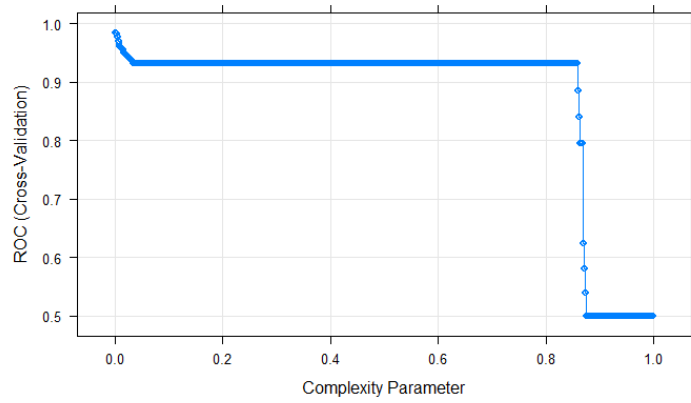
metriği 0.7627, modelin duyarlılığı 0.7453 ve seçiciliği 0.7807 olarak tespit edilmiştir. Yeh, Chi, ve Lin (2014)' in çalışmalarında 220 gözlemin yer aldığı bir veri seti ile eğitilmiş olan radyal tabanlı destek vektör makineleri modelinin tip-I hatası 0.1102 ve tip-II hatası 0.0374 olarak hesaplanmıştır. Çalışmada önerilen model ile karşılaştırma amacı ile eğitilen modele ait tüm metrikler paylaşılmamıştır. C.-F. Tsai, Hsu, ve Yen (2014) çalışmalarında Avustralya, Almanya ve Japonya veri setlerini kullanarak üç farklı destek makinesi modeli eğitmiştir. Avustralya veri seti ile eğitilen modelin doğruluk metriği 0.8563, Almanya veri seti ile eğitilen modelin doğruluk metriği 0.7568 ve Japonya veri seti ile eğitilen modelin doğruluk metriği 0.8637 olarak tespit edilmiştir. Volkov, Benoit, ve Van den Poel (2017) çalışmalarında Belfirst veri tabanından elde edilmiş 2007 ile 2015 yılları arasını kapsayan geniş bir veri seti ile 2011, 2012, 2013, 2014 ve 2015 yılları için beş farklı destek vektör makinesi modeli eğitmiştir. 2011 yılına ait olan en başarılı modelin doğruluk metriği 0.7698 ve modelin AUC skoru 0.8357 olarak tespit edilmiştir. 2012 yılına ait olan modelin doğruluk metriği 0.7578 ve modelin AUC skoru 0.8264 olarak tespit edilmiştir. 2013 yılına ait olan modelin doğruluk metriği 0.7717 ve modelin AUC skoru 0.8352 olarak tespit edilmiştir. 2014 yılına ait olan modelin doğruluk metriği 0.7646 ve modelin AUC skoru 0.8231 olarak tespit edilmiştir. 2015 yılına ait olan modelin doğruluk metriği 0.7705 ve modelin AUC skoru 0.8344 olarak tespit edilmiştir. G. Wang, Chen, ve Chu (2018) çalışmalarında 1726 işletmenin yer aldığı bir veri seti kullanarak, işletmenin riskli olarak sınıflandırıldığı yıldan 3,4 ve 5 yıl önceki veriler ile polinomsal kernel çekirdeği kullanılan destek vektör makinesi modelleri eğitmiştir. Eğitilen modellerden üçü SMOTE ile dengelenmiş veriler kullanılarak eğitilmiştir. SMOTE kullanılmayan modellerin sonuçları sırası ile t-3 dönemine ait olan modelin doğruluk metriği 0.7633, tip-I hatası 0.2267 ve tip-II hatası 0.3605, t-4 dönemine ait olan modelin doğruluk metriği 0.7672, tip-I hatası 0.2271 ve tip-II hatası 0.3034 ve t-5 dönemine ait olan modelin doğruluk metriği 0.7558, tip-I hatası 0.2341 ve tip-II hatası 0.3692 olarak hesaplanmıştır. SMOTE kullanılarak dengelenmiş olan veri seti ile eğitilen t-3 dönemine ait olan modelin doğruluk metriği 0.8806, tip-I hatası 0.1199 ve tip-II hatası 0.1132, t-4 dönemine ait olan modelin doğruluk metriği 0.8649, tip-I hatası 0.1334 ve tip-II hatası 0.1561 ve t-5 dönemine ait olan modelin doğruluk metriği 0.8413, tip-I hatası 0.1545 ve tip-II hatası 0.2107 olarak hesaplanmıştır. Çalışmanın sonuçları veri setinin dengelenmesinin model başarısına katkı sağladığını göstermektedir. Veganzones ve Séverin (2018)' in 1500 gözlemin yer aldığı, mali

oranlardan oluşan veri seti ile eğittikleri radyal tabanlı kernel çekirdeği kullanılan destek vektör makinesi modelinin duyarlılık metriği 0.725, seçicilik metriği 0.913 ve AUC skoru 0.871 olarak tespit edilmiştir. W. Zhang, He, ve Zhang (2019) çalışmalarında Avustralya, Almanya, Japonya, PPDai ve GMSC veri setlerinden faydalanmak sureti ile beş farklı destek vektör makinesi modeli eğitmiştir. Avustralya veri seti ile eğitilen modelin doğruluk metriği 0.8507 ve AUC skoru 0.9183, Almanya veri seti ile eğitilen modelin doğruluk metriği 0.7483 ve AUC skoru 0.6060, Japonya veri seti ile eğitilen modelin doğruluk metriği 0.8493 ve AUC skoru 0.8110, PPDai veri seti ile eğitilen modelin doğruluk metriği 0.8709 ve AUC skoru 0.5424 ve GMSC veri seti ile eğitilen modelin doğruluk metriği 0.9315 ve AUC metriği 0.6288 olarak hesaplanmıştır. G. Wang vd. (2020) çalışmalarında 39 mali oran kullanmıştır. Çalışmada sadece mali oranların yer aldığı veri seti ile eğitilen destek vektör makineleri modelinin AUC skoru 0.8624 olarak hesaplanmıştır. Z. Zhang vd. (2022) 2260 gözlemin yer aldığı çalışmalarında temel sınıflayıcı olarak eğitilen radyal tabanlı destek vektör makineleri modelinin doğruluk metriği 0.701, modelin duyarlılığı 0.701, kesinliği 0.705 ve AUC skoru 0.753 olarak tespit edilmiştir. Literatürde yer alan farklı çalışmalarda eğitilen destek vektör makinesi modellerinin hesaplanan değerlendirme kriterleri ile yürütülen çalışmada yer alan destek vektör makinesi modellerinin değerlendirme metrikleri karşılaştırıldığında, çalışmada eğitilen destek vektör makinesi modellerinin daha iyi bir tahmin performansına sahip olduğu görülmektedir.

#### **4.11.2.3. Sınıflama ve Regresyon Ağacı Modeli**

Karar ağaçları hem sınıflandırma hem de regresyon problemlerinin çözümünde yaygın olarak kullanılan bir makine öğrenmesi metodudur. İşletmenin sürekliliğinin tahmini literatüründe pek çok araştırmacı karar ağaçlarından faydalanmıştır. (Gepp, Kumar, ve Bhattacharya 2010b; M.-Y. Chen 2011a; Chye Koh ve Kee Low 2004a).

Sınıflandırma ve regresyon ağacı modelinin sahip olduğu tek parametre  $cp$  parametresidir. Parametre, ağaç oluşturma süreci sırasında ağaç karmaşıklığı (boyutu) ile doğruluk arasındaki dengeyi düzenlemektedir. Parametre 0.00001 ile 1 aralığında her bir sınamada 0.00001 artırılarak ve ROC skoru ölçüsü kullanılarak sınanmıştır ve en iyi ROC skorunu veren değer seçilmiştir. Buna göre  $cp$  parametresinin 0.00179 olduğu durumda en iyi ROC skoru 0.9842 elde edilmiştir.



**Şekil 4.9.** Sınıflama ve Regresyon Ağacı Modeli Parametre Optimizasyonu  
(Complexity Parameter = Cp)

Şekil 4.9.' da parametre optimizasyonu sürecinde sınanan değerler ve hesaplanan ROC skorlarının yer aldığı grafik gösterilmektedir.

**Tablo 4.14.** Sınıflama ve Regresyon Ağacı Modeline Ait Karmaşıklık Matrisi

		REFERANS		
TAHMİN	158 (GP)	4 (YP)	162	
	5 (YN)	150 (GN)	155	
	163	154		

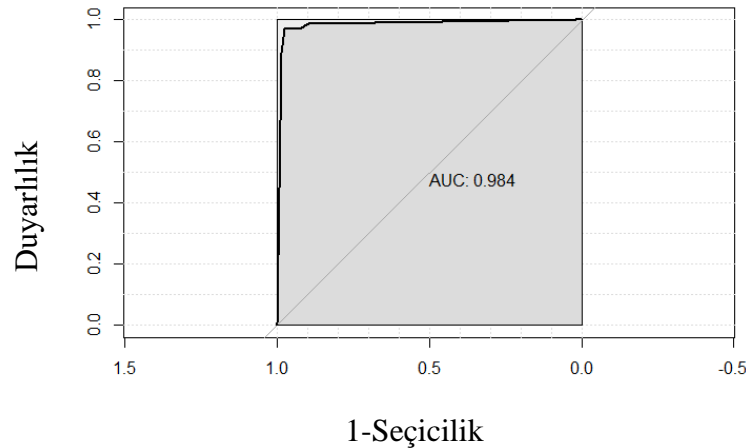
Sınıflandırma ve regresyon ağacı modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.14.' de gösterilmektedir. Sınıflandırma ve regresyon ağacı modeli test veri setinde yer alan toplam 163 riskli gözlemden 158 adedini doğru ve 5 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 150 adedini doğru ve 4 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 162 adedini işletmenin sürekliliği açısından riskli ve 155 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.15.** Sınıflama ve Regresyon Ağacı Modeline Ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9716	0.9693	0.9740	0.0259	0.0306	0.9753

Tablo 4.15.' te sınıflandırma ve regresyon ağacı modeline ait model değerlendirme metrikleri gösterilmektedir. Sınıflandırma ve regresyon ağacı modelinin doğruluk metriği 0.9716

olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, sınıflandırma ve regresyon ağacı modeli test veri seti ile yapılan tahminlerin %97.16' sını doğru tahmin etmiştir. Modelin duyarlılık metriği 0.9693 ve seçicilik metriği 0.9740 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %96.93' ünü doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %97.40' ını doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası 0.0259' dur. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası 0.0306' dır. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığının ölçüsüdür. Modelin kesinlik metriği 0.9753' tür. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %97.53'ünün süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.10.** Sınıflama ve Regresyon Ağacı Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.10.' da gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayırt edilebilirliğinin ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %98.4 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

İşletmenin sürekliliğinin tahmini literatüründe farklı karar ağacı algoritmalarından faydalanılmıştır. Zhou, Lai, ve Yen (2014) 6509 gözlemin yer aldığı çalışmalarında genetik algoritmalar ile öznitelik seçimi uygulamış C4.5 karar ağacı algoritması modelinin doğruluk metriği 0.5067, modelin duyarlılığı 0.9276 ve seçiciliği 0.0858 olarak tespit edilmiştir. C.-F. Tsai, Hsu, ve Yen (2014) çalışmalarında Avustralya, Almanya ve Japonya veri setlerini kullanarak üç farklı karar ağacı modeli eğitmiştir. Avustralya veri seti ile eğitilen modelin doğruluk metriği 0.8491, Almanya veri seti ile eğitilen modelin doğruluk metriği 0.7377 ve Japonya veri seti ile eğitilen modelin doğruluk metriği 0.8637 olarak tespit edilmiştir. G. Wang, Ma, ve Yang (2014) çalışmalarında Pietruszkiewicz (2008) ve Shmueli, Patel, ve Bruce (2010) tarafından derlenmiş olan iki farklı veri setinden faydalanarak karar ağacı modellerinin eğitimi gerçekleştirilmiştir. Birinci modelin doğruluk metriği 0.7163, tip-I hatası 0.2619 ve tip-II hatası 0.2324, ikinci modelin doğruluk metriği 0.7599, tip-I hatası 0.2583 ve tip-II hatası 0.2638 olarak tespit edilmiştir. Yeh, Chi, ve Lin (2014)' in çalışmalarında 220 gözlemin yer aldığı bir veri seti ile eğitilmiş olan C4.5 karar ağacı modelinin tip-I hatası 0.2600 ve tip-II hatası 0.0190 olarak hesaplanmıştır. Çalışmada önerilen model ile karşılaştırma amacı ile eğitilen modele ait tüm metrikler paylaşılmamıştır. C.-F. Tsai (2014) Avustralya, Almanya, Japonya, İflas ve UCcompetiton veri setleri ile beş farklı karar ağacı (CART algoritması) modeli eğitmiştir. Avustralya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8611, tip-I hatası 0.1813 ve tip-II hatası 0.1028 olarak hesaplanmıştır. Almanya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.7812, tip-I hatası 0.0952 ve tip-II hatası 0.4847 olarak hesaplanmıştır. Japonya veri seti ile eğitilmiş olan modelin doğruluk metriği 0.9394, tip-I hatası 0.0612 ve tip-II hatası 0.0593 olarak hesaplanmıştır. İflas veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8593, tip-I hatası 0.1215 ve tip-II hatası 0.1641

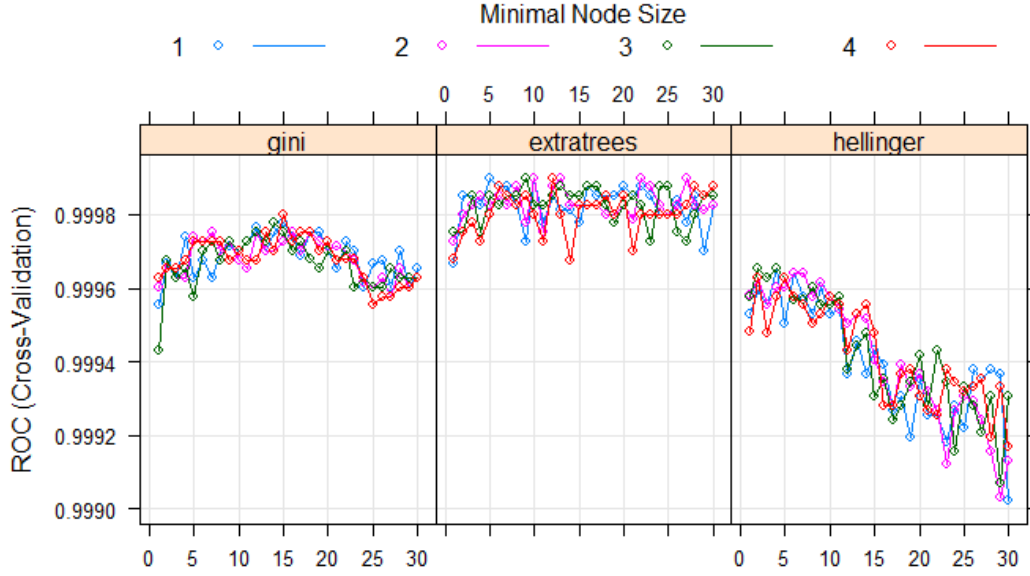
olarak hesaplanmıştır. UCcompetition veri seti ile eğitilmiş olan modelin doğruluk metriği 0.8307, tip-I hatası 0.2035 ve tip-II hatası 0.1331 olarak hesaplanmıştır. W. Zhang, He, ve Zhang (2019) çalışmalarında beş farklı veri seti kullanmıştır. Avustralya veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.8145 ve AUC skoru 0.8124, Almanya veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.6633 ve AUC skoru 0.7753, Japonya veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.8116 ve AUC skoru 0.9190, PPDai veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.7816 ve AUC skoru 0.5793 ve GMSC veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.8915 ve AUC metriği 0.6064 olarak hesaplanmıştır. Z. Zhang vd. (2022) 2007 ile 2020 yılları arasında 2260 gözlemi kapsayan veri seti ile eğitilen karar ağacı modelinin doğruluk metriği 0.799, duyarlılık metriği 0.811, kesinlik metriği 0.793 ve AUC skoru 0.879' dur. Literatürde yer alan farklı çalışmalarda eğitilen karar ağacı modellerinin hesaplanan değerlendirme kriterleri ile yürütülen çalışmada yer alan karar ağacı modelinin değerlendirme metrikleri karşılaştırıldığında, çalışmada eğitilen karar ağacı modelinin daha iyi bir tahmin performansına sahip olduğu görülmektedir.

#### **4.11.2.4. Rastgele Orman Modeli**

Rastgele orman algoritması, gelişmiş bir makine öğrenmesi metodudur. Model, rastgele üretilmiş karar ağaçlarından oluşturulmuş bir topluluktan yaralanarak sınıflandırma yapmaktadır. Rastgele orman algoritması diğer gelişmiş metotlar ile birlikte süreklilik tahmini literatüründe pratik bir değere sahiptir. Rastgele orman algoritmasından süreklilik tahmini çalışmalarında yaygın olarak faydalanılmaktadır (Chandra, Ravi, ve Bose 2009; S. Jones, Johnstone, ve Wilson 2017; Qian vd. 2022; Yeh, Chi, ve Lin 2014b; Barboza, Kimura, ve Altman 2017; Garcia 2022).

Rastgele orman algoritması mtry, splitrule ve min.node.size olmak üzere üç adet parametreye sahiptir. Mtry parametresi, topluluğu oluşturacak olan karar ağaçlarının oluşturulmasında her bir bölmede kaç adet özneliğin dikkate alınacağını tanımlamaktadır. Parametre 1 ile 30 aralığında her bir sınamada 1 artırılarak sınanmıştır. Splitrule parametresi, topluluk içerisindeki her düğümde en iyi bölünmeyi seçmek maksadı kullanılmaktadır. Parametre seçiminde gini, extratrees ve hellinger bölünme kuralları sınanmıştır. Min.node.size parametresi, karar ağacı oluşturma aşamasında algoritmanın davranışını kontrol etmek maksadı ile kullanılmaktadır. Parametre 1 ile 4 aralığında her bir sınamada 1 artırılarak sınanmıştır. Parametre seçimi ROC skoru ölçüsü kullanılarak

yürütülmüştür ve iyi ROC skoru veren parametreler seçilmiştir. Buna göre mtry parametresinin 5, splitrule parametresinin extratrees ve min.node.size parametresinin 1 olduğu durumda en iyi ROC skoru 0.9999 elde edilmiştir.



**Şekil 4.11.** Rastgele Orman Modeli Parametre Optimizasyonu

Şekil 4.11.' de parametre optimizasyonu sürecinde sınanan değerler ve hesaplanan ROC skorlarının yer aldığı grafik gösterilmektedir.

**Tablo 4.16.** Rastgele Orman Modeline Ait Karmaşıklık Matrisi

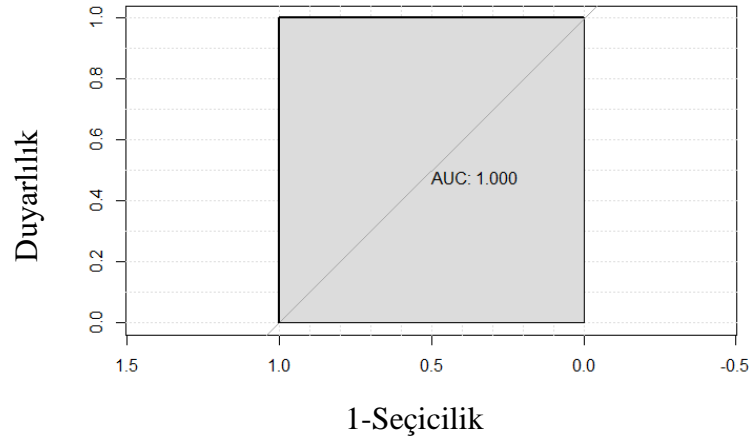
REFERANS			
TAHMİN	163 (GP)	0 (YP)	163
	0 (YN)	154 (GN)	154
	163	154	

Rastgele orman modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.16.' da gösterilmektedir. Rastgele orman modeli test veri setinde yer alan toplam 163 riskli gözlemden 63 adedini doğru sınıflandırmıştır. Model riskli olmayan 154 gözlemin 154 adedini doğru sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 163 adedini işletmenin sürekliliği açısından riskli ve 154 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır. Rastgele orman model test veri seti ile mükemmel tahmin başarısı ortaya koymuştur.

**Tablo 4.17.** Rastgele Orman Modeline Ait Model Değerleme Metrikleri

<b>Metrik</b>	<b>Doğruluk</b>	<b>Duyarlılık</b>	<b>Seçicilik</b>	<b>Tip-I Hata</b>	<b>Tip-II Hata</b>	<b>Kesinlik</b>
<b>Değer</b>	1.000	1.000	1.000	0	0	1.000

Tablo 4.17’ de rastgele orman modeline ait model değerlendirme metrikleri gösterilmektedir. Rastgele orman modelinin doğruluk metriği 1 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, rastgele orman modeli test veri seti ile yapılan tahminlerin %100’ ünü doğru tahmin etmiştir. Modelin duyarlılık metriği 1 ve seçicilik metriği 1 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %100’ nü doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %100’ ünü doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatınlığa sahip olduğunu göstermektedir. Rastgele orman modeli tip-I ve tip-II hataya sahip değildir. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığının ölçüsüdür. Modelin kesinlik metriği 1’ dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %100’ ünün süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.12.** Rastgele Orman Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.12.' de gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayrılabilir ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %100 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

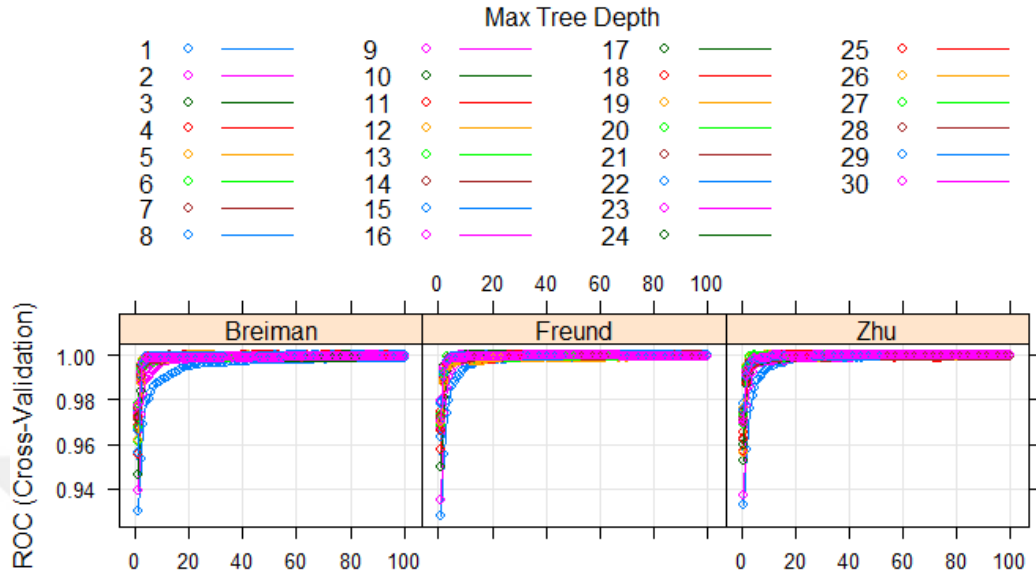
Bir bootstrap aggregating metodu olan rastgele orman algoritması literatürde farklı araştırmacılar tarafından kullanılmıştır. Volkov, Benoit, ve Van den Poel (2017) çalışmalarında Belfirst veri tabanından elde edilmiş 2007 ile 2015 yılları arasını kapsayan geniş bir veri seti ile 2011, 2012, 2013, 2014 ve 2015 yılları için beş farklı rastgele orman modeli eğitmiştir. 2011 yılına ait olan en başarılı modelin doğruluk metriği 0.7887 ve modelin AUC skoru 0.8568 olarak tespit edilmiştir. 2012 yılına ait olan modelin doğruluk metriği 0.7789 ve modelin AUC skoru 0.8458 olarak tespit edilmiştir. 2013 yılına ait olan modelin doğruluk metriği 0.7874 ve modelin AUC skoru 0.8521 olarak tespit edilmiştir. 2014 yılına ait olan modelin doğruluk metriği 0.7835 ve modelin AUC skoru 0.8469 olarak tespit edilmiştir. 2015 yılına ait olan modelin doğruluk metriği 0.7944 ve modelin AUC skoru 0.8590 olarak tespit edilmiştir. Veganzones ve Séverin (2018) 1500 gözlemin yer aldığı, mali oranlardan oluşan veri seti ile eğittikleri rastgele orman modelinin duyarlılık metriği 0.571, seçicilik metriği 0.918 ve AUC skoru 0.807 olarak tespit edilmiştir. Jabeur vd. (2021) tarafından 133 gözlemin yer aldığı ve 2017(Y), 2016(Y-1),

2015(Y-2) ve 2014(Y-3) yıllarına ait verilerin kullanıldığı çalışmada eğitilen rastgele orman modelleri, 2016 yılı için 0.9559 doğruluk ve 0.939 AUC skoru, 2015 yılı için 0.7308 ve 0.813 ve 2014 yılı için 0.7436 ve 0.739 AUC skoruna sahiptir. Qian vd. (2022) çalışmalarında 2001-2019 yılları arasını kapsayan ve 4167 gözlem içeren veri seti ile gözlemin riski kabul edildiği yıldan bir yıl önceki veriler dikkate alınarak eğitilmiş olan rastgele orman modelinin doğruluk metriği 0.8499, duyarlılığı 0.8345, kesinliği 0.8608 ve AUC skoru 0.8248 olarak hesaplanmıştır. Garcia (2022) tarafından Ocak 2010 ile Aralık 2018 döneminin 1824 işletmeye ait 41933 çeyrek dönemlik mali raporları kapsayan veri seti ile eğitilmiş olan rastgele orman ve rastgele orman topluluğu modellerinin duyarlılık metrikleri sırası ile 0.714 ve 1, seçicilik metrikleri sırası ile 0.978 ve 1, AUC skorları sırası ile 0.846 ve 1 olarak tespit edilmiştir. Z. Zhang vd. (2022) 2007 ile 2020 yılları arasında 2260 gözlemi kapsayan veri seti ile eğitilen rastgele orman modelinin doğruluk metriği 0.832, duyarlılık metriği 0.811, kesinlik metriği 0.846 ve AUC skoru 0.893' tür. Sonuçları paylaşılan çalışmalar arasından özellikle Garcia (2022) tarafından geliştirilen rastgele orman topluluğu modeli, farklı bir veri seti ile yürütülen çalışmadaki rastgele orman modelinin sonuçlarını, değerlendirme metrikleri açısından birebir karşılamaktadır. Sonuçları paylaşılan diğer değerli çalışmalar ile yürütülen çalışmadaki modelin sonuçları karşılaştırıldığında, geliştirilen modelin tahmin başarısının daha iyi olduğu görülmektedir.

#### **4.11.2.5. Adaboost Modeli**

Adaboost, Freund ve Schapire (1995) tarafından formüle edilmiş olan bir topluluk öğrenmesi metodudur. Adaboost daha güçlü bir sınıflayıcı elde etmek maksadı ile, temel sınıflayıcılardan bir topluluk oluşturmayı amaçlamaktadır. Metot literatürde yaygın olarak kullanılmaktadır (Alfaro vd. 2008; Heo ve Yang 2014; S. Y. Kim ve Upneja 2014; J. Sun vd. 2017). Adaboost algoritması mfinal, maxdepth ve coeflearn olmak üzere üç adet parametreye sahiptir. Mfinal parametresi, modellemede kullanılacak olan maksimum temel sınıflayıcı sayısını düzenlemektedir. Parametre 1 ile 100 aralığında her bir sınamada 1 artırılarak sınanmıştır. Maxdepth parametresi model derinliğini düzenlemektedir. Parametre 1 ile 30 aralığında her bir sınamada 1 artırılarak sınanmıştır. Coeflearn parametresi kullanılacak olan katsayı öğrencisi tanımlamaktadır. Parametre seçiminde Zhu, Breiman ve Freund methotları sınanmıştır. Parametre seçimi ROC skoru ölçüsü kullanılarak yürütülmüştür ve iyi ROC skoru veren parametreler seçilmiştir. Buna

göre mfinal 20, maxdepth 19 ve coeflearn Zhu olduğu durumda en iyi ROC skoru 0.9999 elde edilmiştir.



Şekil 4.13. Adaboost Modeli Parametre Optimizasyonu

Şekil 4.13.' te parametre optimizasyonu sürecinde sınanan değerler ve hesaplanan ROC skorlarının yer aldığı grafik gösterilmektedir.

Tablo 4.18. Adaboost Modeline Ait Karmaşıklık Matrisi

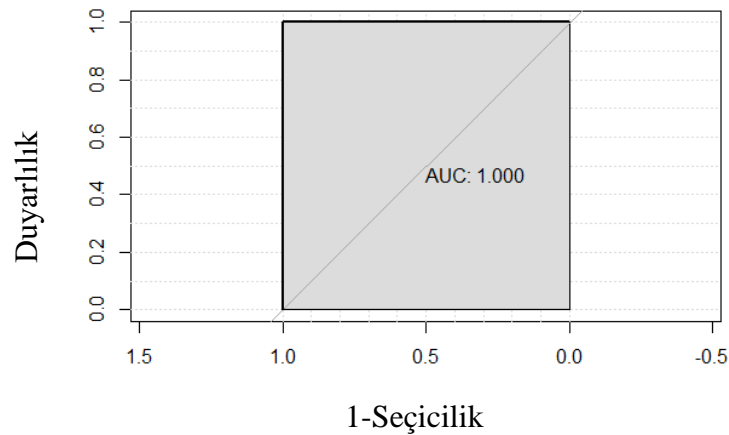
		REFERANS		
TAHMİN	163 (GP)	2 (YP)	165	
	0 (YN)	152 (GN)	152	
	163	154		

Adaboost modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.18.' de gösterilmektedir. Adaboost modeli test veri setinde yer alan toplam 163 riskli gözlemden 163 adedini sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 152 adedini doğru ve 2 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 165 adedini işletmenin sürekliliği açısından riskli ve 152 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

Tablo 4.19. Adaboost Modeline Ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9936	1	0.9870	0.0129	0	0.9878

Tablo 4.19.' da adaboost modeline ait model deęerleme metrikleri gsterilmektedir. Adaboost modelinin doęruluk metrięi 0.9936 olarak hesaplanmıřtır. Doęruluk metrięi modelin genel tahmin bařarısı gstermektedir. Hesaplanan doęruluk metrięine gre, adaboost modeli test veri seti ile yapılan tahminlerin %99.36' sını doęru tahmin etmiřtir. Modelin duyarlılık metrięi 1 ve seicilik metrięi 0.9870 olarak hesaplanmıřtır. Duyarlılık metrięi srekli riski tařıyan gzlemlerden kata kaının doęru olarak tahmin edildięini gstermektedir. Seicilik metrięi ise srekli riski tařımayan gzlemlerden kata kaının doęru sınıflandıęını gstermektedir. Saptanan duyarlılık metrięi, modelin srekli riski tařıyan gzlemlerden %100' n doęru tahmin ettięini ve saptanan seicilik metrięi modelin srekli riski tařımayan gzlemlerden %98.70'ini doęru tahmin ettięini gstermektedir. Duyarlılık ve seicilik metrięi sonuları modelin hem srekli riski tařıyan gzlemleri hem de srekli riski tařımayan gzlemleri doęru sınıflandırmada yksek yatkınlıęa sahip olduęunu gstermektedir. Modelin tip-I hatası 0.0129' dur. Tip-I hata srekli riski tařımayan bir gzlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hesaplanmamaktadır. Karmařıklık matrisi yolu ile hesaplanan son metrik kesinlik metrięidir. Kesinlik metrięi, modelin referans sınıfı ne derece doęru tahmin ettięinin gstergesidir. Bir bařka ifade ile kesinlik srekli riski tařıdığı tahmin edilen gzlemlerin kata kaının gerekte srekli riski tařıdığıının lsdr. Modelin kesinlik metrięi 0.9878' dir. Modelin kesinlik metrięi srekli riski tařıyan gzlem olarak tahmin ettięi rneklerin %98.78'inin srekli riski tařıdığıını gstermektedir. Model, bir gzlemi srekli riski tařıyor olarak sınıflandırmada yksek tahmin kabiliyetine sahiptir.



**řekil 4.14.** Adaboost Modeline Ait ROC Eęrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.14.' te gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayrılabilir ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %100 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

Bir boosting metodu olan adaboost algoritması literatürde hem ağaç temelli algoritmalar ile hem de farklı temel sınıflayıcılar ile süreklilik riskinin tahmininde kullanılmıştır. S. Y. Kim ve Upneja (2014) tarafından 1988-2011 yılları arasını kapsayan dönemde 826 gözlem ile yürütülen çalışmada eğitilen adaboost modelinin doğruluk metriği 0.9769 ve AUC skoru 0.9880 olarak tespit edilmiştir. G. Wang vd. (2020) çalışmalarında 39 mali oran kullanmıştır. Çalışmada sadece mali oranlar öznitelik olarak kullanılarak eğitilen adaboost modelinin AUC skoru 0.8862 olarak hesaplanmıştır. X. Du vd. (2020) tarafından 2017 ve 2018 yıllarını kapsayan ve 7154 gözlemi içeren bir veri setinden faydalanılarak eğitilmiş adaboost modeli duyarlılık metriği 0.9838 ve AUC skoru 0.9458 olarak tespit edilmiştir. Qian vd. (2022) çalışmalarında 2001-2019 yılları arasını kapsayan ve 4167 gözlem içeren veri seti ile gözlemin riski kabul edildiği yıldan bir yıl önceki veriler dikkate alınarak eğitilmiş olan adaboost modelinin doğruluk metriği 0.8410, duyarlılığı 0.8214, kesinliği 0.8548 ve AUC skoru 0.9182 olarak hesaplanmıştır. Literatürde yer alan farklı çalışmalarda eğitilen adaboost modellerinin hesaplanan değerlendirme kriterleri ile yürütülen çalışmada yer alan modelin değerlendirme metrikleri karşılaştırıldığında, çalışmada eğitilen modelin daha iyi bir tahmin performansına sahip olduğu görülmektedir.

#### **4.11.2.6. Gradient Boosting Modeli**

Gradient boosting makine öğrenmesinde en iyi performansı gösteren yöntemlerden birisi olarak kabul edilmektedir. Literatürde farklı çalışmalarda metottan faydalanılmıştır (Son vd. 2019; Wyrobek ve Kluza 2019; Zięba, Tomczak, ve Tomczak 2016). Gradient Boosting n.trees, interaction.depth, shrinkage ve nminobsinnode olmak üzere dört adet parametreye sahiptir. N.trees parametresi, modeldeki ağaç sayısını düzenlemektedir.

Parametre 100 ile 2500 aralığında her bir sınamada 100 artırılarak sınanmıştır. Interaction.depth parametresi temel sınıflayıcı karar ağaçlarının derinliğini düzenlemektedir. Parametre 1 ile 10 aralığında her bir sınamada 1 artırılarak sınanmıştır. Shrinkage topluluktaki her bir ağacın modele katkısını düzenlemektedir. Parametre 0.0005 ile 0.002 aralığında her sınamada 0.0001 artırılarak sınanmıştır. N.minobsinnode ağaç oluşturma sürecinde yeni bir düğüm oluşturmak için gereken minimum gözlem sayısını kontrol etmektedir. Parametre 1 ile 5 aralığında her bir sınamada 1 artırılarak sınanmıştır. Parametre seçimi ROC skoru ölçüsü kullanılarak yürütülmüştür ve iyi ROC skoru veren parametreler seçilmiştir. Buna göre were n.trees 1500, interaction.depth 10, shrinkage 0.0019 and n.minobsinnode 3 olduğu durumda en iyi ROC skoru 0.9996 elde edilmiştir.

**Tablo 4.20.** Gradient Boosting Modeline Ait Karmaşıklık Matrisi

	REFERANS		
TAHMİN	161 (GP)	3 (YP)	164
	2 (YN)	151 (GN)	154
	163	154	

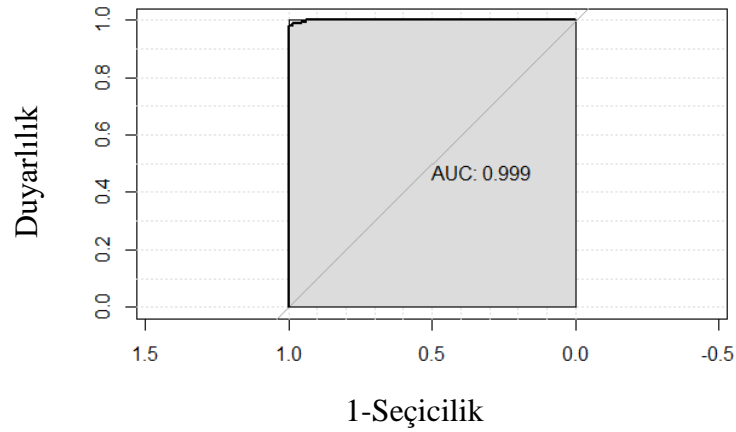
Gradient boosting modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.20.' de gösterilmektedir. Gradient boosting modeli test veri setinde yer alan toplam 163 riskli gözlemden 161 adedini doğru ve 2 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 151 adedini doğru ve 3 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 164 adedini işletmenin sürekliliği açısından riskli ve 154 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.21.** Gradient Boosting Modeline Ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9842	0.9877	0.9806	0.0193	0.0122	0.9817

Tablo 4.21.' de gradient boosting modeline ait model değerlendirme metrikleri gösterilmektedir. Gradient boosting modelinin doğruluk metriği 0.9842 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, gradient boosting modeli test veri seti ile yapılan tahminlerin %98.42' sini doğru tahmin etmiştir. Modelin duyarlılık metriği 0.9877 ve seçicilik metriği 0.9806 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan

gözlemlerden kaçta kaçının doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %98.77' sini doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %98.06' sını doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatkınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası 0.0193' tür. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası 0.0122' dir. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığına ölçüsüdür. Modelin kesinlik metriği 0.9817' dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %98.17'ünün süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.15.** Gradient Boosting Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.15.' te gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayrılabilir ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik

riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %99.9 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

Gradient boosting algoritması literatürde son yıllarda yaygın olarak kullanılan bir topluluk öğrenmesi metodudur. X. Du vd. (2020) tarafından 2017 ve 2018 yıllarını kapsayan ve 7154 gözlemi içeren bir veri setinden faydalanılarak eğitilen gradient boosting modeli duyarlılık metriği 0.9954 ve AUC skoru 0.9644 olarak tespit edilmiştir. Jabeur vd. (2021) tarafından 133 gözlemin yer aldığı ve 2017(Y), 2016(Y-1), 2015(Y-2) ve 2014(Y-3) yıllarına ait verilerin kullanıldığı çalışmada eğitilen gradient boosting modeli ile 2016 yılı için 0.9231 doğruluk ve 0.951 AUC skoru, 2015 yılı için 0.7308 ve 0.733 ve 2014 yılı için 0.7607 ve 0.718 AUC skoru hesaplanmıştır. Qian vd. (2022) çalışmalarında 2001-2019 yılları arasını kapsayan ve 4167 gözlem içeren veri seti ile gözlemin riski kabul edildiği yıldan bir yıl önceki veriler dikkate alınarak eğitilmiş olan gradient boosting modelinin doğruluk metriği 0.8509, duyarlılığı 0.8408, kesinliği 0.8580 ve AUC skoru 0.9274 olarak hesaplanmıştır. Literatürde yer alan farklı çalışmalarda eğitilen gradient boosting modellerinin hesaplanan değerlendirme kriterleri ile yürütülen çalışmada yer alan modelin değerlendirme metrikleri karşılaştırıldığında, çalışmada eğitilen modelin tahmin performansının iyi olduğu görülmektedir.

#### **4.11.2.7. XGBoost Modeli**

Extreme gradient boosting, T. Chen ve Guestrin (2016) tarafından formüle edilmiş gradient boosting metodunun gelişmiş bir versiyonudur. XGBoost yüksek tahmin başarısı yakalama potansiyeli sebebi ile yaygın olarak faydalanılan bir yöntemdir (Volkov, Benoit, ve Van den Poel 2017; Y.-P. Huang ve Yen 2019; Jabeur vd. 2021; Qian vd. 2022). Extreme gradient boosting algoritması nrounds, max\_depth, eta, gamma, colsample\_bytree, min\_child\_weight ve subsample olmak üzere yedi parametreye sahiptir. Nrounds parametresi boosting iterasyon sayısı, max\_depth parametresi temel sınıflayıcıların sahip olacağı maksimum derinliği, eta parametresi modelin her bir iterasyonda aktaracağı veri miktarını, gamma parametresi ağaç oluşturma sürecinde bir yaprak düğümde bölünme yapmak için gerekli minimum kaybı, colsample\_bytree parametresi boosting sırasında her bir ağacı oluştururken rastgele örneklenecek olan

özniteliklerin oranını, min\_child\_weight parametresi ağaç oluşturma sürecinde bir alt düğümde gerekli olan minimum örnek ağırlığı toplamını ve subsample parametresi ağaç oluşturma sırasında rastgele seçilecek ve dahil edilecek olan gözlemler düzenlemektedir. Eğitim aşamasında parametre araması “caret” kütüphanesinin otomatik parametre arama fonksiyonu yardımı ile yürütülmüştür. Parametre seçimi ROC skoru ölçüsü kullanılarak yürütülmüştür ve iyi ROC skoru veren parametreler seçilmiştir. Buna göre nrounds 478, max\_depth 6, eta 0.2970209, gamma 0.5715507, colsample\_bytree 0.6139304, min\_child\_weight 5 ve subsample 0.7726817 olduğu durumda en iyi ROC skoru 0.9988 elde edilmiştir.

**Tablo 4.22.** XGBoost Modeline Ait Karmaşıklık Matrisi

	REFERANS		
TAHMİN	161 (GP)	3 (YP)	164
	2 (YN)	151 (GN)	154
	163	154	

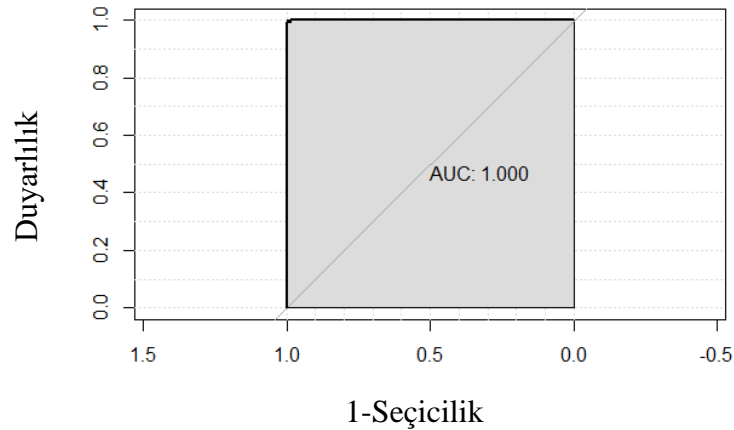
XGBoost modelinin eğitimi sonrasında, test veri seti ile yapılan tahmine ilişkin sonuçları gösteren karmaşıklık matrisi Tablo 4.22.’ de gösterilmektedir. XGBoost modeli test veri setinde yer alan toplam 163 riskli gözlemden 161 adedini doğru ve 2 adedini yanlış sınıflandırılmıştır. Model riskli olmayan 154 gözlemin 151 adedini doğru ve 3 adedini yanlış sınıflandırmıştır. Test veri setinde yer alan toplam 317 gözlemin 164 adedini işletmenin sürekliliği açısından riskli ve 154 adedini işletmenin sürekliliği açısından risksiz olarak sınıflandırmıştır.

**Tablo 4.23.** XGBoost Modeline Ait Model Değerleme Metrikleri

Metrik	Doğruluk	Duyarlılık	Seçicilik	Tip-I Hata	Tip-II Hata	Kesinlik
Değer	0.9842	0.9877	0.9806	0.0193	0.0122	0.9817

Tablo 4.23.’ te XGBoost modeline ait model değerlendirme metrikleri gösterilmektedir. Gradient boosting modelinin doğruluk metriği 0.9842 olarak hesaplanmıştır. Doğruluk metriği modelin genel tahmin başarısı göstermektedir. Hesaplanan doğruluk metriğine göre, XGBoost modeli test veri seti ile yapılan tahminlerin %98.42’ sini doğru tahmin etmiştir. Modelin duyarlılık metriği 0.9877 ve seçicilik metriği 0.9806 olarak hesaplanmıştır. Duyarlılık metriği süreklilik riski taşıyan gözlemlerden kaçta kaçının

doğru olarak tahmin edildiğini göstermektedir. Seçicilik metriği ise süreklilik riski taşımayan gözlemlerden kaçta kaçının doğru sınıflandığını göstermektedir. Saptanan duyarlılık metriği, modelin süreklilik riski taşıyan gözlemlerden %98.77' sini doğru tahmin ettiğini ve saptanan seçicilik metriği modelin süreklilik riski taşımayan gözlemlerden %98.06' sını doğru tahmin ettiğini göstermektedir. Duyarlılık ve seçicilik metriği sonuçları modelin hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri doğru sınıflandırmada yüksek yatkınlığa sahip olduğunu göstermektedir. Modelin tip-I hatası 0.0193' tür. Tip-I hata süreklilik riski taşımayan bir gözlemin, riskli olarak sınıflandırılmasıdır. Modelin tip-II hatası 0.0122' dir. Tip-II hata süreklilik riski taşıyan bir gözlemin, risksiz olarak sınıflandırılmasıdır. Karmaşıklık matrisi yolu ile hesaplanan son metrik kesinlik metriğidir. Kesinlik metriği, modelin referans sınıfı ne derece doğru tahmin ettiğinin göstergesidir. Bir başka ifade ile kesinlik süreklilik riski taşıdığı tahmin edilen gözlemlerin kaçta kaçının gerçekte süreklilik riski taşıdığının ölçüsüdür. Modelin kesinlik metriği 0.9817' dir. Modelin kesinlik metriği süreklilik riski taşıyan gözlem olarak tahmin ettiği örneklerin %98.17'ünün süreklilik riski taşıdığını göstermektedir. Model, bir gözlemi süreklilik riski taşıyor olarak sınıflandırmada yüksek tahmin kabiliyetine sahiptir.



**Şekil 4.16.** XGBoost Modeline Ait ROC Eğrisi ve AUC Skoru

Modelin ROC eğrisi ve AUC skoru Şekil 4.16.' da gösterilmektedir. ROC eğrisi bir olasılık eğrisidir ve AUC skoru ise süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemlerin ayrılabilir ölçüsüdür. Bir başka ifade ile AUC modelin iki sınıfı ne derece ayırt edebildiğini göstermektedir. Hesaplanan AUC skoru modelin süreklilik riski taşıyan gözlemler ile süreklilik riski taşımayan gözlemleri %100 ihtimalle doğru şekilde ayırt edebilme yeteneğine sahip olduğunu göstermektedir. Literatür

incelendiğinde ROC skorunun 0.50' nin üzerinde olması gerektiği ve modelin ROC skorunun 0.90' ın üzerinde olması durumunun modelin mükemmel çalıştığının bir göstergesi olarak kabul edilmektedir (Fawcett 2006).

XGBoost, literatürde son yıllarda yaygın olarak kullanılan bir topluluk öğrenmesi metodudur. Volkov, Benoit, ve Van den Poel (2017) çalışmalarında Belfirst veri tabanından elde edilmiş 2007 ile 2015 yılları arasını kapsayan geniş bir veri seti ile 2011, 2012, 2013, 2014 ve 2015 yılları için beş farklı XGBoost modeli eğitmiştir. 2011 yılına ait olan en başarılı modelin doğruluk metriği 0.7869 ve modelin AUC skoru 0.8506 olarak tespit edilmiştir. 2012 yılına ait olan modelin doğruluk metriği 0.7740 ve modelin AUC skoru 0.8360 olarak tespit edilmiştir. 2013 yılına ait olan modelin doğruluk metriği 0.7773 ve modelin AUC skoru 0.8415 olarak tespit edilmiştir. 2014 yılına ait olan modelin doğruluk metriği 0.7729 ve modelin AUC skoru 0.8337 olarak tespit edilmiştir. 2015 yılına ait olan modelin doğruluk metriği 0.7824 ve modelin AUC skoru 0.8461 olarak tespit edilmiştir. Y.-P. Huang ve Yen (2019) tarafından 2010-2016 dönemini kapsayan ve 64 gözlemi içeren bir veri seti ile yürütülen çalışmada eğitilen XGBoost modeli doğruluk metriği 0.906, duyarlılık metriği 0.875, seçicilik metriği 0.938, tip-I hata 0.0310 ve tip-II hata 0.0630 olarak tespit edilmiştir. X. Du vd. (2020) tarafından 2017 ve 2018 yıllarını kapsayan ve 7154 gözlemi içeren bir veri setinden faydalanılarak eğitilmiş XGBoost modeli duyarlılık metriği 0.9444 ve AUC skoru 0.9321 olarak tespit edilmiştir. Jabeur vd. (2021) tarafından 133 gözlemin yer aldığı ve 2017(Y), 2016(Y-1), 2015(Y-2) ve 2014(Y-3) yıllarına ait verilerin kullanıldığı çalışmada eğitilen XGBoost modeli ile 2016 yılı için 0.9359 doğruluk ve 0.931 AUC skoru, 2015 yılı için 0.8205 ve 0.807 ve 2014 yılı için 0.8162 ve 0.715 AUC skoru hesaplanmıştır. Qian vd. (2022) çalışmalarında 2001-2019 yılları arasını kapsayan ve 4167 gözlem içeren veri seti ile gözlemin riski kabul edildiği yıldan bir yıl önceki veriler dikkate alınarak eğitilmiş olan rastgele orman modelinin doğruluk metriği 0.8542, duyarlılığı 0.8397, kesinliği 0.8647 ve AUC skoru 0.9291 olarak hesaplanmıştır. Garcia (2022) tarafından Ocak 2010 ile Aralık 2018 dönemin 1824 işletmeye ait 41933 çeyrek dönemlik mali raporları kapsayan veri seti ile eğitilmiş olan XGBoost modelinin duyarlılığı 0.571, seçiciliği 0.989 ve AUC skoru 0.780 olarak tespit edilmiştir. Z. Zhang vd. (2022) 2007 ile 2020 yılları arasında 2260 gözlemi kapsayan veri seti ile eğitilen XGBoost modelinin doğruluk metriği 0.835, duyarlılık metriği 0.808, kesinlik metriği 0.853 ve AUC skoru 0.913' tür. Literatürde yer alan farklı çalışmalarda eğitilen XGBoost modellerinin hesaplanan değerlendirme kriterleri ile yürütülen

alıřmada yer alan modelin deęerleme metrikleri karřılařtırıldıęında, alıřmada eęitilen modelin daha iyi bir tahmin performansına sahip olduęu grlmektedir.



## SONUÇ VE ÖNERİLER

Denetim, iktisadi eylemler ya da olaylar hakkındaki iddialar ile önceden belirlenmiş kriterler arasındaki uygunluk derecesini belirlemek maksadı ile bağımsız olarak bu iddialara ilişkin kanıtları nesnel bir şekilde elde etmek ve değerlendirmek ve sonuçları ilgililer iletmeyi içeren sistematik bir süreçtir. Muhasebenin temel kavramlarından birisi olan işletmenin sürekliliği esası, işletmenin sınırsız bir ömre sahip olduğunu ve öngörülebilir bir gelecekte faaliyetlerini sürdüreceği varsayımını ifade etmektedir. Uluslararası denetim standartları, işletme yönetimi ve denetçiyi işletmenin sürekliliğini değerlendirilmesi hususunda sorumlu tutmaktadır. Yapay zekâ teknolojileri işletmeciliğin her alanına etki ettiği gibi denetimi de etki etmiştir. Bu bağlamda yapay zekâ teknolojileri etkinliği artırmak maksadı ile denetim sürecindeki faaliyetler bağlamında değerlendirilmiş ve farklı modeller geliştirilmiştir. Lojistik regresyon, diskriminant analizi, bayes sınıflayıcıları, karar ağaçları, destek vektör makineleri, yapay sinir ağları, rastgele orman algoritması, adaboost, gradient boosting ve extreme gradient boosting gibi gözetimli öğrenme metotlarının yanında self-organizing maps ve k-ortalamlar gibi gözetimsiz makine öğrenmesi metotları da denetimde yapay zekâ ve makine öğrenmesi uygulamalarında yaygın olarak kullanılmaktadır.

İşletmenin sürekliliği tahmini amacı ile yürütülmüş ilk çalışmalarda geleneksel istatistiksel metotlardan faydalanılmıştır (Chye Koh ve Moren Brown 1991; Dopuch, Holthausen, ve Leftwich 1987). Beaver (1966) tarafından yürütülen tek değişkenli istatistiksel yöntemlerden ve E. Altman (1968) tarafında çoklu diskriminant analizden faydalanılan çalışmalar alanda öncü niteliği taşımaktadır. Beaver (1966) ve E. Altman (1968) tarafından yürütülen çalışmaları takip eden yıllarda diskriminant analizi, lojistik regresyon ve probit analiz gibi geleneksel istatistiksel metotlardan faydalanan pek çok araştırmacı literatüre katkı sağlamıştır (Barnes 2004; Kida 1980; E. Altman 1968; Ohlson 1980; Zmijewski 1984). Ancak basit bir yapıya ve daha az parametreye sahip olmalarına ve daha az eğitim süresine ihtiyaç duymalarına karşın normallik, bağımsız değişkenler arasında eşit kovaryans bulunması ve bağımsız değişkenler arasında çoklu bağlanım bulunmaması gibi katı ve kısıtlayıcı varsayımlara sahip olmaları geleneksel metotları sınırlamaktadır. Ayrıca mali nitelikte veriler ile çalışılması sayılan varsayımların sıklıkla ihlal edilmesine neden olmaktadır (J. Sun vd. 2020; Yeh, Chi, ve Lin 2014b). Geleneksel istatistiksel yöntemlerin sahip olduğu katı varsayımlar sebebi ile makine öğrenmesi metotları süreklilik tahmini literatüründe uygulanmaya başlanmıştır (Z. Zhang vd. 2022).

Karar ağaçları (Chye Koh ve Kee Low 2004a; Gepp, Kumar, ve Bhattacharya 2010a; M.-Y. Chen 2011a; Yakut ve Elmas 2013) ve destek vektör makineleri (Hua vd. 2007; Ding, Song, ve Zen 2008; Yeh, Chi, ve Hsu 2010; S. Y. Kim 2011a; M.-Y. Chen 2012; Bae 2012; Hsieh, Hsiao, ve Yeh 2012; S.-C. Huang vd. 2012; Gordini 2014; Goo, Chi, ve Shen 2016; Fallahpour, Lakvan, ve Zadeh 2017a) ile topluluk öğrenmesi metotları olan rastgele orman algoritması (Barboza, Kimura, ve Altman 2017; Terzi 2022; Garcia 2022), adaboost (Alfaro vd. 2008; Heo ve Yang 2014), gradient boosting (S. Jones 2017; E. I. Altman vd. 2020; X. Du vd. 2020; Jabeur vd. 2021) ve XGBoost (Ziëba, Tomczak, ve Tomczak 2016; Y.-P. Huang ve Yen 2019) 1990' lı yıllardan başlayarak farklı arařtırmacılar tarafından yürütölen çalıřmalarda iřletmenin süreklilik riskinin tahminini amaçlayan modellerde kullanılmıřtır.

İřletmenin süreklilięini tahmin etmeyi amaçlayan modellerin temel hedefi bařta iřletme yönetimi ve denetçi olmak üzere tüm iřletme paydařlarının iřletmenin süreklilik kabiliyetini test etmek maksadı ile yaralanabileceęi modeller geliřtirmektir. Bu amaçla makine öğrenmesi metotlarından literatürde yaygın olarak faydalanılmıřtır. Farklı ölkelerin veri setleri ile yürütölmüř olan çalıřmalarda makine öğrenmesi metotlarının süreklilik tahminindeki tutarlılıkları ortaya konulmuřtur. Bu çalıřmada Borsa İstanbul A.ř. pay piyasası imalat ve alt sektörlerin iřlem görmekte olan toplam 201 iřletmeye ait ve 2017, 2018, 2019, 2020 ve 2021 yıllarına ait mali tablolar incelenmiřtir. İřletmelere ait gözlem yıllarının etiketlenmesi amacıyla iřletmenin süreklilik görüřü alıp almamıř olması dikkate alınmıřtır. İřletmenin ilgili yılda süreklilik görüřü almıř olması durumunda gözlem yılı "1" olarak etiketlenmiřtir. Süreklilik görüřü almamıř olan gözlem yılları "0" olarak etiketlenmiřtir. Çalıřmada kullanılan veri setinde 16 iřletmeye ait 43 süreklilik görüřü almıř gözlem yılı ve 198 iřletmeye ait 788 gözlem yılı yer almaktadır. Çalıřmada kapsamlı bir literatür taraması ile süreklilik tahmininde yaygın olarak kullanılan 75 mali oran seçilmiř ve ilk veri seti oluřturulmuřtur.

Veri setinin oluřturulması sonrasında eksik gözlemler özniteliklerin ortalamaları kullanılarak tamamlanmıř ve veri seti min-max normalizasyon metodu kullanılarak normalize edilmiřtir. Çalıřmada kullanılan veri seti dengesiz bir yapıya sahiptir. Dengesiz veri setinin model tahmin bařarısı üzerinde yaratacaęı olumsuz etkinin önüne geçmek maksadı bir oversampling metodu olan SMOTE' tan faydalanılmıřtır. Veri setinin dengelenmesi sonrasında "1" sınıfında 814 ve "0" sınıfında 774 gözlem yılı yer almıřtır. "1" sınıfında yer alan gözlemler süreklilik riski tařımaktadır. "0" sınıfında yer alan

gözlemler süreklilik riski taşımamaktadır. Eğitilen modellerin tahmin başarısının artırılması maksadı ile genetik algoritmalar ile öznitelik seçimi yapılmıştır. Öznitelik seçimi sonrasında ilk veri setinde yer alan 75 öznitelikten 58 adedi seçilmiştir. Genetik algoritmalar ile yürütülen seçim aşamasında her bir iterasyonda farklı öznitelik kombinasyonlarının elde ettiği tahmin başarısı rastgele orman algoritması ile ölçülmüştür.

Çalışmada lojistik regresyon, destek vektör makineleri (doğrusal, polinomsal, radyal tabanlı kernel fonksiyonları), sınıflama ve regresyon ağacı, rastgele orman algoritması, adaboost, gradient boosting ve XGBoost metotları ile toplam dokuz farklı model eğitilmiştir. Eğitimde referans (pozitif sınıf) sınıf "1" olarak tanımlanmış, model değerlendirme aşamasında bu kriter dikkate alınarak karmaşıklık matrisi oluşturulmuş ve değerlendirme metrikleri hesaplanmıştır. Eğitim öncesinde veri seti 80-20 kuralına göre rastgele ayrılmıştır. Veri setinden ayrılan %80 'lik veri eğitim maksadı ile %20 'lik veri test maksadı ile kullanılmıştır. Modellerin eğitiminde 10 katmanlı çapraz doğrulama prosedürüne kullanılmıştır.

### **Bulguların Özeti**

Modellerin değerlendirilmesi maksadı ile doğruluk, duyarlılık, seçicilik, tip-I hata, tip- II hata, kesinlik ve AUC metriklerinden faydalanılmıştır. Eğitilen modeller hesaplanan doğruluk metrikleri açısından karşılaştırıldığında en başarısı sonucu test veri setindeki tüm gözlemleri doğru sınıflayan rastgele orman modeli (%100) elde etmiştir.

Rastgele orman modelini sırası ile radyal tabanlı destek vektör makineleri (%99.36) ve adaboost (%99.36), polinomsal destek vektör makineleri (%98.73), gradient boosting (%98.42) ve XGBoost (%98.42), doğrusal destek vektör makineleri (%97.48), sınıflama ve regresyon ağacı (%97.16) ve lojistik regresyon (%92.43) izlemektedir. Modellerin doğruluk metriği dikkate alınarak topluluk öğrenmesi metotları ve temel sınıflayıcılar arasında yapılan karşılaştırmada, topluluk öğrenmesi modelleri ve destek vektör makinesi modelleri arasında büyük farklılıkları bulunmadığı görülmektedir. Lojistik regresyon modeli doğruluk metriği açısından en başarısız modeldir.

Modeller hesaplanan duyarlılık ve seçicilik metrikleri açısından karşılaştırıldığında en başarılı model rastgele orman modelidir. Model hem süreklilik riski taşıyan gözlemleri hem de süreklilik riski taşımayan gözlemleri sınıflandırmada mükemmel yatınlığa sahiptir. Duyarlılık metriği açısından rastgele orman modeli (1.0000) ve doğrusal destek vektör makineleri (1.0000), adaboost (1.0000) mükemmel tahmin yatınlığına sahiptir.

Modelleri sırası ile polinomsal destek vektör makineleri (0.9938), radyal tabanlı destek vektör makineleri (0.9877), gradient boosting (0.9877) ve XGBoost (0.9877), sınıflama ve regresyon ağacı (0.9693) ve lojistik regresyon (0.9018) izlemektedir. Duyarlılık metriği açısından modeller karşılaştırıldığında topluluk modelleri ve destek vektör makinesi modellerinin arasında büyük farklılıklar olmadığı görülmektedir. Lojistik regresyon modeli süreklilik riski taşıyan gözlemlerin sınıflandırılmasında en düşük yatkinlığa sahiptir. Rastgele orman modeli (1.000) ve radyal tabanlı destek vektör makinesi modeli (1.000) süreklilik riski taşımayan gözlemlerin sınıflandırılmasında mükemmel yatkinlığa sahiptir. Modelleri sırası ile adaboost (0.9870), gradient boosting (0.9806) ve XGBoost (0.9806), polinomsal destek vektör makineleri (0.9805), sınıflama ve regresyon ağacı (0.9740), doğrusal destek vektör makinesi (0.9480) ve lojistik regresyon izlemektedir. Seçicilik metriği açısından topluluk öğrenmesi ve temel sınıflayıcılar karşılaştırıldığında polinomsal ve radyal tabanlı modellerin topluluk modelleri ile yakın sonuçlar ortaya koyduğu görülmektedir. Lojistik regresyon seçicilik metriği açısından en başarısız modeldir.

Rastgele orman modeli (1.000) ve radyal tabanlı destek vektör makinesi modeli (1.000) referans sınıfa yapılan tahminlerde mükemmel sonuçlar ortaya koymuştur. Modelleri sırası ile adaboost (0.9878), polinomsal destek vektör makinesi (0.9818), gradient boosting (0.9817) ve XGBoost (0.9817), sınıflama ve regresyon ağaçları (0.9753), doğrusal destek vektör makinesi (0.9532) ve lojistik regresyon (0.9483) izlemektedir. Kesinlik metriği açısından topluluk öğrenmesi modelleri ile radyal tabanlı DVM modeli ve polinomsal DVM arasında büyük bir fark bulunmamaktadır. Lojistik regresyon kesinlik metriği açısından en başarısız modeldir.

AUC açısından modeller değerlendirdiğinde rastgele orman, adaboost, XGBoost ve radyal tabanlı destek vektör makinelerinin sınıflar arasında doğru ayırım yapma ihtimalinin %100 olduğu görülmektedir. Modelleri sırası ile gradient boosting (0.9990), sınıflama ve regresyon ağacı (0.9840) polinomsal destek vektör makineleri (0.9820), doğrusal destek vektör makineleri (0.9800) ve lojistik regresyon (0.9250) takip etmektedir. AUC skorları açısından modeller karşılaştırıldığında lojistik regresyon dışında tüm modellerin sınıflar arasında ayırım yapma ihtimallerinin %98' in üzerinde olduğu görülmektedir.

### **Bulguların Yorumlanması**

### *Topluluk Öğrenmesi Modellerine ve Temel Sınıflayıcı Modellerine ait Sonuçların Karşılaştırılması*

İşletmenin sürekliliğinin tahmini bağlamında topluluk öğrenmesi metotları ve temel sınıflayıcılar karşılaştırıldığında çeşitli nedenlerden dolayı topluluk öğrenmesi metotlarının temel sınıflayıcılara üstünlük kurmaktadır. Topluluk öğrenmesi metotları birden fazla temel sınıflayıcıya ait çıktıyı birleştirmektedir. Modeller yaygın olarak temel sınıflayıcılardan daha yüksek tahmin başarısı ortaya koymaktadır (Alfaro vd. 2008; Heo ve Yang 2014; Jabeur vd. 2021; Zięba, Tomczak, ve Tomczak 2016; Z. Zhang vd. 2022; Volkov, Benoit, ve Van den Poel 2017; Qian vd. 2022). Topluluk öğrenmesi metotlarının sağladığı bir fayda modellerdeki varyansı azaltmalardır (Valentini ve Dietterich 2004; Anwar, Qamar, ve Muzaffar Qureshi 2014). Bu suretle modellerdeki aşırı uyum riski de azalmaktadır. Süreklilik tahmini literatüründe faydalanılan mali nitelikteki veriler yaygın olarak gürültü ve aykırı gözlem içermektedir. Topluluk öğrenmesi metotları gürültüye karşı daha dayanıklıdır. Sayılan avantajlarının aksine birden fazla temel sınıflayıcıya ait tahmini birleştirmeleri sebebiyle metotlar yerel değerlere karşı daha az duyarlı bir yapıdadır. Çalışmanın ampirik sonuçları incelendiğinde topluluk öğrenmesi modellerinin, radyal tabanlı DVM modeli dışında, temel sınıflayıcılara karşı daha başarılı sonuçlar ortaya koyduğu görülmektedir. Modellerin ortaya koyduğu sonuçlar literatür ile de benzerlik göstermektedir.

### *Topluluk Öğrenmesi Modellerine ait Sonuçların Karşılaştırılması*

Çalışmada eğitilen dört topluluk öğrenmesi modeline ait sonuçlar incelendiğinde, modeller arasından rastgele orman modelinin hesaplanan değerlendirme metriklerine göre en başarılı model olduğu görülmektedir. Rastgele orman algoritmasının ortaya koyduğu sonuçlar da literatür ile de benzerlik göstermektedir (Volkov, Benoit, ve Van den Poel 2017; Jabeur vd. 2021; Garcia 2022). Rastgele orman algoritmasının, diğer topluluk öğrenmesi metotlarından üstün olup olmadığı araştırmanın yapısına ve veri kümesine bağlı olan bir olgudur. Bir başka ifade rastgele orman algoritması belirli durumlarda diğer topluluk öğrenmesi metotlarına üstünlük sağlayabilir. Her metodun kendine özgü avantajları ve dezavantajları bulunmaktadır. Bunun dışında rastgele orman algoritmasının süreklilik tahmini çalışmalarında daha avantajlı bir konumda bulunmasını sağlayan belli başlı özellikleri bulunmaktadır. İlk olarak algoritma diğer topluluk öğrenmesi metotları gibi verideki gürültüye karşı daha az hassaslık göstermektedir. Daha önce belirtildiği gibi, mali nitelikteki veriler yaygın olarak gürültü ve aykırı gözlem içermektedir. Bir topluluk

öğrenmesi metodu olan rastgele orman algoritmasının, diğer topluluk öğrenmesi metotları gibi model varyansını düşürme yeteneği bulunmakta ve bu avantaj algoritmasının aşırı uyum riskini azaltmasını sağlamaktadır. İstatistiksel metotlardan faydalanılan çalışmalara getirilen en önemli eleştiri metotların katı varsayımlara sahip olması ve çalışmalarda kullanılmış olan veri setlerinde bu varsayımların ihlal edilmiş olmasıdır. Verinin doğrusal olmayışı da bu varsayım ihlallerinden birisidir. Rastgele orman algoritması doğrusal olmayan ilişkileri tespit edebilme yeteneğine sahiptir. Algoritmanın bu özelliği mali veriler ile daha başarılı tahmin sonuçları elde etmenin önünü açan bir unsur olarak sayılabilir.

#### *Temel Sınıflayıcı Modellerine ait Sonuçların Karşılaştırılması*

Çalışmada eğitilmiş olan temel sınıflayıcılar bağlamında bir karşılaştırma yapıldığında en yüksek tahmin başarısına sahip olan modelin radyal tabanlı destek vektör makineleri modeli olduğu görülmektedir. Modeli sırası ile polinomsal DVM, doğrusal DVM, CART ve lojistik regresyon izlemektedir. Radyal tabanlı model, topluluk öğrenmesi modelleri ile de yakın tahmin sonuçlarına sahiptir. Destek vektör makinelerinin özellikle mali veriler ile başarı olduğu literatürde belirtilmektedir (J. Sun ve Li 2012). Destek vektör makinelerinin farklı temel sınıflayıcılara kıyasla daha başarılı sonuçlar ortaya koyduğu çalışmalar da literatürde yer almaktadır (Zhou, Lai, ve Yen 2014; Xiao vd. 2012; G. Wang, Ma, ve Yang 2014; Veganzones ve Séverin 2018; C.-F. Tsai, Hsu, ve Yen 2014). Destek vektör makineleri sahip oldukları belirli nitelikler sebebi ile mali veriler ile daha iyi performans ortaya koyma yeteneğine sahiptir. İlk olarak radyal tabanlı ve polinomsal kernel fonksiyonu, doğrusal olamayan veri kümeleri etkili bir şekilde modelleyebilme yeteneğine sahiptir ve destek vektör makinelerinin bu özelliği mali veriler ile daha başarılı sonuçlar ortaya çıkmasına olanak tanımaktadır. Metotlar margin maksimizasyonu prensibi uygulamaktadır. Bir başka ifade ile karar sınırını ya da sınıflar arasında sınırları maksimize etmeyi amaçlamaktadır. Destek vektör makinelerinin bu özelliği sınıflar arasında daha iyi ayırım yapılmasına olanak tanımaktadır. Metodun gürültüye karşı daha az duyarlı oluşu da yaygın olarak gürültü ve aykırı gözlem içeren mali veriler ile başarılı sonuçlar elde edilmesine katkı sağlayabilmektedir.

#### **Araştırmanın Kısıtları**

Araştırmanın bulguları, araştırmanın kısıtları bağlamında yorumlandığında;

- Çalışmada, finansal kurumlar mali yapılarının farklı olması sebebi ile veri seti dışarısında tutulmuş ve sadece imalat işletmelerinin yer aldığı bir veri seti derlenmiştir. Literatür incelendiğinde alanda yürütülmüş olan çalışmaların büyük yüzdesinin imalat işletmelerini kapsadığı görülmektedir. Ancak bankaların, sigorta şirketlerinin ve diğer finans kurumlar ile hizmet işletmelerinin süreklilik kabiliyetinin değerlendirilmesi maksadı ile modellerin geliştirilmesi literatüre büyük katkı sağlayacaktır.

- Çalışmada etiketleme yapmak maksadı ile yasal bir yaklaşım tercih edilmiş ve denetim raporlarından yola çıkılarak gözlemler etiketlenmiştir. Literatür incelendiğinde süreklilik riski taşıyan gözlemler ile risksiz gözlemleri birbirinden ayırt etmek için kabul edilmiş bütünlük bir teorinin yer almadığı görülmektedir. Çalışmanın ampirik sonuçları makine öğrenmesi modellerinin süreklilik tahmininde etkin olduğunu ortaya koysa da çalışmada benimsenen yaklaşım denetçilerin kararları üzerine inşa edilmiştir. Bu yaklaşım denetçinin mesleki tecrübesinin modele dolaysız bir şekilde aktarmaya olanak tanımaktadır. Ancak, işletmelerin etiketlenmesi maksadı ile daha sağlam bir teorinin gerekliliği bulunmaktadır.

- Çalışmada parametre optimizasyonu gerçekleştirilmiş ve her bir model için en uygun parametreler seçilmiştir. Ancak yürütülen kapsamlı parametre optimizasyonu süreci sonucu seçilen parametrelerin her durum ve koşulda mükemmel sonucu vereceği garanti edilmemektedir.

- Çalışmada ilk olarak belirlenen 75 adet mali oran içerisinde bir sarmalama öznelik seçimi metodu olan genetik algoritmalar ile öznelik seçimi yapılmıştır. Seçilen özneliklerin her durumda aynı modeller ile aynı tahmin sonuçlarını ortaya koyması mümkün değildir. Makine öğrenmesi metotları veriye bağımlıdır ve veri kalitesi modellerin tahmin başarısı üzerinde önemli bir etkiye sahiptir.

### **Gelecek Çalışmalar Öneriler**

Araştırmanın kısıtları ve bulguları bağlamında gelecek araştırmalar için aşağıdaki öneriler sunulabilir;

- Süreklilik tahmini araştırmalarında, bütünlük bir teorinin bulunmayışı, karşılaştırma yapmanın önüne geçmektedir. Ayrıca hangi durumlarda süreklilik riskinden söz edilebileceği hakkında genel kabul görmüş bir tanımın bulunması denetçinin kararını da destekleyecek bir unsurdur. Bu bağlamda bütünlük bir teorinin oluşturulması ve süreklilik riskinin hangi durumlarda ortaya çıkacağı hakkında genel bir tanımın ortaya koyulması alana önemli katkı sağlayacaktır.

- Yapısal veriler yaygın olarak süreklilik tahmini çalışmalarında kullanılmaktadır. Ancak yapısal olmayan veriler, yapısal verilere kıyasla daha fazla bilgi içermektedir. Bu bağlamda yapısal veriler ile birlikte yapısal olmayan verilerin ya da yalnızca yapısal olmayan verilerin, metin madenciliği ve doğal dil işleme teknikleri yardımı ile modellere aktarılmasının, modellerin tahmin başarısı üzerinde olumlu etkiye sahip olacağı literatürde belirtilmektedir. Bu bağlamda gelecekte yürütülecek çalışmalarda yapısal olmayan verilere yer verilmesi önerilmektedir.

- Süreklilik tahmini çalışmalarında faydalanılan mali oranlar ve sayıları literatürde farklılaşmaktadır. Ancak genel kabul, çalışmalarda yaygın olarak kullanılan mali oranlardan faydalanılması yönündedir. Bu bağlamda pek çok araştırmacı araştırmalarında yer verdikleri oranları detaylı ve kapsamlı literatür taraması yardımı ile belirlemiştir. Literatürde faydalanılacak olan mali oranlar hakkında genel bir kabulün bulunmayışı, karşılaştırma yapmanın önüne geçmektedir. Ayrıca veri setinin derlenmesinin ve modellemenin maliyetlerini artırabilmektedir. Bu bağlamda gelecek çalışmalarda örüntü tanımlama metotlarının süreklilik tahmininde kullanımı araştırma konusu olarak dikkate alınabilir.

Araştırma bulguları bir bütün olarak incelendiğinde çalışma da şu temel sonuca ulaşılmıştır. İşletmenin sürekliliğinin tahmininde, makine öğrenmesi metotlarının, doğru bir süreç tasarımı ile etkin oldukları ve denetim sürecinde işletmenin süreklilik kabiliyetinin değerlendirmesi maksadı ile denetçinin karar sürecini destekleme yeteneğine sahip oldukları sonucuna ulaşılmıştır.

## KAYNAKLAR

- “A Statement of Basic Accounting Concepts”. 1973. American Accounting Association.
- Abbott, Dean. 2014. *Applied predictive analytics: principles and techniques for the professional data analyst*. Indianapolis, IN: Wiley.
- Adnan Aziz, M., ve Humayon A. Dar. 2006. “Predicting Corporate Bankruptcy: Where We Stand?” *Corporate Governance: The International Journal of Business in Society* 6 (1): 18-33. <https://doi.org/10.1108/14720700610649436>.
- Agustia, Dian, Nur Pratama Abdi Muhammad, ve Yani Permatasari. 2020. “Earnings Management, Business Strategy, and Bankruptcy Risk: Evidence from Indonesia”. *Heliyon* 6 (2): e03317. <https://doi.org/10.1016/j.heliyon.2020.e03317>.
- Ahmed, Mohiuddin, ve Al-Sakib Khan Pathan, ed. 2019. *Data analytics: concepts, techniques and applications*. Boca Raton: CRC Press, Taylor & Francis Group.
- Ahn, Hyunchul, ve Kyoung-jae Kim. 2009. “Bankruptcy Prediction Modeling with Hybrid Case-Based Reasoning and Genetic Algorithms Approach”. *Applied Soft Computing* 9 (2): 599-607. <https://doi.org/10.1016/j.asoc.2008.08.002>.
- Akbiyık, Sedat. 2005. *Vergi uygulamaları yönüyle: denetim ve raporlama*. Bursa: Ekin Kitabevi.
- Akdoğan, Nalan. 2020. “Borsa İstanbul (BİST) Şirketlerinin 2017 ve 2018 Yıllarındaki Denetim Görüşleri ile Bağımsız Denetçi Raporlarında Verilen Görüşün Dayanağının İncelenmesine Yönelik Bir Araştırma”. *Muhasebe ve Denetime Bakış* 19 (59): 1-18. <https://dergipark.org.tr/en/pub/mdbakis/issue/63886/967122>.
- Aksela, Matti, ve Jorma Laaksonen. 2006. “Using Diversity of Errors for Selecting Members of a Committee Classifier”. *Pattern Recognition* 39 (4): 608-23. <https://doi.org/10.1016/j.patcog.2005.08.017>.
- Aksoy, Tamer. 2005. “Yeni Türk Ticaret Kanunu Tasarısı Işığında Denetim ve Muhasebe Yasal Düzenleme Çerçevesine İlişkin Analitik Bir Değerlendirme” Eylül (16): 30.
- Aktaş, Ramazan, Mete Doğanay, ve Birol Yıldız. 2003. “Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırılması”. *Ankara Üniversitesi SBF Dergisi* 58 (4): 1. [https://doi.org/10.1501/SBFder\\_0000001691](https://doi.org/10.1501/SBFder_0000001691).
- Alam, Nurul, Junbin Gao, ve Stewart Jones. 2021. “Corporate Failure Prediction: An Evaluation of Deep Learning vs Discrete Hazard Models”. *Journal of*

- International Financial Markets, Institutions and Money* 75 (Kasım): 101455.  
<https://doi.org/10.1016/j.intfin.2021.101455>.
- Alfaro, Esteban, Matías Gámez, ve Noelia García. 2013. “**Adabag**: An R Package for Classification with Boosting and Bagging”. *Journal of Statistical Software* 54 (2).  
<https://doi.org/10.18637/jss.v054.i02>.
- Alfaro, Esteban, Noelia García, Matías Gámez, ve David Elizondo. 2008. “Bankruptcy Forecasting: An Empirical Comparison of AdaBoost and Neural Networks”. *Decision Support Systems* 45 (1): 110-22.  
<https://doi.org/10.1016/j.dss.2007.12.002>.
- Altman, Edward. 1968. “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”. *The Journal of Finance* 23 (4): 589-609.  
[https://www.jstor.org/stable/2978933?casa\\_token=Xcd5kk07FlgAAAAA%3AsdOqN4rLqOOkO2PQPchbBFQZiHomEEsATcAQdWWAgkNoGH4Pkl9did74CPMnCefih9XuUxQeV9TkXOX-Mt7vAJsYP9sMaSr\\_yBA\\_1sn1kJnhE7qedAZ5EA&seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/2978933?casa_token=Xcd5kk07FlgAAAAA%3AsdOqN4rLqOOkO2PQPchbBFQZiHomEEsATcAQdWWAgkNoGH4Pkl9did74CPMnCefih9XuUxQeV9TkXOX-Mt7vAJsYP9sMaSr_yBA_1sn1kJnhE7qedAZ5EA&seq=1#metadata_info_tab_contents).
- . 1982. “Accounting Implications of Failure Prediction Models”. *Journal of Accounting, Auditing & Finance* 6: 4-31.
- Altman, Edward I, ve Edith Hotchkiss. 2006. *Corporate Financial Distress and Bankruptcy Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt*. Hoboken, N.J.: Wiley. <http://www.books24x7.com/marc.asp?bookid=13995>.
- Altman, Edward I., Małgorzata Iwanicz-Drozdowska, Erkki K. Laitinen, ve Arto Suvas. 2020. “A Race for Long Horizon Bankruptcy Prediction”. *Applied Economics* 52 (37): 4092-4111. <https://doi.org/10.1080/00036846.2020.1730762>.
- Altman, Edward, ve T. P. McGough. 1974. “Evaluation of a Company as a Going Concern”. *Journal of Accountancy*, sy December 1974: 59-67.
- Amani, Farzaneh A., ve Adam M. Fadlalla. 2017. “Data Mining Applications in Accounting: A Review of the Literature and Organizing Framework”. *International Journal of Accounting Information Systems* 24 (Şubat): 32-58.  
<https://doi.org/10.1016/j.accinf.2016.12.004>.
- Anandarajan, M, ve A Anandarajan. 1999. “A Comparison of Machine Learning Techniques with a Qualitative Response Model for Auditor’s Going Concern

- Reporting”. *Expert Systems with Applications* 16 (4): 385-92.  
[https://doi.org/10.1016/S0957-4174\(99\)00014-7](https://doi.org/10.1016/S0957-4174(99)00014-7).
- Anandarajan, Murugan, Picheng Lee, ve Asokan Anandarajan. 2001. “Bankruptcy Prediction of Financially Stressed Firms: An Examination of the Predictive Accuracy of Artificial Neural Networks”. *International Journal of Intelligent Systems in Accounting, Finance & Management* 10 (2): 69-81.  
<https://doi.org/10.1002/isaf.199>.
- Anwar, Hina, Usman Qamar, ve Abdul Wahab Muzaffar Qureshi. 2014. “Global Optimization Ensemble Model for Classification Methods”. *The Scientific World Journal* 2014: 1-9. <https://doi.org/10.1155/2014/313164>.
- Appiah, Kingsley Opoku, Amon Chizema, ve Joseph Arthur. 2015. “Predicting Corporate Failure: A Systematic Literature Review of Methodological Issues”. *International Journal of Law and Management* 57 (5): 461-85. <https://doi.org/10.1108/IJLMA-04-2014-0032>.
- Arens, Alvin A., Randal J. Elder, ve Mark S. Beasley. 2012. *Auditing and Assurance Services: An Integrated Approach*. 14th ed. Boston: Prentice Hall.
- Arruñada, Benito. 1999. *The Economics of Audit Quality*. Boston, MA: Springer US.  
<https://doi.org/10.1007/978-1-4757-6728-5>.
- Ataman, Başak. 2010. “Türkiye’ de Kamu Denetimi ve Kamu Denetçilerine Genel Bakış”. *Maliye ve Finans Yazıları* 1 (87): 17-26.  
<https://dergipark.org.tr/tr/pub/mfy/issue/16293/170826>.
- Ataman, Ümit, Rüstem Hacırüstemoğlu, ve Nejat Bozkurt. 2001. *Muhasebe Denetim Uygulamaları*. İstanbul: ALFA Basım Yayım Dağıtım San. ve Tic. Ltd Şti.
- Awad, Mariette, ve Rahul Khanna. 2015. *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. The expert’s voice in machine learning. Apress Open.
- Bae, Jae Kwon. 2012. “Predicting Financial Distress of the South Korean Manufacturing Industries”. *Expert Systems with Applications* 39 (10): 9159-65.  
<https://doi.org/10.1016/j.eswa.2012.02.058>.
- Balcaen, Sofie, ve Hubert Ooghe. 2006. “35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems”. *The British Accounting Review* 38 (1): 63-93.  
<https://doi.org/10.1016/j.bar.2005.09.001>.

- Baldwin, Amelia A., Carol E. Brown, ve Brad S. Trinkle. 2006. "Opportunities for Artificial Intelligence Development in the Accounting Domain: The Case for Auditing". *Intelligent Systems in Accounting, Finance and Management* 14 (3): 77-86. <https://doi.org/10.1002/isaf.277>.
- Ball, Ray. 2006. "International Financial Reporting Standards (IFRS): Pros and Cons for Investors". *Accounting and Business Research* 36 (sup1): 5-27. <https://doi.org/10.1080/00014788.2006.9730040>.
- Barboza, Flavio, Herbert Kimura, ve Edward Altman. 2017. "Machine Learning Models and Bankruptcy Prediction". *Expert Systems with Applications* 83 (Ekim): 405-17. <https://doi.org/10.1016/j.eswa.2017.04.006>.
- Barnes, Paul. 2004. "The Auditor's Going Concern Decision and Types I and II Errors: The Coase Theorem, Transaction Costs, Bargaining Power and Attempts to Mislead". *Journal of Accounting and Public Policy* 23 (6): 415-40. <https://doi.org/10.1016/j.jaccpubpol.2004.10.003>.
- Barth, Mary E., Wayne R. Landsman, ve Mark H. Lang. 2008. "International Accounting Standards and Accounting Quality". *Journal of Accounting Research* 46 (3): 467-98. <https://doi.org/10.1111/j.1475-679X.2008.00287.x>.
- Başpınar, Ahmet. 2005. "Türkiye' de ve Dünyada Denetim Standartlarının Oluşumuna Genel Bir Bakış". *Maliye Dergisi*, sy 148: 35-62.
- Beaver, William H. 1966. "Financial Ratios As Predictors of Failure". *Journal of Accounting Research* 4: 71. <https://doi.org/10.2307/2490171>.
- Beaver, William H., Maureen F. McNichols, ve Jung-Wu Rhie. 2005. "Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy". *Review of Accounting Studies* 10 (1): 93-122. <https://doi.org/10.1007/s11142-004-6341-9>.
- Behn, Bruce K., Steven E. Kaplan, ve Kip R. Krumwiede. 2001. "Further Evidence on the Auditor's Going-Concern Report: The Influence of Management Plans". *AUDITING: A Journal of Practice & Theory* 20 (1): 13-28. <https://doi.org/10.2308/aud.2001.20.1.13>.
- Bell, Jason. 2015. *Machine learning: hands-on for developers and technical professionals*. Indianapolis, Indiana: Wiley.

- Bell, Timothy B., ve Richard H. Tabor. 1991. "Empirical Analysis of Audit Uncertainty Qualifications". *Journal of Accounting Research* 29 (2): 350. <https://doi.org/10.2307/2491053>.
- Bellovary, Jodi L., Don E. Giacomino, ve Michael D. Akers. 2007. "A Review of Bankruptcy Prediction Studies: 1930 to Present". *Journal of Financial Education*, sy 33: 1-42. <http://www.jstor.org/stable/41948574>.
- Bilir, Hakan. 2015. "Finansal Sıkıntının Tanımı ve Piyasa Odaklı Çözümleri: Borç Yapılandırma, Varlık Satışı ve Yeni Sermaye Enjeksiyonu". *Sosyoekonomi* 2015 (1): 11-24.
- Blay, Allen D., Marshall A. Geiger, ve David S. North. 2011. "The Auditor's Going-Concern Opinion as a Communication of Risk". *AUDITING: A Journal of Practice & Theory* 30 (2): 77-102. <https://doi.org/10.2308/ajpt-50002>.
- Bose, Indranil. 2006. "Deciding the Financial Health of Dot-Coms Using Rough Sets". *Information & Management* 43 (7): 835-46. <https://doi.org/10.1016/j.im.2006.08.001>.
- Bose, Indranil, ve Raktim Pal. 2006. "Predicting the Survival or Failure of Click-and-Mortar Corporations: A Knowledge Discovery Approach". *European Journal of Operational Research* 174 (2): 959-82. <https://doi.org/10.1016/j.ejor.2005.05.009>.
- Bozkurt, Nejat. 2018. *Muhasebe Denetimi*. İstanbul: ALFA Basım Yayım Dağıtım San. ve Tic. Ltd Şti.
- Brabazon, Anthony, ve Michael O'Neill. 2004. "Diagnosing Corporate Stability Using Grammatical Evolution". *International Journal of Applied Mathematics and Computer Science* 14 (3): 363-74. <http://eudml.org/doc/207703>.
- Breiman, Leo, ed. 1998. *Classification and Regression Trees*. 1. CRC Press repr. Boca Raton, Fla.: Chapman & Hall/CRC.
- . 2001. "Random Forests". *Machine Learning* 45 (1): 5-32. <https://doi.org/10.1023/A:1010933404324>.
- Bunyaminu, Alhassan, ve Mohammed Issah. 2012. "Predicting Corporate Failure of UK's Listed Companies: Comparing Multiple Discriminant Analysis and Logistic Regression", sy 94: 18.

- Buuren, Stef Van, ve Karin Groothuis-Oudshoorn. 2011. “**Mice** : Multivariate Imputation by Chained Equations in *R*”. *Journal of Statistical Software* 45 (3). <https://doi.org/10.18637/jss.v045.i03>.
- Carcello, Joseph V., ve Terry L. Neal. 2000. “Audit Committee Composition and Auditor Reporting”. *The Accounting Review* 75 (4): 453-67. <https://doi.org/10.2308/accr.2000.75.4.453>.
- Cecchini, Mark, Haldun Aytug, Gary J. Koehler, ve Praveen Pathak. 2010. “Making Words Work: Using Financial Text as a Predictor of Financial Events”. *Decision Support Systems* 50 (1): 164-75. <https://doi.org/10.1016/j.dss.2010.07.012>.
- Chandra, D. Karthik, V. Ravi, ve I. Bose. 2009. “Failure Prediction of Dotcom Companies Using Hybrid Intelligent Techniques”. *Expert Systems with Applications* 36 (3): 4830-37. <https://doi.org/10.1016/j.eswa.2008.05.047>.
- Charalambous, Chris, Andreas Charitou, ve Froso Kaourou. 2000. “Comparative Analysis of Artificial Neural Network Models: Application in Bankruptcy Prediction”. *Annals of Operations Research* 99 (1/4): 403-25. <https://doi.org/10.1023/A:1019292321322>.
- Charitou, Andreas, Evi Neophytou, ve Chris Charalambous. 2004. “Predicting Corporate Failure: Empirical Evidence for the UK”. *European Accounting Review* 13 (3): 465-97. <https://doi.org/10.1080/0963818042000216811>.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, ve W. P. Kegelmeyer. 2002. “SMOTE: Synthetic Minority Over-sampling Technique”. *Journal of Artificial Intelligence Research* 16 (Haziran): 321-57. <https://doi.org/10.1613/jair.953>.
- Chen, Mu-Yen. 2011a. “Predicting Corporate Financial Distress Based on Integration of Decision Tree Classification and Logistic Regression”. *Expert Systems with Applications* 38 (9): 11261-72. <https://doi.org/10.1016/j.eswa.2011.02.173>.
- . 2011b. “Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a Comparison of Evolutionary Computation Approaches”. *Computers & Mathematics with Applications* 62 (12): 4514-24. <https://doi.org/10.1016/j.camwa.2011.10.030>.
- . 2012. “Visualization and Dynamic Evaluation Model of Corporate Financial Structure with Self-Organizing Map and Support Vector Regression”. *Applied Soft Computing* 12 (8): 2274-88. <https://doi.org/10.1016/j.asoc.2012.03.046>.

- Chen, Tianqi, ve Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System". İçinde *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-94. San Francisco California USA: ACM. <https://doi.org/10.1145/2939672.2939785>.
- Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, vd. 2022. "xgboost: Extreme Gradient Boosting". R. <https://CRAN.R-project.org/package=xgboost>.
- Chen, Wei-Sen, ve Yin-Kuan Du. 2009. "Using Neural Networks and Data Mining Techniques for the Financial Distress Prediction Model". *Expert Systems with Applications* 36 (2): 4075-86. <https://doi.org/10.1016/j.eswa.2008.03.020>.
- Cheng, Ching-Hsue, Chia-Pang Chan, ve Yu-Jheng Sheu. 2019. "A Novel Purity-Based k Nearest Neighbors Imputation Method and Its Application in Financial Distress Prediction". *Engineering Applications of Artificial Intelligence* 81 (Mayıs): 283-99. <https://doi.org/10.1016/j.engappai.2019.03.003>.
- Cho, Sungbin, Hyojung Hong, ve Byoung-Chun Ha. 2010. "A Hybrid Approach Based on the Combination of Variable Selection Using Decision Trees and Case-Based Reasoning Using the Mahalanobis Distance: For Bankruptcy Prediction". *Expert Systems with Applications* 37 (4): 3482-88. <https://doi.org/10.1016/j.eswa.2009.10.040>.
- Choi, Hyunchul, Hyojoo Son, ve Changwan Kim. 2018. "Predicting Financial Distress of Contractors in the Construction Industry Using Ensemble Learning". *Expert Systems with Applications* 110 (Kasım): 1-10. <https://doi.org/10.1016/j.eswa.2018.05.026>.
- Chye Koh, Hian, ve Chan Kee Low. 2004a. "Going Concern Prediction Using Data Mining Techniques". *Managerial Auditing Journal* 19 (3): 462-76. <https://doi.org/10.1108/02686900410524436>.
- . 2004b. "Going Concern Prediction Using Data Mining Techniques". *Managerial Auditing Journal* 19 (3): 462-76. <https://doi.org/10.1108/02686900410524436>.
- Chye Koh, Hian, ve Robert Moren Brown. 1991. "Probit Prediction of Going and Non-going Concerns". *Managerial Auditing Journal* 6 (3). <https://doi.org/10.1108/02686909110004914>.

- Cleofas-Sánchez, L., V. García, A.I. Marqués, ve J.S. Sánchez. 2016. "Financial Distress Prediction Using the Hybrid Associative Memory with Translation". *Applied Soft Computing* 44 (Temmuz): 144-52. <https://doi.org/10.1016/j.asoc.2016.04.005>.
- Constand, Richard L., ve Rassoul Yazdipour. 2011. "Firm Failure Prediction Models: A Critique and a Review of Recent Developments". İçinde *Advances in Entrepreneurial Finance*, kitap editörü Rassoul Yazdipour, 185-204. New York, NY: Springer New York. [https://doi.org/10.1007/978-1-4419-7527-0\\_10](https://doi.org/10.1007/978-1-4419-7527-0_10).
- Cömert, Nuran, Seval Selimoğlu Kardeş, Şaban Uzay, ve Süleyman Uyar. 2013. *Uluslararası Denetim Standartları Kapsamında Bağımsız Denetim*. 1. bs. C. 1. Sakarya: Sakarya Üniversitesi Sürekli Eğitim Merkezi.
- Çalgan, Erkan, Erkan Menteşe, Fatma Işıloğlu, M. Vefa Toroslu, Mukaddes Özcan, Nurgül Gülmez Kanpak, Şükrü Savaş, Zeliha Jülide Onur, ve Zekiye Ağca. 2008. *Muhasebe Denetimi*. İstanbul: İstanbul Serbest Muhasebeci Mali Müşavirler Odası (İSMMMO).
- Çalış, Yıldırım Ercan. 2013. *Aracı kurumlarda iç denetim uygulaması*. 1. basım. İstanbul: Yalın.
- Çömlekçe, Ferruh, Münevver Yılandı, Nurten Erdoğan, Saime Önce, Seval Kardeş Selimoğlu, ve Ergun Kaya. 2004. *Muhasebe Denetimi ve Mali Analiz*. 1. bs. Eskişehir: Anadolu Üniversitesi Yayınları.
- Dai, Jun, ve Miklos A. Vasarhelyi. 2016. "Imagining Audit 4.0". *Journal of Emerging Technologies in Accounting* 13 (1): 1-15. <https://doi.org/10.2308/jeta-10494>.
- Dal Pozzolo, Andrea, Olivier Caelen, ve Gianluca Bontempi. 2015. "unbalanced: Racing for Unbalanced Methods Selection". <http://mlg.ulb.ac.be>.
- Dalak, Güldem. 2000. "Denetim ve Kalite Denetimi". *Muğla Üniversitesi SBE Dergisi* 1 (1): 65-79.
- De Bock, Koen W. 2017. "The Best of Two Worlds: Balancing Model Strength and Comprehensibility in Business Failure Prediction Using Spline-Rule Ensembles". *Expert Systems with Applications* 90 (Aralık): 23-39. <https://doi.org/10.1016/j.eswa.2017.07.036>.
- Deakin, Edward B. 1972. "A Discriminant Analysis of Predictors of Business Failure". *Journal of Accounting Research* 10 (1): 167. <https://doi.org/10.2307/2490225>.

- Debreceeny, Roger S. 2011. "Betwixt and Between? Bringing Information Systems and Accounting Systems Research Together". *Journal of Information Systems* 25 (2): 1-9. <https://doi.org/10.2308/isys-10140>.
- DeFond, Mark L., K. Raghunandan, ve K.R. Subramanyam. 2002. "Do Non-Audit Service Fees Impair Auditor Independence? Evidence from Going Concern Audit Opinions". *Journal of Accounting Research* 40 (4): 1247-74. <https://doi.org/10.1111/1475-679X.00088>.
- Demir, Berna. 2010. "Muhasebe Bilgi Sistemlerinde Bilgi Kalitesi". *Muhasebe ve Finansman Dergisi*, sy 48: 142-53. <https://dergipark.org.tr/en/pub/mufad/issue/35627/396017>.
- Demir, Berna. 2015. "Muhasebe ve Denetim Mesleğinde Etik". *Eğitim ve Öğretim Araştırmaları Dergisi* 4 (4): 341-52. <http://www.jret.org/FileUpload/ks281142/File/034.demir.pdf>.
- Ding, Yongsheng, Xinpeng Song, ve Yueming Zen. 2008. "Forecasting Financial Condition of Chinese Listed Companies Based on Support Vector Machine". *Expert Systems with Applications* 34 (4): 3081-89. <https://doi.org/10.1016/j.eswa.2007.06.037>.
- Dopuch, Nicholas, Robert W. Holthausen, ve Richard W. Leftwich. 1987. "Predicting Audit Qualifications with Financial and Market Variables". *The Accounting Review* 62 (3): 431-54. <http://www.jstor.org/stable/247571>.
- Dönmez, Adnan, P. Başak Berberoğlu, ve Ayten Ersoy. 2005. "Ülkemiz Bağımsız Dış Denetim Standartlarının ABD Genel Kabul Görmüş Denetim Standartları-AB Sekizinci Yönergesi ve Uluslararası Denetim Standartlarıyla Karşılaştırılması". *Akdeniz İ.İ.B.F. Dergisi*, sy 9: 52-78. <https://kutuphane.dogus.edu.tr/mvt/pdf.php>.
- Du, Peijun, Junshi Xia, Wei Zhang, Kun Tan, Yi Liu, ve Sicong Liu. 2012. "Multiple Classifier System for Remote Sensing Image Classification: A Review". *Sensors* 12 (4): 4764-92. <https://doi.org/10.3390/s120404764>.
- Du, Xudong, Wei Li, Sumei Ruan, ve Li Li. 2020. "CUS-Heterogeneous Ensemble-Based Financial Distress Prediction for Imbalanced Dataset with Ensemble Feature Selection". *Applied Soft Computing* 97 (Aralık): 106758. <https://doi.org/10.1016/j.asoc.2020.106758>.

- Du nez-Guzm n, Edgar A., ve Michael D. Vose. 2013. “No Free Lunch and Benchmarks”. *Evolutionary Computation* 21 (2): 293-312. [https://doi.org/10.1162/EVCO\\_a\\_00077](https://doi.org/10.1162/EVCO_a_00077).
- El Naqa, Issam, ve Martin J. Murphy. 2015. “What Is Machine Learning?” İinde *Machine Learning in Radiation Oncology*, edit r Issam El Naqa, Ruijiang Li, ve Martin J. Murphy, 3-11. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-18305-3\\_1](https://doi.org/10.1007/978-3-319-18305-3_1).
- Elliott, Robert K. 1992. “The Third Wave Breaks on the Shores of Accounting”. *Accounting Horizons* 6 (2): 26.
- Erdođan, Melih, Nurten Erdođan, Nuran C mert, Ali Kamil Uzun, ve M nevver Yılcıncı. 2012. *Denetim*. C. 2. Eskiřehir: Anadolu  niversitesi Aık đretim Fak ltesi.
- Erdođan, Murat. 2002. “Muhasebe, Denetim ve Bađımsız Denetimin Gerekliiđi”. *Dođuř  niversitesi Dergisi*, sy 5: 51-63.
- Erol, Mikail, ve Muhsin Aslan. 2017. “Uluslararası Muhasebe ve Denetim Standartlarının Geliřmesi”. *Muhasebe ve Finans Tarihi Arařtırmaları Dergisi*, sy 12: 55-86. <https://dergipark.org.tr/en/pub/muftad/issue/30117/324916>.
- Ertař, Abdulaziz, ve Tuđrul Kandemir. 2020. “Bađımsız Denetim Kapsamında T rkiye’de Yapılan Yasal D zenlemelerin Bađımsız Denetim  zerine Etkileri”. *Muhasebe ve Denetime Bakıř* 20 (61): 83-104. <https://dergipark.org.tr/tr/download/article-file/1164130>.
- Fallahpour, Saeid, Eisa Norouziyan Lakvan, ve Mohammad Hendijani Zadeh. 2017a. “Using an Ensemble Classifier Based on Sequential Floating Forward Selection for Financial Distress Prediction Problem”. *Journal of Retailing and Consumer Services* 34 (Ocak): 159-67. <https://doi.org/10.1016/j.jretconser.2016.10.002>.
- . 2017b. “Using an Ensemble Classifier Based on Sequential Floating Forward Selection for Financial Distress Prediction Problem”. *Journal of Retailing and Consumer Services* 34 (Ocak): 159-67. <https://doi.org/10.1016/j.jretconser.2016.10.002>.
- Fawcett, Tom. 2006. “An Introduction to ROC Analysis”. *Pattern Recognition Letters* 27 (8): 861-74. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Fedorova, Elena, Evgenii Gilenko, ve Sergey Dovzhenko. 2013. “Bankruptcy Prediction for Russian Companies: Application of Combined Classifiers”. *Expert Systems with Applications* 40 (18): 7285-93. <https://doi.org/10.1016/j.eswa.2013.07.032>.

- Fernandez, Alberto, Salvador Garcia, Francisco Herrera, ve Nitesh V. Chawla. 2018. "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary". *Journal of Artificial Intelligence Research* 61 (Nisan): 863-905. <https://doi.org/10.1613/jair.1.11192>.
- Fernández-Gámez, Manuel Ángel, Juan Antonio Campos Soria, José António C. Santos, ve David Alaminos. 2020. "European Country Heterogeneity in Financial Distress Prediction: An Empirical Analysis with Macroeconomic and Regulatory Factors". *Economic Modelling* 88 (Haziran): 398-407. <https://doi.org/10.1016/j.econmod.2019.09.050>.
- Financial Accounting Standards Board. 1978. "Statement of Financial Accounting Concepts No. 1". Financial Accounting Standards Board. [https://www.fasb.org/Page/Document?pdf=con1.pdf&title=CON%201%20\(AS%20ISSUED\)](https://www.fasb.org/Page/Document?pdf=con1.pdf&title=CON%201%20(AS%20ISSUED)).
- Fleak, Sandra K., ve Earl R. Wilson. 1994. "The Incremental Information Content of the Going-Concern Audit Opinion". *Journal of Accounting, Auditing & Finance* 9 (1): 149-66. <https://doi.org/10.1177/0148558X9400900109>.
- Foster, Benjamin P., ve Trim Shastri. 2016. "Determinants of Going Concern Opinions and Audit Fees for Development Stage Enterprises". *Advances in Accounting* 33 (Haziran): 68-84. <https://doi.org/10.1016/j.adiac.2016.05.001>.
- Francis, Jere R. 2004. "What Do We Know about Audit Quality?" *The British Accounting Review* 36 (4): 345-68. <https://doi.org/10.1016/j.bar.2004.09.003>.
- Freund, Yoav, ve Robert E. Schapire. 1995. "A decision-theoretic generalization of on-line learning and an application to boosting". İçinde *Computational Learning Theory*, editör Paul Vitányi, 904:23-37. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-59119-2\\_166](https://doi.org/10.1007/3-540-59119-2_166).
- Frydman, Halina, Edward I. Altman, ve Duen-Li Kao. 1985. "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress". *The Journal of Finance* 40 (1): 269-91. <https://doi.org/10.1111/j.1540-6261.1985.tb04949.x>.
- Gaeremynck, Ann, ve Marleen Willekens. 2003. "The Endogenous Relationship between Audit-Report Type and Business Termination: Evidence on Private Firms in a

- Non-Litigious Environment”. *Accounting and Business Research* 33 (1): 65-79.  
<https://doi.org/10.1080/00014788.2003.9729632>.
- Garcia, John. 2022. “Bankruptcy Prediction Using Synthetic Sampling”. *Machine Learning with Applications* 9 (Eylül): 100343.  
<https://doi.org/10.1016/j.mlwa.2022.100343>.
- Geng, Ruibin, Indranil Bose, ve Xi Chen. 2015. “Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining”. *European Journal of Operational Research* 241 (1): 236-47.  
<https://doi.org/10.1016/j.ejor.2014.08.016>.
- Gepp, Adrian, Kuldeep Kumar, ve Sukanto Bhattacharya. 2010a. “Business Failure Prediction Using Decision Trees”. *Journal of Forecasting* 29 (6): 536-55.  
<https://doi.org/10.1002/for.1153>.
- . 2010b. “Business Failure Prediction Using Decision Trees”. *Journal of Forecasting* 29 (6): 536-55. <https://doi.org/10.1002/for.1153>.
- Ghatak, Abhijit. 2017. *Machine Learning with R*. Singapore: Springer Singapore.  
<https://doi.org/10.1007/978-981-10-6808-9>.
- Goo, Yeung-Ja James, Der-Jang Chi, ve Zong-De Shen. 2016. “Improving the Prediction of Going Concern of Taiwanese Listed Companies Using a Hybrid of LASSO with Data Mining Techniques”. *SpringerPlus* 5 (1): 539.  
<https://doi.org/10.1186/s40064-016-2186-5>.
- Gopal, M. 2018. *Applied Machine Learning*. New Delhi: McGraw Hill Education (India).
- Gordini, Nicolò. 2014. “A Genetic Algorithm Approach for SMEs Bankruptcy Prediction: Empirical Evidence from Italy”. *Expert Systems with Applications* 41 (14): 6433-45. <https://doi.org/10.1016/j.eswa.2014.04.026>.
- Göçen, Ceren Ayça. 2010. “Kurumsal Yönetim, İç Kontrol ve Bağımsız Denetim: Parmalat Vakası”. *Mali Çözüm Dergisi* 20 (97): 107-29.
- Greenwell, Brandon, Bradley Boehmke, Jay Cunningham, ve GBM Developers. 2022. “gbm: Generalized Boosted Regression Models”. R. <https://CRAN.R-project.org/package=gbm>.
- Güney, Selami, ve Salim Sercan Sarı. 2016. “Muhasebe Denetiminin Etkinliğini Sağlamada İç Kontrolün Rolü”. *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, sy 5: 62-80. <https://dergipark.org.tr/en/download/article-file/160436>.

- Güngör, Nevzat, Arif Ayluçtarhan, ve Emre Selçuk Sari. 2022. “BDS 570 İşletmenin Sürekliliği Standardının Denetim Raporlarına Etkisi: BİST Şirketlerinde Bir İnceleme”. *Sayıştay Dergisi*, Aralık. <https://doi.org/10.52836/sayistay.1202437>.
- Güredin, Ersin. 2014. *Denetim ve Güvence Hizmetleri SMMM ve YMM'lere Yönelik İlkeler ve Teknikler*. İstanbul: Türkmen Kitabevi.
- Haftacı, Vasfi. 2011. *Muhasebe Denetimi*. 2. bs. İzmit: Umuttepe Kitabevi.
- Hagendorff, Thilo. 2021. “Linking Human And Machine Behavior: A New Approach to Evaluate Training Data Quality for Beneficial Machine Learning”. *Minds and Machines* 31 (4): 563-93. <https://doi.org/10.1007/s11023-021-09573-8>.
- Hand, D. J., Heikki Mannila, ve Padhraic Smyth. 2001. *Principles of data mining. Adaptive computation and machine learning*. Cambridge, Mass: MIT Press.
- Harris, Terry. 2015. “Credit Scoring Using the Clustered Support Vector Machine”. *Expert Systems with Applications* 42 (2): 741-50. <https://doi.org/10.1016/j.eswa.2014.08.029>.
- Hastie, Trevor, Robert Tibshirani, ve J. H. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer series in statistics. New York, NY: Springer.
- He, Hongliang, Wenyu Zhang, ve Shuai Zhang. 2018. “A Novel Ensemble Method for Credit Scoring: Adaption of Different Imbalance Ratios”. *Expert Systems with Applications* 98 (Mayıs): 105-17. <https://doi.org/10.1016/j.eswa.2018.01.012>.
- Heo, Junyoung, ve Jin Yong Yang. 2014. “AdaBoost Based Bankruptcy Forecasting of Korean Construction Companies”. *Applied Soft Computing* 24 (Kasım): 494-99. <https://doi.org/10.1016/j.asoc.2014.08.009>.
- Honaker, James, Gary King, ve Matthew Blackwell. 2011. “**Amelia** II: A Program for Missing Data”. *Journal of Statistical Software* 45 (7). <https://doi.org/10.18637/jss.v045.i07>.
- Hsieh, Tsung-Jung, Hsiao-Fen Hsiao, ve Wei-Chang Yeh. 2012. “Mining Financial Distress Trend Data Using Penalty Guided Support Vector Machines Based on Hybrid of Particle Swarm Optimization and Artificial Bee Colony Algorithm”. *Neurocomputing* 82 (Nisan): 196-206. <https://doi.org/10.1016/j.neucom.2011.11.020>.
- Hua, Zhongsheng, Yu Wang, Xiaoyan Xu, Bin Zhang, ve Liang Liang. 2007. “Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and

- Logistic Regression”. *Expert Systems with Applications* 33 (2): 434-40.  
<https://doi.org/10.1016/j.eswa.2006.05.006>.
- Huang, Shian-Chang, Yu-Cheng Tang, Chih-Wei Lee, ve Ming-Jen Chang. 2012. “Kernel Local Fisher Discriminant Analysis Based Manifold-Regularized SVM Model for Financial Distress Predictions”. *Expert Systems with Applications* 39 (3): 3855-61. <https://doi.org/10.1016/j.eswa.2011.09.095>.
- Huang, Yu-Pei, ve Meng-Feng Yen. 2019. “A New Perspective of Performance Comparison among Machine Learning Algorithms for Financial Distress Prediction”. *Applied Soft Computing* 83 (Ekim): 105663. <https://doi.org/10.1016/j.asoc.2019.105663>.
- Huang, Zan, Hsinchun Chen, Chia-Jung Hsu, Wun-Hwa Chen, ve Soushan Wu. 2004. “Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study”. *Decision Support Systems* 37 (4): 543-58. [https://doi.org/10.1016/S0167-9236\(03\)00086-1](https://doi.org/10.1016/S0167-9236(03)00086-1).
- Hull, John. 2020. *Machine Learning in Business: An Introduction to the World of Data Science*. Second edition. Leipzig: Amazon Distribution.
- “İç denetim Terimleri Sözlüğü”. 2011. Türkiye İç Denetim Enstitüsü. <https://www.tide.org.tr/file/documents/pdf/IcDenetimTerimlerSozlugu.pdf>.
- Jabeur, Sami Ben, Cheima Gharib, Salma Mefteh-Wali, ve Wissal Ben Arfi. 2021. “CatBoost Model and Artificial Intelligence Techniques for Corporate Failure Prediction”. *Technological Forecasting and Social Change* 166 (Mayıs): 120658. <https://doi.org/10.1016/j.techfore.2021.120658>.
- James, Gareth, Daniela Witten, Trevor Hastie, ve Robert Tibshirani, ed. 2013. *An introduction to statistical learning: with applications in R*. Springer texts in statistics 103. New York: Springer.
- Jansen, Stefan. 2018. *Hands-on Machine Learning for Algorithmic Trading: Design and Implement Investment Strategies Based on Smart Algorithms That Learn from Data Using Python*. Birmingham, UK: Packt Publishing.
- Jardin, Philippe du. 2009. “Bankruptcy Prediction Models: How to Choose the Most Relevant Variables?” *Bankers, Markets & Investors*, sy 98: 39-46. <https://core.ac.uk/download/pdf/211606512.pdf>.

- . 2015. “Bankruptcy Prediction Using Terminal Failure Processes”. *European Journal of Operational Research* 242 (1): 286-303. <https://doi.org/10.1016/j.ejor.2014.09.059>.
- . 2016. “A Two-Stage Classification Technique for Bankruptcy Prediction”. *European Journal of Operational Research* 254 (1): 236-52. <https://doi.org/10.1016/j.ejor.2016.03.008>.
- . 2021. “Forecasting Bankruptcy Using Biclustering and Neural Network-Based Ensembles”. *Annals of Operations Research* 299 (1-2): 531-66. <https://doi.org/10.1007/s10479-019-03283-2>.
- Jayasekera, Ranadeva. 2018. “Prediction of Company Failure: Past, Present and Promising Directions for the Future”. *International Review of Financial Analysis* 55 (Ocak): 196-208. <https://doi.org/10.1016/j.irfa.2017.08.009>.
- Jiang, Hui. 2021. *Machine learning fundamentals: a concise introduction*. United Kingdom ; New York, NY: Cambridge University Press.
- Jones, Frederick L. 1996. “The Information Content of the Auditor’s Going Concern Evaluation”. *Journal of Accounting and Public Policy* 15 (1): 1-27. [https://doi.org/10.1016/0278-4254\(95\)00062-3](https://doi.org/10.1016/0278-4254(95)00062-3).
- Jones, Stewart. 2017. “Corporate Bankruptcy Prediction: A High Dimensional Analysis”. *Review of Accounting Studies* 22 (3): 1366-1422. <https://doi.org/10.1007/s11142-017-9407-1>.
- Jones, Stewart, ve David A. Hensher, ed. 2008. *Advances in credit risk modelling and corporate bankruptcy prediction*. Quantitative methods for applied economics and business research. Cambridge, UK ; New York: Cambridge University Press.
- Jones, Stewart, David Johnstone, ve Roy Wilson. 2017. “Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks”. *Journal of Business Finance & Accounting* 44 (1-2): 3-34. <https://doi.org/10.1111/jbfa.12218>.
- Jung, Alexander. 2022. *Machine Learning: The Basics*. Machine Learning: Foundations, Methodologies, and Applications. Singapore: Springer Singapore. <https://doi.org/10.1007/978-981-16-8193-6>.
- Kamu Gözetim Kurumu. 2012. “Bağımsız Denetim Yönetmeliği”. Kamu Gözetim Kurumu.

- <https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/Mevzuat/BDY/Gun cel%20BDY.pdf>.
- . 2018a. *TMS 1 - Finansal Tabloların Sunuluşu*.  
[https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TMS\\_TFRS\\_Setleri/2022/Kirmizi\\_Kitap/tms/TMS%201.pdf](https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TMS_TFRS_Setleri/2022/Kirmizi_Kitap/tms/TMS%201.pdf).
- . 2018b. *TMS 10 - Raporlama Döneminden Sonraki Olaylar*.  
[https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TMS\\_TFRS\\_Setleri/2022/Mavi\\_Kitap/TMS%2010.pdf](https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TMS_TFRS_Setleri/2022/Mavi_Kitap/TMS%2010.pdf).
- . 2020. *BDS 570 - İşletmenin Sürekliliği*.  
[https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TDS/TDS\\_2021\\_Seti/BDS%20570.pdf](https://www.kgk.gov.tr/Portalv2Uploads/files/Duyurular/v2/TDS/TDS_2021_Seti/BDS%20570.pdf).
- Kamu Gözetimi, Muhasebe ve Denetim Standartları Kurumunun Teşkilat ve Görevleri Hakkında Kanun Hükmünde Kararname*. 2011. C. 660.  
<https://www.mevzuat.gov.tr/MevzuatMetin/4.5.660.pdf>.
- Kantardzic, Mehmed. 2011. *Data Mining Concepts, Models, Methods, and Algorithms*. 2. bs. New Jersey: Wiley.
- Kapusuzoğlu, Ayhan. 2006. “Bağımsız Denetime Yönelik Genel Bir Bakış”. *Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 2006-2 (13): 61-73.  
<https://doi.org/10.11616/AbantSbe.173>.
- Karacan, Sami, ve Rahmi Uygun. 2012. *Tekdüzen muhasebe sistemi ve uluslararası denetim standartları ile uyumlu denetim ve raporlama*. 1. baskı. İzmit: Umuttepe Yayınları.
- Karanfiloğlu, A. Yasin. 1999. “Muhasebe Denetimi Kavram İlke ve Yöntemleri”. *Sayıştay Dergisi*, sy 35: 28-37.
- Karatzoglou, Alexandros, Alex Smola, Kurt Hornik, ve Achim Zeileis. 2004. “**Kernlab** - An S4 Package for Kernel Methods in R”. *Journal of Statistical Software* 11 (9).  
<https://doi.org/10.18637/jss.v011.i09>.
- Kardeş Selimoğlu, Seval. 1999. “Faaliyet Denetimi”. *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 15 (1): 193-206.  
<https://dergipark.org.tr/en/pub/adoluibfd/issue/48501/615261>.
- Kaval, Hasan, Aydın Karapınar, Aydın Bayırlı, Adem Altay, ve Salih Torun. 2015. *Türkiye Denetim Standartları Uygulama ve Yorumları*. Ankara: Gazi Kitabevi.

- Kavut, Lerzan. 2000. "Genel Kabul Görmüş Denetim Standartları ve Türkiye'de ki Durumu". *Muhasebe Bilim Dünyası Dergisi* 2 (4): 9-28.
- . 2001. "Bağımsız Denetçiler ile Denetim Hizmeti Alan İşletmelerin Yöneticilerinin Denetim Kalitesine İlişkin Görüş ve Tutumları". *İMKB Dergisi* 5 (20): 1-30.
- Kavut, Lerzan, Oktay Taş, ve Tuğba Şavlı. 2009. *Uluslararası Denetim Standartları Kapsamında Bağımsız Denetim*. İstanbul Serbest Muhasebeci Mali Müşavirler Odası.
- Kayrak, Musa. 2015. "Denetim Başarısızlığı Kavramı: Nedenleri, Sonuçları ve Çözüm Önerileri". *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 8 (1): 91-103. <https://dergipark.org.tr/tr/download/article-file/185156>.
- Kelleher, John D., Brian Mac Namee, ve Aoife D'Arcy. 2020. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Second edition. Cambridge, Massachusetts: The MIT Press.
- Kepekçi, Celal. 2004. *Bağımsız Denetim*. 5. bs. İstanbul: Avcıol Basım Yayın.
- Kılıç, Yunus, ve H. İbrahim Seyrek. 2012. "Finansal Başarısızlık Tahmininde Yapay Sinir Ağlarının Kullanılması: İmalat Sektöründe Bir Uygulama". *İçinde* , 677-89. [https://www.researchgate.net/publication/296705043\\_Finansal\\_Basarisizlik\\_Tahmininde\\_Yapay\\_Sinir\\_Aglarinin\\_Kullanilmasi\\_Imalat\\_Sektorunde\\_Bir\\_Uygulama](https://www.researchgate.net/publication/296705043_Finansal_Basarisizlik_Tahmininde_Yapay_Sinir_Aglarinin_Kullanilmasi_Imalat_Sektorunde_Bir_Uygulama).
- Kida, Thomas. 1980. "An Investigation into Auditors' Continuity and Related Qualification Judgments". *Journal of Accounting Research* 18 (2): 506. <https://doi.org/10.2307/2490590>.
- Kim, Kyoung-jae, Kichun Lee, ve Hyunchul Ahn. 2018. "Predicting Corporate Financial Sustainability Using Novel Business Analytics". *Sustainability* 11 (1): 64. <https://doi.org/10.3390/su11010064>.
- Kim, Myoung-Jong, ve Dae-Ki Kang. 2010. "Ensemble with Neural Networks for Bankruptcy Prediction". *Expert Systems with Applications* 37 (4): 3373-79. <https://doi.org/10.1016/j.eswa.2009.10.012>.
- Kim, Soo Y. 2011a. "Prediction of Hotel Bankruptcy Using Support Vector Machine, Artificial Neural Network, Logistic Regression, and Multivariate Discriminant Analysis". *The Service Industries Journal* 31 (3): 441-68. <https://doi.org/10.1080/02642060802712848>.

- . 2011b. “Prediction of Hotel Bankruptcy Using Support Vector Machine, Artificial Neural Network, Logistic Regression, and Multivariate Discriminant Analysis”. *The Service Industries Journal* 31 (3): 441-68. <https://doi.org/10.1080/02642060802712848>.
- Kim, Soo Y., ve Arun Upneja. 2014. “Predicting Restaurant Financial Distress Using Decision Tree and AdaBoosted Decision Tree Models”. *Economic Modelling* 36 (Ocak): 354-62. <https://doi.org/10.1016/j.econmod.2013.10.005>.
- Knox, Steven W. 2018. *Machine Learning: A Concise Introduction*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons.
- Koh, Hian C., ve Larry N. Killough. 1990. “The Use of Discriminant Analysis in the Assesment of the Going-Concern Status of the Audit Client”. *Journal of Business Finance & Accounting* 17 (2): 179-92. <https://doi.org/10.1111/j.1468-5957.1990.tb00556.x>.
- Koh, Hian Chye. 1991. “Model Predictions and Auditor Assessments of Going Concern Status”. *Accounting and Business Research* 21 (84): 331-38. <https://doi.org/10.1080/00014788.1991.9729848>.
- Krishnan, Gopal V., ve Changjiang Wang. 2015. “The Relation between Managerial Ability and Audit Fees and Going Concern Opinions”. *AUDITING: A Journal of Practice & Theory* 34 (3): 139-60. <https://doi.org/10.2308/ajpt-50985>.
- Kuhn, Max. 2008. “Building Predictive Models in R Using the **Caret** Package”. *Journal of Statistical Software* 28 (5). <https://doi.org/10.18637/jss.v028.i05>.
- Larose, Daniel T., ve Chantal D. Larose. 2015. *Data mining and predictive analytics*. Second edition. Wiley series on methods and applications in data mining. Hoboken, New Jersey: John Wiley & Sons Inc.
- Lee, Sanghoon, Colton Farley, Simon Shim, Wook-Sung Yoo, Yanjun Zhao, ve Wookjin Choi. 2020. “Unsupervised Learning of Deep-Learned Features from Breast Cancer Images”. İçinde *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 740-45. Cincinnati, OH, USA: IEEE. <https://doi.org/10.1109/BIBE50027.2020.00126>.
- Lee, Sangjae, ve Wu Sung Choi. 2013. “A Multi-Industry Bankruptcy Prediction Model Using Back-Propagation Neural Network and Multivariate Discriminant Analysis”. *Expert Systems with Applications* 40 (8): 2941-46. <https://doi.org/10.1016/j.eswa.2012.12.009>.

- Levitan, A, ve J Knoblett. 1985. "Indicators of Exceptions to the Going Concern Assumption". *Auditing: A Journal of Practice and Theory*, sy 5: 26-39.
- Li, Hui, Diego Andina, ve Jie Sun. 2013. "Multiple Proportion Case-Basing Driven CBRE and Its Application in the Evaluation of Possible Failure of Firms". *International Journal of Systems Science* 44 (8): 1409-25. <https://doi.org/10.1080/00207721.2012.659686>.
- Li, Hui, ve Jie Sun. 2008. "Ranking-Order Case-Based Reasoning for Financial Distress Prediction". *Knowledge-Based Systems* 21 (8): 868-78. <https://doi.org/10.1016/j.knosys.2008.03.047>.
- . 2011. "Predicting Business Failure Using Support Vector Machines with Straightforward Wrapper: A Re-Sampling Study". *Expert Systems with Applications* 38 (10): 12747-56. <https://doi.org/10.1016/j.eswa.2011.04.064>.
- Liang, Deron, Chia-Chi Lu, Chih-Fong Tsai, ve Guan-An Shih. 2016. "Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study". *European Journal of Operational Research* 252 (2): 561-72. <https://doi.org/10.1016/j.ejor.2016.01.012>.
- Liang, Deron, Chih-Fong Tsai, Hung-Yuan (Richard) Lu, ve Li-Shin Chang. 2020a. "Combining Corporate Governance Indicators with Stacking Ensembles for Financial Distress Prediction". *Journal of Business Research* 120 (Kasım): 137-46. <https://doi.org/10.1016/j.jbusres.2020.07.052>.
- . 2020b. "Combining Corporate Governance Indicators with Stacking Ensembles for Financial Distress Prediction". *Journal of Business Research* 120 (Kasım): 137-46. <https://doi.org/10.1016/j.jbusres.2020.07.052>.
- Liang, Deron, Chih-Fong Tsai, ve Hsin-Ting Wu. 2015. "The Effect of Feature Selection on Financial Distress Prediction". *Knowledge-Based Systems* 73 (Ocak): 289-97. <https://doi.org/10.1016/j.knosys.2014.10.010>.
- Lisic, Ling Lei, Sabatino (Dino) Silveri, Yanheng Song, ve Kun Wang. 2015. "Accounting Fraud, Auditing, and the Role of Government Sanctions in China". *Journal of Business Research* 68 (6): 1186-95. <https://doi.org/10.1016/j.jbusres.2014.11.013>.
- Loh, Wei-Yin, ve Yu-Shan Shih. 1997. "Split Selection Methods for Classification Trees". *Statistica Sinica* 7 (4): 815-40. [https://www.jstor.org/stable/24306157?casa\\_token=lhkFAyQzSy4AAAAA%3A](https://www.jstor.org/stable/24306157?casa_token=lhkFAyQzSy4AAAAA%3A)

- 0EowjNCzDvPT7QvuoLkrAkH3ypPyAlrNCuQvIH5013xrgA4OMRUgBjhSjvs  
oNb24hW1-nqMzrs65Hodz4WaJCvnLVRgRW1QefAfZMRrz5-vpsL6f8rIvUg.
- Lombardi, Danielle Rose, Rebecca Bloch, ve Miklos A Vasarhelyi. 2014. “The Future of Audit”. *Journal of Information Systems and Technology Management* 11 (1): 21-32. <https://doi.org/10.4301/S1807-17752014000100002>.
- Loudder, Martha L., Inder K. Khurana, Roby B. Sawyer, Cindy Cordery, Carol Johnson, Jordan Lowe, ve Robert Wunderle. 1992. “The Information Content of Audit Qualifications”. *AUDITING: A Journal of Practice & Theory* 11 (1): 69-82.
- Lukason, Oliver, ve María-del-Mar Camacho-Miñano. 2019. “Bankruptcy Risk, Its Financial Determinants and Reporting Delays: Do Managers Have Anything to Hide?” *Risks* 7 (3): 77. <https://doi.org/10.3390/risks7030077>.
- Luo, Jiaxin, Qingjun Meng, ve Yan Cai. 2018. “Analysis of the Impact of Artificial Intelligence Application on the Development of Accounting Industry”. *Open Journal of Business and Management* 06 (04): 850-56. <https://doi.org/10.4236/ojbm.2018.64063>.
- Maheshwari, Anil K. 2020. *Data Analytics Made Accessible*.
- Manita, Riadh, Najoua Elommal, Patricia Baudier, ve Lubica Hikkerova. 2020. “The Digital Transformation of External Audit and Its Impact on Corporate Governance”. *Technological Forecasting and Social Change* 150 (Ocak): 119751. <https://doi.org/10.1016/j.techfore.2019.119751>.
- Manyika, James, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, ve Martin Dewhurst. 2017. “A Future That Works: Automation, Employment and Productivity”. McKinsey&Company. <https://www.mckinsey.com/~media/mckinsey/featured%20insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Executive-summary.ashx>.
- Marsland, Stephen. 2015. *Machine Learning: An Algorithmic Perspective*. Second edition. Boca Raton, FL: CRC Press.
- Martens, David, Liesbeth Bruynseels, Bart Baesens, Marleen Willekens, ve Jan Vanthienen. 2008. “Predicting Going Concern Opinion with Data Mining”. *Decision Support Systems* 45 (4): 765-77. <https://doi.org/10.1016/j.dss.2008.01.003>.

- Martin, Roger D. 2000. "Going-Concern Uncertainty Disclosures and Conditions: A Comparison of French, German, and U.S. Practices". *Journal of International Accounting, Auditing and Taxation* 9 (2): 137-58. [https://doi.org/10.1016/S1061-9518\(00\)00029-X](https://doi.org/10.1016/S1061-9518(00)00029-X).
- McCarthy, John, ve Edward A. Feigenbaum. 1990. "In Memoriam: Arthur Samuel: Pioneer in Machine Learning". *AI Magazine* 11 (3): 2. <https://doi.org/10.1609/aimag.v11i3.840>.
- McClendon, Lawrence, ve Natarajan Meghanathan. 2015. "Using Machine Learning Algorithms to Analyze Crime Data". *Machine Learning and Applications: An International Journal* 2 (1): 1-12. <https://doi.org/10.5121/mlaij.2015.2101>.
- Menon, Krishnagopal, ve Kenneth B. Schwartz. 1987. "An Empirical Investigation of Audit Qualification Decisions in the Presence of Going Concern Uncertainties". *Contemporary Accounting Research* 3 (2): 302-15. <https://doi.org/10.1111/j.1911-3846.1987.tb00640.x>.
- Messier, William F., Steven M. Glover, ve Douglas F. Prawitt. 2016. *Auditing & assurance services: a systematic approach*. Tenth edition. New York, NY: McGraw-Hill Education.
- Meyer, David, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, ve Friedrich Leisch. 2023. "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien". R. <https://CRAN.R-project.org/package=e1071>.
- Meyer, Paul A., ve Howard W. Pifer. 1970. "Prediction of Bank Failures". *The Journal of Finance* 25 (4): 853-68. <https://doi.org/10.1111/j.1540-6261.1970.tb00558.x>.
- Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill series in computer science. New York: McGraw-Hill.
- Mohri, Mehryar, Afshin Rostamizadeh, ve Ameet Talwalkar. 2018. *Foundations of machine learning*. Second edition. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press.
- Moreira, João, André Carlos Ponce de Leon Ferreira Carvalho, ve Tomáš Horváth. 2018. *A general introduction to data analytics*. Hoboken, NJ: John Wiley & Sons.
- Mselmi, Nada, Amine Lahiani, ve Taher Hamza. 2017. "Financial Distress Prediction: The Case of French Small and Medium-Sized Firms". *International Review of Financial Analysis* 50 (Mart): 67-80. <https://doi.org/10.1016/j.irfa.2017.02.004>.

- Murphy, Kevin P. 2012. *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. Cambridge, MA: MIT Press.
- Murthy, Sreerama K. 1998. "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey". *Data Mining and Knowledge Discovery* 2 (4): 345-89. <https://doi.org/10.1023/A:1009744630224>.
- Mutchler, Jane F. 1984. "Auditor's Perceptions of the Going-Concern Opinion Decision". *Auditing: A Journal of Practice & Theory* 3 (2): 17-29.
- . 1985. "A Multivariate Analysis of the Auditor's Going-Concern Opinion Decision". *Journal of Accounting Research* 23 (2): 668. <https://doi.org/10.2307/2490832>.
- Mutchler, Jane F., William Hopwood, ve James M. McKeown. 1997. "The Influence of Contrary Information and Mitigating Factors on Audit Opinion Decisions on Bankrupt Companies". *Journal of Accounting Research* 35 (2): 295. <https://doi.org/10.2307/2491367>.
- Nanni, Loris, ve Alessandra Lumini. 2009. "An Experimental Comparison of Ensemble of Classifiers for Bankruptcy Prediction and Credit Scoring". *Expert Systems with Applications* 36 (2): 3028-33. <https://doi.org/10.1016/j.eswa.2008.01.018>.
- Nasta, Laura Nicoleta, ve Călin Tănase Ladar. 2015. "Convergences and Divergences between Internal and External Audit on International Context". *AGORA International Journal of Administration Sciences*, sy 3: 46-55. <https://univagora.ro/jour/index.php/aijas/article/view/2593>.
- Ohlson, James A. 1980. "Financial Ratios and the Probabilistic Prediction of Bankruptcy". *Journal of Accounting Research* 18 (1): 109. <https://doi.org/10.2307/2490395>.
- Olson, David L., Dursun Delen, ve Yanyan Meng. 2012a. "Comparative Analysis of Data Mining Methods for Bankruptcy Prediction". *Decision Support Systems* 52 (2): 464-73. <https://doi.org/10.1016/j.dss.2011.10.007>.
- . 2012b. "Comparative Analysis of Data Mining Methods for Bankruptcy Prediction". *Decision Support Systems* 52 (2): 464-73. <https://doi.org/10.1016/j.dss.2011.10.007>.
- O'Regan, David. 2004. *Auditor's dictionary: terms, concepts, processes, and regulations*. Electronic resource. Hoboken, N.J: Wiley. <http://www.myilibrary.com?id=26509>.

- Özbirecikli, Mehmet, ve Cemil Süslü. 2005. “Bağımsız Denetim Firmalarının Yolsuzluk Riski Faktörlerini Değerleme Uygulamaları ve Türkiye’deki Bağımsız Denetim Firmaları Üzerine Karşılaştırmalı Bir Araştırma-1”. *Muhasebe ve Finansman Dergisi*, sy 27: 67-85.
- Parker, Susan, Gary F. Peters, ve Howard F. Turetsky. 2005. “Corporate Governance Factors and Auditor Going Concern Assessments”. *Review of Accounting and Finance* 4 (3): 5-29. <https://doi.org/10.1108/eb043428>.
- Pham Vo Ninh, Binh, Trung Do Thanh, ve Duc Vo Hong. 2018. “Financial Distress and Bankruptcy Prediction: An Appropriate Model for Listed Firms in Vietnam”. *Economic Systems* 42 (4): 616-24. <https://doi.org/10.1016/j.ecosys.2018.05.002>.
- Pietruszkiewicz, Wieslaw. 2008. “Dynamical systems and nonlinear Kalman filtering applied in classification”. İçinde *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, 1-6. London, UK: IEEE. <https://doi.org/10.1109/UKRICIS.2008.4798948>.
- Qian, Hongyi, Baohui Wang, Minghe Yuan, Songfeng Gao, ve You Song. 2022. “Financial Distress Prediction Using a Corrected Feature Selection Measure and Gradient Boosted Decision Tree”. *Expert Systems with Applications* 190 (Mart): 116202. <https://doi.org/10.1016/j.eswa.2021.116202>.
- Quinlan, J. R. 1986. “Induction of Decision Trees”. *Machine Learning* 1 (1): 81-106. <https://doi.org/10.1007/BF00116251>.
- . 1993. *C4.5: programs for machine learning*. The Morgan Kaufmann series in machine learning. San Mateo, Calif: Morgan Kaufmann Publishers.
- “R Mevcut Paketler”. t.y. <https://Cran.r-Project.Org/Web/Packages/>. Erişim 14 Temmuz 2023. <https://cran.r-project.org/web/packages/>.
- Raphael, Jon. 2015. “How Artificial Intelligence Can Boost Audit Quality”. <https://www.cfo.com/Auditing/2015/06/Artificial-Intelligence-Can-Boost-Audit-Quality/>. 2015. <https://www.cfo.com/auditing/2015/06/artificial-intelligence-can-boost-audit-quality/>.
- Raschka, Sebastian. 2015. *Python Machine Learning: Unlock Deeper Insights into Machine Learning with This Vital Guide to Cutting-Edge Predictive Analytics*. Community Experience Distilled. Birmingham Mumbai: Packt Publishing open source.

- Ravi Kumar, P., ve V. Ravi. 2007. "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review". *European Journal of Operational Research* 180 (1): 1-28. <https://doi.org/10.1016/j.ejor.2006.08.043>.
- Reynolds, J.Kenneth, ve Jere R Francis. 2000. "Does Size Matter? The Influence of Large Clients on Office-Level Auditor Reporting Decisions". *Journal of Accounting and Economics* 30 (3): 375-400. [https://doi.org/10.1016/S0165-4101\(01\)00010-6](https://doi.org/10.1016/S0165-4101(01)00010-6).
- Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, ve Markus Müller. 2011. "pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves". *BMC Bioinformatics* 12 (1): 77. <https://doi.org/10.1186/1471-2105-12-77>.
- Rokach, Lior. 2019. *Ensemble Learning: Pattern Classification Using Ensemble Methods*. Second edition. Series in Machine Perception and Artificial Intelligence, vol. 85. New Jersey London Singapore Beijing Shanghai Hong Kong Taipei Chennai: World Scientific.
- Runkler, Thomas A. 2020. *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-29779-4>.
- Sammut, Claude, ve Geoffrey I. Webb, ed. 2017. *Encyclopedia of Machine Learning and Data Mining*. Second edition. Springer Reference. New York, NY: Springer.
- Samuel, A. L. 1959. "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development* 3 (3): 210-29. <https://doi.org/10.1147/rd.33.0210>.
- Sánchez-Serrano, José Ramón, David Alaminos, Francisco García-Lagos, ve Angela M. Callejón-Gil. 2020. "Predicting Audit Opinion in Consolidated Financial Statements with Artificial Neural Networks". *Mathematics* 8 (8): 1288. <https://doi.org/10.3390/math8081288>.
- Saygılı, Arıkan Tarık. 2007. "Bağımsız Denetim Firmalarında Kalite Yönetimi". *Mevzuat Dergisi* 10 (110).
- Sermaye Piyasası Kurumu. 2011. "Sermaye Piyasasında Bağımsız Denetim Standartları Hakkında Tebliğ (Seri: X, No: 22)". Sermaye Piyasası Kurulu. <https://www.mevzuat.gov.tr/anasayfa/MevzuatFihristDetayIframe?MevzuatTur=9&MevzuatNo=10384&MevzuatTertip=5>.
- Sevilengül, Orhan. 2016. *Genel Muhasebe*. 2016. bs. Ankara: Gazi Kitabevi.

- Shin, Kyung-Shik, Taik Soo Lee, ve Hyun-jung Kim. 2005. "An Application of Support Vector Machines in Bankruptcy Prediction Model". *Expert Systems with Applications* 28 (1): 127-35. <https://doi.org/10.1016/j.eswa.2004.08.009>.
- Shmueli, Galit, Nitin R. Patel, ve Peter C. Bruce. 2010. *Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. Hoboken, N.J: Wiley.
- Singh, Pradeep. 2022. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*. Hoboken Beverly: Wiley Scrivener publishing.
- Smolarski, Jan, Neil Wilner, ve Weifang Yang. 2011. "The Use of Financial Information by Private Equity Funds in Evaluating New Investments". *Review of Accounting and Finance* 10 (1): 46-68. <https://doi.org/10.1108/14757701111113811>.
- Son, H., C. Hyun, D. Phan, ve H.J. Hwang. 2019. "Data Analytic Approach for Bankruptcy Prediction". *Expert Systems with Applications* 138 (Aralık): 112816. <https://doi.org/10.1016/j.eswa.2019.07.033>.
- Sree, Dhana K., ve Shoba Bindu. 2018. "Data Analytics: Why Data Normalization". *International Journal of Engineering & Technology*, sy 7: 209-13. <https://www.academia.edu/download/74142822/9582.pdf>.
- Sterling, Robert R. 1968. "The Going Concern: An Examination". *The Accounting Review* 43 (3): 481-502. <http://www.jstor.org/stable/244071>.
- Sun, Jie, Hamido Fujita, Peng Chen, ve Hui Li. 2017. "Dynamic Financial Distress Prediction with Concept Drift Based on Time Weighting Combined with Adaboost Support Vector Machine Ensemble". *Knowledge-Based Systems* 120 (Mart): 4-14. <https://doi.org/10.1016/j.knosys.2016.12.019>.
- Sun, Jie, Hamido Fujita, Yujiao Zheng, ve Wenguo Ai. 2021. "Multi-Class Financial Distress Prediction Based on Support Vector Machines Integrated with the Decomposition and Fusion Methods". *Information Sciences* 559 (Haziran): 153-70. <https://doi.org/10.1016/j.ins.2021.01.059>.
- Sun, Jie, ve Hui Li. 2009. "Financial Distress Prediction Based on Serial Combination of Multiple Classifiers". *Expert Systems with Applications* 36 (4): 8659-66. <https://doi.org/10.1016/j.eswa.2008.10.002>.

- . 2012. “Financial Distress Prediction Using Support Vector Machines: Ensemble vs. Individual”. *Applied Soft Computing* 12 (8): 2254-65. <https://doi.org/10.1016/j.asoc.2012.03.028>.
- Sun, Jie, Hui Li, Hamido Fujita, Binbin Fu, ve Wenguo Ai. 2020. “Class-Imbalanced Dynamic Financial Distress Prediction Based on Adaboost-SVM Ensemble Combined with SMOTE and Time Weighting”. *Information Fusion* 54 (Şubat): 128-44. <https://doi.org/10.1016/j.inffus.2019.07.006>.
- Sun, Ting, ve Miklos A Vasarhelyi. 2017. “How an Evolving Technology Could Transform Analysis and Improve Judgment”. *The CPA Journal* 87 (6): 24-29. <https://www.cpapjournal.com/2017/06/19/deep-learning-future-auditing/>.
- Taylor, Donald H., ve G. William Glezen. 1991. *Auditing, Integrated Concepts and Procedures*. 5th ed. New York: Wiley.
- Terzi, Serkan. 2022. “İşletmelerin Sürekliliğinin Değerlendirilmesinde Makine Öğrenmesi Algoritmalarının Kullanımı: Türkiye Örneği”. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, Mart. <https://doi.org/10.46928/iticusbe.991788>.
- Therneau, Terry, ve Beth Atkinson. 2022. “rpart: Recursive Partitioning and Regression Trees”. R. <https://CRAN.R-project.org/package=rpart>.
- Tierney, Nicholas, ve Dianne Cook. 2023. “Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations”. *Journal of Statistical Software* 105 (7). <https://doi.org/10.18637/jss.v105.i07>.
- Toroslu, M. Vefa. 2012. *Yeni Türk Ticaret Kanunu Kapsamında Finansal Tablolar Denetimi*. 1. bs. İstanbul: Seçkin Yayıncılık.
- Tosun, Aytaç. 2021. “Muhasebe Denetimi ve Denetim Türleri”. *Balkan ve Yakın Doğu Sosyal Bilimler Dergisi*, sy 7 (Special Issue): 101-7. [http://ibaness.org/bnejss/2021\\_07\\_special\\_issue/11\\_Tosun.pdf](http://ibaness.org/bnejss/2021_07_special_issue/11_Tosun.pdf).
- Tsai, C, ve J Wu. 2008. “Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring”. *Expert Systems with Applications* 34 (4): 2639-49. <https://doi.org/10.1016/j.eswa.2007.05.019>.
- Tsai, Chih-Fong. 2014. “Combining Cluster Analysis with Classifier Ensembles to Predict Financial Distress”. *Information Fusion* 16 (Mart): 46-58. <https://doi.org/10.1016/j.inffus.2011.12.001>.

- Tsai, Chih-Fong, Yu-Feng Hsu, ve David C. Yen. 2014. "A Comparative Study of Classifier Ensembles for Bankruptcy Prediction". *Applied Soft Computing* 24 (Kasım): 977-84. <https://doi.org/10.1016/j.asoc.2014.08.047>.
- Tuan, Koray. 2009. "Bağımsız Dış Denetim Sürecinde İç Kontrol Sisteminin İncelenmesi ve Değerlendirilmesi". *Çukurova Üniversitesi İİBF Dergisi* 13 (2): 1-15.
- . 2015. "Bağımsız Denetimde Rotasyon Uygulaması: Sistemin Artıları-Eksileri Üzerine Tartışmalar". *Akademik Sosyal Araştırmalar Dergisi* 3 (16): 291-309. <https://doi.org/10.16992/ASOS.767>.
- Türk Ticaret Kanunu*. 2011. C. 6102. <https://www.mevzuat.gov.tr/mevzuatmetin/1.5.6102.pdf>.
- Türker, Masum. 2006. "Uluslararası Denetim Standartları'na Yakınsama ve Türkiye Denetim Standartlarının Oluşturulması". *Muhasebe ve Denetime Bakış*, sy 19: 87-98. <https://dergipark.org.tr/en/pub/mdbakis/issue/63880/967073>.
- Türker, Masum, Recep Pekdemir, Yakup Selvi, ve Yılmaz Fatih. 2003. "Sınırlı Uygunluk Denetimi (SUD)". 2003-1. Ankara: TÜRMOB.
- Uzay, Şaban. 2004. "Türkiye' de Denetçi Bağımsızlığı". İçinde *The 1st Annual International Accounting Conference*, 69-95. The 1st Annual International Accounting Conference.
- Valentini, Giorgio, ve Thomas G. Dietterich. 2004. "Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods". *Journal of Machine Learning Research* 5: 725-75. <http://www.jmlr.org/papers/volume5/valentini04a/valentini04a.pdf>.
- Van der Aalst, Wil. 2016. *Process mining: data science in action*. 2nd edition. New York, NY: Springer Berlin Heidelberg.
- Vapnik, V.N. 1999. "An overview of statistical learning theory". *IEEE Transactions on Neural Networks* 10 (5): 988-99. <https://doi.org/10.1109/72.788640>.
- Vargün, Hakan, ve Berk Yıldız. 2021. "Bağımsız Denetim Standardı (BDS) 570 Kapsamında İşletme Sürekliliğinin Değerlendirilmesi: BİST KOBİ Sanayi Endeksi Üzerine Ekonometrik Bir Analiz". *Muhasebe Bilim Dünyası Dergisi*, Eylül. <https://doi.org/10.31460/mbdd.844439>.
- Veganzones, David, ve Eric Séverin. 2018. "An Investigation of Bankruptcy Prediction in Imbalanced Datasets". *Decision Support Systems* 112 (Ağustos): 111-24. <https://doi.org/10.1016/j.dss.2018.06.011>.

- Verikas, Antanas, Zivile Kalsyte, Marija Bacauskiene, ve Adas Gelzinis. 2010. "Hybrid and Ensemble-Based Soft Computing Techniques in Bankruptcy Prediction: A Survey". *Soft Computing* 14 (9): 995-1010. <https://doi.org/10.1007/s00500-009-0490-5>.
- Volkov, Andrey, Dries F. Benoit, ve Dirk Van den Poel. 2017. "Incorporating Sequential Information in Bankruptcy Prediction with Predictors Based on Markov for Discrimination". *Decision Support Systems* 98 (Haziran): 59-68. <https://doi.org/10.1016/j.dss.2017.04.008>.
- Wagner, Johannes Martin. 2016. "Continuous Auditing - The Future of Internal Audit?" İçinde , 3244-52.
- Wang, Gang, Gang Chen, ve Yan Chu. 2018. "A New Random Subspace Method Incorporating Sentiment and Textual Information for Financial Distress Prediction". *Electronic Commerce Research and Applications* 29 (Mayıs): 30-49. <https://doi.org/10.1016/j.elerap.2018.03.004>.
- Wang, Gang, Jian Ma, ve Shanlin Yang. 2014. "An Improved Boosting Based on Feature Selection for Corporate Bankruptcy Prediction". *Expert Systems with Applications* 41 (5): 2353-61. <https://doi.org/10.1016/j.eswa.2013.09.033>.
- Wang, Gang, Jingling Ma, Gang Chen, ve Ying Yang. 2020. "Financial Distress Prediction: Regularized Sparse-Based Random Subspace with ER Aggregation Rule Incorporating Textual Disclosures". *Applied Soft Computing* 90 (Mayıs): 106152. <https://doi.org/10.1016/j.asoc.2020.106152>.
- Wang, Lu, ve Chong Wu. 2017. "Business Failure Prediction Based on Two-Stage Selective Ensemble with Manifold Learning Algorithm and Kernel-Based Fuzzy Self-Organizing Map". *Knowledge-Based Systems* 121 (Nisan): 99-110. <https://doi.org/10.1016/j.knosys.2017.01.016>.
- White, Halbert. 2000. "A Reality Check for Data Snooping". *Econometrica* 68 (5): 1097-1126. <https://doi.org/10.1111/1468-0262.00152>.
- Wickham, Hadley. 2011. "The Split-Apply-Combine Strategy for Data Analysis". *Journal of Statistical Software* 40 (1). <https://doi.org/10.18637/jss.v040.i01>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Golemund, vd. 2019. "Welcome to the Tidyverse". *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

- Wickham, Hadley, ve Jennifer Bryan. 2023. “readxl: Read Excel Files”. R. <https://CRAN.R-project.org/package=readxl>.
- Wickham, Hadley, Romain François, Lionel Hnery, Kirill Müller, ve Davis Vaughan. 2023. “dplyr: A Grammar of Data Manipulation”. R. <https://CRAN.R-project.org/package=dplyr>.
- Williams, Densil A. 2016. “Can Neural Networks Predict Business Failure? Evidence From Small High Tech Firms in the UK”. *Journal of Developmental Entrepreneurship* 21 (01): 1650005. <https://doi.org/10.1142/S1084946716500059>.
- Williams, Graham. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4419-9890-3>.
- Wirth, Rüdiger, ve Jochen Hipp. 2000. “CRISP-DM: Towards a Standard Process Model for Data Mining”. İçinde . <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>.
- Wright, Marvin N., ve Andreas Ziegler. 2017. “**Ranger**: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R”. *Journal of Statistical Software* 77 (1). <https://doi.org/10.18637/jss.v077.i01>.
- Wu, Desheng, Xiyuan Ma, ve David L. Olson. 2022. “Financial Distress Prediction Using Integrated Z-Score and Multilayer Perceptron Neural Networks”. *Decision Support Systems* 159 (Ağustos): 113814. <https://doi.org/10.1016/j.dss.2022.113814>.
- Wu, Desheng(Dash), Liang Liang, ve Zijiang Yang. 2008a. “Analyzing the Financial Distress of Chinese Public Companies Using Probabilistic Neural Networks and Multivariate Discriminate Analysis”. *Socio-Economic Planning Sciences* 42 (3): 206-20. <https://doi.org/10.1016/j.seps.2006.11.002>.
- . 2008b. “Analyzing the Financial Distress of Chinese Public Companies Using Probabilistic Neural Networks and Multivariate Discriminate Analysis”. *Socio-Economic Planning Sciences* 42 (3): 206-20. <https://doi.org/10.1016/j.seps.2006.11.002>.
- Wyrobek, Joanna, ve Krzysztof Kluza. 2019. “Efficiency of Gradient Boosting Decision Trees Technique in Polish Companies’ Bankruptcy Prediction”. İçinde *Information Systems Architecture and Technology: Proceedings of 39th*

- International Conference on Information Systems Architecture and Technology – ISAT 2018*, editör Zofia Wilimowska, Leszek Borzemski, ve Jerzy Świątek, 854:24-35. *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-99993-7\\_3](https://doi.org/10.1007/978-3-319-99993-7_3).
- Xiao, Zhi, Xianglei Yang, Ying Pang, ve Xin Dang. 2012. “The Prediction for Listed Companies’ Financial Distress by Using Multiple Prediction Methods with Rough Set and Dempster–Shafer Evidence Theory”. *Knowledge-Based Systems* 26 (Şubat): 196-206. <https://doi.org/10.1016/j.knosys.2011.08.001>.
- Xu, Xiaoyan, ve Yu Wang. 2009a. “Financial Failure Prediction Using Efficiency as a Predictor”. *Expert Systems with Applications* 36 (1): 366-73. <https://doi.org/10.1016/j.eswa.2007.09.040>.
- . 2009b. “Financial Failure Prediction Using Efficiency as a Predictor”. *Expert Systems with Applications* 36 (1): 366-73. <https://doi.org/10.1016/j.eswa.2007.09.040>.
- Yakut, Emre, ve Bekir Elmas. 2013. “İşletmelerde Finansal Başarısızlığının Veri Madenciliği ve Diskriminant Analizi Modelleri ile Tahmin Edilmesi”. *Afyon Kocatepe Üniversitesi İİBF Dergisi* 15 (1): 261-80. <https://dergipark.org.tr/en/pub/akuiibfd/issue/1619/20289?publisher=aku>.
- Yang, Zijiang, Wenjie You, ve Guoli Ji. 2011. “Using Partial Least Squares and Support Vector Machines for Bankruptcy Prediction”. *Expert Systems with Applications* 38 (7): 8336-42. <https://doi.org/10.1016/j.eswa.2011.01.021>.
- Yeh, Ching-Chiang, Der-Jang Chi, ve Ming-Fu Hsu. 2010. “A Hybrid Approach of DEA, Rough Set and Support Vector Machines for Business Failure Prediction”. *Expert Systems with Applications* 37 (2): 1535-41. <https://doi.org/10.1016/j.eswa.2009.06.088>.
- Yeh, Ching-Chiang, Der-Jang Chi, ve Yi-Rong Lin. 2014a. “Going-Concern Prediction Using Hybrid Random Forests and Rough Set Approach”. *Information Sciences* 254 (Ocak): 98-110. <https://doi.org/10.1016/j.ins.2013.07.011>.
- . 2014b. “Going-Concern Prediction Using Hybrid Random Forests and Rough Set Approach”. *Information Sciences* 254 (Ocak): 98-110. <https://doi.org/10.1016/j.ins.2013.07.011>.
- Yılandı, Münevver, Birol Yıldız, ve Murat Kiracı. 2013. *Muhasebe Denetimi*. 1. bs. Ankara: Detay Yayıncılık.

- Yıldız, Birol. 2001. “Finansal Başarısızlığın Öngörülmesinde Yapay Sinir Ağı Kullanımı ve Halka Açık Şirketlerde Ampirik Bir Uygulama”. *İMKB Dergisi* 5 (17): 52-67.
- Yılmaz, Zeliha, ve Alpaslan Yaşar. 2018. “Bağımsız Denetim Raporlarının İşletmenin Sürekliliği Denetim Standardı 570 Çerçevesinde İncelenmesi: BİST İmalat Şirketleri Üzerine Bir Araştırma”. *Mali Çözüm Dergisi* 28 (145): 55-90. <https://www.acarindex.com/pdfs/324376>.
- Zdolšek, Daniel, Timotej Jagrič, ve Iztok Kolar. 2022. “Auditor’s Going-Concern Opinion Prediction: The Case of Slovenia”. *Economic Research-Ekonomska Istraživanja* 35 (1): 106-21. <https://doi.org/10.1080/1331677X.2021.1888766>.
- Zeng, Sen, Yaqin Li, Wanjun Yang, ve Yanru Li. 2020. “A Financial Distress Prediction Model Based on Sparse Algorithm and Support Vector Machine”. Editör Chen Chen. *Mathematical Problems in Engineering* 2020 (Kasım): 1-11. <https://doi.org/10.1155/2020/5625271>.
- Zhang, Guoqiang, Michael Y Hu, B Eddy Patuwo, ve Daniel C Indro. 1999. “Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis”. *European Journal of Operational Research*, 17. [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4).
- Zhang, Liang, Lingling Zhang, Weili Teng, ve Yibing Chen. 2013. “Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning”. *Procedia Computer Science* 17: 695-703. <https://doi.org/10.1016/j.procs.2013.05.090>.
- Zhang, Wenyu, Hongliang He, ve Shuai Zhang. 2019. “A Novel Multi-Stage Hybrid Model with Enhanced Multi-Population Niche Genetic Algorithm: An Application in Credit Scoring”. *Expert Systems with Applications* 121 (Mayıs): 221-32. <https://doi.org/10.1016/j.eswa.2018.12.020>.
- Zhang, Zijiao, Chong Wu, Shiyu Qu, ve Xiaofang Chen. 2022. “An Explainable Artificial Intelligence Approach for Financial Distress Prediction”. *Information Processing & Management* 59 (4): 102988. <https://doi.org/10.1016/j.ipm.2022.102988>.
- Zhao, Shuping, Kai Xu, Zhao Wang, Changyong Liang, Wenxing Lu, ve Bo Chen. 2022. “Financial Distress Prediction by Combining Sentiment Tone Features”. *Economic Modelling* 106 (Ocak): 105709. <https://doi.org/10.1016/j.econmod.2021.105709>.

- Zhou, Ligang, ve Kin Keung Lai. 2017. "AdaBoost Models for Corporate Bankruptcy Prediction with Missing Data". *Computational Economics* 50 (1): 69-94. <https://doi.org/10.1007/s10614-016-9581-4>.
- Zhou, Ligang, Kin Keung Lai, ve Jerome Yen. 2012. "Empirical Models Based on Features Ranking Techniques for Corporate Financial Distress Prediction". *Computers & Mathematics with Applications* 64 (8): 2484-96. <https://doi.org/10.1016/j.camwa.2012.06.003>.
- . 2014. "Bankruptcy Prediction Using SVM Models with a New Approach to Combine Features Selection and Parameter Optimisation". *International Journal of Systems Science* 45 (3): 241-53. <https://doi.org/10.1080/00207721.2012.720293>.
- Zhou, Ligang, Dong Lu, ve Hamido Fujita. 2015. "The Performance of Corporate Financial Distress Prediction Models with Features Selection Guided by Domain Knowledge and Data Mining Approaches". *Knowledge-Based Systems* 85 (Eylül): 52-61. <https://doi.org/10.1016/j.knosys.2015.04.017>.
- Zhou, Ligang, Yain-Whar Si, ve Hamido Fujita. 2017. "Predicting the Listing Statuses of Chinese-Listed Companies Using Decision Trees Combined with an Improved Filter Feature Selection Method". *Knowledge-Based Systems* 128 (Temmuz): 93-101. <https://doi.org/10.1016/j.knosys.2017.05.003>.
- Zięba, Maciej, Sebastian K. Tomczak, ve Jakub M. Tomczak. 2016. "Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction". *Expert Systems with Applications* 58 (Ekim): 93-101. <https://doi.org/10.1016/j.eswa.2016.04.001>.
- Zmijewski, Mark E. 1984. "Methodological Issues Related to the Estimation of Financial Distress Prediction Models". *Journal of Accounting Research* 22: 59. <https://doi.org/10.2307/2490859>.



