



**KREDİ ANALİZİNDE MAKİNE ÖĞRENMESİ
KULLANIMI: TARIMSAL KREDİLERDE
UYGULAMA ÖRNEĞİ**

Mehmet Akif BULUT

(Doktora Tezi)

Eskişehir, 2019

**KREDİ ANALİZİNDE MAKİNE ÖĞRENMESİ
KULLANIMI:
TARIMSAL KREDİLERDE UYGULAMA ÖRNEĞİ**

Mehmet Akif BULUT

T.C.

Eskişehir Osmangazi Üniversitesi

Sosyal Bilimler Enstitüsü

İşletme Ana Bilim Dalı

İşletme Bilim Dalı

DOKTORA TEZİ

Eskişehir

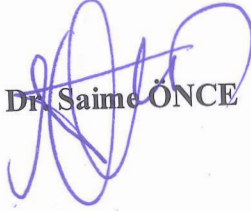
2019

T.C.

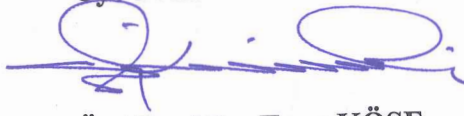
ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Mehmet Akif BULUT tarafından hazırlanan "Kredi Analizinde Makine Öğrenmesi Kullanımı: Tarımsal Kredilerde Uygulama Örneği" başlıklı bu çalışma 3/1/2019 tarihinde Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Enstitüsü Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili maddesi uyarınca yapılan savunma sınavı sonucunda başarılı bulunarak, Jürimiz tarafından İşletme Ana Bilim Dalında doktora tezi olarak kabul edilmiştir.

Başkan Prof. Dr. Saim ÖNCE



Üye Prof. Dr. Özcan DAĞDEMİR



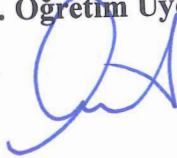
Üye Prof. Dr. Tunç KÖSE



Üye Prof. Dr. Birol YILDIZ (Danışman)



Üye Dr. Öğretim Üyesi Alper BİLGE



ONAY

.../.../2019

Enstitü Müdürü

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin Eskişehir Osmangazi Üniversitesi Bilimsel Araştırma ve Yayın Etiği Yönergesi hükümlerine göre hazırlandığını; bana ait, özgün bir çalışma olduğunu; çalışmanın hazırlık, veri toplama, analiz ve bilgilerin sunumu aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Osmangazi Üniversitesi tarafından kullanılan bilimsel intihal tespit programıyla taranmasını kabul ettiğimi ve hiçbir şekilde intihal içermediğini beyan ederim. Yaptığım bu beyana aykırı bir durumun saptanması halinde ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

Mehmet Akif BULUT

ÖZET

KREDİ ANALİZİNDE MAKİNE ÖĞRENMESİ KULLANIMI: TARIMSAL KREDİLERDE UYGULAMA ÖRNEĞİ

BULUT, Mehmet Akif

Doktora-2019

İşletme Ana Bilim Dalı

Danışman: Prof. Dr. Birol YILDIZ

Tarımsal kredi limitlerinin tespit edilmesinde kullanılan değişkenler ile kredi kullanan şahıslara ait bazı demografik özelliklerin, kredilerin vadesinde tahsil edilmesine ne ölçüde etkide bulunduğunu tespit etmek, çalışmamızın hareket noktasını oluşturmaktadır.

Bu çalışmada, yurt genelinde faaliyet gösteren Tarım Kredi Kooperatiflerinden, farklı bölgelerde bulunan 19 tarım kredi kooperatifinin 2012-2016 dönemine ait verileri kullanılarak kredilerin vadesinde ödenmesine etki eden faktörlerin saptanması amaçlanmıştır. Söz konusu amaç doğrultusunda bazı makine öğrenmesi algoritmaları kullanılarak, kredi tahsisinde kullanılan değişkenleri içeren bir model geliştirilmesi ve bu modelin kredilerin vadesinde tahsil edilip edilemeyeceğini en iyi şekilde öngörmesi hedeflenmiştir.

Tezimizde, ilgili Kurum tarafından kooperatif ortaklarına kullandırılan kredilerin büyüklüğünün tespitinde kullanılan değişkenler ile kredi kullanan şahıslara ait bazı demografik unsurlar kullanılarak kredilerin vadesinde ödenip ödenmeyeceğini tahmin eden bir model geliştirilmeye çalışılmıştır. Böyle bir modelin geliştirilmesi için WEKA yazılımı kullanılarak söz konusu değişkenler farklı makine öğrenmesi algoritmalarının kullanımına sunulmuştur. Her bir algoritma tarafından geliştirilecek modelin performansının değerlendirilmesinde, ZeroR algoritması eşik değeri belirleyen algoritma olarak kullanılmıştır. Doğrusal Ayırma Analizi ve Lojistik

Regresyon algoritmaları tarafından geliştirilen modellerin, eşik değerin altında doğru sınıflandırma yapabildiği görülmüştür. K en yakın komşu, Naive Bayes ve Yapay Sinir Ağları algoritmaları tarafından ortaya konulan modellerin eşik değerin üzerinde ve eşit düzeyde doğru sınıflandırma yapabildiği gözlenmiştir. En yüksek düzeyde doğru sınıflandırma yapabilen tahmin modeli, Karar Ağacı algoritması geliştirilmiştir. En yüksek düzeyde doğru sınıflandırma başarısı sağlayan modelin, kredi kullanan şahısların yetiştirdikleri tarımsal ürünün kg fiyatı ile sahip oldukları büyükbaş hayvanların değerini esas alarak tahminlerini şekillendirdiği algoritma çıktısından anlaşılmıştır. Bu bağlamda anılan değişkenlerin, kredilerin vadesinde tahsil edilmesinde önem arz eden değişkenler olduğu gözlenmiştir.

Her ne kadar çalışmamızda kullanılan algoritmaların çoğunluğu eşik değerin üzerinde doğru sınıflandırma yapabilen tahmin modelleri geliştirmiş olsa da vadesinde tahsil edilemeyecek kredilerin daha yüksek düzeyde tespit edilebilmesi için analiz kapsamının geliştirilmesi gerektiği anlaşılmıştır. Tarımsal kredilerin daha riskli olmasına sebep olan çok sayıda etmen olduğu dikkate alındığında, diğer kredi türlerindeki analizlerde olduğu gibi, vadesinde ödenmeyecek kredilerin daha yüksek düzeyde önceden tespit edilebilmesi gerek tasarruf sahipleri gerekse kredi kurumlarınca önem arz etmektedir. Bu sebeple, tarımsal kredilere dair veri setinin genişletilmesi ve daha fazla sayıda değişkenin modellemeye dahil edilmesi ile vadesinde ödenmeyen kredilere ilişkin daha yüksek düzeyde doğru sınıflandırma yapabilecek bir model geliştirilebileceği kanaatine varılmıştır.

ABSTRACT

USE OF MACHINE LEARNING IN CREDIT ANALYSIS: APPLICATION IN AGRICULTURAL LOANS

BULUT, Mehmet Akif

Doctoral Degree-2019

Department of Business Administration

Adviser: Professor Dr. Birol YILDIZ

In addition to the factors used in deciding the credit limits, some of the demographic characteristics of the credit users can be collected in credit analysis. The starting point of our study is to determine which factors are valuable to detect overdue loans.

In this study, it was aimed to determine the factors that affect the payment of the credits on the maturity of the loans by using the data of the agricultural credit cooperatives in different regions in the 2012-2016 period. For this purpose, it is aimed to develop a model which processes the variables used in credit granting by using some machine learning algorithms and to determine whether this model can be collected in detecting of overdue loans.

In our thesis, a model has been tried to be developed to estimate whether the loans will be paid or not by using the variables which are used to determine the size of credits by the credit institution and some demographic elements belonging to the credit users. For the development of such a model, these variables were introduced to the use of different machine learning algorithms using WEKA software. In evaluating the performance of the model to be developed by each algorithm, ZeroR algorithm is used as the algorithm to determine the threshold value. Linear Discriminant Analysis and Logistic Regression algorithms developed models which were able to make a correct classification below the threshold value. It was observed that the models introduced

by K nearest neighbors, Naive Bayes and Artificial Neural Networks algorithms were able to make an accurate classification above the threshold value. The estimation model, which can make the correct classification at the highest level, was developed by Decision Tree algorithm. It is understood from the model which has the highest correct classification level, that the price of agricultural crops and cattle were based in forming the estimates. In this context, it is observed that the variables mentioned are important variables for the collection of loans.

Although the majority of the algorithms used in our study have developed prediction models that can make accurate classification on the threshold value, it has been understood that the scope of analysis should be developed in order to determine the loans that cannot be collected on date at a higher level. Considering that there are many factors that cause agricultural loans to be more risky, as in the analyzes in other types of loans, it is important for the credit agencies and account owners to detect overdue credits in advance. For this reason, it was concluded that the data set and the number of variables on agricultural loans could be expanded and then a more accurate model could be developed to obtain higher level in estimating overdue credits.

İÇİNDEKİLER

ÖZET.....	v
ABSTRACT	vii
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ	xiv
EKLER LİSTESİ	xv
ÖNSÖZ	xvi
GİRİŞ	1

1. BÖLÜM KREDİ ANALİZİ

1.1 KREDİ ANALİZİ KAVRAMI VE GEREKLİLİĞİ.....	3
1.2 KREDİ ANALİZİNDE DİKKATE ALINAN FAKTÖRLER	5
1.2.1. Kişisel Nitelikler.....	5
1.2.2. Mali Faktörler.....	7
1.2.3. Ekonomik Faktörler.....	9
1.2.4. Yönetim Performansı	10
1.2.5. Hukuki Faktörler	11
1.2.6. Krediyeye İlişkin Teminatlar	12
1.3 KREDİ RİSKİ	13
1.3.1. Kredi Riskinin Ölçülmesi.....	15
1.3.1.1. Kredi Notlaması.....	15
1.3.1.2. Kredi Derecelendirmesi	16
1.4 KREDİ ANALİZ SİSTEMLERİ.....	18
1.4.1. Kalitatif Analiz	18
1.4.2. Kantitatif Analiz	21
1.4.3. Kredi Notlama Sistemi	21
1.4.3.1. Kredi Notlama Modelleri.....	23

1.4.3.1.1. Lineer Olasılık Modeli	23
1.4.3.1.2. Logit Modeli	24
1.4.3.1.3. Probit Model	24
1.4.3.1.4. Lineer Diskriminant Modeli.....	25
1.4.3.2. Kredi Notlama Modellerinin Zayıf Yönleri.....	26
1.5 TARIMSAL KREDİLER	27

2. BÖLÜM

MAKİNE ÖĞRENMESİ

2.1 VERİ İŞLEMENİN GEREKLİLİĞİ.....	30
2.2 VERİ MADENCİLİĞİ.....	31
2.3 MAKİNE ÖĞRENMESİNE İLİŞKİN TEMEL KAVRAMLAR	33
2.4 MAKİNE ÖĞRENMESİNDE KULLANILAN ALGORİTMALAR	37
2.4.1. K-NN Algoritması (K En Yakın Komşu Algoritması)	37
2.4.2. Naive Bayes Algoritması.....	44
2.4.3. Karar Ağacı Algoritması	52
2.4.4. Doğrusal Ayırma Analizi Algoritması	61
2.4.5. Lojistik Regresyon Algoritması	64
2.4.6. Yapay Sinir Ağları Algoritması	67
2.4.7. ZeroR Algoritması.....	79
2.5 KREDİ ANALİZİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMINA İLİŞKİN LİTERATÜR İNCELEMESİ.....	80

3. BÖLÜM

UYGULAMA ÖRNEĞİ

3.1 ARAŞTIRMANIN KONUSU	84
3.2 ARAŞTIRMANIN AMACI.....	84
3.3 ARAŞTIRMANIN KAPSAMI	85
3.4 ARAŞTIRMANIN YÖNTEMİ.....	85

3.5 HİPOTEZ GELİŞTİRME	85
3.6 VERİ SETİ.....	86
3.7 DEĞİŞKENLERİN TANIMLANMASI.....	87
3.8 UYGULAMA ÖRNEĞİNDE BAZ ALINACAK PERFORMANS ÖLÇÜTLERİ	92
3.8.1. Kappa Katsayısı (Kappa Statistics)	92
3.8.2. Hata Oranları	93
3.8.2.1. Ortalama Mutlak Hata (Mean Absolute Error-MAE).....	93
3.8.2.2. Ortalama Hata Karelerinin Karekökü (Root Mean Squared Error- RMSE)	94
3.8.2.3. Nispi Mutlak Hata (Relative Absolute Error-RAE)	94
3.8.2.4. Nispi Hata Karelerinin Karekökü (Root Relative Squared Error-RRSE)	94
3.8.3. Doğru Sınıflandırmaya İlişkin Detaylı Oranlar	95
3.9 UYGULAMA SONUÇLARI.....	99
3.9.1. ZeroR Algoritması Uygulama Çıktısı	100
3.9.2. K En Yakın Komşu Algoritması Uygulama Çıktısı.....	102
3.9.3. Naive Bayes Algoritması Uygulama Çıktısı	106
3.9.4. Karar Ağacı Algoritması Uygulama Çıktısı	109
3.9.5. Doğrusal Ayırma Analizi Algoritması Uygulama Çıktısı	114
3.9.6. Lojistik Regresyon Algoritması Uygulama Çıktısı	118
3.9.7. Yapay Sinir Ağları Algoritması Uygulama Çıktısı	121
3.10 UYGULAMA SONUÇLARININ KARŞILAŞTIRILMASI.....	125
SONUÇ	129
KAYNAKÇA	131
EKLER	147

TABLolar LİSTESİ

Tablo 1: Veri Madenciliği Temel İşlevler.....	33
Tablo 2: K En Yakın Komşu Algoritmasının Güçlü ve Zayıf Yönleri.....	38
Tablo 3: Örnek Frekans ve İhtimal Tabloları.....	48
Tablo 4: Naive Bayes Algoritmasının Güçlü ve Zayıf Yönleri.....	48
Tablo 5: Karar Ağacı Algoritmasının Güçlü ve Zayıf Yönleri.....	54
Tablo 6: Karar Ağacı Algoritması Örnek Kredi Sınıflandırması.....	57
Tablo 7: Kovaryans Matrisi Örnek Veri Seti.....	62
Tablo 8: Lojistik Regresyon Algoritmasının Avantaj ve Dezavantajları.....	66
Tablo 9: Geri Yayılım Algoritmasının Avantaj ve Dezavantajları.....	77
Tablo 10: Araştırmanın Hipotezleri.....	86
Tablo 11: Uygulama Örneğindeki Bağımlı Değişken.....	88
Tablo 12: Uygulama Örneğinde Kullanılan Bağımsız Değişkenler.....	88
Tablo 13: Kappa Katsayısı Uyum Düzeyleri.....	93
Tablo 14: Sınıflandırmaya İlişkin Örnek Çıktı Tablosu.....	95
Tablo 15: ZeroR Algoritması Program Çıktı Özeti.....	100
Tablo 16: ZeroR Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	101
Tablo 17: ZeroR Algoritması Hata Matrisi.....	101
Tablo 18: K En Yakın Komşu Algoritması Program Çıktı Özeti.....	103
Tablo 19: K En Yakın Komşu Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	105
Tablo 20: K En Yakın Komşu Algoritması Hata Matrisi.....	106
Tablo 21: Naive Bayes Algoritması Program Çıktı Özeti.....	107

Tablo 22: Naive Bayes Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	109
Tablo 23: Naive Bayes Algoritması Hata Matrisi.....	109
Tablo 24: Karar Ağacı Algoritması Program Çıktı Özeti.....	111
Tablo 25: Karar Ağacı Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	113
Tablo 26: Karar Ağacı Algoritması Hata Matrisi.....	113
Tablo 27: Doğrusal Ayırma Analizi Algoritması Program Çıktı Özeti.....	115
Tablo 28: Doğrusal Ayırma Analizi Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	117
Tablo 29: Doğrusal Ayırma Analizi Algoritması Hata Matrisi.....	117
Tablo 30: Lojistik Regresyon Algoritması Program Çıktı Özeti.....	118
Tablo 31: Lojistik Regresyon Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	120
Tablo 32: Lojistik Regresyon Algoritması Hata Matrisi.....	121
Tablo 33: Yapay Sinir Ağları Algoritması Program Çıktı Özeti.....	122
Tablo 34: Yapay Sinir Ağları Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları.....	124
Tablo 35: Yapay Sinir Ağları Algoritması Hata Matrisi.....	124
Tablo 36: Algoritma Sonuçlarının Karşılaştırılması.....	127
Tablo 37: Kappa İstatistik Değerlerinin Karşılaştırılması.....	128

ŞEKİLLER LİSTESİ

Şekil 1: Veri Tabanlarında Bilgi Keşfi.....	31
Şekil 2: Ortak Olasılık Örneği.....	46
Şekil 3: Örnek Karar Ağacı.....	59
Şekil 4: Lojistik Fonksiyon-Sigmoid Fonksiyon Görseli.....	66
Şekil 5: Sinir Hücresi Yapısı.....	68
Şekil 6: Yapay Nöron.....	69
Şekil 7: Birim Aktivasyon Fonksiyonu.....	70
Şekil 8: Sigmoid Aktivasyon Fonksiyonu.....	71
Şekil 9: Alternatif Aktivasyon Fonksiyonları.....	72
Şekil 10: Tek Katmanlı Ağ.....	74
Şekil 11: Çok Katmanlı Ağ.....	74
Şekil 12: Çoklu Çıktı Döngüleri.....	75
Şekil 13: Çoklu Gizli Katmanlar.....	75
Şekil 14: Geri Beslemeli Ağ.....	76
Şekil 15: ROC-Karar Vericinin Etkinliği Eğrisi.....	97
Şekil 16: Örnek Kesinlik-Hassasiyet Eğrileri.....	99
Şekil 17: Uygulama Örneğindeki Karar Ağacı.....	110

EKLER LİSTESİ

Ek 1: Çalışma Evrenindeki Tarım Kredi Kooperatifleri.....	147
Ek 2: Örnek Ortak Beyannamesi.....	148
Ek 3: Örnek Ortak Kartı.....	149
Ek 4: ZeroR Algoritması Detaylı Program Çıktısı.....	150
Ek 5: K En Yakın Komşu Algoritması Detaylı Program Çıktısı.....	153
Ek 6: Naive Bayes Algoritması Detaylı Program Çıktısı.....	156
Ek 7: Karar Ağacı Algoritması Detaylı Program Çıktısı.....	165
Ek 8: Doğrusal Ayırma Analizi Algoritması Detaylı Program Çıktısı.....	168
Ek 9: Lojistik Regresyon Algoritması Detaylı Program Çıktısı.....	172
Ek 10: Yapay Sinir Ağları Algoritması Detaylı Program Çıktısı.....	182

ÖNSÖZ

Kredi analizinde, kredilerin vadesinde tahsil edilip edilmeyeceğine etki eden çok sayıda faktör bulunmakla birlikte tarımsal krediler bünyesinde barındırdığı farklı riskler nedeniyle analizi diğer kredi türlerine kıyasla daha zorlu bir süreç gerektiren kredi türleridir. Kredi analizinde çok farklı istatistiksel ve ekonometriye dayalı yöntemler kullanılmakla birlikte gelişen ve değişen teknolojik imkânlar, bilgisayarların bu alanda kullanımını hızlandırmakta, analiz sürecinde gözden kaçabilecek ancak sonuca etki edebilecek unsurların tamamının bu sürece dâhil edilmesi sağlanmaya çalışılmaktadır. Çalışmamızda, yaşamın her alanında kullanılmaya başlanan ve ciddi imkânlar sunan makine öğrenmesi araçları kullanılarak, analiz sürecinde daha farklı değişkenleri barındıran tarımsal kredilerin vadesinde tahsil edilip edilemeyeceğini önceden tahmin edebilecek bir model geliştirilmesi amaçlanmıştır.

Bu çalışmamda, öncelikli olarak sonsuz sabrı ve ilgisini beden esirgemeyen, açmış olduğu yeni ufuklarla akademik hayatımın yanı sıra mesleki hayatımı ve geleceğe dair planlarımı büyük ölçüde şekillendiren, yapıcı eleştirileriyle hatalarımı ve eksikliklerimi daha farklı açılardan değerlendirmemi sağlayan değerli danışman hocam Prof. Dr. Birol YILDIZ'a sonsuz teşekkürlerimi sunarım. Doktora eğitimi süresince, görüş ve destekleriyle bana yardımcı olan Araştırma Görevlisi Dr. Şafak AĞDENİZ'e teşekkür ederim. Yoğun çalışma programı içerisinde, doktora eğitimimi sürdürme ve çalışmalarımı tamamlama fırsatının yanı sıra, tezimin uygulama bölümüne esas teşkil eden verileri çalışmamda kullanma imkânını bana sunan Türkiye Tarım Kredi Kooperatifleri Merkez Birliği Genel Müdürlüğü Rehberlik ve Teftiş Kurulu Başkanlığı'na katkı ve yardımlarından ötürü teşekkürü borç bilirim.

Akademik hayatım süresince benden desteklerini hiçbir zaman esirgemeyen anne ve babama, çalışmalarım süresince yeterince zaman ayıramadığım eşim ve kızıma, göstermiş oldukları fedakârlık ve anlayışlarından ötürü müteşekkirim.

Mehmet Akif BULUT

GİRİŞ

Gelişmekte olan ülkelerin, ekonomik büyüme ve kalkınmasının önündeki en önemli kısıtlardan birisi, ihtiyaç duyulan fon kaynaklarına sınırlı erişim imkânı bulunmasıdır. Ülkemizin de yer aldığı gelişmekte olan ülkeler grubunda, zaten sınırlı olan kaynakların etkin bir şekilde kullanılması elzemdir. Söz konusu fonların kullanımında riskin tamamen ortadan kaldırılması mümkün olmamakla birlikte, mümkün olduğunca risklerin tespit edilerek bu risklere karşı önceden tedbir alınması gerekmektedir.

Fonları ihtiyaç sahipleriyle buluşturan kredi kurumlarının büyük çoğunluğunu bankalar oluşturmakla birlikte, sektör özelinde kredi kullandıran farklı yapılar da bulunabilmektedir. Ülkemizde, bu farklı yapıların faaliyetleri arasında yer alan tarımsal kredilerin, vadesinde tahsil edilip edilemeyeceğini tahmin edebilecek bir model geliştirilmesi çalışmamızın konusunu oluşturmaktadır.

Kredi kurumlarınca kredi riskinin belirlenmesi ve yönetilmesi için farklı yaklaşımlar benimsenmiştir. Farklı yaklaşımlar bulunsada dahi yapılan analizlerde bazı standart değişkenler ve yöntemler de bulunmaktadır. Hangi yaklaşım benimsenirse benimsensin, hangi yöntem kullanılırsa kullanılsın, kredi riskinin tamamen ortadan kaldırılması mümkün olamamaktadır. Ancak, riskin daha iyi analiz edilmesi neticesinde, kredi kararlarının daha etkin verilmesi sağlanabilmektedir.

Kredi analiz süreçleri için özellikle bankalar, ileri düzeyde teknikler ve yöntemler kullanmaktadır. Çalışmamıza konu olan tarımsal kredi uygulamalarında, üreticilerin tarımsal gelirlerine etki eden unsurlar değerlendirilerek kredi kararı verilmekte, ancak yapılan değerlendirmenin, bankalar tarafından yapılan analizlere kıyasla daha basit düzeyde gerçekleştirildiğini söylemek mümkün olabilmektedir. Diğer sektörlerle kıyasla tarım sektörünün maruz kaldığı risklerin daha fazla ve farklı olması sebebiyle bu sektörde kullanılan kredilerin analiz süreçleri daha zorlu bir çalışma gerektirmektedir.

Teknolojik gelişmelerle birlikte insanlık, hayatın birçok alanında bilgisayarları kullanarak çok zorlu analizleri yapabilir, uzun zaman alacak işlemleri çok kısa sürede gerçekleştirilebilir hale gelmiştir. Ayrıca son yıllarda bilgisayarlara biçilen rol, geçtiğimiz yıllara kıyasla bambaşka bir boyut kazanmıştır. İnsan beyninin çalışma prensibi ancak genel hatlarıyla belirlenmiş olsa da belirlenen bu yapıya benzer şekilde

bilgisayarların, bir diğere deyişle makinelerin, çalıřma prensiplerinin oluřturulması amaçlanmıř, böylelikle insan doęasının kısıtlarından arınmıř, ancak insan beyninin sahip olduęu kompleks veri iřleme yeteneęine haiz makine öğrenmesi yöntemleri geliřtirilmiřtir. Bu yöntemlerde, veri setinin özelliklerine ve eriřilmek istenen amaca göre farklı teknikler (farklı algoritmalar) kullanılarak, karar veren modeller geliřtirilmesi amaçlanmaktadır.

Ülkemizin son dönemde tarım sektöründe yařadığı dar boęazlar da göz önüne alındığında, bu sektörde geliřimin ve büyümenin saęlanması için önemli bir etken olan tarımsal kredilerin en iyi şekilde yönetilmesi büyük önem arz etmektedir. Tarımsal gelirlere etki eden faktörlerin çeřitlilięi ve bu faktörlerin analiz edilmesinin zorluęu göz önüne alındığında, daha doęru kredi kararlarının verilebilmesi için makinelerin kullanılabilir hale getirilmesi saęlanmalıdır. Bu sebeple, üreticilerin tarımsal gelirlerine etki eden ve tarımsal kredi tutarlarının tespitinde kullanılan unsurların, kredilerin vadesinde tahsil edilmesine olan etkilerini deęerlendiren bir model geliřtirilmesi amaçlanmıřtır. Böylelikle, hali hazırda kredi tespitinde kullanılan deęiřkenlerin, makine öğrenmesi algoritmalarının doęru karar verebilmelerine ne denli katkı saęlayabildięinin anlaşılması, tarımsal kredilerin analizine eklenebilecek başka faktörlere ihtiyaç olup olmaması arařtırma sonuçlarında deęerlendirilmiřtir. Ayrıca, farklı makine öğrenmesi algoritmaları kullanılarak, mevcut veri setinin özelliklerine göre en iyi sınıflandırma kararı verebilecek algoritmanın belirlenmesi ve bu algoritmanın ortaya koyduęu modelin daha sonra geliřtirilerek, kredi kararlarında iyi neticeler verebilecek baz bir model ortaya konulması hedeflenmiřtir.

Söz konusu amaç ışığında, çalıřmamızın birinci bölümünde kredi analizi, kredi riski ve analizde kullanılan yöntemler açıklanmıřtır. İkinci bölümde uygulama örneğimizde kullandığımız makine öğrenmesi algoritmalarının çalıřma prensiplerine dair teorik açıklamalar yapılmıřtır. Üçüncü bölümde ise tarımsal kredilerin vadesinde tahsil edilip edilemeyeceęini öngören model geliřtirilmesine yönelik, ikinci bölümde açıklanan algoritmalar kullanılmak suretiyle, uygulama ařaması gerçekteřirilmiřtir.

1. BÖLÜM

KREDİ ANALİZİ

1.1 KREDİ ANALİZİ KAVRAMI VE GEREKLİLİĞİ

Kredi sisteminde yer alan gerçek ya da tüzel kişilerin, kredi değerlilikleri hususunda bir karara ulaşabilmek için gerekli görülen çeşitli bilgi ve etmenlerin değerlendirilmesi, ayrıca kredi riskinin yönetimi kredi analizinin konusunu oluşturmaktadır (Çabuk ve Lazol, 2014: 155). Söz konusu kapsama, kredi kullanan tarafların ödeme gücü içine düşmesi durumunda ortaya çıkacak kriz yönetimi de dâhil edilmektedir (Özden, 2010: 60). Kredi analizi, sadece kredi kullanan tarafların mali verilerinin incelenmesini kapsamamakta, ilgili kişilerin geçmiş, güncel ve gelecekteki durumlarını etkileyebilecek tüm unsurları içermektedir (Vaidya, 14.06.2018).

Kredi analizinde amaç, kredi talep eden tarafa sınırlı bir süre için verilecek belirli tutardaki kredinin taşıdığı riskin saptanmasıdır (Cardoso vd., 2016: 71). Kredi faaliyetleri iki temel riski bünyesinde barındırmaktadır (Selimler, 2015: 137-138). İlki, kredinin verildiği tarih ile geri tahsil edildiği tarih arasında paranın değerinde oluşabilecek kayıplardan kaynaklı enflasyon riskidir. Söz konusu risk, ekonomideki diğer sorunlarla birlikte daha geniş anlamda değerlendirildiğinde sistematik risk olarak da adlandırılabilir. Çok önemli değişimlerin/krizlerin beklenmediği ya da öngörülemeyen durumlarda enflasyon riski kredi veren kuruluşlar tarafından çok önemsenmemektedir. Kredi kurumlarınca daha çok dikkate alınan ve mümkün olduğunca düşük seviyede tutulmaya çalışılan diğer risk ise geri ödenmeme riskidir. Her kredinin özünde geri ödenmeme riski taşıdığı kabul edilmekle birlikte bu riski doğuran başlıca etmenler aşağıdaki şekilde sıralanabilir (Türe, 2015: 28-29):

- Kredinin vadesine de bağlı olarak geleceğin sağlıklı bir şekilde tahmin edilememesi,
- Yangın, sel, deprem vb. doğal afetler,
- Kredi talep eden kişinin üretmiş olduğu mal ve hizmetlere olan talepteki olumsuz değişimler,
- Teknolojik değişim ve gelişmelere yeterince ayak uyduramayan kredi borçlusunun, faaliyet gösterdiği sektördeki rekabet gücünü yitirmesi,

- Yerel ya da global ölçekte ekonomide meydana gelen kriz ortamlarının satışlar ve satışlardan elde edilen kâra olumsuz etkisi,
- Kredi talep eden kişilerin, kendi alacaklarını tahsil etmede yaşadıkları olumsuzluklar,
- Kredi borçlusunun faaliyet gösterdiği sektöre has sorunların borçlunun ticari faaliyetlerine olumsuz etkisi,
- Kredi kullanan kişinin faaliyetlerini sürdürdüğü sektörün, monopol ya da oligopol piyasalara doğru evrilmesi,
- Hükümetlerin ekonomik faaliyetlere olumsuz etki edebilecek kararlar alması ve bu yönde politikalar uygulaması,
- Mevsimsel dalgalanmaların kredi kullanan tarafların gelirlerinde meydana getirdiği değişiklikler,
- Kredi alan tarafların, faaliyetlerine ilişkin hatalı politika tercihleri,
- Kredi borçlusunu yükümlülük altına sokan kefalet ve garantiler,
- Kredi alan kişinin, borcunu ödememe eğilimi göstermesi,
- Özellikle tüzel kişi kredi borçlularının üst yönetimlerindeki uyumsuzluklar,
- Uluslararası anlaşmazlıklar ve ülke riskleri sebebiyle kredi borçlularının zarara uğramaları.

Yukarıda sıralanan etmenlere yenilerini eklemek mümkün olmakla birlikte, gerek bireysel gerekse ticari kredilerde geri ödenmeme riskini artıran en önemli etkenin yönetim hatalarından kaynaklandığı ifade edilmektedir (Yüksel, 2011: 7).

Kredi analizinin temel amacı, kredi talep eden tarafın, kredi sözleşmesi kapsamında, krediyi geri ödeme kapasitesine ve isteğine sahip olup olmadığını belirleyerek kredi riskinin mümkün olduğunca azaltılmasıdır (Ştefa ve Pelin, 2016: 247).

Kredi kuruluşlarının faaliyetlerine devam edebilmeleri, kârlı ancak riskli olan kredi işlemlerini, geri ödenmeme riskinden olabildiğince korumalarına bağlıdır. Üçüncü kişilere ait olan kaynakları, kredi ihtiyacı bulunan taraflarla buluşturan kredi kuruluşlarında, geri dönmeyen krediler sadece kredi kuruluşunun varlığını tehlikeye sokmakla kalmamakta, geri ödenmeme riski tasarruf/fon sahiplerinin birikimlerinin

yok olmasına neden olarak ekonomik sistemin etkinliğine ciddi ölçüde zarar vermektedir. Daha önce de değinildiği üzere kredi analizinin temel amacı kredi riskinin azaltılması olmakla birlikte, kredi talep eden tarafın talebine uygun cevap verilebilmesi kredi analizinin bir diğer önemli amacıdır. Söz konusu uygun cevap, kredinin tutarı, vadesi ve koşullarının uygun bir şekilde oluşturulmasından ibarettir (Tunç, 2012: 33). Ancak, kredi analizinin ilk ve en temel adımı, kredi talep eden tarafın hangi gerekçe ile krediyi talep ettiğinin anlaşılmasıdır. Kabul edilebilir ekonomik bir nedene dayanmayan ve kredi kuruluşunun kredi politikalarıyla örtüşmeyen taleplerin, analizin diğer aşamalarına ulaşması düşünülemez (Yıldırım, 2007: 43). Özellikle tüzel kişilerin ticari nitelikteki kredi taleplerinin değerlendirilmesi esnasında, salt mali tablolarla yetinmek kredi analizinin temel amacından uzaklaşılmasına neden olabilmektedir. Bu bağlamda kredi kuruluşlarındaki görevliler iş yeri ziyaretleri gerçekleştirmekte, firma sahipleri ve çalışanlarla görüşmeler yapmakta, mali tablolara ya da yazılı dokümanlara aktarılmamış gerçeklere/sebeplere bu yöntemle ulaşabilmektedirler.

1.2 KREDİ ANALİZİNDE DİKKATE ALINAN FAKTÖRLER

Kredi başvurularının değerlendirilmesinde kredi kurumları önceleri başvuranın karakteri, girişim yeteneği ve sermayesi gibi sınırlı sayıda faktörü dikkate alırken ilerleyen zaman içerisinde söz konusu faktörlerin sayısında artış olmuştur. Bu faktörler; kredinin amacı ve tutarı, kredi talep eden kişinin kişisel nitelikleri, mali faktörler, ekonomik faktörler, yönetim performansı, kredinin güvencesi ve kredi talep eden tarafın hukuki yapısı gibi temel başlıklar altında toplanabilmektedir (Peprah, Agyei ve Oteng, 2017: 78-79).

1.2.1. Kişisel Nitelikler

Geçmişte ve günümüzdeki uygulamalarda, kredi analizinde en fazla önem verilen faktörlerin başında kuşkusuz kredi talep eden kişinin ahlaki özellikleri gelmektedir (Tokel, 2004: 9). Borcunu geri ödeme niyetinden yoksun kişilerden alınan maddi teminatlar, kredinin geri ödenme olasılığını istenilen ölçüde artıramamaktadır.

Kredi kurumlarınca, talepleri uygun görülecek kişilerin dürüst, borçlarına sadık, kanun dışı uygulamalara başvurmayan şahıslardan oluşması beklenmektedir.

Yasa dışı faaliyetler içerisinde yer alan ya da mevzuata aykırı iş ve işlemleri gerçekleştirmeyi içselleştirmiş şahısların, orta vadede yükümlülüklerini yerine getirebilmeleri oldukça düşük bir olasılığa sahiptir. Ayrıca, başvuru sahibinin kredi süreci boyunca, kendisine ait olumlu unsurları sunabildiği kadar olumsuz etmenleri de ortaya koyabilmesi, kredi kuruluşunca olumsuz olarak değerlendirilebilecek durumları gizleme çabası içerisinde olmaması gerekmektedir (Akgüç, 2014: 15).

Kredi talep eden kişinin, kullanmak istediği krediyi değerlendireceği alanda yeterli bilgi sahibi olması beklenmektedir. Örneğin, ticari kredi başvurusunda bulunan tarafın, faaliyet gösterdiği sektörün özelliklerini, sektördeki rakiplerini, ekonomik yaşamda sektörü etkileyebilecek gelişmeleri bilmesi ve takip etmesi gerekmektedir (Kök ve Aksu, 2013: 168).

Çoğu kredi kurumu, müşterilerine mali danışmanlık hizmeti de vermektedir. Bu hizmetin sunulma amacı, gelecekte kredinin geri ödemelerine sağlıklı bir şekilde kaynak sağlanması, alınabilecek yanlış kararlar neticesinde geri ödenmeme riskinin ortaya çıkmasına engel olunma çabasıdır. Bu bağlamda, kredi kurumlarındaki uzmanlar tarafından kredi kullanan kişilere yapılan tavsiyelerin dikkate alınması ve gerekli düzeltmelerin zaman kaybetmeksizin yerine getirilmesi beklenmektedir.

Kredi riski açısından, kredi talep eden kişinin yetenekli, yeniliğe açık ve girişimci ruha sahip olması önem arz eden diğer kişisel faktörlerdir. Özellikle ticari kredilerde, firma sahip ya da yöneticilerinin yönetim fonksiyonlarını gereği gibi yerine getirememesi geri ödenmeme riskini artıran en önemli unsur olarak görülmektedir (Tokel, 2004: 36).

Bireysel kredilerde, başvuru sahibinin bazı özellikleri dikkate alınarak, şahsın sağlık sorunlarından ötürü geri ödenmeme riskinin ortaya çıkması, çeşitli sigorta uygulamalarıyla (kredi hayat sigortası gibi) aşmaya çalışılmaktadır. Ancak tüzel kişiliklerin kullandığı ticari kredilerde, firmanın sevk ve idaresini sağlayan şahsın sağlık sorunlarından yönetimden çekilmesi ve bu şahsın yerini doldurabilecek birisinin bulunamaması, firmanın geleceğini, haliyle kredinin geleceğini ciddi anlamda riske sokacaktır (Oral, 2015: 43).

Yukarıda sıralanan hususlar değerlendirilirken, ekonomik koşulların kredi karakteri üzerine etkisi dikkate alınmalı, ekonomik dar boğazlarda

şahısların/şirketlerin takındığı tutumlar kredi karakterinin belirlenmesinde göz önünde bulundurulmalıdır (Yiğitbaş, 2012: 37-38).

1.2.2. Mali Faktörler

Kredi talep edenin mali durumu temelde iki başlık altında ele alınabilir. Öncelikle, kredinin geri ödenmesinde birinci derecede önem arz eden mali etkenin kaynak yaratma kapasitesi olduğunu ifade etmek mümkündür. Bir diğer etken, özellikle tüzel kişilikler için, öz kaynak yeterliliğidir (Akgüç, 2014: 17-19).

Kredi borçlarının geri ödenmesinde genel olarak üç farklı yol ayrı ayrı ya da birlikte izlenebilir. Kredi talep eden taraf, varlıklarını satarak, yeniden farklı bir kaynaktan borçlanarak ya da hali hazırda yürütmekte olduğu faaliyetleri neticesinde elde ettiği fonlar ile borcunu ödeyebilmektedir (Tuna, 18.06.2018).

Kredinin maddi güvencesini oluşturan teminatların paraya çevrilmesi ile geri ödemenin yapılması, varlıkların satılması yöntemi kapsamında değerlendirilebilir. Ancak bu yöntem, gerek kredi veren gerekse kredi talep edenlerce arzulanan bir yöntem olmamakla birlikte, söz konusu yöntemin bazı sakıncaları bulunmaktadır (Bektöre, Çömlekçi ve Sözbilir, 2013: 238). Gerek maddi teminatların paraya çevrilmesinde gerekse hali hazırda kredi talep edenin mülkiyetinde olan varlıkların satılması aşamasında, söz konusu varlıkların likiditesinin düşük olması nedeniyle zaman kaybı olacak ve varlıklara likidite kazandırmak için satış işlemleri varlığın değerinin altında gerçekleşecektir. Ayrıca varlık satışı, kredi talep edenin mali yapısında ciddi bozulmalara neden olabilecektir. Söz konusu varlıkların fon sağlayıcı niteliğe sahip olduğu durumlarda bu sonuç kaçınılmazdır. Sıralanan hususlardan daha önemlisi, kredi kuruluşunun kredi taleplerini yeterince iyi değerlendirmedeği, akabinde başarısız kişi ya da firmaların fonlanarak kısıtlı fonların israf edildiği düşüncesinin ortaya çıkmasıdır. Çünkü kredinin geri ödemesi, kredi kullanan tarafın olağan faaliyetleri neticesinde elde ettiği fonlar ile gerçekleştirilmelidir. Varlık satışlarına giden kredi kuruluşlarının, piyasada olumsuz izlenim bırakması da bu yöntemin bir diğer olumsuz sonucudur (Akgüç, 2014: 17).

Borcun borçla finansmanı, kredi talep eden tarafın aslında geri dönüşü oldukça zor olan bir kısır döngüye girdiğinin göstergesidir. Kullanılmış olan kredinin hali hazırda işlemiş faiz, komisyon vb. unsurları da dikkate alındığında borç toplamı

ödenmesi zorlaşan bir seviyeye ulaşmış olmaktadır. Bu seviyede yeniden borçlanmak daha yüksek tutarda geri ödemeleri gerektirdiğinden, ilk kredi tahsil edilmiş gibi görünse de ikinci krediyi ister aynı ister başka bir kredi kuruluşu sağlasın, fonların israfına doğru sürüklenmesi çoğu zaman kaçınılmaz olmaktadır.

Kredi borçlusu kişi ya da firmaların faaliyetleri neticesinde elde ettikleri kaynaklar farklı etmenlere bağlı olmaktadır. Gerçek kişiler elde ettikleri ücret gelirleri ya da bireysel ticari faaliyetleri neticesinde fon sağlayabilmekte, tüzel kişiler ise faaliyet konularında mal/hizmet satışı ile gelir elde etmektedirler. Bireysel kredilerde şahsın unvanı, eğitim durumu, sahip olduğu özel bilgi, beceri ve tecrübeler ile çalıştığı kurum ya da firmalar gelecekte elde edilecek geliri şekillendirmektedir. Ticari nitelikteki kredilerde ise faaliyetlerin yürütüldüğü yerler, üretime konu mal ya da hizmetin kalitesi, ürün çeşitliliği, sektördeki rekabet koşulları, pazarlama politikaları ve personelin yetenekleri gibi çok sayıda unsur satış hacmini etkilemektedir. Ancak satış hacmi ya da net satışlar gelecekte elde edilmesi planlanan fon tutarı için tek başına yeterli bir ölçüt olmayacaktır. Kişi ya da firmaların giderleri de dikkate alınmak kaydıyla net gelirin hesaplanması gerekecektir. Bireylerin harcama alışkanlıkları ve sabit giderleri kişiden kişiye değişebileceği gibi firmaların mal/hizmet üretimlerinde katlandıkları maliyetler de ciddi farklılıklar gösterebilmektedir. Nihai olarak gelecekte elde edilmesi planlanan kaynaklar değerlendirilirken, salt gelir-gider dengesinin ötesinde kredi talep edenlerin diğer yükümlülüklerinin de hesaba katılması gerekir. Bu nedenle, yaratılması planlanan kaynaklardan, diğer yükümlükler çıkarıldıktan sonra ve belirli bir tutarda emniyet payı bırakılarak kredi talep eden tarafın fon yaratma becerisi değerlendirilmelidir (Akgüç, 2014: 18).

Kişi ya da firmaların fon yaratma becerisi değerlendirilirken, ekonomik hayattaki değişimler de göz önünde bulundurulmalıdır. Ekonomideki genel seyrin olumlu olduğu, piyasada yatırım konusundaki isteğin fazla olduğu dönemler ile ekonomik durgunluğun ve haliyle yatırımlarda düşüşün hâkim olduğu dönemlerin, kredi analizine birlikte dahil edilmesi, her iki durumda kredi talep eden tarafın fon yaratma becerisinin değerlendirilmesi için daha gerçekçi sonuçlar elde edilmesini sağlayacaktır (Tunay, 2016: 26).

Mali faktörlerin ikincisi olarak nitelendirilen öz kaynak yeterliliği, iktisadi faaliyetlerin sürdürülmesi için kaynak sağlayabildiği gibi firmanın alacaklıları için de bir güvence oluşturmaktadır (Reis ve Kötüoğlu, 2016: 101). Bu noktadan hareketle, öz

kaynakları güçlü olan firmanın, krediye daha kolay ve daha düşük maliyetle erişebildiği söylenebilmektedir.

1.2.3. Ekonomik Faktörler

Kredi borçlusunun fon yaratma kapasitesini etkilemekle birlikte, borçlunun kontrolünde olmayan faktörler bu başlık altında değerlendirilmektedir. Ekonomideki değişimlerin kişi ya da firmalar üzerindeki etkisi farklılık gösterebilmektedir. Özel sektörde ya da herhangi bir kamu kurumunda çalışan şahısların, kriz dönemlerinde elde edebilecekleri ücret gelirlerindeki değişim tutarları ciddi ölçüde farklılık gösterebilecektir. Diğer yandan, talebin gelir esnekliğinin yüksek olduğu mal ya da hizmetleri üreten firmaların gelirleri, ekonomideki dalgalanmalara paralel bir seyir izlemektedir (Akbaş, Bilgiç ve Miran, 2008: 55). Firmalar açısından, ekonomideki dalgalanmalar üretilen mal ya da hizmetin fiyatını etkileyebileceği gibi salt satış hacmi açısından da değişimlere sebebiyet verebilmektedir. Fiyatların yapışkan olduğu sektörlerde, ekonomik dalgalanmalar fiyatlarda ciddi değişimlere neden olmaz iken firmaların satış hacimlerinde daralmaya sebep olabilmesi nedeniyle elde edilen gelir toplamında azalmalar görülebilmektedir (Akgüç, 2014: 20). Bu durumdan ötürü, kredi talep eden kişi ya da firmanın gelir sağladığı sektörün, ekonomik değişimlere ne ölçüde duyarlı olduğu gerek fiyatlar seviyesi gerekse sektör satış hacmi açısından dikkate alınmalıdır.

Kredi analizi yapılırken, kredi talep edenin faaliyette bulunduğu sektörün/endüstrinin geniş çerçevede ele alınması gerekmektedir. Endüstrideki ürünlerin karakteristik özellikleri, müşteri profili, sektörün büyüme hızı, rekabet şekli, sektöre has teknolojik gelişmeler, sektördeki kâr oranları ve bu oranların artış hızı vb. unsurlar analiz çalışmaları esnasında, kredi talep edenin gelecek fon akımlarının değerlendirilmesi için kullanılmaktadır (İskender: 2014: 15-16). Hızla gelişen yeni sektörlerde yüksek oranlarda kâr edilmesinin mümkün olması gibi, rekabete karşı korunmuş endüstrilerde, kredi kuruluşları gelecekteki nakit akımlarından daha emin bir şekilde kredi onay süreçlerini yürütebilmektedirler. Öte yandan faaliyet konusunu oluşturan endüstride teknolojinin hızla gelişmesi ve değişmesinin yanı sıra rakip ikame malların ortaya çıkması, geleceğe ilişkin farklı risklerin doğmasına da sebep olabilecektir (Karacaoğlu, 2009: 167).

Ekonomik faktörleri büyük ölçüde şekillendiren, hükümetlerin ekonomiye ilişkin karar ve uygulamaları olmaktadır. Vergi kanunlarında yapılan değişiklikler başta olmak üzere, dış ticaret rejimlerindeki uygulamalar, teşvik kararları, taban-tavan fiyat uygulamaları, faiz hadlerinin indirilmesi ya da yükseltilmesi gibi ekonominin genelini ilgilendiren kararlar, endüstri bazında olumlu ya da olumsuz gelişmeleri tetikleyebilmektedir (Eşsiz ve Karabulut, 2018: 662). Bu durumdan ötürü, hali hazırda hükümetler tarafından yürütülmekte olan ekonomi politikalarının, genel ekonomiye yansımaları ile sektör bazında ortaya çıkaracağı olumlu ya da olumsuz gelişmelerin de kredi analizinde göz önünde bulundurulmasında fayda bulunmaktadır.

1.2.4. Yönetim Performansı

Bireysel anlamda şahısların sosyal ve ekonomik planlamalarının, gelecekte elde edecekleri kaynakları etkilemesi söz konusu olduğu gibi kurumsal anlamda yapılan planlamaların işletmelerin gelecekteki nakit akımlarını belirlediği bir gerçektir. Dolayısıyla kredi kuruluşlarınca söz konusu tercih ve planlamaların daha detaylı ele alınması önem arz etmektedir. Planlar, firma yöneticileri tarafından şekillendirildiğinden, yöneticilerin değerlendirilmesi yapılmaktadır. Yönetici değerlendirmeleri daha çok subjektif öğeler içermekle birlikte bu konuda genel kabul görmüş standartların olduğunu söylemek mümkün görünmemektedir. Ancak yönetim kalitesini ve performansını etkileyen bazı unsurları aşağıdaki şekilde sıralamak mümkündür (Akman, Özkan ve Eriş, 2008: 97-98):

- Yönetim kadrosunun yeterli sayıda olması ve bu kadroya dâhil olan kişilerin, konularının gereklerine uygun eğitim, bilgi ve tecrübeye sahip olmaları,
- Kısa vadeli günü kurtarma planlarından ziyade orta ve uzun vadeli bakış açısına sahip planların varlığı,
- Değişen ve gelişen sektöre paralel olarak yeni ürün ve hizmet arayışında olunması,
- Araştırma ve geliştirme faaliyetlerine önem verilmesi, bu faaliyetler için yeterli miktarda kaynak ayrılması,
- Kayıt dışı işlemlerin bulunmaması, mali verilerin uluslararası standartlara uygun bir şekilde gerçeği yansıtır şekilde hazırlanması,

- İç denetim bölümünün, denetim faaliyetlerini firmaya değer katacak şekilde yürütebilmesi için gerekli ortamın sağlanması,
- Etkin bir finansal yönetim ile gelecekte karşılaşılabilecek muhtemel risklere karşı mümkün olan finansal araçların kullanılması,
- Firmanın sosyal sorumluluk anlayışı çerçevesinde, projeler içerisinde yer alması ya da bu konudaki projeleri desteklemesi,
- Kurumsal yönetim prensiplerinin, en üst yönetimden en alt birimde çalışan personele kadar tüm firma çalışanlarıncaya anlaşılmasının sağlanması ve bu prensiplerin uygulanması,
- Firma çalışanlarında aidiyet duygusunun güçlü olması,
- İnsan kaynağının nitelikli hale getirilmesi ve sürekli geliştirilmesi için yatırım yapılması,
- Firmanın gelecekteki kaynak akışını etkileyebilecek ve firmayı bazı risklere maruz bırakabilecek tüm konularda, alınan kararlar için etkili kontrol sistemlerinin kurulmuş olması.

1.2.5. Hukuki Faktörler

Bireysel kredi taleplerinde, şahısların fiil ehliyetlerinin yanı sıra tarafı oldukları hukuki süreçler kredi analizinde dikkate alınması gereken hususlardır. Şahıslar hakkında yürütülen icra takipleri, maddi varlıkları dolayısıyla gelecekteki kaynak akışını etkileyebilecek hususlarda açılmış olan davalar kredi analistlerince başvurunun değerlendirilmesinde göz önünde bulundurulur (Geçer, 2014: 29). Mali konularda işlemekte olan hukuki süreçlerin yanı sıra şahsın sosyal hayatına ilişkin yürütülmekte olan davalar da yukarıda kişisel nitelikler başlığı altında temas edilen unsurlara ilişkin fikir verebilmektedir.

Kredi talep eden firma olduğunda, firmanın hukuki yapısı bazı avantaj ve dezavantajlar doğurmaktadır. Türk Ticaret Kanunu'na göre ülkemizde adi şirketlerin yanı sıra kolektif, komandit, limited ve anonim şirketler faaliyet göstermektedir. Firma sahipliğinin hukuki yapısına göre söz konusu avantaj ve dezavantajları aşağıdaki şekilde sıralamak mümkündür (Akgüç, 2014: 24):

- Tek sahipli firmalarda (özellikle adi şirketlerde) yönetim giderleri diğer şirket türlerine göre daha azdır.

- Karar alma süreçleri küçük şirketlerde daha hızlı işlemektedir.
- Şahıs firmalarında vergi yükü daha az olduğundan kredinin geri ödenmesinde kullanılacak fon büyüklüğü daha fazladır.
- Şahıs firmalarında, firmanın hukuki yapısında değişim daha kolay gerçekleştirilmektedir.
- Şahıs firmalarında, firmanın borçlarından ötürü firma sahibi tüm mal varlığıyla sınırsız bir şekilde sorumlu tutulmuştur.
- Firmanın öz kaynaklar için sermaye elde etme imkânı ortaklıklarda daha fazladır.
- Şahıs şirketlerinde hissedar ve yönetici ayrımı genelde bulunmamaktadır. Bu durum firma yönetiminin performansını olumsuz yönde etkileyebileceği gibi firmanın ömrü de tek bir şahsın durumuna göre şekillenmektedir. Ortaklıklarda ise daha profesyonel yönetim anlayışı benimsenebilmekte, firma ömrünün sınırsız olduğu varsayımı ile yönetim faaliyetleri gerçekleştirilmektedir.
- Şahıs şirketlerinde firma sahibinin kişisel yetenek ve becerileri başarıyı büyük ölçüde belirlerken, ortaklıklarda başarı çok sayıda kişinin karar alma süreçlerine katılımı ile elde edilmektedir.
- Yetenekli çalışanlar daha çok ortaklıklarda çalışmayı tercih etmekte, şahıs şirketlerinin sahip olduğu dezavantajlar onları daha az tercih edilebilir kılmaktadır.
- Şahıs firmalarının elden çıkarılması, sermayesi paylara bölünmüş şirketlere kıyasla daha zordur.

Firmaların kredi başvurularında, analizi yapan şahısların, firmanın hukuki statüsünden kaynaklı yukarıda sıralanan olumlu ve olumsuz durumları değerlendirmesi, ayrıca bireysel kredilerde olduğu gibi şirket lehine ya da aleyhine açılmış davalar ve yürütülen takipleri de dikkatle incelemesi gerekmektedir.

1.2.6. Krediyeye İlişkin Teminatlar

Kullandırılan kredilerin geri ödenmesinde arzulanan durumun, gelecekteki olağan faaliyetler neticesinde elde edilen kazançlar ile yapılması olduğu daha önceki bölümlerde ifade edilmişti. Gelecekteki belirsizliklerden kaynaklanan riskler, fonların

geri dönüşünün sağlanması için, çoğu zaman maddi ya da maddi olmayan teminatlar alınarak kredi kuruluşlarınca azaltılmaya çalışılmaktadır (Ertaş, Kaban ve Sobacı, 2016: 129). Talep edilen teminatlara ilişkin mevzuat, her bir kredi kuruluşuna göre farklılık göstermektedir. Ancak, ortak nokta alınan teminatların rayiç değerinin, kullanılan krediyi faiz ve diğer tüm masraflarıyla birlikte karşılaması ve teminatın likiditesinin mümkün olduğunca yüksek olmasıdır. Bu noktadan hareketle kredi kurumları günümüzde maddi teminatların rayiç değerlerini profesyonelce belirlemeye çalışmakta, çoğu zaman taşınmazlara ilişkin değerlemeler SPK lisanslı gayrimenkul değerlendirme uzmanlarına yaptırılmaktadır. Özellikle teminat olarak kabul edilecek taşınmazlar üzerinde (haciz, ipotek gibi) hak sınırlayıcı işlemlerin olup olmadığı ya da gelecekte taşınmazın satılması durumunda satış kabiliyetini sınırlandırıcı bazı şartların bulunup bulunmadığı analistler tarafından dikkatle incelenmesi gereken bir konu olarak karşımıza çıkmaktadır (BDDK, 2016: 15).

1.3 KREDİ RİSKİ

Kişi ya da firmaların belirli bir amaç doğrultusunda kullanmak istedikleri fonların, fon fazlasına sahip taraflarca belirli bir ücret karşılığında kullanılması, genel anlamda kredinin tanımını oluşturmaktadır (Gümrah, 2009: 31). Fazla fonların ihtiyaç sahiplerince etkin bir şekilde kullanılabilmesi ve bu işlemlerin belirli düzenlemeler altında yapılabilmesi için kredi kuruluşları oluşturulmuştur. Kredi kuruluşlarının faaliyetleri de düzenleyici ve denetleyici otoriteler tarafından gözetim altında tutularak hem fon arz edenlerin hem de fon talep edenlerin menfaatlerinin korunması amaçlanmıştır (Dayan ve Karğın, 2013: 5435). Söz konusu kredi kuruluşları, faaliyetlerini gerçekleştirirken, fon talep edenlere sunmuş oldukları imkânlar ölçüsünde belirli risklere maruz kalmakta, bu riskler de kredi talep eden taraflarca üstlenilmesi gereken maliyetleri şekillendirmektedir (Koçyiğit ve Demir, 2014: 224).

Bugünden herhangi bir yatırımın yapılabilmesi ya da herhangi bir mala sahip olunabilmesi karşılığında, gelecekte bazı yükümlülüklerin üstlenilmesi söz konusu olmaktadır. Gelecek çoğu zaman belirsizlik ve risk sözcükleriyle eş anlamlı olarak değerlendirildiğinden, gerek fon arz edenler gerekse fonları ihtiyaç sahipleriyle buluşturan kredi kuruluşları, bu belirsizlikleri mümkün olduğunca aydınlatma,

gelecekteki riskleri olabildiğinde tespit ederek gerekli önlemleri alma eğilimindedirler. Kredi işlemlerine ilişkin risklerin farklı çeşitleri bulunmakla birlikte, kredi riski denildiğinde ilk akla gelen kredi temerrüt riski olmaktadır (Şimşek, 2007: 5-6). Kredi temerrüt riski (credit default risk), borçlunun vadesinde borcun faizini ve/veya anaparasını ödeyememesi olarak tanımlanmaktadır (Oktay ve Temel, 2007: 166).

Kredi analizini etkileyen faktörler ışığında, kredi riskinin çok farklı boyutlarının olabildiği, tüm boyutlarda söz konusu riskin ortadan kaldırılabilmesinin mümkün olmadığı aşikârdır. Bu noktada önemli olan, kredinin vadesinde ödenmesine engel olabilecek faktörlerden kaynaklı risklerin mümkün olduğunca düşük seviyeye indirilmesi ve bu risklerin kontrol altına alınabilmesidir.

Risk yönetim sürecinde temelde aşağıdaki adımlar takip edilmektedir (Tanç ve Altun, 2016: 236):

- Kredi fiyatlandırılırken iki ana unsur fiyatı belirlemektedir. İlki fon arz edenlerin bekledikleri getiri yani fon maliyeti, diğer ise risk primidir. Kredinin geri ödenme riski düştüğünde, söz konusu risk primi de düşecek böylelikle fonlar kredi talep edene daha düşük fiyattan sunulabilecektir. Tam tersi durumda, kredinin temerrüt riskini artıran unsurlar tespit edildiğinde, kredi fiyatı daha yüksek olacaktır. Ancak, ülkemizde özellikle bireysel kredilerde, söz konusu yaklaşıma, yani temerrüt riskine göre kredi fiyatlaması yapılması uygulamasına çok fazla yer verilmediğini belirtmek mümkün olacaktır.
- Sahip olunan risk durumuna göre kredi talep edenlere bir kredi limiti belirlenmektedir. Risk derecesi düştükçe fon talep edenlere tahsis edilen kredi limiti yükselmektedir.
- Uygulamada kredinin temerrüt riskiyle doğrudan ilişkiye sahip olan unsur, kredi için talep edilecek teminatlardır. Kredi riski oldukça düşük olan tarafa sadece imza karşılığı kredi kullandırılabilmesi mümkün iken, riskin yükselmesiyle birlikte kefalet ve maddi teminatlar talep edilebilmektedir.
- Kredilerin tahsis edilmesinden sonra, kredi kuruluşlarına ciddi maliyetler yükleyebilecek bir diğer konu ise kredi izleme işlemleridir. Düşük kredi riskine sahip tarafların izleme faaliyetleri için oldukça az miktarda kaynak ayrılması söz konusu iken, yüksek kredi riskine sahip tarafların izlenmesi yüksek maliyetler içerebilmektedir. Nitekim borçlunun temerrüde düşmesi

ve yasal takip başlatılması halinde bu konuya ayrılacak/ayrılması gereken kaynak daha da yükselecektir.

Kredi riskinin ölçümü ve bu ölçüm neticesinde üstlenilen riskleri karşılar tutarda yeterli sermayenin bulundurulması bankalar için belirli yasal düzenlemelerle zorunluluk haline getirilmiştir (Horasan, 2012: 213). Ayrılan karşılıklar, risklerin gerçekleşmesi halinde bankaların fon arz edenlere karşı sahip oldukları sorumlulukları yerine getirebilmelerini sağlamaktadır.

Sonuç olarak kredi riskinin ölçülmesi ile kredi portföyü sağlam bir yapıya sahip olacak, bankaların sermaye yeterliliği riski karşılayabilecek düzeyde korunacak, fon arz edenlerin tasarrufları ve bekledikleri getiriler güvence altına alınacak, fonların daha etkin kullanımı sağlanarak ekonomik kalkınmanın sürdürülmesine katkı sağlanacaktır (Balkaş, 2004: 7-8).

1.3.1. Kredi Riskinin Ölçülmesi

Kredi riskinin ölçülmesinde çok çeşitli yöntemler bulunmakla birlikte söz konusu yaklaşımları temelde iki ana başlık altında toplamak mümkündür. Kredi notlaması (credit scoring) ve kredi derecelendirme (credit rating) olarak adlandırılan bu yöntemler aşağıda sırasıyla ele alınacaktır.

1.3.1.1. Kredi Notlaması

Kredi notlaması, kredi talep eden tarafların analizde değerlendirmeye tabi tutulacak özelliklerinin dikkate alınmasıyla, kredinin temerrüt riskini temsil eden kredi notunun hesaplandığı ve yapılan bu hesaplamaların matematiksel modellere dayandığı yöntem olarak tanımlanmaktadır (Özden, 2010: 135).

Kredi notlaması grubunda değerlendirilen matematiksel modellerin tamamının ortak noktası temerrüt riski olasılığının (probability of default- PD) hesaplanmasıdır. Özellikle bankaların sermaye yeterliliklerinin değerlendirilmesinde dikkate alınan ve temerrüt riski olasılığının yer aldığı beklenen kayıp (expected loss- EL) denklemi aşağıdaki şekilde gösterilmektedir (TBB Çalışma Grubu, 2006: 33).

$$EL = PD * LGD * EAD$$

Denkleimde “EL-Expected Loss” beklenen kaybı, “PD-Probability of Default” temerrüt riski olasılığını, “LGD- Loss Given Default” temerrüt halinde zararı ve “EAD-Exposure at Default” temerrüt anında risk bakiyesini ifade etmektedir.

1.3.1.2. Kredi Derecelendirmesi

Borçlunun, borcunu geri ödeme isteği ve kapasitesi hakkında uzman kuruluşlarca bağımsız görüş verilmesi, kredi derecelendirmesi olarak tanımlanmaktadır (Haspolat, 2015: 6). Ortak finansal bir dilin kullanıldığı piyasalarda, kredi riskinin standart ve anlaşılır kılınması amacıyla derecelendirme işlemleri neticesinde belirlenen riskler bazı sembollere dönüştürülmüştür. Söz konusu semboller A, B, C ve bu harflerin + ya da – işaretleri kullanılarak gösterimiyle oluşturulmuştur (Bayar, 2015: 42).

Bağımsız kredi derecelendirme kuruluşlarınca gerçekleştirilen derecelendirme görüşleri, kamuya açık bilgiler ile borçlanan tarafın sunduğu özel bilgiler kullanılarak yapılan kredi analizleri neticesinde şekillendirilmektedir (Akbulak, 2012: 179).

Kredi derecelendirme daha yaygın olarak, menkul kıymet ihraç etmek suretiyle borçlanan tarafların kredi değerliliğini ölçmek için kullanılmakta, bu süreçte derecelendirmesi yapılan enstrümanlar başta tahviller, varlığa dayalı menkul kıymetler ve yatırım fonları olmak üzere birçok farklı enstrümanı içermektedir (Toraman ve Yürük, 2014: 128).

Para ve sermaye piyasalarının genişlemiş olması sebebiyle, küresel ölçekte yatırım fırsatları ortaya çıkmıştır. Birbirinden uzak coğrafyalarda, farklı ekonomik ve yasal düzenlemeler çerçevesinde faaliyet gösteren işletmelerin kredi değerliliğinin profesyonelce ortaya konulması yatırımcılar için önem arz etmektedir. Daha önce ifade edildiği üzere kamuya açık bilgiler bulunmakla birlikte özellikle araştırılması/incelenmesi ve uzmanlarca değerlendirilmesi gereken bilgiler bulunmaktadır. Söz konusu bilgilerin farklı işletmeler için eş zamanlı ve profesyonelce değerlendirilmesi ve kredi riskinin bizzat yatırımcılar tarafından ölçülmesi günümüz ekonomik dünyasında mümkün olamamaktadır. Bu durumdan ötürü, kredi derecelendirme faaliyetleri yoğun bilgi akışında kredi riskini ölçerek, yatırımcılara alacakları yatırım kararlarında ışık tutmaktadır (Demir, 2014: 1).

Ekonomilerde düzenleyici ve denetleyici kurumlar, kredi derecelendirme faaliyetlerini esas almak suretiyle belirli standartlar getirmekte, uluslararası yatırım fonları yine kredi derecelendirme görüşlerini dikkate alan yatırım kararları vermektedir. Borçlanmak isteyen taraflar için verilen notlar, “yatırım yapılabilir” ve “yatırım yapılamaz” şeklinde iki ana kategoriye ayrılmaktadır. Söz konusu kategoriler, az önce belirtildiği üzere, yatırım kararlarının verilmesinde sınırlayıcı olabilmektedir. Borçluların yer aldıkları kategori içerisinde de gruplandırmalar bulunmaktadır. Harf notlarına “+” ya da “-“ eklenerek kredi notunun rating görünümü belirlenmektedir. Söz konusu görünümler; pozitif, durağan, negatif ve gelişmekte olan şeklinde dört ana gruba ayrılmaktadır (Coşkun, 2016: 263-264). Görünüm olarak adlandırılan bu gruplandırmalar, borçlunun notunun zaman içerisinde hangi yönde değişeceği hakkında yatırımcılara fikir vermektedir.

Esas olan, asimetrik bilginin meydana getireceği kayıp ve zararları minimize ederek piyasalarda şeffaf bir ortamın oluşturulması ve ekonomik kaynakların etkin bir şekilde kullanımının sağlanmasıdır. Nitekim şeffaflığın yüksek olduğu piyasalar, yatırımcılar için daha cazip yatırım merkezleri haline gelmektedir. Ayrıca fon talep edenler, şeffaf bir piyasa ortamında rekabet içerisinde oldukları diğer fon talep edenlerin seviyesine erişmek adına daha iyi yönetim stratejileri belirlemek zorunda kalacak, bu durum da makro düzeyde ekonomik kaynakların daha etkin kullanılmasına zemin hazırlayacaktır.

Kredi derecelendirme işlemlerinin, tüzel kişilikler ya da ülkeler için yapılmakta olduğu dikkate alındığında, kredi derecelendirme, hangi kuruluş tarafından yapılırsa yapılsın temelde iki alana odaklanılmaktadır (Kayalı ve Çümen, 2011: 117).

Öncelikle, şirketlerin faaliyetlerine ilişkin güçlü ve zayıf yönlerin belirlenmesiyle derecelendirme çalışmalarına başlanmaktadır. Şirketlerin yer aldığı sektörün genel durumu, sektördeki rekabet durumu ve şirketin bu rekabet ortamında pazardan alabilmiş olduğu pay, şirketin yürütmekte olduğu işe ilişkin riskin değerlendirilmesinde kullanılmaktadır.

Şirketlerin mali durumlarının değerlendirilmesinde kullanılan finansal oranlar aracılığıyla şirketin finansal riskinin belirlenmesi, derecelendirme faaliyetlerinin diğer odak noktasıdır. Finansal oranlar, fon talep eden taraf hakkında sayısal veriler sağlıyor olmaları sebebiyle, benzer durumda olan borçluların risk açısından sıralanmasına

olanak sağlamaktadır. Söz konusu finansal oranlar, ilerleyen bölümlerde detaylı olarak ele alınacaktır.

Şirketlerin kredi derecelerini dolaylı olarak etkileyen bir diğer unsur ise ülke riskidir. Herhangi bir şirketin kredi derecelendirme notu, o şirketin faaliyet gösterdiği ülkenin kredi derecelendirme notuna göre şekillenmektedir (Ovalı, 2014: 66).

1.4 KREDİ ANALİZ SİSTEMLERİ

Gerek kredi notlaması gerekse kredi derecelendirme yaklaşımında asıl amaç, fon talebinde bulunan tarafın geri ödeme kapasitesinin tespit edilerek kredi kurumlarının ya da daha geniş anlamda fon arz edenlerin üstleneceği riskin minimize edilmesi ve ihtiyaç sahibi tarafın uygun tutar ve vadede kredilendirilmesinin sağlanmasıdır (Desterci, 2009: 34). Ülkemiz uygulamalarında, bireysel kredi talepleri için daha basit ve standart analiz yöntemleri ile karşılaşılması mümkün iken (Ertaş, Kaban ve Sobacı, 2016: 134), ticari kredi talepleri için daha kompleks ve kredi kurumları arasında farklılık arz eden yöntemlerin bulunduğunu söylemek mümkündür (Girginer, 2008: 135).

Kredi analiz sistemleri için farklı sınıflandırmalar bulunmakla birlikte, temelde bu sistemler üç ana başlık altında toplanmaktadır. Söz konusu sistemler sırasıyla, Niteliksel ve Niceliksel Sistemler ve Kredi Notlama Sistemleridir (Vurucu ve Arı, 2017: 342).

1.4.1. Kalitatif Analiz

Öznel bir yaklaşımı içerdiği adından da anlaşılın bu yöntem, finansal verilerden çok yukarıda 1.2.1.-1.2.6. bölümlerinde sıralanan faktörleri esas almaktadır. Söz konusu faktörler uluslararası bankacılık sisteminde “5-C”, ülkemiz uygulamalarında ise “5-K” olarak adlandırılmakta, kalitatif analiz aşağıda sıralanan bu 5 kritere göre yapılmaktadır (Poyraz, 2010: 34).

- Karakter (Character): Kredi isteklisinin güvenilirliği hakkında en önemli göstergelerden biri olan karakter, bizzat kredi kurumunun kendisinin sahip olduğu ya da piyasadan edindiği bilgiler ışığında fon talep edenin geri ödeme isteğinin ortaya konulmasını ifade etmektedir. Karakter, kredi

talebinde bulunan kiři ya da iřletmenin saygınlıđı, ticari gemiři, borlarına karřı duyarlılıđı ve deme alışkanlıđı gibi sbjektif faktrleri iermektedir.

- Kapasite (Capacity/Capability): Fon talep eden tarafların, geri demelerini tam ve zamanında ifa etmeleri, gelir yaratma ve bu gelir edinimine srekliyet kazandırma yetenekleriyle ilgilidir. Dolayısıyla, kredi isteklisinin olumlu bir karaktere sahip olması tek başına yeterli sayılmaz. Geri deme isteđini somut bir řekilde destekleyecek fon elde etme faaliyetlerinin de bulunması gerekmektedir. Daha nce de ifade edildiđi zere farklı řekillerde fon elde edilmesi mmkndr. Ancak geri demenin olađan faaliyetler sonucunda elde edilen kazançlar ile yapılması arzulanan durumdur. Ayrıca gelir elde etme gcnn istikrarının, geri deme planlamasıyla uyuřması da kredi kurumlarınca dikkate alınan bir durumdur.
- zsermaye (Capital): zsermaye firma sahipleri tarafından iřletmeye konulan deđerler ile firmanın zaman ierisindeki faaliyetleri sonucunda elde ettiđi ve iřletmede bırakılmıř kr toplamlarından oluřmaktadır. İřletmenin zsermayesi ile toplam borları arasındaki dengeyi gsteren kaldıra oranı, iřletmenin borlanma kapasitesinin en nemli gstergesi olarak kabul edilmektedir. Aktiflerini byk lde borla finanse eden bir iřletme, kredi kurumlarınca daha riskli olarak deđerlendirileceđinden, kredi temin etmede zorlanacaktır.
- Karřılıklar (Collaterals): Kredi kuruluřları, borlu taraftan farklı nitelikte teminat talep edebilmektedir. Sz konusu teminatların niteliđi kredi kuruluřlarının kredi politikalarına ve bu kuruluřların tabi oldukları mevzuat hkmlerine gre deđiřebildiđi gibi kredi isteklisinin sahip olduđu zelliklere gre de deđiřebilmektedir. Borlunun temerrde dřmesi halinde sz konusu teminatlara iřlerlik kazandırılarak kredinin faiz ve diđer masraflarıyla birlikte tahsil edilmesi amalanmaktadır. Bu noktada fon sađlayan taraf iin nemli olan husus, teminatın likiditesi yani mmkn olan en hızlı řekilde ve deđer kaybetmeksizin nakde evrilebilmesidir. Bu niteliđe sahip teminat sunabilen kredi isteklisinin talebinin olumlu sonulanma ihtimali daha fazla olabileceđi gibi kredinin maliyetinin de daha dřk olması beklenebilmektedir.

- Koşullar (Conditions): Kredi borçlusunun faaliyet gösterdiği ülke ekonomisi ve sektör ile uluslararası piyasalardaki gelişmeler, borçlunun gelir yaratma kapasitesini etkileyeceğinden, söz konusu unsurlar dikkatle takip edilmesi gereken kriterlerdir. Ekonomi genelinde olumlu bir seyir mevcut olsa dahi yasal düzenlemeler ya da ekonomik politikalar neticesinde sektör bazında gelir yaratma kapasitesi önemli ölçüde, olumlu ya da olumsuz yönde, değişebilmektedir. Ekonominin önemli parametrelerinden birisi olan faiz hadleri de kredi kuruluşlarının kredi politikalarını etkileyen önemli bir konudur. İlk bakışta yüksek faiz oranları kredi kurumları için yüksek kâr gibi değerlendirilse de aslında durum bundan farklıdır. Yüksek faiz oranı, kredi kullanan taraf için yüksek maliyet anlamına geldiğinden borçlu, daha kazançlı, bir diğer deyişle daha riskli projelere yatırım yapmayı tercih etmek durumunda kalacaktır. Ayrıca yüksek maliyetli kredileri üstlenmeyi daha çok riski yüksek taraflar kabul edeceğinden bankaların müşteri portföyü riski yüksek kişi ya da kuruluşlardan oluşacaktır. Sayılan durumlardan ötürü kredi kurumları faiz hadlerinin yüksek olduğu koşullarda kredi arzlarını kısımladıkları.

Literatürde 5-C olarak değerlendirilen kriterlerin sayısında, değişen koşullarla birlikte artış yaşanmış, 7-C prensibi olarak değişen kriterlerin arasına aşağıdaki unsurlar eklenmiştir (Vurucu ve Arı, 2017: 343).

- Rekabet (Competition): Kredi isteklisi firmaların, üretmiş oldukları mal ya da hizmetlerin kalitesi, fiyatı, tasarımı gibi rekabet edebilirliğini etkileyecek hususlar bulunmaktadır. Geri ödemelerin tam ve zamanında olağan faaliyetler neticesinde elde edilmiş gelire yapılabilmesi bu hususlardan önemli ölçüde etkilenmekte, firmaların rekabetçi avantajları kredi kurumlarınca dikkate alınmaktadır.
- Müşteri İlişkileri (Customer Relations): Firmadan satın almış olduğu mal ve hizmet neticesinde memnun kalan müşteriler, bir sonraki satın alma kararlarını verirken daha önce duymuş oldukları memnuniyeti göz önünde bulundurmakta, ayrıca bu memnuniyeti diğer tüketicilerle paylaşarak olumlu dışsallık sağlamaktadırlar. Gelişen pazarlama teknikleri ve yöntemleri dikkate alınsa dahi, olumlu dışsallığın satışlar üzerindeki etkisinin daha fazla olacağı aşikârdır. Müşteri ilişkileri tüm bu etkenlerle

birlikte değerlendirildiğinde, memnun müşteri kitlesine sahip olan ve bu durumu korumak ve geliştirmek için ciddi politikaları bulunan firmaların, kredi elde etmek için bir adım önde olduklarını söylemek mümkündür. Olumlu dışsallığın satışlar üzerindeki pozitif etkisinin yanı sıra tüketiciler arasındaki fisıltı pazarlamanın, firmaya pazarlama ve reklam harcamaları için ayıracağı bütçede tasarruf etme imkânı sağlayacağı beklenmektedir.

1.4.2. Kantitatif Analiz

Kişi ya da kurumların kredi taleplerini değerlendirirken, finansal verilerin esas alındığı yöntemler kantitatif analiz başlığı altında toplanmaktadır. Söz konusu yöntemler mali tablolar üzerinden yapılan analizler ile diğer sayısal analiz tekniklerini içermektedir (Vurucu ve Arı, 2017: 347). Bu başlık altında değerlendirilecek analiz türleri, kapsamına göre ve yönetime göre analiz türleri olmak üzere iki temel başlık altında toplanmaktadır. Kapsamına göre analiz türleri statik ve dinamik analiz yaklaşımlarını içermektedir. Yönetime göre analiz türlerinde ise karşılaştırmalı tablolar analizi, dikey analiz, trend analizi ve oran analizi kullanılmaktadır. Likidite oranları, kaldıraç oranları, mali yapı oranları, aktivite oranları ve kârlılık oranları, oran analizinde kullanılmakla birlikte kantitatif analiz başlığı altında yer alan analiz türleri, çalışmamızın uygulama bölümüne konu edilmediğinden bahse konu analiz türlerine dair detaylı açıklamalara tezimizde yer verilmemiştir.

1.4.3. Kredi Notlama Sistemi

Kredi notlamasında esas olan amacın, kredi kullanan tarafın temerrüt olasılığının hesaplanması olduğu daha önce ifade edilmişti. Kredi notlamasındaki modeller, borçluların sahip oldukları özellikler değerlendirilerek, temerrüt olasılığını hesaplamak ya da borçluları farklı temerrüt gruplarına ayırmak için kullanılmaktadır (Gurny, 2013: 164). Söz konusu modeller aracılığıyla, temerrüt riskine etki eden faktörlerin sayısal olarak belirlenmesi ve bu etkilerin birbiriyle karşılaştırılması mümkün olmaktadır. Daha somut ve sayısal olarak ifade edilebilen bu etkiler, kredilerin fiyatlandırılmasında ve kayıp riski çok yüksek olan kredi taleplerinin belirlenmesinde kullanılmaktadır.

Bireysel kredi talepleri için kredi notlama modellerinin dikkate aldığı özellikler arasında, düzenli gelir seviyesi, toplam mal varlığı, daha önce kullanılmış kredilerin durumu, yaş, medeni hal ve mesleki unvan gibi unsurlar yer alabilmektedir. Ticari kredi taleplerinin değerlendirilmesinde ise oran analizi bölümünde yer alan birçok oran söz konusu modeller tarafından değerlendirilmekle birlikte kalitatif bir takım unsurlar nümerik hale dönüştürülerek bu modeller tarafından kullanılabilir. Gerek bireysel gerekse ticari kredi taleplerinde, analize esas olacak veriler belirlendikten sonra, istatistiksel tekniklerle temerrüt olasılığı hesaplanmaya çalışılır.

Ülkemiz bankacılık uygulamalarında, standart bir kredi skorlama modeli henüz bulunmamaktadır (Vurucu ve Arı, 2017: 368). Her banka, ilgili düzenleyici kuruluşun belirlediği çerçeve içerisinde, kendi notlama modelini oluşturmaktadır. Ticari krediler için yukarıda kalitatif ve kantitatif sistemler başlıkları altında değerlendirilen yöntemler, banka bazında birkaç değişikliğe uğramakla birlikte genel olarak uygulanmaktadır. Bireysel kredi talepleri için, bankacılık sistemine entegre olmuş kredi kurumlarının erişebileceği ve standart hale getirilmiş kredi notu sistemi bulunmaktadır. Söz konusu sistem, “Kredi Kayıt Bürosu Bireysel Kredi Notu” olarak adlandırılmaktadır. Bireysel kredi notu, bireyin, kredi kayıt bürosu (KKB) üyesi olan kuruluştan aldığı ya da alacağı kredileri, tam ve zamanında geri ödemesini ne ölçüde yapabileceğini ortaya koyan risk göstergesi olarak tanımlanmaktadır (Kredi Kayıt Bürosu, 31.07.2018). Söz konusu bireysel kredi notu, şahsa ait olumlu ya da olumsuz bilgilerin tek bir puanda ifade edilmiş şeklindedir. Bu puan temel olarak bireylerin geçmişte kullandıkları kredileri tam ve zamanında ödeme durumlarını belirli bir skor olarak göstermektedir. Söz konusu notlama belirli aralıklarla kademelendirilmiş, böylelikle kredi kuruluşunun uygun riske uygun fiyatlandırma yapmasına ve temerrüt olasılığını en az düzeye indirebilmesine imkân tanımıştır. Genel olarak bireysel kredi notu uygulamasında, şahısların geçmiş kredi davranışları esas alınarak, gelecekteki kredi davranışları öngörülme çalışılır. Ancak hali hazırda uygulamada, diğer özellikleri istenilen biçimde olsa dahi, daha önce hiç kredi geçmişi olmayan şahıslar için, notlama sistemi ihtiyatlı davranarak bireysel kredi notunda bu durumu olumsuz bir özellik olarak değerlendirmektedir. Bireysel kredi notu sadece yeni kredi talepleri için değil, mevcut kredi borçlarının temerrüt olasılığının izlenmesinde de kullanılabilir.

1.4.3.1. Kredi Notlama Modelleri

Borçluların ya da kredi talep edenlerin, gözlemlenebilen özelliklerini baz alarak temerrüt olasılığının hesaplanmasında kullanılan kredi notlama modelleri izleyen bölümde açıklanmıştır. Söz konusu modellerde, kredinin bireysel ya da ticari niteliği dikkate alınarak, olasılık hesabında kullanılacak değişkenler kredi kurumlarınca belirlenmektedir.

1.4.3.1.1. Lineer Olasılık Modeli

Lineer olasılık modelinde, kredi isteklisinin geçmiş verileri kullanılarak, kredinin geri ödenmesinde kullanılacak değişkenlerin önem dereceleri belirlenmeye çalışılmaktadır. Geçmişte kullandırılmış olan krediler iki gruba ayrılmaktadır. Y bağımlı değişkeni, temerrüde düşen krediler için “ $Y_i=1$ ”, zamanında ödenen krediler için “ $Y_i=0$ ” şeklinde belirlenir. Söz konusu iki değeri alan Y bağımlı değişkenleri ile geçmiş kredi verileri (X_{ij}), lineer regresyon kullanılarak ilişkilendirilir (Kasapoğlu, 2009: 17). Bu temelde oluşturulan model, matematiksel olarak aşağıdaki şekilde ifade edilebilir.

$$Y_i = \sum b_j * X_{ij} + hata$$

Denklemden b_j ifadesi, j değişkeninin kredinin geri ödenmesindeki önem derecesini göstermektedir. Söz konusu denklem sayısal olarak örneklendirilirse, ticari kredi talebinde bulunan firmanın finansal borçlar/aktifler oranının ve borç yapısı oranının, kredinin geri ödenmesinde önem arz eden iki değişken olduğu kabul edilsin. Bu durumda yukarıdaki formül aşağıdaki şekilde gösterilebilecektir.

$$Y_i = 0,3 * (Finansal Borçlar/Aktifler) + 0,4 * (Borç Yapısı Oranı)$$

Söz konusu gösterimden, her iki oranın yüksek olmasının, temerrüt olasılığını artırdığı anlaşılabilir. Birlikte, her bir özellik için belirlenen katsayılar, daha önce de ifade edildiği üzere lineer regresyon aracılığıyla geçmiş veriler kullanılarak elde edilmektedir. Ancak bu gösterimle birlikte, modelin sınırları ortaya çıkmaktadır. Temerrüt olasılığını düşüren değişkenlerin, yani b_j 'nin negatif olduğu durumlarda olasılık yani Y_i 'nin hesaplanması mümkün olmayacaktır. Çünkü olasılık teorisine göre Y_i 'nin mutlak surette 0-1 aralığında değer alması gerekmektedir.

1.4.3.1.2. Logit Modeli

Söz konusu model lineer olasılık modeline benzer bir yaklaşım sergilemekle birlikte, temerrüt olasılığının fonksiyonel bir dağılım gösterdiğini kabul etmektedir. Logit model, bağımsız değişkenin sonsuza gittiği durumda, bağımlı değişkenin 1'e yakınsadığı matematiksel bir model olarak ifade edilebilir (Kaptan, 2011: 96). Söz konusu olasılık hesabı, aşağıdaki iki denklem aracılığıyla yapılmaktadır (Demirci ve Astar, 2011: 123).

$$Z_i = \alpha + \beta X_i$$

$$P_i = \frac{1}{(1 + e^{-Z_i})}$$

Z_i , lineer olasılık modeli kullanılarak regresyon aracılığıyla hesaplanan değeri ifade etmekte, P_i ise lojistik dağılım özelliğine sahip, kredinin kümülatif temerrüt olasılığını göstermektedir. Z_i ister negatif ister pozitif değer alsın, P_i her halükarda pozitif olacağından, lineer olasılık modelinin dezavantajı geride bırakılmış olarak, olasılık teorisine uygun bir şekilde $0 \leq P_i \leq 1$ koşulu elde edilmiş olacaktır. Logit dağılım fonksiyonu olarak Z_i , $-\infty$ ile ∞ arasında değerler alırken, P_i , 0 ile 1 arasında değerler almakta, söz konusu iki değer arasındaki ilişki doğrusal olmamaktadır (Albright, 16.07.2018).

1.4.3.1.3. Probit Model

Logit modelinden farklı olarak söz konusu modelde temerrüt olasılığı normal dağılım göstermektedir (Kliestik, Kocisova ve Misankova, 2015: 853). Probit modelin, çok benzer olduğu logit modelden temel farkı, birikimli dağılım fonksiyonunun normal dağılıma sahip olmasıdır (Kaptan, 2011: 97). Probit modeli, belirli bir bağımlı değişken oranına ulaşmak için gerekli olan bağımsız değişken etkisinin tahmin edilmesini sağlamaktadır (Seyitoğlu, 17.07.2018).

Probit modelinde, temerrüt olasılığını belirleyen değişkenlerin doğrusal fonksiyonu (Z), logit modeldekine benzer bir yöntemle aşağıdaki şekilde hesaplanmaktadır.

$$Z = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Z'nin herhangi bir deęeri için temerrüt olasılıęı, standart hale getirilmiş birikimli normal daęılıma sahip F(Z) fonksiyonu aracılıęıyla hesaplanmakta, $p_i=F(Z_i)$ şeklinde ifade edilen fonksiyonun normal hale getirilmesi ařaęıda gsterilen probit modeli ile saęlanmaktadır (elik, 2013: 41).

$$P(y = 1) = \phi (X_i\beta) = \int_{-\infty}^{X_i\beta} \frac{1}{2\pi} \exp\left(\frac{-z^2}{2}\right) dz$$

1.4.3.1.4. Lineer Diskriminant Modeli

Bu model, kredi talep eden tarafların zelliklerine gre, bařvuruları temerrüt riskinin yksek ve dřk olmasına gre sınıflandıran istatistiksel yntem olarak deęerlendirilmektedir. Modelin skortlama amacıyla kullanımı, Edward Altman'ın yapmıř olduęu analizde, borsada iřlem gren řirketlerin temerrüt olasılıęının hesabına dayanmaktadır. Altman yapmıř olduęu analizde 5 farklı mali oran kullanarak ok deęiřkenli diskriminant fonksiyonu elde etmiřtir (Kaptan, 2011: 98). Sz konusu fonksiyonda temerrüt olasılıęı baęımlı deęiřken olarak "Y" şeklinde belirlenmiř, řirketlerin gemiř verileri deęerlendirilerek baęımsız deęiřken olan mali oranlar iin katsayılar ařaęıdaki şekilde hesaplanmıřtır (Kasapoęlu, 2009: 18).

$$Y = 1,2X_1 + 1,4X_2 + 3,3X_3 + 0,6X_4 + 1X_5$$

X_1 = Net alıřma Sermayesi/Toplam Varlıklar

X_2 = Daęıtılmamıř Kârlar/Toplam Varlıklar

X_3 = Faiz ve Vergi ncesi Kâr/Toplam Varlıklar

X_4 = Hisse Senetlerinin Piyasa Deęeri/Toplam Borcun Defter Deęeri

X_5 = Satıřlar/Toplam Varlıklar

Sz konusu fonksiyonda, baęımlı deęiřken olarak belirlenen Y'nin almıř olduęu deęer ykseldike, firmanın temerrüt olasılıęının dřtę kabul edilmiřtir. Geliřtirilen modelde Y'nin alabileceęi deęerler iin sınıflandırma yapılmıřtır. Bu modelde $Y > 2,99$ dřk risk grubunu, $2,99 > Y > 1,81$ ortalama risk grubunu, $Y < 1,81$ yksek risk grubunu ifade etmektedir (Kaptan, 2011: 99).

Altman'ın yapmıř olduęu alıřmaya benzer şekilde, temerrüt olasılıęını en iyi şekilde yansıtabilecek baęımsız deęiřkenler belirlendikten sonra bu deęiřkenler iin

lineer diskriminant analizi ile katsayılar belirlenerek temerrüt risk gruplarının kredi analiz sürecinde oluşturulabilmesi mümkündür.

1.4.3.2. Kredi Notlama Modellerinin Zayıf Yönleri

Önceki bölümlerde sıralanan modeller, kredi riskinin değerlendirilmesinde analiz yapan taraflara belirli ölçüde ışık tutmuş olsa da söz konusu modellerin bazı eksiklikleri bulunmaktadır.

Kredi notlama için kullanılan modellerin öne çıkan en önemli eksikliği, modellerin temerrüde düşme ya da tam ve zamanında krediyi ödeme gibi ikili sonuçla ilgilenmesidir. Günümüz kredi işlemlerinde, borçluların kısmi ödemeler yaptıkları, bazı durumlarda sadece anapara ya da faiz geri ödemelerini yapabildikleri görülebilmektedir. Bu tarz durumlardan ötürü borçluların temerrüt riski açısından daha doğru bir şekilde sınıflandırılabilmesi için, analizde kullanılacak değişkenlerin sayısında ve seçiminde farklı yaklaşımlar benimsenmesi gerekmektedir (Kaptan, 2011: 99).

Söz konusu modellere getirilen bir diğer eleştiri ise modellerin değişkenler için belirlediği ağırlıkların sabit kalması ve söz konusu değişkenlerin birbirinden bağımsız olduğu kabulünün yapılmış olmasıdır. Modellerin üzerine inşa edildiği değişkenlerin önem derecesinde, zaman içerisinde değişimler olabileceği gibi söz konusu değişkenlerin birbirlerinden tamamen bağımsız olarak kabul edilmesi çok gerçekçi bir yaklaşım olarak kabul görmemektedir (Kasapoğlu, 2009: 18-19).

Ortaya çıkan bir diğer sorun ise, temerrüt riskinin belirlenmesinde önem arz etmesine rağmen, ölçümü zor olduğu için bazı değişkenlere bu modellerde yer verilmemesidir. Kalitatif analiz bölümünde değinilen borçlunun kişisel özelliklerinin yanı sıra bazı makroekonomik göstergelere modellerde yer verilmemesi, temerrüt riskinin hesaplanmasında bir eksiklik olarak değerlendirilebilmektedir (Şimşek, 2007: 47-48).

Kredi kurumlarınca sağlıklı bir model oluşturulabilmesinin önündeki bir diğer engel, işlenebilir veri setlerinin bulunmamasıdır (Şimşek, 2007: 48). Ancak söz konusu eksikliğin gelişen teknolojik alt yapı ve veri madenciliğinin artan önemi neticesinde büyük ölçüde ortadan kalktığını söylemek mümkündür.

1.5 TARIMSAL KREDİLER

Dünya genelinde ekonomik sistemlerde tarımın toplam üretim içerisindeki göreceli payının azalması, bir diğer deyişle sanayi işletmelerinin üretim içindeki paylarının hızla artması ile birlikte, tarımsal kredilerin toplam kredi portföyü içerisindeki payının azaldığı söylenebilir. Tarım kredilerinin, diğer kredilere kıyasla daha düşük bir oranda kullanıldığının bir diğer sebebi ise söz konusu kredilerin risk düzeyinin daha yüksek olmasıdır. Ülkemizde örnekleri görülmekle birlikte, tarım kredileri genelde devletlerin desteklediği bazı kredi kurumları, kooperatif bankaları ve bunlara benzer kredi kurumlarınca kullanılmakta, üreticilerin tarımsal girdileri bu kurumlarca finanse edilmektedir (Adanacıoğlu, Artukoğlu ve Güneş, 2017: 195-196).

Tarım kredilerinin diğer kredi türlerine kıyasla daha riskli olmasına neden olan bazı etmenler bulunmaktadır (Bayramoğlu, 30.07.2018). Tarım ürünlerinin birçoğunda, maliyet-fiyat ilişkisindeki bozukluk, en önemli risk etmeni olarak ortaya çıkmaktadır. Üretimin arttığı sezonlarda, söz konusu ürünlere olan talep aynı hızda artmadığından tarımsal ürünlerin piyasa fiyatı düşmekte, bu durumda üreticinin gelirinde (kârında) azalma olduğundan kredilerin geri ödenmesinde sıkıntılar yaşanmaktadır. Diğer yandan çiftçilerin fiyat düşüşüne karşı verebilecekleri tepki diğer sektörlerle kıyasla oldukça sınırlı olmaktadır. Örneğin sanayi işletmeleri, fiyatlar düştüğünde üretim planlamalarında hızla değişiklik yapabilmelerine rağmen, tarımsal üretim yapan kişi ya da firmalar, üretim faaliyetlerini büyük ölçüde değiştirememektedir. Çünkü tarımsal üretim uzun bir süreç olmakla birlikte esneklikten yoksundur. Tarlasını herhangi bir ürün için hazırlamış, bu ürüne göre girdilerini kullanmış, ürün alma aşamasına yaklaşan bir çiftçinin, sarf ettiği tüm bu enerjiyi ve maliyetleri yok sayıp yeniden farklı bir ürün için üretim çabasına girmesi nerdeyse imkânsızdır. Bu konuda coğrafi ve iklimsel özellikler, bazı yörelerde sezonda birden fazla ürün için hasat yapma imkânı sağladığından çiftçiler üretimlerini, dolayısıyla gelirlerini çeşitlendirebilmektedir. Ancak söz konusu durum, tarımsal üretimin gerçekleştirildiği sınırlı bölgelerde ortaya çıkabilmektedir. Netice itibariyle, tarım ürünlerinin arz ve talebinin esnek olmamasından ötürü ortaya çıkan gelir dengesizlikleri üreticilerin mali imkânlarında da dalgalanmalara sebebiyet verdiği için tarımsal kredilerin riski daha fazla olmaktadır. Söz konusu gelir dalgalanmalarının önüne geçmek adına, devletler bazı destekleme uygulamalarında bulunmakta, taban fiyat uygulamaları, belirli fiyattan alım garantisi gibi olanaklar sunarak genel fiyatlar

seviyesini düzenlemeye çalışmakta, böylelikle üreticilerin maliyetleri ve gelirleri arasındaki makasın üreticiler aleyhine daralmasını sınırlandırmaya çalışmaktadırlar (Hayran ve Gül, 2018: 272).

Tarımsal üretimin kredilendirilmesini riskli kılan bir diğer durum, söz konusu faaliyetlerin büyük ölçüde kontrolü mümkün olmayan hava koşullarına bağlı olmasıdır. Mekanizasyon, bitki besleme ve koruma ürünlerinin kullanılması, sulama olanaklarında artış gerçekleşmesine rağmen üretim büyük ölçüde hava koşullarına bağlı olmakta, söz konusu koşullar üretimi büyük sekteye uğratabileceği gibi üretimin zamanlamasında da değişime sebep olabileceğinden gelir dalgalanmaları kaçınılmaz olmaktadır. Yine büyük ölçüde hava şartlarına bağlı olmakla birlikte bazı doğal afetlerin tarımsal üretime etkisi yıkıcı olabilmektedir. Kuraklık, don, aşırı yağışlar ve sonrasında yaşanan sel ve su baskınları gibi doğal yıkımlar üretime konu olan arazilerden gelir elde edilmesini olanaksız hale getirmektedir. Hava koşullarına bağlı olumsuzlukların öngörülmesinde yaşanan güçlükler, kredi kurumlarının tarımsal kredilerin analizinde karşılaştıkları en önemli zorluk olarak değerlendirilebilmektedir (Ünlüer ve Güneş, 2013: 92).

Ülkemiz uygulamalarında sıklıkla karşılaşıldığı üzere tarımsal üretimde küçük aile işletmeleri yaygındır. Bu durumda, kullandırılan kredilerin, tarımsal faaliyetlerin finansmanı yerine ailenin tüketim harcamaları için kullanılması söz konusu olabilmektedir. Tarım işletmelerinde yeterli ve düzenli bir kayıt sisteminin bulunmaması, toplanan verilerin ise amaca hizmet eder şekilde işlenmemesi gibi sebeplerden ötürü tarımsal kredilerin izlenmesinde de zorluklar yaşanmakta, kullandırılan kredilerin hangi amaçla üreticiler tarafından değerlendirildiği sağlıklı bir şekilde belirlenememektedir (Yurtoğlu, 2015: 28). Tarımsal kredilerin, üretimi ya da geliri artırıcı faaliyetler dışında kullanılmış olması, kredinin tam ve zamanında geri ödenme ihtimalini oldukça düşürecektir.

Kredi analizi yapılabilmesi için borçlunun gözlemlenebilen ve kredinin ödenmesine etki edebilecek tüm özelliklerinin analize dahil edilmesi gerektiği daha önce belirtilmişti. Bu analizin tarımsal kredi kullanmak isteyen çiftçiler için yapılabilmesini zorlaştıran bazı uygulamalar bulunmaktadır. Çoğu ülkede çiftçiler düzenli olarak kayıt tutmamakta, mali tablo düzenleme zorunluluğu bulunmamakta, birçok yasal çerçeve de bu duruma zemin hazırlamaktadır (Günden ve Miran, 2008: 74). Söz konusu şartlar, çiftçilerin sosyal alışkanlıkları ile birleştiğinde, üreticilerin

gelir yaratan faaliyetlerinin tam manasıyla değerlendirilebilmesi çok sık bir şekilde gerçekleştirilebilmektedir.

Kredi başvurusunda bulunan birçok taraf, kredi sözleşmelerinin gereklerini tam ve zamanında yerine getirme konusunda bilinçli olmakla birlikte söz konusu yükümlüklerin yerine getirilmemesi halinde karşılaşılabilecekleri yaptırımlar hakkında endişe taşımaktadır. Ancak çiftçiler, mali olanakları imkân verse dahi kredi kurumlarına olan borçlarını tam ve zamanında ifa etmekte özensiz davranabilmekte, tarımsal faaliyetler neticesinde elde ettikleri gelirleri öncelikli olarak şahsi harcamalarda kullanabilmektedirler. Bu genel tutum, çiftçilerin geçmişteki ödeme alışkanlıklarının değerlendirilmesini oldukça önemli kılmaktadır.

Diğer kredi uygulamalarında da olduğu gibi tam ve zamanında ödenmeyen krediler için yasal yaptırımlar bulunmaktadır. Ancak bu yaptırımların uygulanması, politik ve bazı toplumsal gerekçeler nedeniyle ötelenebilmekte, bu durum da çiftçilerin ödeme alışkanlıklarında ciddi bozulmalara sebep olduğu gibi gelecekteki gelir akımlarında geri dönüşü oldukça zor bozulmalara neden olmaktadır. Ayrıca, koşulları gerçekleşmemiş ya da kısmen gerçekleşmiş olduğu halde, yasal alt yapısı hazırlanarak devlet eliyle gerçekleştirilen erteleme ve uzun vadeli yapılandırma uygulamaları kredilerin geri ödenmesinde olumsuzluklara neden olmaktadır.

Tarımsal kredilerin, özel bankaların yanı sıra, devlet destekli bazı kredi kurumlarınca kullandırıldığı ifade edilmişti. Söz konusu kurumların yönetiminde, kredi kuruluşu olmanın gerektirdiği disiplinden uzaklaşılması, politik müdahalelere bu tarz kurumların açık olması, gelişen ve değişen ekonomik ve teknolojik sisteme adapte olunamaması gibi nedenlerle anılan kurum ve kuruluşlar tarımın finansmanında yeterince etkin olamamaktadır. Bu nitelikteki işleyiş, tarımın kredilendirilmesi/finansmanı için ayrılan fonların israf edilmesine, bunun neticesinde tarım sektörünün giderek küçülmesine neden olmaktadır.

2. BÖLÜM

MAKİNE ÖĞRENMESİ

2.1 VERİ İŞLEMENİN GEREKLİLİĞİ

Sanayi devrimiyle birlikte insanoğlunun, üretim miktar ve hızında meydana gelen olağanüstü artış, hayatın tüm alanlarını takip eden zaman içerisinde etkilemiş, makine kullanımını yaşamın hemen hemen her alanı için zaruri bir hal almıştır. Başlangıçta üretimin niceliği ve niteliği üzerine kaydedilen önemli gelişmelerle insan merakı ve zekâsı kısa süreliğine tatmin edilmiş olsa da, ilerleyen zaman içerisinde, doğasının bir gereği olarak, insanlığın aklını cevaplandırılması gereken yeni sorular ve ulaşılmaması gereken yeni amaçlar meşgul etmeye devam etmiştir. Söz konusu merak ile daha fazlasını bilme ya da daha iyisini yapma arayışı, insan zekâsının ve becerisinin ufuklarının keşfedilmesini ya da en azından bu ufuklar üzerine düşünülme başlanmasını sağlamıştır. Kendi zekâ ve beceri hudutlarında en iyisini yapmaya çalışan insanlık, yine doğasının bir parçası olan kısıtlarından ötürü önüne çıkan setleri aşmak için kendisini bu setlerden kurtaracak daha farklı bir düşünsel dünyaya adım atmıştır. Kapıları aralanan bu farklı düşünsel dünyada, insanoğlu en kompleks matematiksel hesaplamaları yapabildiğini, ancak bu hesaplamaları aynı anda ve kısa sürede yapmadan, arzuladığı amaca ulaşamayacağını yani zayıflıklarını fark etmiştir. Dolayısıyla insan, kendisine benzer bir çalışma sistematiği olan ancak kendi zayıflıklarını bünyesinde barındırmayan bir makine yapma düşüncesini bu yeni dünyada geliştirmiştir. İmal edilen bu makineler, düşünce dünyası ve onun sonuçlarını öncelikle basit seviyede kopyalayabilmişse de ilerleyen zaman içerisinde bilgisayar olarak adlandırılan makineler geliştirilmiş, bu yeni makineler 20. YY.'ın sonuna doğru, insanlığı farklı düşünsel dünyalara sokacak sanal âlemle birlikte daha fazla ve daha farklı işlevler barındırmaya başlamıştır.

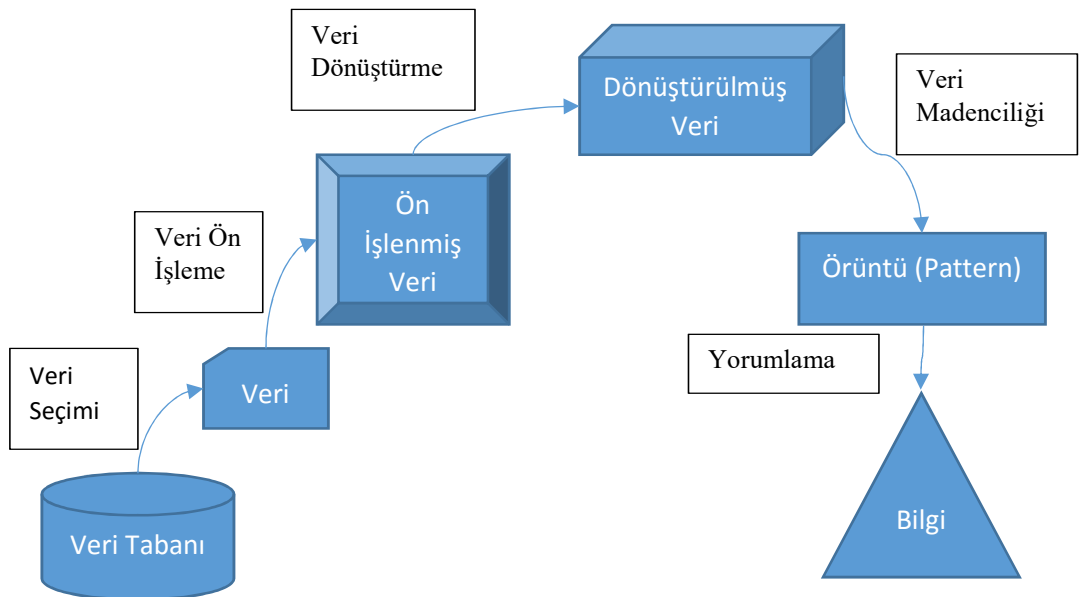
Sanal âlemle birlikte, dünyanın hemen hemen her yeri birbirine bağlanmış, hayatın tüm alanları bir şekilde kesişir hale gelmiştir. Böylesine devasa bir ilişkiler âlemi haliyle yeni sorular ve yeni fikirlerin doğmasına sebebiyet vermiştir. Hayatın her alanında daha iyisini daha kolay yapmayı amaçlayan insan, karşısında bir veri yağmuru bulmuş, söz konusu verileri anlamlandırma yani bilgiye dönüştürme hususunda yine darboğaza girmiştir.

Veri, herhangi bir işleme maruz tutulmadan, gözlem ve ölçüm yapılarak elde edilmiş değer olarak tanımlanmaktadır (Şeker, 2013: 22). Ancak veri dağınık, karmaşık ve kaotik bir yapıda bulunduğundan, söz konusu yapıya bir düzen kazandırılması, bilgiye ulaşılmasının ön koşuludur (Gürsakal, 2014: 35). Bu kaotik yapının anlamlı bir hale dönüştürülmesi için, insanlığın bu kez makinelerin hudutları üzerine düşünmesi ve makineleri bu hudutları aşacak şekilde geliştirmesi gerekmiştir. Geline nokta, bilimsel gelişmelerin ve teknolojik alt yapının insanlığa sağladığı avantajlar, verilerin anlamlandırılması yolunda insanlığa ve makinelere büyük hız kazandırmıştır. Ancak farklı türdeki verilerle birlikte veri dünyasının kaotik yapısının henüz keşfedilmemiş noktaları, gerek insanların gerekse makinelerin çalışmalarını ve gelişmelerini devam ettirmelerindeki temel motivasyon olarak karşımıza çıkmaktadır. Söz konusu çalışmalar veri setinden, amaca hizmet eder, daha etkin ve farklı sonuçlar çıkarabilmek için sürdürülmektedir.

2.2 VERİ MADENCİLİĞİ

Sahip olunması arzulanan bilgi için, gelişen teknoloji ve sanal dünya platformu ile birlikte, çok sayıda işleme tabi tutulması gereken veri her gün üretilmektedir. Çok farklı disiplinlerde ve çok farklı amaçlarla bilgiye ulaşmak isteyen tarafların, veriyi bilgiye dönüştürme aşaması temel olarak aşağıdaki Şekil 1’de tasvir edilmiştir (Şeker, 2013: 27).

Şekil 1: Veri Tabanlarında Bilgi Keşfi



Veri tabanlarında bilgi keşfi olarak adlandırılan bu süreçte öncelikle verilerin tutulduğu veri tabanından, ulaşılmak istenen bilgiyle ilgisi olduğu düşünülen veriler seçilerek işleme başlanılır. Söz konusu seçim, bilgisayarın tespit edeceği doğrudan ilişkilere göre yapılabileceği gibi buna ilave olarak uzman kişilerin, kendi bilgi ve tecrübeleri ışığında belirlediği kriterlere göre de yapılabilmektedir. Ancak şunu eklemekte fayda vardır ki bilgisayarların veri-bilgi arasındaki ilişkilere göre yapmaya çalıştığı “objektif” seçimler ile uzman kişilerin “sübjektif” seçimleri, istenilen bilgiye ulaşılmak için gerekli tüm verileri içermeyebilir.

Veri seçiminin ardından, veriler arasında yer alabilecek bazı olumsuzlukların düzeltilmesi gerekebilmektedir. Örneğin, kredi başvurusunda bulunmuş bazı şahıslara ait teminat bilgilerinin girilmemiş olması ya da kredi başvurusunda bulunmuş şahsın evli ve 3 çocuklu olduğu belirtilmesine rağmen yaşının veri setinde 8 olması gibi durumlarda, söz konusu kayıp ve hatalı verilerin tespit edilmesi, mümkünse bunların düzeltilmesi gerekmektedir. Düzeltme işleminden ziyade doğrudan bu verilerin silinmesi, ulaşılmak istenen bilgiye giden yolda, geliştirilecek modelin (örüntünün) şekillendirilmesinde bazı hatalara ya da noksanlıklara neden olabilecektir.

Üçüncü olarak, düzeltilmiş veri setinde gereken dönüşüm işlemleri yapılarak veri setinin, veri madenciliğine uygun hale getirilmesi sağlanır.

Çok sayıda verinin, bu veri seti içerisindeki örüntüleri ve kuralları keşfetmek ve analiz etmek amacıyla işlenmesi olarak adlandırılan veri madenciliği sürecinde ise veri madenciliğinde kullanılan bazı yöntemler ile ulaşılmak istenen bilgiye hizmet edecek anlamlı sonuçların ortaya konulması hedeflenir (Berry ve Linoff, 2004: 7).

Son aşamada, çıkarılan örüntünün anlamı sorgulanmaktadır (Wong ve Leung, 2002: 2). Satış ve pazarlama örnekleri incelenecek olursa, belirli bir ürünü (örneğin kolay) satın alanların çoğunun, başka bir ürünü (patates cipsini) satın aldıkları gibi anlamlı bir sonuç elde edilebilmektedir. Ancak gazlı içecek alanların duvar boyası da aldıkları yönünde bir örüntü ortaya çıkmış ise daha önceki aşamaların tekrar gözden geçirilmesi gerekmektedir. Fakat varılan bu sonuç kendi içerisinde anlamlı olabilir. Daha önceki aşamalarda herhangi bir hata ya da eksiklik tespit edilememiş ise ilk bakışta anlamsız görünen bu sonucun anlaşılması için sonuca dair daha detaylı araştırmalar yapılabilir. Böylelikle daha önce hiç düşünülmemiş ya da fark edilmemiş gerçeklerin (anlamaların) ortaya çıkması sağlanabilir.

Bilgiye giden yolun tasvir edildiği yukarıdaki aşamalarda, temel işlem örüntünün ortaya çıkarılmasını sağlayan veri madenciliğidir. Farklı disiplinlerde farklı bilgiye ulaşmak amacıyla temelde 6 farklı işlem kullanılmakta, bu işlevlerden bazıları doğrudan, bazıları dolaylı bazıları ise hem doğrudan hem de dolaylı şekilde veri madenciliği ile ilgili olmaktadır (Berry ve Linoff, 2004: 8-12). Söz konusu 6 işlem aşağıdaki Tablo 1’de kısaca özetlenmiştir.

Tablo 1: Veri Madenciliğindeki Temel İşlevler

İşlevin Adı	Açıklama
Sınıflandırma	Yeni örneğin, söz konusu örneğin özelliklerine göre, önceden belirlenmiş sınıflardan uygun olana atanmasıdır. Eldeki örneklerle geliştirilen model sayesinde yeni örneğin sınıflandırması yapılmaktadır.
Tahmin	Sınıflandırmanın aksine, tahminde sürekli değerlere sahip çıktılar elde edilmekte, girdi verileri esas alınarak bilinmeyen sürekli değişken için bir değer tahmini yapılmaktadır.
Öngörü	Öngörü, sınıflandırma ve tahmin işlevlerine benzese de örnekler bazı muhtemel gelecekteki davranışlara ya da değerlere göre sınıflandırılmaktadır. Doğru sınıflandırmayı kontrol etmenin tek yolu bekleyip görmektir.
Benzerlik Gruplaması	Bu işlevin temel amacı, hangi örneklerin hangileriyle birlikte hareket ettiğini tespit etmektir.
Kümeleme	Heterojen yapıya sahip veri setini, homojen alt kümelere bölmeyi amaçlayan işlevdir.
Profil Çıkarma	Belirli bir sınıfa ait örneklerin sahip oldukları karakteristik özelliklerin belirlenerek, bu özelliklere sahip birimlerin neden o sınıfta birleştiklerinin ortaya konulmasıdır.

2.3 MAKİNE ÖĞRENMESİNE İLİŞKİN TEMEL KAVRAMLAR

Bilgisayar teknolojisindeki hızlı değişim ve ilerlemelerle birlikte yüksek miktarda verinin güvenilir bir şekilde depolanma imkânı elde edilmiştir. Daha önceki

bölümde de ifade ettiğimiz üzere, oluşturulan sanal dünya, depolanan bu verilere fiziksel şartlar gözetilmeksizin ulaşım imkânı sağlamıştır. En kısa zaman diliminde dahi dünya üzerinde üretilen veri katlanarak artmakta, gözlemlenen verinin altında yatan bir süreç bulunmakta, bu sürecin tamamen ortaya çıkarılması mümkün olamamakla birlikte, süreci anlamaya yaklaşacak bir model inşası hedeflenmektedir (Alpaydın, 2004: 1-2). Söz konusu sürece dair yakınsamalar, veri içerisindeki modelleri ve düzenleri anlayabilmek için önem arz etmektedir. Verinin akış düzenini sağlayan yapıya ait bu model ve düzenlere dair edinilen fikirler ve yakınsamalar geleceğe dair öngöründe bulunulmasına imkân sağlamaktadır.

Verinin anlaşılabilir aksiyonlara dönüştürülmesini sağlayan, bilgisayar algoritmalarının geliştirilmesine dayalı çalışma alanı, makine öğrenmesi olarak tanımlanmaktadır (Cui, Wong ve Lui, 2006: 598). Söz konusu çalışma alanı; ulaşılabilir veri, istatistiksel metotlar ve gelişen hızlı hesaplama gücü olmak üzere bu 3 temel sac ayak üzerinde yükselmektedir (Lantz, 2013: 7). Söz konusu 3 temel dayanak, birbirinin gelişimine bağlı olarak beraber büyümektedir.

Makine öğrenmesiyle çok benzer temeller üzerine oturmuş veri madenciliğinde, büyük veri setlerinden yeni ya da farklı anlayışlar elde edilmesi amaçlanmakta, çoğu zaman bu iki benzer alanın ne denli örtüştüğü tartışılmaktadır. Temelde bu iki alanın ayrıldığı nokta; makine öğrenmesi tanımlanmış bir görevin başarılmasına odaklanırken, veri madenciliği veri içerisinde gizli olan ve o verinin sınırlarını barındıran bilgilere ulaşmayı hedeflemektedir (Gerlein, McGinnity, Belatreche ve Coleman, 2016: 193). Örneğin, makine öğrenmesi ile bir robota nasıl araba kullanılacağı öğretilirken, veri madenciliği ile hangi arabaların daha güvenli olduğu bulunabilmektedir (Lantz, 2013: 7). Söz konusu örnekten ve yapılan açıklamalardan da anlaşılacağı üzere makine öğrenmesi bir veri tabanı problemi olmaktan çok, yapay zekanın bir parçası olarak değerlendirilmelidir. Zekâdan söz edilebilmesi için zekâ sahibinin çevresindeki dünyayı algılayabilmesi ve bu dünyaya ilişkin rutinleri öğrenebilmesi gerekmektedir (Russell ve Norvig, 2003: 3). Çevrenin algılanması ve öğrenilmesiyle birlikte, karşılaşılan yeni durumlara dair tahlil yapılması ve çıkarımda bulunulması mümkün olmaktadır.

Makine öğrenmesi, örnek veri seti ya da geçmiş veriler üzerinden bir performans kriterinin optimize edilmesi için bilgisayarları programlamaktadır (Alpaydın, 2004: 3). Söz konusu örnek ya da geçmiş verilere ait parametreleri optimize

eden bir model geliştirilmekte, söz konusu model tahminler yapılabilmesi için öngörü temelli ya da veri hakkında bilgi sağlamak adına tanımlama temelli, bazen her ikisini de kapsayıcı olabilmektedir (Canepa, 2016: 4).

Öğrenme işlemini gerçekleştirecek olan ister insan olsun ister makine, süreç aşağıdaki üç temel adımda gerçekleşmektedir (Shwartz ve David, 2014: 19-20):

- Girdi veriler: Bu aşamada verilerin gözlenmesi, hafızaya alınması ve daha sonraki değerlendirme için gerekli olan anımsama sağlanır.
- Çıkarım: Verinin daha geniş anlamda sunuma imkân sağlayacak şekilde dönüştürülmesidir.
- Genelleştirme: Aksiyon alınmasına temel oluşturacak çıkarım yapılmış verilerin kullanılmasıdır.

Makine öğrenmesinde geliştirilecek model makinenin kararına bırakılmamakta, bunun yerine öğrenilmesi amaçlanan görev ve analize konu olan verinin yapısı bu hususu şekillendirmektedir (Barnes, 2015: 14).

Veri setine uygun bir model geliştirilmesi süreci eğitim olarak isimlendirilmekte, öğrenme ise eğitim sürecinde belirlenen modelin gelecekteki verilere uygulanması ile neticelenmektedir (Witten, Frank ve Hall, 2011: 28). Geçmişteki verilerin ya da mevcut veri setinin bir kısmı üzerinde eğitime tabi tutulan makine, veri içerisindeki yapıyı kullanılan algoritmanın çalışma prensibine bağlı olarak, anlamaya çalışmakta ve çıkarım yapmakta, oluşturulan bu model ışığında kendisine sunulacak test verileri (gelecekteki veriler) için bu modeli sınyarak sonuca ulaşmaktadır.

Spesifik bir amaca ulaşmak için makine öğrenmesi aşağıda sıralanan adımların takip edilmesi ile uygulanmaktadır (Zocca, Spacagna, Slater ve Roelants, 2017: 9-10):

- Verinin Toplanması: Analize tabi tutulacak verinin elektronik ortama uygun bir şekilde aktarılması gerekmektedir. Söz konusu veri, makinenin üzerinde öğrenme gerçekleştireceği temeli oluşturmaktadır.

- Verinin İncelenmesi ve Hazırlanması: Herhangi bir makine öğrenmesi algoritmasının kalitesi, verinin kalitesine dayanmaktadır. Bu aşama büyük ölçüde insan müdahalesi gerektirmektedir. Analiz yapmak isteyen kişinin veriyi ve yapısını anlaması, gerektiği takdirde düzeltmeler yapması zaruridir.

- Modelin Veri Üzerinde Eğitilmesi: Öncelikle veri setinden ne öğrenilmek istendiği açıkça belirlenmelidir. Makine öğrenmesiyle ulaşılması hedeflenen amaç, kullanılacak algoritmayı belirlemekte, kullanılan algoritmaya göre veriye uygun bir model bu algoritma tarafından sunulmaktadır.

- Modelin Performansının Değerlendirilmesi: Kullanılan modelin türüne bağlı olarak, test verileri kullanılmak suretiyle modelin ne kadar isabetli kararlar verdiği sınıdır.

- Modelin Performansının Artırılması: Test verileriyle sınanan model, beklenen performansı sergileyemediği takdirde, modelin değiştirilmesi, ilave veriler eklenmesi ya da her ikisini de kapsayacak değişiklikler yapılması gerekebilmektedir.

Eğitim metodolojisi ve veri setinin karakteristik yapısına göre makine öğrenmesi üç temel kategoriye ayrılmaktadır (Canepa, 2016: 4):

- Gözetimli Öğrenme (Supervised Learning): Makine eğitilirken işaretlenmiş verileri kullanmakta, her bir örnek girdi ve çıktı çiftlerinden oluşmaktadır. Makine girdi ve çıktı arasındaki ilişkiyi öğrenmektedir. Bu kategoride amaç, önceden işaretlenmiş veri setini esas alarak tahmin yapılması ya da sınıflandırma kararı verilmesidir (Russell ve Norvig, 2003: 650).

- Gözetimsiz Öğrenme (Unsupervised Learning): Makine etiketlenmemiş verileri kullanmaktadır. Bu kategoride amaç, benzer özelliklere göre elementlerin gruplandırılmasıdır. Bu tarz gruplar küme olarak adlandırılmakla birlikte, bu noktada amaç spesifik ve tek bir doğruya yakınsama elde edilmesi değildir. Sonuçların isabeti, yapılan kümeleme işleminde, her bir birimin dahil edildiği gruba ne ölçüde benzediği ve diğer gruplardan ne ölçüde farklı olduğu ile ölçülmektedir (Berry ve Linoff, 2004: 57).

- Destekli Öğrenme (Reinforcement Learning): Gözetimsiz öğrenmede olduğu gibi eğitim verisi etiketlenmemiş olmakla birlikte, bu kategoride öğrenme bir nevi ödül-ceza prensibine dayanmaktadır. Öğrenme esnasında, yapılan tercihin neticesi olumlu ise o tercih yüksek skorlu olarak işaretlenmekte, tam tersi durumda tercih düşük skorlu olarak belirlenmektedir (Kaelbling, Littman ve Moore, 1996: 238-240).

Makine öğrenmesinde, yukarıda da açıklandığı üzere farklı yaklaşımlar ve yöntemler (algoritmalar) bulunmakla birlikte, makinenin öğrenmesine esas olacak unsurların temelde veri setinin özelliklerine ve öğrenilmesi amaçlanan konuya göre

değiştirdiğini ifade etmek mümkündür. Çalışmamızın bir sonraki bölümünde, makine öğrenmesinde kullanılan ve tezimizin uygulama örneğinde sonuçları sunulan bazı algoritmaların çalışma prensip ve yöntemleri izah edilmiştir. Söz konusu algoritmalar, çalışmamızın uygulamasında kullandığımız bilgisayar programında yer almakta olup, ikili (binary) sınıflandırma kararına uygun yöntemlerdir. Ayrıca, tezimizin amacı temelde kredi analizi yapabilecek bir model ortaya koymak olduğundan, literatürdeki birçok kredi analizi uygulamasında, aşağıdaki bölümde yer alan algoritmaların kullanıldığı görülmüştür. Netice itibariyle, gerek uygulama örneğindeki veri setine uygunluğu gerekse sınıflandırma kararının niteliğinden ötürü takip eden bölümde yer alan algoritmalar tez çalışmamız için tercih edilmiştir.

2.4 MAKİNE ÖĞRENMESİNDE KULLANILAN ALGORİTMALAR

2.4.1. K-NN Algoritması (K En Yakın Komşu Algoritması)

K en yakın komşu algoritması, ilk olarak 1950'li yılların başında tanımlanmış, eğitim veri setlerinin büyümesiyle birlikte söz konusu algoritma emek yoğun bir hal almış, 1960'lı yıllarda hesaplama gücünde artış kaydedilene dek popüler bir kullanım alanına sahip olamamıştır (Han ve Kamber, 2006; 348).

Veri kümesi üzerinde tanımlı olan sınıflar arasında veriyi uygun şekilde dağıtmayı amaçlayan k-NN algoritması, diğer sınıflandırma algoritmalarında da olduğu gibi sunulan eğitim kümesinden, sınıfların dağılım şeklini öğrenerek daha sonra sınıfı belli olmayan test verileri sunulduğunda, öğrendiği dağılım şekliyle bu yeni verileri doğru bir şekilde sınıflandırmayı amaçlamaktadır. Sınıflandırma esnasında, test verisinin, daha önceki sınıflandırılmış birimlerden k tanesine yakınlığına bakılmakta ve sınıflandırma kararı buna göre verilmektedir (Şeker, 2013: 161). Ayrıca, k-NN algoritması, değeri bilinmeyen değişken için, en yakın komşuların ortalamasını alarak, değer öngörüsünde de bulunabilmektedir (Han ve Kamber, 2006; 349).

Kümeleme analizinde, kullanılabilir yöntemler arasında yer alan K-NN algoritması, her bir örneğin özellikleri baz alınarak kaydıyla, benzer özelliklere sahip, bir diğer deyişle birbirine benzeyen örnekleri bir arada konumlandırmaktadır (Barnes, 2015: 170).

En yakın komşu sınıflandırmaları, henüz sınıflandırılmamış örnekleri, kendilerine en yakın karakteristik özelliklere sahip sınıflara atamalarıyla tanımlanmaktadır (Gaganis, Pasiouras, Spathis ve Zopounidis, 2007; 30). Algoritmanın temelinde yatan bu prensip basit görünse de en yakın komşu algoritmaları:

- Fotoğraf ve videolardaki optik karakter tanıma ve yüz tanıma gibi görsel bilgisayar uygulamalarında,
- Otomatik ses/konuşmacı tanıma sistemlerinde,
- Herhangi bir kişinin, kendisine tavsiye edilen filmi beğenip beğenmeyeceği gibi tercih öngörülerinde,
- Özellik arz eden hastalıkların tanımlanması için esas alınan genetik verilerden model çıkarımlarında,

Başarıyla kullanılmaktadır (Lantz, 2013: 66; Nahayo ve Arı, 2016: 2).

Veri seti içerisindeki birimlerin özellikleri arasındaki ilişkilerin ve hedeflenen sınıf sayısının çok sayıda ve komplike olduğu ya da anlaşılmasının zor olduğu durumlara en yakın komşu sınıflandırıcıları uygun algoritmalar olmakla birlikte, bu tarz sınıflandırıcılar aynı sınıftaki öğelerin oldukça homojen olmasını sağlama eğilimindedirler. Diğer taraftan, gruplar arasında net farklılıkların bulunmadığı durumlarda, en yakın komşu algoritmasının başarısı, sınıflar arasındaki sınırların tanımlanmasında kısıtlı olmaktadır (Lantz, 2013: 66).

En yakın komşu yaklaşımının kullanıldığı k-nn algoritmasının güçlü ve zayıf yönlerini Tablo 2'deki şekilde sıralamak mümkündür (Lantz, 2013: 66; Webb, 2002: 98).

Tablo 2: K-NN Algoritmasının Güçlü ve Zayıf Yönleri

Güçlü Yönleri	Zayıf Yönleri
- Basit ve efektiftir.	- Özellikler arasındaki ilişkilerde, yeni çıkarımlar bulma kabiliyetini sınırlayan bir model geliştirmez.

- Verinin dağılımına ilişkin herhangi bir varsayımda bulunmaz.	- Sınıflandırma safhası yavaştır.
- Öğrenme safhası hızlıdır.	- Yüksek düzeyde işlemci hafızasına ihtiyaç duyar.
	- Nominal özellikler ve kayıp veriler için ilave işlem gerektirir.

K-NN algoritması, birçok kategoriye ayrılmış ve sınıflandırılmış örneklerle eğitim (öğrenme) sürecine başlamaktadır. Öğrenme sürecinin ardından, test veri setindeki her bir kayıt için, k-NN algoritması, test verisine en çok benzerlik gösteren k tane eğitim verisini getirmektedir. Test aşamasında kullanılan k sayısı, tam sayı olmakla birlikte test aşamasından önce belirlenmelidir. Test aşamasının başında, henüz sınıflandırılmamış olan örnek, kendisine en çok benzerlik gösteren k tane sınıflandırılmış örnekten, çoğunluğun ait olduğu sınıfa atanmaktadır (Ledolter, 2013: 116).

K-NN algoritması, k tane benzer örneği seçerken bir uzaklık (*distance*) fonksiyonu kullanmakta, bu fonksiyon vasıtasıyla iki örnek arasındaki uzaklığı hesaplamaktadır (Vukovic, Delibasic, Uzelac ve Suknovic, 2012: 8391). Uzaklığı hesaplamak için farklı yaklaşımlar bulunmakla birlikte, geleneksel olarak K-NN, Öklid mesafesini kullanmaktadır (Murphy, 2012: 17-18). Öklid mesafesi, iki noktayı cetvelle birleştiren doğrunun uzunluğu olarak ifade edilebilmektedir. En kısa (kuş uçuşu) rota olarak da tanımlanabilen Öklid mesafesinin formülünü, aşağıdaki iki şekilde göstermek mümkündür. Söz konusu formülde, p ve q karşılaştırılacak örnekleri, n her bir özelliği göstermektedir. Özellik sayısındaki yani n 'deki artış, modelin oluşturulacağı uzaydaki boyut sayısının artması, dolayısıyla yapılacak hesaplamaların bir o kadar karmaşık hale gelmesi anlamını taşımaktadır (Han ve Kamber, 2006: 348). Örneğin, p_1, p örneğinin 1. özellikteki değerini, q_1 ise q örneğinin 1. özellikteki değerini göstermektedir (Zaccone, 2016: 81).

$$dist(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

$$distance = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Uzaklık formülünde, her bir özellik için hesaplama yapılmaktadır. Yapılan hesaplamalar neticesinde, p 'ye olan uzaklık değeri en düşük olandan başlamak üzere k tane en yakın komşu belirlenir. $K=1$ olması durumunda, algoritma 1NN olarak adlandırılmakta (Shwartz ve David, 2014: 259), böylelikle en yakın komşu hangi sınıfa dahilse, sınıflandırılmak istenen p de o sınıfa dahil edilmektedir (Ledolter, 2013: 116; Lantz, 2013: 70). Veri setindeki özellik sayısının, yani n 'in yüksek olmasının, yapılacak çalışmayı zorlaştırabileceği göz önüne alındığında, özellik sayısını azaltma, bir diğer deyişle özellik silme yapılarak, modelin üzerinde çalıştığı uzayın boyut sayısını azaltmak mümkün olabilmektedir (Raschka, 2016: 96). Ayrıca, k -nn algoritmasında en yaygın biçimde kullanılan uzaklık ölçme yöntemi Öklid mesafesi olsa da fazla sayıda boyuta sahip (n 'in yüksek olduğu) veri setlerinde, Öklid mesafesi yönteminden ziyade farklı yöntemlerin uzaklık ölçümünde kullanılması daha yararlı olabilmektedir (Zafarani, Abbasi ve Liu, 2014: 150).

K sayısı, yani modelde esas alınacak en yakın komşu sayısı, modelin gelecekteki (test aşamasındaki) veriler için ne kadar iyi genelleme yapabileceğini belirleyen en önemli unsurlardan birisidir (Filiz ve Öz, 2017: 122). Büyük bir k sayısı belirlendiğinde, veri setindeki anlamsız örneklerden kaynaklanan varyansın etkisi azaltılmakta, diğer yandan aynı tutum, küçük ancak önemli düzenlerin ihmal edilmesine sebebiyet verebilmektedir (Ledolter, 2013: 117). K sayısının yüksek belirlendiği, örneğin tüm veri setindeki gözlemlerin sayısı kadar olduğu kabul edildiğinde, geliştirilecek model çoğunluğun ait olduğu sınıfa yönelik öngörü yapacak, sınıflandırılmak istenen birime en yakın komşuları dikkate almayacaktır. Böylesi bir durum, geliştirilmesi hedeflenen model için katlanılan maliyetleri anlamsız ve karşılıksız kılacaktır. K sayısının oldukça düşük belirlendiği tam tersi bir senaryoda ise, örneğin $k=1$ olarak kabul edildiğinde, anlamsız ve aykırı örneklerin model üzerine etkisi olması gerekenin çok üzerinde gerçekleşebilecektir. Sınıflandırılmak istenen birime en yakın komşunun anlamsız ya da aykırı bir örnek olduğu kabul edildiğinde, model yanlış öngörüde bulunacak, hâlbuki k sayısı çok düşük de olsa 1'den farklı bir sayı olarak kabul edildiğinde modelin çok daha iyi sonuçlar üretebileceği çözümler

elde edilebilecektir (Lantz, 2013: 71). Her iki senaryodan çıkarılabilecek sonuç, ideal k sayısının iki uç nokta arasında olduğudur (Bishop, 2006: 125). Küçük k değerlerinin, sınıflar arasında kompleks sınırların bulunduğu durumlarda daha iyi sonuçlar verdiğini söylemek mümkün olmakla birlikte, sınıflar arasındaki sınırın ne denli kompleks olduğu ya da bu sınırın lineer bir doğru yerine fonksiyonu belirlenemeyen bir eğriyle ayrılıp ayrılmadığının ortaya konulamaması sorun teşkil etmektedir (Lantz, 2013: 71). Her ne kadar eğrinin fonksiyonu belirlenemese de k -nn algoritmasının öne çıkan avantajlarından birisi, lineer bir şekilde ayrılmamış sınıflar için de sınıflandırma yapabilmesidir (Zaccone, 2016: 81).

Uygulamada k sayısının belirlenmesi, öğrenilmek istenen yapının zorluğuna ve öğrenme verisindeki birim sayısına bağlı olmaktadır. K sayısı genellikle 3 ile 10 arasında bir değer olarak belirlenmektedir (Lantz, 2013: 72; Derelioğlu ve Gürgen, 2011: 9315). Ancak, k sayısının tek sayı olarak belirlenmesinde büyük yarar bulunmaktadır (Ledolter, 2013: 117). Örneğin, $k=4$ olarak belirlendiğinde ve en yakın 4 komşudan ikisinin X sınıfına, diğer ikisinin ise Y sınıfına dahil olduğu durumlarda, algoritma işlevini yitirecek ve sonuç üretemeyecektir. Bir diğer yaygın uygulama ise eğitim aşamasında kullanılacak verilerin karekökünün, k sayısı olarak seçilmesidir (Lantz, 2013: 72). Ayrıca, çok yaygın olmamakla birlikte, büyük bir k sayısı belirlenerek sınıflandırılmak istenen birime yakınlığına göre komşular için farklı ağırlıklar esas alınabilmekte, böylelikle en yakın komşunun tesirinin modeldeki karar aşamasında daha fazla olması amaçlanmaktadır (Vukovic, Delibasic, Uzelac ve Suknovic, 2012: 8391; Han ve Kamber, 2006: 349). Ancak, en iyi sonucu veren k sayısı için herhangi bir standart getirilememiştir. Bu durumda izlenebilecek bir diğer yol, farklı k sayılarının sırayla veri seti üzerinde denenmesi, elde edilen sınıflandırma başarı oranlarının oluşturduğu eğrinin maksimum noktasını veren k sayısının model için esas alınmasıdır. Fakat bu noktada belirtmek gerekir ki farklı k sayılarının sırayla denenmesi, uzun zaman alacağı gibi, her bir k değeri için elde edilen sınıflandırma başarı oranlarının göstereceği dağılım net bir şekilde belirlenemeyebilir. Diğer yandan, anlamsız verilerin bulunmadığı, veri setinin küçük olduğu ve öğrenme verilerinin geneli temsil kabiliyetinin yüksek olduğu durumlarda, k sayısının belirlenmesi daha az önemli bir hal almaktadır (Lantz, 2013: 72).

Algoritmanın üzerinde çalıştığı veri setindeki her bir özellik farklı aralıklara sahip olabilmekte, örneğin veri setindeki kişilerin sahip oldukları arazinin dekar değeri

1.000-25.000 TL arasında deęişirken, dekardan alınan ürünün birim kg fiyatı 0,50-5 TL arasında deęişebilmektedir. K-NN algoritması uygulanmadan önce, genellikle özellikler standart bir aralıęa dönüştürülmektedir (Han ve Kamber, 2006: 348-349; Ledolter, 2013: 116). Bu tarz bir işlemin yapılması, uzaklık hesaplamasının (formülünün), özelliklerdeki ölçü aralıklarına bağımlılıęından kaynaklanmaktadır. Bazı özellikler dięerlerine kıyasla daha yüksek deęerlere sahip olduęunda, uzaklık hesaplamasında yüksek deęerlere sahip bu özellikler ağır basacaktır (Berry ve Linoff, 2004: 275). Bu noktada yapılması gereken, her bir özellięin yeniden ölçeklendirilerek, uzaklık hesaplamasına, ilk duruma nazaran, daha eęit etki etmesini saęlamaktır. K-NN algoritması için yeniden ölçeklendirmede geleneksel metot min-maks normalizasyonu olarak isimlendirilmekte, bu süreçte herhangi bir özellięe ait tüm deęerler 0 ve 1 arasına indirgenmektedir (Berry ve Linoff, 2004: 275). Söz konusu normalleştirme işlemine ait formül ařaęıda gösterilmiş olmakla birlikte, formülde X özellięine ait her bir deęerden X'in minimum deęeri çıkarılmakta, elde edilen deęer X özellięinin deęer aralıęına bölünmektedir.

$$X_{yeni} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

K-NN algoritmasındaki yaygın olan bir dięer dönüştürme yaklařımı ise z-skor standardizasyonu olarak adlandırılmaktadır. Bu yöntemde, X özellięine ait her bir deęerden, o özellikteki deęerlerin ortalaması çıkarılmakta, elde edilen sonuç X özellięindeki deęerlerin standart sapmasına bölünmektedir (Lantz, 2013: 73).

$$X_{yeni} = \frac{X - \mu}{\sigma} = \frac{X - Ort(X)}{StdSap(X)}$$

K-NN algoritmasında yaygın kullanıma sahip Öklid uzaklık formülü, nominal veriler için tasarlanmadıęından, veri setindeki nominal verilerin sayısal verilere dönüştürülmesi gerekmektedir (Berry ve Linoff, 2004: 276). Bu duruma çözüm getirebilmek maksadıyla "kukla kodlama (dummy coding)" adıyla anılan yöntem uygulanmakta, kategoriler 0 ve 1 olarak gösterilmektedir (Han ve Kamber, 2006: 349). Örneęin cinsiyet deęişkeni için ařaęıdaki řekilde bir kodlama yapılmaktadır.

$$erkek = \begin{cases} 1 & \text{eğer } X = erkek \\ 0 & \text{diğeri} \end{cases}$$

Kukla kodlama ile iki kategoriye sahip cinsiyet değişkeni, tek bir değişkene indirilmekte, *kadın* seçeneği için ayrı bir özellik oluşturulmasına gerek kalmamakta, kategorilerden birisinin bilinmesinin yeterli olacağı görülmektedir.

Kukla kodlama, kategori sayısının 2'den fazla olduğu özelliklerde de kullanılabilir. Bu tarz durumlarda, kategori sayısının (n), bir eksiği (n-1) kadar kodlama yapılmaktadır (Lantz, 2013: 73). Örnek olarak veri setindeki kişilerin eğitim durumlarının, ilkokul, ortaokul ve lise olduğu kabul edilsin. Böyle bir senaryoda, yapılacak modelleme aşağıdaki şekilde olacaktır:

$$ilkokul = \begin{cases} 1, & \text{eğer } x = ilkokul \\ 0, & \text{diğeri} \end{cases}$$

$$ortaokul = \begin{cases} 1, & \text{eğer } x = ortaokul \\ 0, & \text{diğeri} \end{cases}$$

Yukarıdaki durumların her ikisinde de “0” değeri söz konusu olduğunda, kişinin eğitim durumunun lise olduğu anlaşılacak, eğitim özelliği için üçüncü bir niteliğe ihtiyaç olmayacaktır.

En yakın komşu algoritmaları tembel öğrenme algoritması (lazy learning algorithm) olarak değerlendirilmekte (Boschetti ve Massaron, 2016: 209), bu durum da çıkarım elde edilememesinden kaynaklanmaktadır (Raschka, 2016: 92). Tembel öğrenme algoritması, aslında öğrenmekten ziyade eğitim verisini harfi harfine biriktirmekte, bu durumdan ötürü eğitim aşaması hızla tamamlanmasına rağmen tahmin aşaması yani test verilerinin değerlendirildiği aşama daha yavaş tamamlanabilmektedir (Boschetti ve Massaron, 2016: 210; Lantz, 2013: 75). Tembel öğrenme algoritması tam olarak bir model inşa etmemektedir. Veriler hakkında herhangi bir parametre öğrenilmediğinden bu tarz algoritmalar parametrik olmayan öğrenme metotları arasında sayılmaktadır. Veri seti hakkında teori üretilmediğinden, parametrik olmayan bu tarz yaklaşımlar, sınıflandırma algoritmasının veriyi nasıl kullandığının anlaşılmasını zorlaştırmaktadır. Ancak diğer taraftan bu durum öğrencinin veri setindeki doğal yapıyı bulmasını sağlamakta, verinin yargıya dayalı bir forma uydurulmasının önüne geçmektedir (Lantz, 2013: 75).

2.4.2. Naive Bayes Algoritması

Naive Bayes algoritması, geçmiş olaylara ait verileri kullanarak gelecekteki olayların olasılıklarını tahmin etmeyi amaçlayan yöntemdir (Boschetti ve Massaron, 2016: 206). Algoritmanın temelini oluşturan teknik, 18. yy matematikçisi Thomas Bayes tarafından geliştirilmiş, söz konusu teknikle vakaların gerçekleşme olasılıklarının tanımlanması, aynı zamanda eklenen bilgilerle bu olasılıkların nasıl gözden geçirilmesi gerektiği ortaya konulmuştur (Anderson, Sweeney, Williams, Freeman ve Shoesmith, 2009: 143; Han ve Kamber, 2006; 310; Bell, 2015: 73-74).

Herhangi bir olayın gerçekleşme olasılığının %0 ile %100 arasında değişebileceği, olasılığın %0'a yaklaştıkça bu olayın meydana gelme ihtimalinin düştüğü, %100'e yaklaştıkça söz konusu olayın gerçekleşme ihtimalinin arttığı kabul edilmektedir (Newbold, 2008: 91). %0 olasılık olayın kesinlikle gerçekleşmeyeceğini, %100 olasılık ise olayın mutlak surette gerçekleşeceğini ifade etmektedir (Newbold, 2008: 91).

Bayes metotlarına dayalı sınıflandırma algoritmaları, eğitim verisinden faydalanarak ve veri setindeki özelliklerin değerlerini esas olarak her bir sınıf için gözlenmiş olasılıkları hesaplamaktadır (Han ve Kamber, 2006; 310). Sınıflandırıcı yöntem (algoritma) daha sonra sınıflandırılmamış veri üzerinde çalışırken, yeni verilere ait özellikler için, daha önce hesaplanmış gözlenmiş olasılıkları baz alarak sınıflandırma tahminini gerçekleştirmektedir (Nabiyev, 2003: 450). Bayes algoritması basit bir fikre dayansa da daha komplike algoritmalarından elde edilen sonuçlarla eşit düzeyde sınıflandırma başarısı sunabilmektedir (Cichosz, 2015: 118; Barber, 2012: 242). Bayes yöntemini esas alan sınıflandırıcılar:

- Spam niteliğindeki e-postaların filtrelenmesi, yazarın kimliğinin saptanması ya da konuya göre kategori tanımlanması gibi metin sınıflandırma uygulamalarında,
- Bilgisayar ağlarındaki izinsiz girişlerin ya da anormal işlemlerin tespit edilmesinde,
- Gözlenmiş semptomları esas alarak tıbbi teşhislerin konulmasında,
- Sigortacılık, bankacılık ve yatırım endüstrilerinde tahmin yapılmasında,

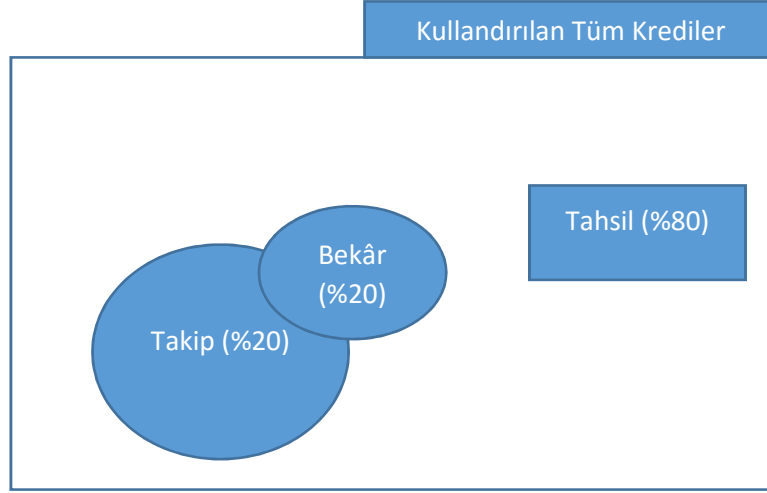
Kullanılabilmektedirler (Bell, 2015: 69-70).

Bayes yöntemini esas alan sınıflandırıcılar, çıktılarına ait olasılıkların tahmin edilmesinde çok sayıda özelliğin aynı anda değerlendirildiği problemlere en iyi şekilde uygulanabilen yöntemler arasında değerlendirilmektedir (Hastie, Tibshirani ve Friedman, 2001: 184-185). Çoğu algoritma, zayıf etkisi olan özellikleri dikkate almaz iken, Bayes metodu tüm bulguları dikkate alarak tahminlerde incelikle değişiklik yapmaktadır (Lantz, 2013: 90). Nispeten daha az etkiye sahip çok sayıdaki özellik, bir arada değerlendirildiğinde, bu özelliklerin kombine etkisinin yüksek olabilmesi mümkündür (Lantz, 2013: 90).

Bayes yöntemine ait temel kavramlara bakıldığında, durum (sonuç) gerçekleşme olasılığı bulunan çıktı olarak tanımlanmaktadır. Örneğin, kişinin kullanmış olduğu krediyi vadesinde ödeyip ödememesi olası iki durumu ifade etmektedir. Deneme ise durumun gerçekleşme olanağı şeklinde ifade edilmektedir. İlk örneğimiz dikkate alındığında, kredinin kullanılması deneme kavramı için örnek olarak gösterilebilir. Bayes yönteminin esas unsurlarından birisi olan olasılık kavramı, herhangi bir durumun gerçekleştiği deneme sayısının (örneğin kredisini vadesinde ödemeyen kişi sayısının), tüm deneme sayılarına (kredi kullanan kişi sayısına) bölünmesiyle elde edilen oranı ifade etmektedir (Newbold, 2008: 91; Barber, 2012: 239). Tüm durumlara ait olası çıktıların gerçekleşme olasılıklarının toplamının %100 olması gerekmektedir. Örneğimizde olduğu gibi, denemelerin iki sonucunun olduğu ve sonuçların aynı anda gerçekleşme imkânının bulunmadığı şartlarda, bir sonucun olasılığının bilinmesi, aynı zamanda diğer sonucun bilinmesine de imkân sağlamaktadır. Şöyle ki kredinin vadesinde ödenmeme olasılığı $P(\text{takip})=0,20$ şeklinde ifade edilirse, vadesinde ödenen kredilerin olasılığını $P(\text{tahsil})= 1-P(\text{takip})= 1-0,20=0,80$ şeklinde göstermek mümkün olacaktır. Söz konusu ifade şekli, bu iki durumun ayrık olaylar olması (iki durumun aynı anda gerçekleşmemesi) ve bu iki durumdan başka sonuç olmamasından ötürü geçerli olmaktadır (Newbold, 2008: 105).

Yukarıda verilen örneklerin aksine tamamen ayrışık olmayan durumlar da çalışmanın konusunu oluşturabilmektedir. Bazı durumlar, çalışmanın hedefindeki durum ile birlikte gerçekliği takdirde, bu durumlar da yapılacak tahminlerde kullanılabilir. Yukarıdaki kredi örneğini baz alınarak, ayrışık olmayan durumlar aşağıdaki şekilde gösterilebilir.

Şekil 2: Ortak Olasılık Örneği



Şekil 2 incelendiğinde, kredi kullanan bekâr şahıslara ait dairenin, bir kısmının takipteki kredilerle, bir kısmının ise tahsil edilmiş kredilerle kesiştiği görülmektedir. Bu durumda bekâr olan şahıslardan bazılarının kredilerini vadesinde ödediği, bazılarının ise borçlarını vadelerinde ödemedikleri anlaşılmaktadır. Bekâr şahıslara ait daire ile takipteki kredilere ait dairenin kesiştiği alan, literatürde Venn diyagramı olarak adlandırılmakta, söz konusu diyagramlarda dairelerin büyüklüğü ve kesişim derecesi önem arz etmemekle birlikte, asli unsurun tüm olası durumlara ait kombinasyonlar için olasılık hesabı yapılması olduğu ifade edilmektedir (Blangiardo ve Cameletti, 2015: 53).

Şekil 2’de görselleştirilen örnekten, kullandırılan tüm kredilerin %20’sinin vadesinde ödenmediği ve tüm kredilerden %20’sinin bekâr şahıslara kullandırıldığı anlaşılmaktadır. Bu aşamada yapılması gereken, bu her iki durumun örtüştüğü alana tekabül eden olasılığın, yani hem bekâr olup hem de kredi borcunu vadesinde ödemeyen şahıslara ait olasılığın hesaplanmasıdır. Diğer bir deyişle, $P(\text{takip} \cap \text{bekâr})$ şeklinde yazılabilen olasılığın tahmin edilmesi amaçlanmaktadır. Söz konusu kesişim olasılığının (ortak olasılığın) hesaplanması, bir durumun olasılığının diğer durumun olasılığına ne şekilde bağlı olduğuyla ilgilidir (Murphy, 2012: 82). Eğer, kişilerin medeni halleri ile kredilerini vadelerinde ödemeleri olayları arasında herhangi bir bağ yok ise bu iki durum bağımsız durum (independent events) olarak tanımlanmaktadır. Tüm durumların birbirinden bağımsız olduğu koşullarda, herhangi bir durumdan elde edilen veri ile bir diğerine ilişkin netice alınması imkânsız hale gelmektedir (Lantz, 2013: 93). Ancak tam aksi durumlarda, yani durumların birbirinden bağımsız olmadığı

(dependent events) koşullarda, tahmine yönelik modeller geliştirilmesi mümkün olmaktadır (Duda, Hart ve Stork, 2001: 22-23).

Kişilerin medeni halleri ile kredi borçlarını vadesinde ödemeleri durumları, birbirinden bağımsız durumlar olarak kabul edildiğinde, Şekil 2'deki kesişime konu alanın olasılığını $P(\text{takip} \cap \text{bekâr}) = P(\text{takip}) * P(\text{bekâr}) = 0,20 * 0,20 = 0,04$ şeklinde hesaplamak mümkün olacaktır (Newbold, 2008: 110). Söz konusu iki durumun bağımsız olaylar kabul edildiği koşullarda, yapılan hesaplama göre kullanılan tüm kredilerin %4'ünün bekâr şahıslara kullanılmış ve vadesinde ödenmemiş krediler olduğu söylenebilecektir. Ancak, kişilerin medeni halleri ile kredi borçlarını zamanında ödemeleri durumlarının birbirinden tamamen bağımsız olduğunu ifade etmek, gerçek yaşamda pek mümkün olmamakla birlikte, bu tarz koşullarda her iki durum için daha farklı bir hesaplamanın yapılması gerekmektedir.

Birbirine bağımlı durumlar, Bayes Teorisinin konusunu oluşturmakta ve bu durumlara ilişkin olasılık hesaplaması aşağıdaki formül vasıtasıyla yapılmaktadır. Formüldeki $P(A | B)$ şeklindeki gösterim, B durumunun gerçekleştiği koşullarda A durumunun gerçekleşme olasılığını ifade etmektedir (Anderson, Sweeney, Williams, Freeman ve Shoesmith, 2009: 145). Söz konusu gösterim koşullu olasılık olarak adlandırılmakta, A durumunun olasılığının, B durumunun olasılığına bağımlı olduğu gösterilmektedir.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

Kredi örneğindeki iki durumun birbirine bağımlı olaylar olduğu, yani medeni hal ile krediyi vadesinde ödeme durumlarının gerçekleşme olasılıkları arasında ilişki bulunduğu kabul edildiğinde, bekâr şahısların kredilerini vadesinde ödememeleri olasılığının hesaplanması aşağıdaki şekilde formüle edilebilir.

$$P(\text{takip}|\text{bekâr}) = \frac{P(\text{bekâr}|\text{takip}) * P(\text{takip})}{P(\text{bekâr})}$$

Kredi örneğine ilişkin formülde, $P(\text{takip})$, kullanılan tüm kredilerde vadesinde ödenmeyen kredilerin olasılığını ifade etmekte ve öncel olasılık şeklinde isimlendirilmektedir. $P(\text{bekâr}|\text{takip})$, ihtimal olarak nitelendirilmekte, vadesi geçen

kredilerde bekâr şahısların kullandığı kredilerin olasılığını ifade etmektedir. $P(\text{bekâr})$ ise tüm kullanılan kredilerde bekâr şahıslarca kullanılan kredilerin olasılığını göstermekte ve marjinal ihtimal olarak adlandırılmaktadır. Bayes teorisini uygulayarak kredinin vadesinde ödenmeme olasılığını, $P(\text{takip}|\text{bekâr})$ ardıl olasılığını, hesaplamak mümkün olacaktır. Kredi örneğine ilişkin Bayes denklemindeki, ihtiyaç duyulan olasılıkların bir kısmı zaten bilinmekle birlikte özellikle ihtimal olarak nitelendirilen $P(\text{bekâr}|\text{takip})$ olasılığı, veri setindeki bu iki durum için mevcut olan gerçekleşme sayılarının frekans tablosuna dökülmesi, ardından bu tablonun ihtimal tablosuna dönüştürülmesi ile hesaplanabilmektedir (Tuffery, 2011: 493-494). Kredi örneği için aşağıdaki şekilde frekans ve ihtimal tabloları oluşturulabilir.

Tablo 3: Örnek Frekans ve İhtimal Tabloları

	Bekâr				Bekâr		
Frekans	Evet	Hayır	Toplam	İhtimal	Evet	Hayır	Toplam
Takip	60	20	80	Takip	60/80	20/80	80
Tahsil	20	300	320	Tahsil	20/320	300/320	320
Toplam	80	320	400	Toplam	80/400	320/400	400

Kredi örneği için, bekar şahısların kredilerini vadesinde ödememesi durumunun olasılığı, yukarıdaki formül baz alındığında, $((60/80)*(80/400))/80/400=$ %75 olarak hesaplanır.

Naive Bayes algoritması, Bayes Teorisi'ni esas alan sınıflandırma uygulamaları arasında yer almakla birlikte, özellikle metin sınıflandırmasında en çok kullanılan algoritma türüdür (Boschetti ve Massaron, 2016: 206). Naive Bayes algoritmasının güçlü ve zayıf yönleri aşağıdaki şekilde sıralanabilir (Lantz, 2013: 93; Han ve Kamber, 2006: 310; Cichosz, 2015: 118).

Tablo 4: Naive Bayes Algoritmasının Güçlü ve Zayıf Yönleri	
Güçlü Yönleri:	Zayıf Yönleri:
- Basit, hızlı ve efektiftir.	- Özelliklerin birbirinden bağımsız olduğu ve eşit

	derecede öneme haiz olduğu gibi genellikle yanlış bir ön kabul yapmaktadır.
- Anlamsız ve kayıp verilerin bulunduğu veri setlerinde başarılıdır.	- Nümerik verilerin fazla olduğu veri setleri için ideal bir yöntem değildir.
- Göreceli olarak eğitim aşamasında az sayıda örnek veriye ihtiyaç duyar. Ancak çok sayıda örneğin bulunduğu veri setlerinde de başarı gösterir.	- Tahmin edilen olasılıklar, öngörülen sınıflardan daha az güvenilirdir.
- Öngörü yapılması için gerekli olan tahmini olasılığın elde edilmesi kolaydır.	

Naive Bayes algoritmasının, “naive (naif, saf)” şeklinde isimlendirilmesinin nedeni, algoritmanın veri hakkında yapmış olduğu bir takım ön kabullerden kaynaklanmaktadır. Söz konusu algoritma, veri setindeki tüm özelliklerin eşit derecede öneme haiz olduğunu ve özelliklerin birbirinden bağımsız olduğunu farz etmektedir (Zocca, Spacagna, Slater ve Roelants, 2017: 20; Idris, 2014: 302). Bu tarz kabuller, gerçek dünyadaki uygulamalarda çok nadiren karşılık bulabilmektedir (Cichosz, 2015: 122). Her ne kadar birçok uygulamada, algoritmanın ön kabullerine aykırı veri setleri bulunsa da Naive Bayes yeterince iyi performans gösterebilmektedir (Witten, Frank ve Hall, 2011: 99). Özellikler arasında yüksek düzeyde bağımlılığın bulunduğu ekstrem koşullarda dahi Naive Bayes algoritması, çok yönlülüğü ve çeşitli koşullardaki isabetliliği ile sınıflandırma algoritmaları arasında en güçlü adaylardan birisi olarak öne çıkmaktadır (Hastie, Tibshirani ve Friedman, 2001: 184-185). Naive Bayes algoritmasının yapmış olduğu, gerçeği tam manasıyla yansıtmayan ön kabuller birçok tartışmaya yol açmıştır. Ancak, öngörülen sınıf doğru olduğu sürece tahmini olasılığın önem arz etmeyeceği, şöyle ki eğer şahıs kredisini vadesinde ödemeyecek

şeklinde sınıflandırıldıysa, tahmin edilen olasılığın %51 ya da %99 olmasının modelin başarısını değiştirecek bir unsur olmadığı ifade edilmektedir (Cichosz, 2015: 122).

Yukarıdaki kredi örneğinde, kişinin medeni hali ile kredisini vadesinde ödememesi durumları için Bayes Teorisi'ne göre olasılık hesaplaması yapılmış idi. Naive Bayes algoritmasının başlangıçta yapmış olduğu ön kabulün, kompleks ve çok sayıda hesaplama gerektiren senaryolar için durumu daha basite indirgediğini ifade etmek mümkün olacaktır (Bishop, 2006: 381). Kredi örneğine geri dönecek olursa, yeni senaryoda medeni halin, cinsiyetin ve yaşın 65'ten yüksek olmasının, kredinin takip hesaplarına düşmesi olasılığını nasıl etkilediğinin araştırıldığı kabul edildiğinde, aşağıdaki şekilde Bayes formülü karşımıza çıkacaktır.

$$\begin{aligned} & P(\text{takip}|\text{bekar} \cap \text{bayan} \cap \text{yas} > 65) \\ &= \frac{P(\text{bekar} \cap \text{bayan} \cap \text{yas} > 65|\text{takip})P(\text{takip})}{P(\text{bekar} \cap \text{bayan} \cap \text{yas} > 65)} \end{aligned}$$

İlave özelliklerin eklenmesiyle birlikte, yeni formül için tüm muhtemel kesişim noktalarının olasılık hesabı çok büyük işlemci hafızası gerektirecektir. Bayes Teorisi'nin zorlandığı bu noktada çözüm Naive Bayes algoritmasının başlangıçtaki ön kabulünden gelmektedir. Bu ön kabulde, özelliklerin birbirinden bağımsız olduğu ve eşit düzeyde öneme sahip oldukları ifade edilmişti. Özelliklerin birbirinden bağımsız olduğunun kabulü, herhangi iki bağımsız özellik için aşağıdaki şekilde hesaplama yapılmasına imkân sağlamaktadır (Anderson, Sweeney, Williams, Freeman ve Shoesmith, 2009: 140).

$$P(A \cap B) = P(A) * P(B)$$

Bu durumda eklenen özelliklerle birlikte kredi örneğindeki hesaplama aşağıdaki şekilde yapılabilecektir.

$$\begin{aligned} & P(\text{takip}|\text{bekar} \cap \text{bayan} \cap \text{yas} > 65) \\ &= \frac{P(\text{bekar}|\text{takip})P(\text{bayan}|\text{takip})P(\text{yas} > 65|\text{takip})P(\text{takip})}{P(\text{bekar})P(\text{bayan})P(\text{yas} > 65)} \end{aligned}$$

Kredi örneğinde sadece 3 farklı özellik ele alınmakla birlikte, çok sayıda özelliğin bulunduğu veri setleri için Naive Bayes algoritmasınca kullanılan formül aşağıdaki şekilde ifade edilebilir (Lantz, 2013: 97). Formüldeki C_L , sınıfın L

seviyesindeki olasılığını, F şeklindeki gösterimler olasılık hesabında kullanılacak her bir özelliği, $1/Z$ ise sonucu olasılığa çeviren ölçekleme katsayısını ifade etmektedir.

$$P(C_L|F_1, \dots, F_n) = \frac{1}{Z} P(C_L) \prod_{i=1}^n P(F_i|C_L)$$

Yukarıdaki kredi örneğinde, sadece medeni hal özelliğinin yer aldığı frekans ve ihtimal tablolarında, her bir senaryo için sıfırdan farklı bir olasılık düzeyi bulunmaktaydı. Ancak, olasılık hesabı yapılmak istenen senaryoda, herhangi bir ya da birkaç özellik için olasılık “0 (sıfır)” şeklinde gerçekleştiği takdirde, hesaplamada yer alan çarpma işlemi nedeniyle işlemin sonucu sıfır olarak hesaplanacaktır. Bu durumda, algoritmanın yanlış sınıflandırma yapmasına neden olunacaktır. Söz konusu duruma engel olmak adına Laplace katsayısı (Laplace estimator) yöntemi kullanılarak frekans tablosundaki her bir hesaba küçük bir sayı eklenmekte ki bu sayı genelde 1 olmakta, böylelikle her bir özellik için sıfırdan farklı bir olasılık hesabı yapılabilmektedir (Han ve Kamber, 2006; 314). Netice itibariyle, her bir sınıf-özellik kombinasyonunun en az bir kere hesaplamada yer alması sağlanmaktadır.

Naive Bayes algoritması frekans tabloları vasıtasıyla veri setini öğrenmekte, her bir özelliğin kategorik olması sayesinde sınıflar ve özellikler arasında kombinasyonlar yapılarak matris oluşturulabilmektedir (Bishop, 2006: 381). Ancak nümerik özellikler kategorik değerlere sahip olmadığından, algoritmanın nümerik verilerle doğrudan çalışması mümkün olmamaktadır (Ledolter, 2013: 126). Algoritmanın bu sorunu aşması için nümerik özelliklerin kategorik kümelere dönüştürülmesi gerekmektedir. Naive Bayes algoritmasıyla çalışılmasının genel koşulu olan çok sayıda eğitim verisinin bulunması durumlarında, söz konusu çevirme işleminin kullanılması uygun olmaktadır (Lantz, 2013: 100). Nümerik verilerin dönüştürülmesi için birçok metot bulunmakla birlikte en yaygın olanı verilerin dağılımındaki geçiş noktalarının (kesim noktalarının) kullanılması yöntemidir (Maslove, Podchiyska ve Lowe, 2013: 544-546). Verilerin dağılımına göre söz konusu kesim noktaları belirlenmekte, böylelikle nümerik veriler için kategoriler oluşturulmaktadır. Verilerin karşılık geldiği kategorilere göre algoritma çalışmasını tamamlamaktadır. Ancak bu noktada altı çizilmesi gereken husus, dönüştürme işleminin bilgi silinmesine sebebiyet vermiş olmasıdır. Dolayısıyla, kategori sayısı

belirlenirken bir denge gözetilmeli, az sayıdaki kategorinin veri setindeki önemli trendlerin gölgede kalmasına, çok sayıdaki kategorinin ise frekans tablolarında küçük hesaplar bulunmasına neden olacağı göz önünde bulundurulmalıdır (Yang ve Webb, 2001: 567-568).

Öğrenme aşamasında, tüm muhtemel durumlar için olasılık hesabı yapıldıktan sonra (Berry ve Kogan, 2010: 40), test aşamasındaki verilere ait özelliklere göre olasılık hesabı yapılmakta, yeni veriler oluşturulan modele göre sınıflandırılmaktadır (Cichosz, 2015: 123-124).

2.4.3. Karar Ağacı Algoritması

Karar ağacı algoritmasının sınıflandırma işlemi için geliştireceği model, ağaç formunda olduğundan, söz konusu algoritma bu şekilde adlandırılmaktadır. Oluşturulan model, akış şemalarındakine benzer bir şekilde bir takım mantıksal kararlardan oluşmakta, model içerisindeki akış, herhangi bir özelliğe ilişkin belirli kararların verildiği düğüm noktalarından geçerek nihai sonuca ulaşmaktadır (Zocca, Spacagna, Slater ve Roelants, 2017: 16). Ağacın düğüm noktalarındaki dallanma, yani akışın yönü, söz konusu noktalardaki tercihlere (alınan değerlere) göre şekillenmektedir (Billari, Fürnkranz ve Prskawetz, 2006: 42). Modelin oluşturduğu ağaç, yaprak noktalarında bitmekte, yaprak uçlarında, o noktaya ulaşılan dek verilen kararların kombinasyonunun sonucu görülmektedir (Williams, 2011: 206).

Ağacın kök noktasında bekleyen ve sınıflandırılmak istenen veri, kendi özelliklerinin sahip olduğu değerlere göre ağaç üzerindeki akış şemasında ilerleyerek, ilgili sınıfa ait yaprak noktasına ulaşacak, böylelikle oluşturulan karar ağacı modeli tarafından sınıflandırılmış olacaktır (Witten, Frank ve Hall, 2011: 99; Zafarani, Abbasi ve Liu, 2014: 146-147).

Temelinde karar ağacı bir akış şeması olduğundan, sınıflandırma mekanizmasının şeffaf olmasına ihtiyaç duyulan karar alma durumlarında, bu algoritma kullanılmaya elverişlidir (Williams, 2011: 205). Bu bağlamda söz konusu yöntemin:

- Başvuru sahiplerinin reddedilme nedenlerinin açıkça ifade edilmesine gerek duyulan kredi skor modellerinde,
- Finansal kurumlarca yapılan opsiyon fiyatlama çalışmalarında,

- Yönetim ve pazarlama acenteleriyle paylaşılacak müşteri şikâyet ve memnuniyetlerini içeren pazarlama çalışmalarında,
- Laboratuvar ölçümleri, semptomlar ya da hastalığın ilerleme oranına dayalı tıbbi teşhislerde,

Kullanılabildiğini söylemek mümkündür (Bell, 2015: 46; Lantz, 2013: 120). Aslında, karar ağaçlarının makine öğrenmesi teknikleri arasında, en geniş kullanım alanına sahip teknik olduğu, söz konusu tekniğin nerdeyse herhangi bir türde veri içeren tüm veri setlerine uygulanabildiği ifade edilebilir (Kuhnert ve Venables, 2005: 259). Her ne kadar geniş bir uygulama alanına sahip olsa da bazı durumlar karar ağaçları için ideal olmamaktadır. Çok sayıda nominal özelliğin olduğu ya da çok sayıda numerik özelliklerin bulunduğu veri setleri ideal olmayan durumlara örnek olarak gösterilebilir (Lantz, 2013: 120). Bu gibi özelliklere sahip veri setleriyle çalışan modelin çok sayıda karar (düğüm) noktasına sahip olması ve aşırı derecede kompleks bir yapısının olması kaçınılmaz olacaktır (Onan, 2015: 11). Karar ağacının kompleksliği, düğüm sayısı, yaprak sayısı, ağacın derinliği ve kullanılan özellik sayısına göre değerlendirilmektedir (Mesaric ve Sebalj, 2016: 375).

Karar ağaçları, yinelemeli bölümlenme (*recursive partitioning*) şeklinde adlandırılan, böl ve ele geçir esasına dayalı bir yaklaşım üzerine inşa edilmekte, özelliklerin sahip olduğu değerlere göre veri seti mümkün olduğunca küçük parçalara bölünmektedir (Williams, 2011: 209, 211; Fan, Ong ve Koh, 2006: 2304).

Ağacın kök noktasında, yani tüm veri setinin bulunduğu noktada, algoritma sınıflandırma için en yüksek düzeyde öngörü kapasitesine sahip özelliği belirlemektedir. Belirlenen bu özelliğe göre örnekler, sahip oldukları değerlere göre bölümlendirilmekte, böylelikle ağacın ilk dalları şekillendirilmiş olmaktadır. Algoritma böl ve ele geçir mantığını devam ettirmekte, düğüm noktalarında en iyi öngörüye sahip özelliği tespit etmek suretiyle, durdurma kriterlerine ulaşıncaya dek bu işlemi devam ettirmektedir (Cichosz, 2015: 72). Bahse konu durdurma kriterleri:

- Erişilen düğüm noktasındaki tüm örneklerin (ya da nerdeyse tüm örneklerin) aynı sınıfa tabi olması,
- Örnekler arasında bölümlenme yapabilecek başka bir özelliğin kalmamış olması,
- Ağacın, daha önceden belirlenmiş limit büyüklüğüne ulaşması,

Şeklinde ortaya çıkabilmektedir (Gorakala ve Usuelli, 2015: 21; Cichosz, 2015: 80).

Karar ağacı yöntemine dayalı algoritmanın güçlü ve zayıf yönleri Tablo 5'te gösterilmiştir (Şeker, 2013: 158-160; Tuffery, 2011: 313, 327-329; James, Witten, Hastie ve Tibshirani, 2013: 315-316).

Tablo 5: Karar Ağacı Algoritmasının Güçlü ve Zayıf Yönleri	
Güçlü Yönleri:	Zayıf Yönleri:
- Birçok farklı problemde, birçok farklı ihtiyaca cevap verebilen bir sınıflandırıcıdır.	- Algoritma, çok sayıda seviyeye sahip özelliklere karşı yanlı davranabilmektedir.
- Yüksek düzeyde otomatik öğrenme sürecine sahip olmakla birlikte, numerik ve nominal özelliklerin yanı sıra kayıp verilerle başa çıkabilmektedir.	- Algoritma ile çalışılırken yapılabilecek bir hata, ağaca özel amaçlara yönelik bir şekil verebilmekte, bu durumda aşırı uyum (overfitting), tam aksi durumda ise düşük uyum (underfitting) söz konusu olabilmektedir.
- Sadece en önemli özellikleri kullanır.	- Eksenlere paralel bölme yöntemi nedeniyle veri setindeki bazı ilişkilerin modellenmesinde güçlükler yaşanabilmektedir.
- Göreceli olarak az sayıda ya da çok sayıda eğitim verisinin olduğu veri setinde kullanılabilir.	- Eğitim verisindeki küçük değişiklikler, karar mantığında büyük değişikliklere sebebiyet verebilmektedir.
- Modelin sonuçlarını anlamak için matematiksel geçmiş	- Büyük ve geniş ağaçların anlaşılır kılınması zor

gerekmemektedir. Modelin anlaşılması ve yorumlanması basittir.	olmakla birlikte, ortaya çıkan sonuç beklentinin aksi ya da mantığa aykırı bir durumda olabilir.
- Diğer kompleks modellere kıyasla daha etkilidir.	

Algoritmanın karşılaştığı ilk zorluk hangi özelliğin bölümlenmede esas alınacağıdır (Kuhnert ve Venables, 2005: 261). Daha önce algoritmanın çalışma prensibi izah edilirken, böl ve yönet yaklaşımını benimsediği, durdurma kriterlerine erişene dek bu yaklaşımını devam ettirdiği ifade edilmişti. Gerçek verilere dayalı uygulamalarda karşılaşılabilecek en olası durdurma kriterinin, erişilen düğüm noktasındaki tüm örneklerin (ya da neredeyse tüm örneklerin) aynı sınıfa tabi olması olduğu belirtilebilir. Tamamen saf ya da homojen (*pure*) bir dal elde etmek çok zor olmakla birlikte, ağacın dalındaki ögelerin büyük çoğunluğunun aynı sınıfa dahil olması durumunda, o dalda durdurma kriterinin gerçekleştiği varsayılmakta ve algoritma o dal için çalışmasına son vermektedir (Berry ve Linoff, 2004: 177). Aksi halde, tamamen saf (homojen) bir dal elde edilene değin algoritmanın çalışmasına devam ettiği kabul edildiğinde, işlem maliyeti yükselmekle birlikte, elde edilecek sonucun değişmemesi durumunda, fazladan maliyete katlanılmış olacağı ortaya çıkacaktır.

Ağacın bölümlendirmesinde kullanılmak üzere farklı saflık (*purity*) ölçü yöntemleri bulunmakla birlikte, en çok kullanılan yöntem entropi hesabı olarak adlandırılmaktadır (Tuffery, 2011: 317). Örnekleme konu olan verilerin entropisi, bu verilerin ne kadar karışık sınıflara ait olduklarını göstermekle birlikte, entropi değeri “0” olduğunda, örnekleme konu olan verilerin tamamen homojen (saf) bir yapıya sahip olduğu, söz konusu değer “1” olduğunda ise örnekleme tamamen karışık, farklı sınıfların bulunduğu ifade edilmektedir (Williams, 2011: 212). Entropinin matematiksel tanımını aşağıdaki şekilde göstermek mümkün olmakla birlikte, söz konusu formülde geçen S , verinin incelemeye tabi olan belirli bir kısmını, c , farklı sınıf sayısını, p_i ise i sınıfının incelemeye konu veriler arasındaki (S içerisindeki) oranını

temsil etmektedir (Witten, Frank ve Hall, 2011: 104; Nguyen, Tran, Ma ve Sharma, 2017: 71).

$$Entropi(S) = \sum_{i=1}^c -p_i \log_2(p_i)$$

Homojenlik (saflık) derecesine ilişkin entropi hesabı yapıldıktan sonra, algoritmanın ağacın dallarını şekillendirirken hangi özelliği baz alarak bölümlene yapacağına karar vermesi gerekmektedir. Bunun tespiti için, bölümlenmede kullanılacak her bir özellik baz alındığında, entropi hesabı aracılığıyla homojenlik derecesindeki değişim hesaplanmaktadır (Cichosz, 2015: 39). Söz konusu hesaplama bilgi kazanımı (*information gain*) olarak adlandırılmakta ve aşağıdaki şekilde yapılmaktadır (Bell, 2015: 52). F özelliği için bilgi kazanımı, bu özelliğe göre yapılan bölümlenmeden önceki (S_1) entropi ile bölümlenmeden sonraki (S_2) entropi arasındaki fark olarak gösterilmektedir.

$$InfoGain(F) = Entropi(S_1) - Entropi(S_2)$$

Bölümlenme yapıldığı takdirde, veri seti birden fazla alt bölüme ayrılmakta, bu durumda Entropi (S_2) hesabında tüm bölümlerin toplam entropisinin hesaplanması gerekmektedir (Lantz, 2013: 127). Söz konusu hesaplamada, her bir bölümlenme, o bölüme düşen kayıtların oranına göre, ağırlık değeriyle çarpılmaktadır. Bahse konu hesaplama aşağıdaki formülde gösterilmiştir (Duda, Hart ve Stock, 2001: 398-399; Williams, 2011: 213).

$$Entropi(S) = \sum_{i=1}^n w_i Entropi(P_i)$$

Bölümlenme işlemi neticesinde toplam entropi, n adet bölümün, her bir bölümün entropisinin, o bölümdeki örneklerin oranına göre bulunan ağırlık ölçüsüyle (w_i) çarpımı neticesinde hesaplanan entropilerin toplamı olarak ifade edilmektedir.

Bölümlenme işleminde baz alınan özellik, bölümlenme işleminde ne denli homojen bir grup oluşmasını sağlıyorsa, söz konusu özelliğin sağladığı bilgi kazanımı da o denli yüksek olacaktır (Tuffery, 2011: 313). Bilgi kazanımının sıfır olarak hesaplanması, bölümlenme işlemindeki özelliğin entropide azalmaya sebep olmadığı manasına gelmektedir. Diğer yandan, maksimum bilgi kazanımı, bölümlenme işleminden önceki entropiye eşit olmaktadır. Netice itibariyle, bölümlenme işlemi

neticesinde hesaplanan entropinin sıfır olması, karara esas olan özelliğin tamamen homojen bir grup oluşturduğunun göstergesi olmakta (Hackeling, 2014: 103; Meenakshi ve Venkatachalam, 2015: 3191), bilgi kazanım oranı en yüksek olan özellik bölümlenmede kullanılacak özellik olarak seçilmektedir (Li, Jiang, Yao, Wang, Yiu, Huang, 2017: 7).

Yukarıdaki formülde nominal özellikler esas alınsa da karar ağaçları nümerik özelliklere dayalı bölümlenmeler neticesinde elde edilen bilgi kazanımını da kullanmaktadır (Cheng, Chen ve Liu, 2009: 91). Nümerik uygulamalarda yaygın bir şekilde kullanılan yöntem, belirli bir eşik değerinin altında ya da üstünde kalan değerlere göre bölümlenme yapılmasıdır (Williams, 2011: 208; Berry ve Linoff, 2004: 173). Bu tarz bir yaklaşım nümerik özelliği iki kategoriye sahip bir özelliğe indirgemekte, böylelikle bilgi kazanımı kolaylıkla hesaplanabilmektedir (Witten, Frank ve Hall, 2011: 193-194). Nümerik özellikler için kategoriye indirgeme yapılmadığı takdirde daha fazla hesaplama gerekebilmektedir.

Yukarıda açıklanan entropi hesabı yaklaşımını, kredilerin vadesinde ödenme durumunu simgeleyen bir örnek ile aşağıdaki şekilde izah etmek mümkün olacaktır. Söz konusu örnekte, 3 özellik (eğitim durumu, ekilen arazi büyüklüğü ve medeni hal) bulunmakta ve sınıflandırma kararı da kredinin vadesinde ödenip ödenmediği olmaktadır. Özelliklerine göre kredilerin durumu da Tablo 6'de gösterilmiştir.

Tablo 6: Karar Ağacı Algoritması Örnek Kredi Sınıflandırması

Özellik 1	Özellik 2	Özellik 3	Sınıf
<i>Eğitim Durumu</i>	<i>Ekilen Arazi Büyüklüğü</i>	<i>Medeni Hal</i>	<i>Kredinin Vadesinde Ödenmesi</i>
İlkokul	70	Bekar	Ödendi
İlkokul	90	Bekar	Ödenmedi
İlkokul	85	Evli	Ödenmedi
İlkokul	95	Evli	Ödenmedi
İlkokul	70	Evli	Ödendi
ortaokul	90	Bekar	Ödendi
ortaokul	78	Evli	Ödendi
ortaokul	65	Bekar	Ödendi
ortaokul	75	Evli	Ödendi

lise	80	Bekar	Ödenmedi
lise	70	Bekar	Ödenmedi
lise	80	Evli	Ödendi
lise	80	Evli	Ödendi
lise	96	Evli	Ödendi

Entropi hesabı için yukarıda belirtilen formül göz önüne alınarak öncelikle, sınıflandırma kararının yani kredinin vadesinde ödenip ödenmediği özelliğinin entropisinin aşağıdaki şekilde hesaplamak gerekecektir.

$$Entropi (S) = -9/14 * \log_2(9/14) - 5/14 * \log_2(5/14) = 0,940$$

Ağacın bölümlenmeye hangi özellikten başlayacağına karar verebilmesi için her bir özelliğin entropisi ve buna bağlı olarak her bir özellikten elde edilecek bilgi kazanımı aşağıdaki şekilde hesaplanır. Sadece eğitim durumu özelliğine ait entropi hesabının detayı verilmiş olmakla birlikte, diğer özelliklere ait entropi de benzer şekilde hesaplanmaktadır.

$$Entropi (S_1) = 5/14(-2/5 * \log_2(2/5) - 3/5 * \log_2(3/5)) + 4/14(-4/4 * \log_2(4/4) - 0/4 * \log_2(0/4)) + 5/14(-3/5 * \log_2(3/5) - 2/5 * \log_2(2/5)) = 0,694$$

$$Entropi (S_2) = 0,536413$$

$$Entropi (S_3) = 0,892158$$

$$InfoGain_1 = 0,94 - 0,694 = 0,246$$

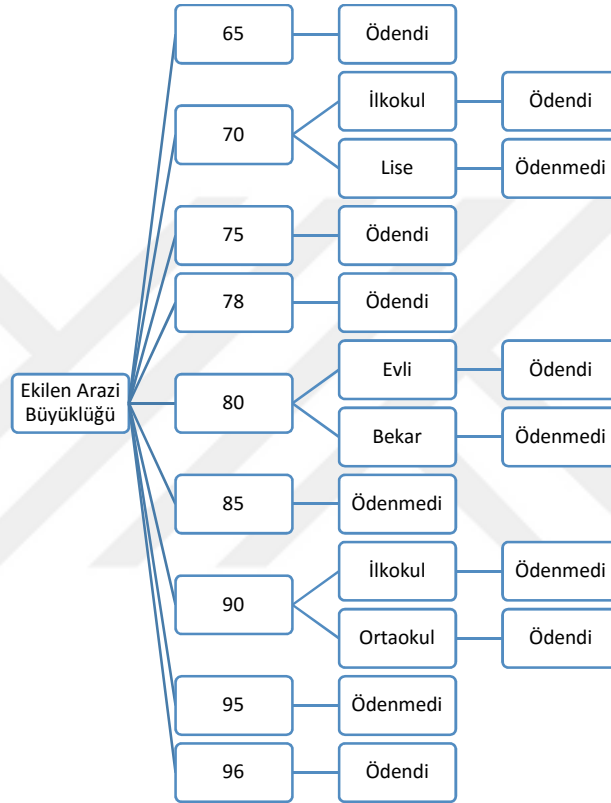
$$InfoGain_2 = 0,94 - 0,536413 = 0,403587$$

$$InfoGain_3 = 0,94 - 0,892158 = 0,047842$$

Yapılan hesaplamalar neticesinde, 2. özelliğin, yani ekilen arazi büyüklüğünün sınıflandırma kararı için daha fazla bilgi sağladığı görülmekte, netice itibariyle ilk düğüm noktasında bu özellik baz alınarak ağacın şekli ilk aşamada aşağıda gösterildiği şekilde oluşmaktadır. Bilgi kazanımının en fazla olduğu “ekilen arazi büyüklüğü” özelliğinden başlanarak ağacın dalları oluşturulduğunda, arazi büyüklüğünün 65, 75, 78, 85, 95 ve 96 dekar olduğu durumlarda, kullanılan kredilerin vadesinde ödenip ödenmeyeceğine dair doğrudan netice alınabilmektedir. Arazi büyüklüğü için geri kalan değerlerde, şahısların eğitim durumları ya da medeni hallerine göre söz konusu

model çıkarımında bulunmaktadır. Ağacın ikincil dalları oluşturulurken bölümlenme baz alınacak özelliğin (medeni hal ya da eğitim durumu) belirlenmesinde benzer şekilde entropi hesabı yapılarak en fazla bilgi sağlayan özellik kullanılacaktır. Nitekim arazi büyüklüğünün 70 dekar olduğu durumda şahısların eğitim durumları dikkate alınırken, arazi büyüklüğünün 80 dekar olduğu durumda kişilerin medeni halleri temel alınmıştır.

Şekil 3: Örnek Karar Ağacı



Yukarıdaki örnek karar ağacında, nümerik değerler içeren “*ekilen arazi büyüklüğü*” özelliği, olduğu gibi modele dahil edildiğinden ağacın oldukça yatay bir genişliğe sahip olduğu gözlenmektedir. Esasen karar ağacı algoritmasında, küçük boyutlu ve az derinlikli ağaçlar oluşturulması amaçlanmaktadır (Onan, 2015: 11). Diğer taraftan, daha önce de belirtildiği üzere nümerik değerler içeren özellik için kategori oluşturulması tercih edildiği takdirde, ağacın bambaşka bir yapıya bürünmesi mümkün olabilecektir. Örneğin, ekilen arazi büyüklüğü için 78 dekardan küçük ve 78 dekardan büyük değerler şeklinde iki kategori oluşturulduğunda, entropi hesaplaması neticesinde özelliklerin sağlayacağı bilgi kazanımı değerleri değişeceğinden, ağaç farklı bir özelliği, örneğin eğitim durumunu, bölümlenme ilk sırada kullanabilecektir.

Karar ağaçları, bilgi kazanımı yüksek özellikleri esas alarak daha küçük birçok dala ayrılabilir. Bu durum, dalın ucunda mükemmel sınıflandırmanın elde edilmesine ya da algoritmanın kullanabileceği farklı bir özelliğin kalmamasına kadar devam edebilmektedir (Duda, Hart ve Stock, 2001: 395-396). Ancak ağaç çok fazla büyüdüğünde, modelin vereceği sınıflandırma kararları çok spesifik olmakla birlikte, modelin eğitim verisine aşırı uyumu da söz konusu olabilecektir (Raschka, 2016: 81). Bu tarz sakıncaların önüne geçmek adına, karar ağaçlarında budama (*pruning*) yapılarak ağaçların çok fazla büyümesinin önüne geçilmekte ve test verileri için daha iyi genellemeler yapılması sağlanmaktadır (Tuffery, 2011: 319).

Karar ağaçlarının aşırı büyümesi durumunda karşılaşılan problemleri aşmak adına benimsenen çözümlerden ilki, ağaç dalının belirli sayıdaki karara ulaşması ya da düğüm noktasının sınırlı sayıda örnek içermesi durumunda ağacın büyümesine nihayet verilmesidir (Shwartz ve David, 2014: 254). Bu yöntem, erken durdurma ya da erken budama (*pre-pruning*) olarak adlandırılmaktadır (Alpaydın, 2004: 182). Söz konusu yaklaşımda, lüzumsuz gibi görülen çabadan kaçınmak cazip olsa da ağacın budanan kısmında ince fakat önemli bir ayrıntının gözden kaçırılmış olma ihtimalinin bulunduğu da bir gerçek olarak karşımıza çıkmaktadır (Witten, Frank ve Hall, 2011: 195).

Karar ağaçlarının aşırı büyümesi durumunda kabul gören ikinci yaklaşım ise geç budama (*post-pruning*) olarak isimlendirilmekte, bu yöntemde öncelikle ağacın büyümesine izin verilmekte, daha sonra düğüm noktalarındaki hata oranlarına göre ağaç uygun bir seviyeye kadar budanmaktadır (James, Witten, Hastie ve Tibshirani, 2013: 308). Bu yöntem, erken budamaya kıyasla daha efektif olmakta, ağaç büyümeden optimum derinliğin ne olması gerektiğine erken budamayla karar verilmesi zorluk teşkil etmektedir (Lantz, 2013: 128). Geç budama yaklaşımı algoritmaya, veri seti içerisindeki yapıların tamamını keşfetme imkânı sunmaktadır.

Karar ağacı algoritmasında, oluşturulan hiyerarşik veri yapısı ile böl ve yönet stratejisi benimsenmekte (Kokol, Pohorec, Stiglic ve Podgorelec, 2012: 241), parametrik olmayan metotlar arasında sayılan bu algoritma hem sınıflandırmada hem de regresyon çalışmalarında kullanılabilir (Alpaydın, 2004: 173).

2.4.4. Doğrusal Ayırma Analizi Algoritması

Doğrusal ayırma analizi, veri setindeki özelliklere göre yapılan çıkarımlar neticesinde parçalara bölünemeyen veri kümesini farklı şekilde bölmek için kullanılmaktadır (Boschetti ve Massaron, 2016: 141). Söz konusu algoritmanın kısaltması literatürde LDA (*Linear Discriminant Analysis*) şeklinde ifade edilse de algoritmayı 1936 yılında biyoloji alanındaki çalışmasıyla literatüre kazandıran kişi Ronald A. Fisher olduğundan, algoritma FLDA (*Fisher's Linear Discriminant Analysis*) şeklinde de ifade edilmektedir (Raschka, 21.02.2018). Söz konusu algoritma, iki sınıflı barındıran veri setleri için uygun olsa da ikiden fazla sınıflı barındıran durumlarda da bu algoritmanın kullanılması mümkündür (Tuffery, 2011: 58). C. R. Rao tarafından 1948 yılında yapılan çalışmada söz konusu algoritma, çoklu sınıflar barındıran biyolojik veri setleri için ilk kez kullanılmıştır (Raschka, 21.02.2018). Çok sayıda (ikiden fazla) sınıf barındıran veri setleri için yapılacak sınıflandırma çalışmalarında, sınıfların ikişerli gruplar halinde değerlendirmeye tabi tutulması gerekmektedir (Murphy, 2012: 271-273). Doğrusal ayırma analizi algoritmasının kullanılabilmesi için, veri setindeki özelliklere ait değerlerin sayısal olması, veri setinde kayıp veri bulunmaması diğer koşullar olarak sıralanabilir (Şeker, 2016: 65).

Birçok algoritma sınıflandırma yaparken, eksenlerdeki, yani her bir özellikteki, değerlere göre her bir örneği sınıflandırmaya çalışmaktadır. Doğrusal Ayırma Analizi Algoritması'nın çalışma prensibi, özellik sayısına bağlı boyut sayısında azaltma sağlayarak sınıflandırma kararına etki edecek kritik özellikleri belirlemektir (Boschetti ve Massaron, 2016: 141). Ancak bazı durumlarda, örneklerin eksenlerdeki özelliklere göre aldıkları değerler birbirlerine o denli yakın olmaktadır ki bu noktada herhangi iki örnek arasında ayırım yapmak mümkün olamamaktadır (Raschka, 2016: 139). Bu durumda doğrusal ayırma analizi veri setindeki özelliklerin istatistiksel dağılım özelliklerini baz alarak bir ayırım yapma yoluna gitmektedir (Buglear, 2014: 114). Doğrusal Ayırma Analizi Algoritması, dağılımların sergiledikleri eğim farklılıklarına göre sınıflandırma yapmakta, istatistiksel dağılımların belirlenmesi için de kovaryans matrislerinin oluşturulması gerekmektedir (Ledolter, 2013: 153).

Kovaryans matrisi, n adet rastgele değişkenden oluşan bir vektörde, her bir değişkenin diğer tüm değişkenlerle olan kovaryansını ortaya koyan matris olarak ifade edilmektedir (Rojas, 21.02.2018). Kovaryans, herhangi iki değişkenin birbirine ne

denli bağı olduğunu, içlerinden birindeki değişime diğerinin ne ölçüde duyarlı olduğunu ifade eden istatistiksel ölçü birimi olarak tanımlanmaktadır (Orloff ve Bloom, 21.02.2018).

Kovaryans matrisinin, veri setinde yer alan özellikler arasındaki ilişkiyi nasıl ortaya koyduğu aşağıdaki örnekte kısaca açıklanmıştır. Tablo 7’de, tarımsal kredi kullanan çiftçilerin, ekip biçtikleri arazi büyüklükleri, sahip oldukları büyükbaş ve küçükbaş hayvan sayıları verilmiştir. Söz konusu veriler ışığında, adım adım kovaryans matrisi oluşturularak, sayılan bu üç özellik arasındaki ilişki belirlenmeye çalışılacaktır.

Tablo 7: Kovaryans Matrisi Örnek Veri Seti

<i>Adı</i>	<i>Arazi Büyüklüğü (dekar)</i>	<i>Büyükbaş Hayvan Sayısı</i>	<i>Küçükbaş Hayvan Sayısı</i>
Ahmet	90	60	90
Ömer	90	90	30
Rıza	60	60	60
Ayşe	60	60	90
Fatih	30	30	30

İlk olarak tabloda sunulan veriler aşağıdaki şekilde matrise çevrilmektedir.

$$\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

Sonrasında, oluşturulan bu matris kullanılarak sapma matrisi aşağıdaki formülde gösterilen şekilde hesaplanır (Şeker, 2016: 67). Formülde A şeklindeki gösterim yukarıda oluşturulan matrisi, 1A gösterimindeki 1, A matrisinin genişliğine sahip 1 matrisini, n ise matrisin satır sayısını ifade etmektedir.

$$a = A - 1A\left(\frac{1}{n}\right)$$

Örnek tablodaki verilerle oluşturulan matris, formülde yerine konulduğunda aşağıdaki a matrisi elde edilir.

$$a = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} \left(\frac{1}{5}\right) =$$

$$= \begin{bmatrix} 24 & 0 & 30 \\ 24 & 30 & -30 \\ -6 & 0 & 0 \\ -6 & 0 & 30 \\ -36 & -30 & -30 \end{bmatrix}$$

“a” matrisinin hesaplanmasından sonra, bu matris, transpozu ile çarpılır ve aşağıdaki matris elde edilir.

$$a'a = \begin{bmatrix} 24 & 24 & -6 & -6 & -36 \\ 0 & 30 & 0 & 0 & -30 \\ 30 & -30 & 0 & 30 & -30 \end{bmatrix} \begin{bmatrix} 24 & 0 & 30 \\ 24 & 30 & -30 \\ -6 & 0 & 0 \\ -6 & 0 & 30 \\ -36 & -30 & -30 \end{bmatrix}$$

$$= \begin{bmatrix} 2520 & 1800 & 900 \\ 1800 & 1800 & 0 \\ 900 & 0 & 3600 \end{bmatrix}$$

Varyans matrisinin hesaplanabilmesi için, yukarıda elde edilen matrisin, ilk matristeki satır sayısına bölünmesi gerekmektedir (Şeker, 2016: 67).

$$V = \frac{a'a}{n} = \begin{bmatrix} 2520/5 & 1800/5 & 900/5 \\ 1800/5 & 1800/5 & 0/5 \\ 900/5 & 0/5 & 3600/5 \end{bmatrix} = \begin{bmatrix} \mathbf{504} & 360 & 180 \\ 360 & \mathbf{360} & 0 \\ 180 & 0 & \mathbf{720} \end{bmatrix}$$

Elde edilen nihai matriste, köşegende yer alan değerler, veri setindeki özelliklerin kendileri ile olan ilişkilerini, kendi varyanslarını ortaya koymaktadır (Maindonald, 2008: 53). Diğer bir deyişle, en yüksek varyansa sahip olan küçükbaş hayvan sayısı, en fazla hareketliliğe yani ortalama değerlerden en fazla sapmaya sahip olan özellik olarak karşımıza çıkmaktadır.

Nihai matriste, köşegen haricindeki değerler, özelliklerin birbirleri ile olan ilişkilerini, yani kovaryanslarını, ortaya koymaktadır (Murphy, 2012: 275). Matristeki kovaryans değerleri incelendiğinde, arazi büyüklüğü ile büyükbaş hayvan sayısı arasındaki ilişkinin diğer ikili ilişkilere göre daha güçlü olduğunu ifade etmek mümkün olmaktadır. Sonuç itibariyle, doğrusal ayırma analizi algoritması, özellikler

arasında bu şekilde hesaplanan kovaryans değerlerini baz alarak veri seti için sınıflandırma kararları vermektedir (Şeker, 2016: 68).

2.4.5. Lojistik Regresyon Algoritması

Regresyon yaklaşımı, lineer yani geçmişteki verilerin doğrusal bir dağılım sergilediği durumlara uyarlanarak, yaygın bir şekilde gelecekteki durumlara dair tahmin yapılmasında kullanılmaktadır (Prajapati, 2013: 150). Matematiksel olarak lineer regresyonu aşağıdaki şekilde göstermek mümkündür (Prajapati, 2013: 151).

$$y = ax + e$$
$$a = \frac{N\sum xy - (\sum x)(\sum y)}{N\sum x^2 - (\sum x)^2}$$
$$e = \frac{(\sum y - a(\sum x))}{N}$$

Doğrusal regresyona ait yukarıdaki formülde, “y” tahmin edilmesi hedeflenen değişkeni (bağımlı değişkeni), “ax” açıklayıcı değişkeni (bağımsız değişkeni), “e” ise regresyon doğrusunun y eksenini kestiği noktayı ifade etmektedir. “a” ve “e” değerlerinin hesaplandığı formüllerdeki x ve y değerleri veri setindeki değişkenleri, N veri setindeki gözlem sayısını ifade etmektedir. X değişkenin, y’nin tahmin edilmesindeki etkisini a katsayısı (regresyon doğrusunun eğimi) şekillendirmektedir. Söz konusu formülde sadece tek değişken olsa da uygulamada çok sayıda açıklayıcı değişken ($a_1x_1, a_2x_2, a_3x_3\dots$) yapılacak tahmin çalışmalarında kullanılmaktadır. Geçmiş verilere dayanılarak regresyon denklemi oluşturulduktan sonra, yeni “x” değerleri için tahminler yapılacaktır.

Lineer regresyon çok basit bir şekilde uygulansa da gerçek hayatta, verilerin dağılım özelliğine pek uyduğu söylenemez. Lineer regresyon, değişkenlerin normal dağılıma sahip olduğunu, dolayısıyla veri setinin simetrik bir yapıda olduğunu varsaymaktadır (Hackeling, 2014: 81). Genellikle “x” ve “y” değerleri arasındaki ilişki, doğrusal bir denklem ile modellenebilecek kadar basit olmayabilmektedir. Bu durumlarda logaritmik bir yaklaşım benimsenerek olasılık hesaplamasına dayalı bir sınıflandırma yöntemi kullanılmakta, söz konusu yöntem de lojistik regresyon olarak adlandırılmaktadır (Prajapati, 2013: 157). Bağımlı değişkenin, bağımsız değişkeni sürekli olarak karşılayabilecek nitelikte olduğu durumlar lojistik regresyon için söz

konusu olabilmekte, lineer regresyonda bağımlı deęişken için bu durum geçerli olamamaktadır (Buglear, 2014: 84).

Lojistik eşitlik, 1838 yılında nüfustaki artışa ilişkin geliştirilen modelde literatürde yerini almış olsa da çok sayıda kompleks hesaplama içermesi sebebiyle daha çok istatistikçiler tarafından kullanılagelmiş, ancak sahip olduęu niteliklerden ötürü son dönemde sınıflandırma problemlerinde tercih edilen bir algoritma olmuştur (Tuffery, 2011: 437).

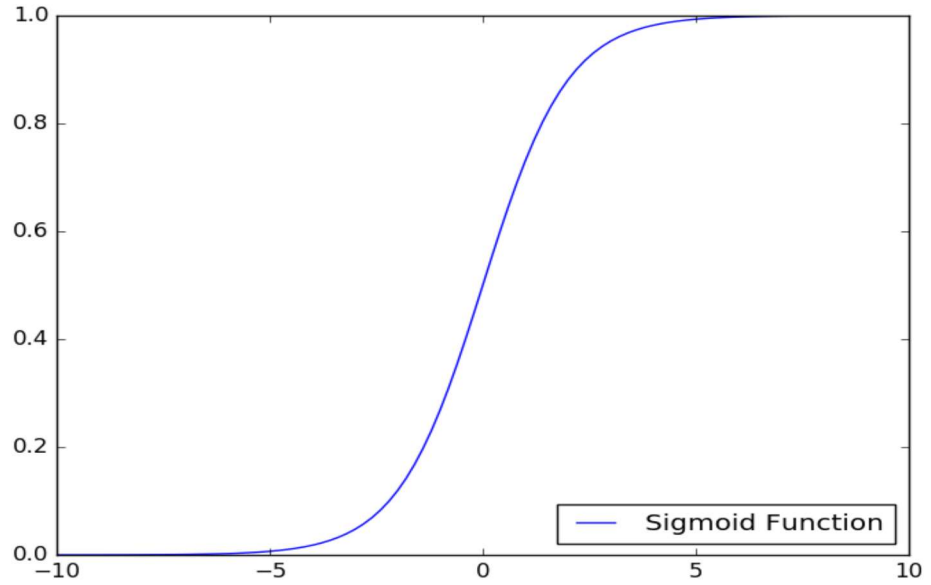
Lojistik regresyon yöntemi birçok farklı disiplinde kullanılmakla birlikte, iki bağımlı deęişkene (binomial) ya da çok sayıda bağımlı deęişkene (multinomial) sahip durumlarda uygulama alanı bulmaktadır (Prajapati, 2013: 157). Özellikle çok boyutlu yani çok sayıda özellięi barındıran veri setlerinde, problemi “binary (ikili)” yapıdaki bağımlı deęişkene göre şekillendirmek gerekmektedir (Ledolter, 2013: 83). Çok sayıda sınıfın olduęu durumlarda ise ikili sınıf yapısı baz alınarak seçilen sınıf ile geri kalan tüm sınıflar karşılaştırılmaktadır (Hackeling, 2014: 98). Lojistik regresyona dair temel lojistik fonksiyonlar aşağıda gösterilmiş olup, sınıflandırmaya ilişkin logaritmik oranlar ve olasılık hesabı, bu fonksiyonlar aracılığı ile yapılmaktadır (Zhao, 2013: 47).

$$\text{logit}(p) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$$
$$p = e^{\text{logit}(p)} / 1 + e^{\text{logit}(p)}$$

“logit (p)”, x açıklayıcı deęişkenlerinin lineer fonksiyonu olduęundan, söz konusu fonksiyonun çıktısı (0, 1) aralığında gerçekleşecektir (Ledolter, 2013: 84). Bu duruma istinaden (0, 1) aralığında olasılık belirlenecektir. Çoęu durumda olasılığın 0,5’ten yüksek olması durumunda, söz konusu olasılık 1 olarak kabul edilmekte ve sınıflandırma kararı bu eşik deęere göre verilmektedir (Prajapati, 2013: 158). Herhangi bir örnek için hesaplanan olasılık 0 (sıfır) olduęunda, söz konusu örnek A sınıfına, 1 olduęunda ise B sınıfına atanmış olacaktır (McClure, 2017: 79). Esasen algoritmanın temel odak noktası bağımsız deęişkenlerin katsayıları olan β deęerlerini, doęru sınıflandırma sağlayacak ölçüde belirleyebilmektir (Tattar, 2013: 207).

Bağımlı deęişkene ilişkin tahmin yapılırken lojistik fonksiyon, bir dięer deyişle sigmoid fonksiyon kullanılmakta, söz konusu fonksiyon Şekil 4’de gösterildięi şekilde bir yol izlemektedir (Zaccone, 2016: 99). Fonksiyonun izledięi yol göz önüne alındığında, bağımlı deęişkenin 0-1 aralığında deęerler alacaęı görülmektedir (Murphy, 2012: 21).

Şekil 4: Lojistik Fonksiyon-Sigmoid Fonksiyon Görseli



Lojistik regresyon algoritmasının avantajları ve dezavantajları Tablo 8’de sıralanmıştır (Tuffery, 2011: 477-478).

Tablo 8: Lojistik Regresyon Algoritmasının Avantajları ve Dezavantajları

Avantajları	Dezavantajları
<ul style="list-style-type: none">- Ayrık, kalitatif ve sürekli bağımsız değişkenlerle çalıştırılabilir.	<ul style="list-style-type: none">- Açıklayıcı değişkenler doğrusal olarak bağımsız olmalıdır.
<ul style="list-style-type: none">- Sıralı ya da nominal bağımlı değişkenlerle çalıştırılabilir.	<ul style="list-style-type: none">- Bazı durumlarda yaklaşık bir sayısal değerden öteye geçememektedir.
<ul style="list-style-type: none">- Doğrusal ayırma analizine kıyasla daha az kısıtlayıcı koşula sahiptir.	<ul style="list-style-type: none">- Sürekli değişkenlerde, kayıp veriler bulunduğu takdirde algoritma işlemez hale gelmektedir.
<ul style="list-style-type: none">- Önerdiği modellerin doğru sınıflandırma başarısı genelde yüksektir.	<ul style="list-style-type: none">- Uç değerlere sahip sürekli değişkenlere karşı duyarlıdır.

- Az sayıda örnekle iyi sonuçlar verebilmektedir.	
- Bağımsız değişkenler arasında etkileşimi mümkün kılmaktadır.	
- Doğrudan olasılık modellemesi yapabilmektedir.	
- Sonuçlar için güven aralığı sunabilmektedir.	

2.4.6. Yapay Sinir Ağları Algoritması

Yapay sinir ağları, bir takım girdi sinyalleri ve çıktı sinyalleri arasındaki ilişkiyi, biyolojik beynin algısal girdilere verdiği tepkilere benzer şekilde, modellemektedir (Witten, Frank ve Hall, 2011: 233). Beynin, birbirine bağlı nöron olarak adlandırılan hücreleri kullanması gibi, yapay sinir ağları da yapay nöronlar ya da düğüm noktalarından oluşan ağı kullanarak öğrenme problemlerini çözmektedir (Tuffery, 2011: 217).

Sinir ağının ilk örneği 1957 yılında Frank Rosenblatt tarafından geliştirilmiştir (Zocca, Spacagna, Slater ve Roelants, 2017: 23). Basit yapay sinir ağları 50 yılı aşkın bir süredir, insan beyninin problem çözmedeki yaklaşımını kopyalamak için kullanılmaktadır (Alpaydın, 2004: 229). Başlangıçta, bu çalışmalar “ve (AND)”, “ya da (OR)” gibi basit mantıksal öğrenme fonksiyonlarını içermekteydi (Witten, Frank ve Hall, 2011: 233). Söz konusu çalışmalar, biyolojik beynin fonksiyonlarının modellenmesi için kullanılmıştır (Nabiyev, 2003: 575). Ancak geçen zaman zarfında, bilgisayarların daha da güçlenmesi ile birlikte, yapay sinir ağları daha kompleks bir yapıya bürünmüştür. Öyle ki yapay sinir ağları:

- Sesli mesaj kaydetme servisleri ve posta sıralama makinelerince kullanılan konuşma ve el yazısı tanımlama programlarında,
- Ofis binaları çevre kontrol sistemleri, kendi kendine giden arabalar ve kendi kendine uçan hava taşıtları gibi akıllı cihazların otomasyon sistemlerinde,

- Hava ve iklim düzenlerinin sofistike modellenmesinde, gerilme dayanımı hesaplanmasında, akışkanlar dinamiğinde ve diğer birçok bilimsel, sosyal ve ekonomik olaylarda,

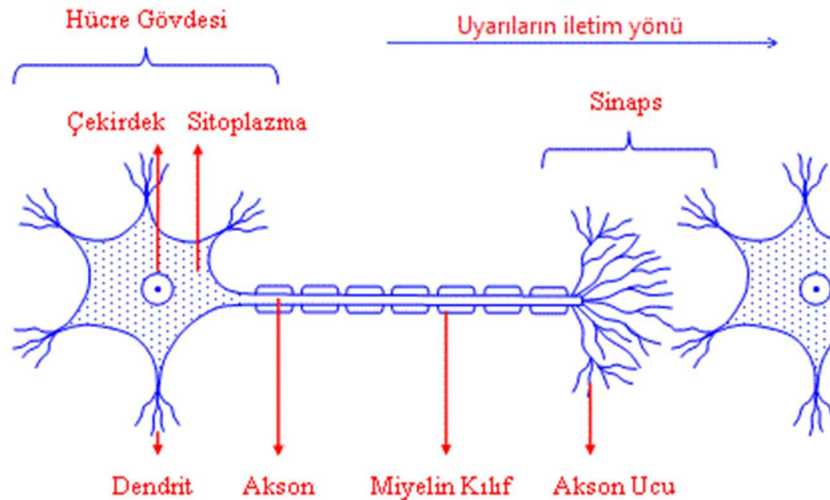
Kullanılmaktadır (Tuffery, 2011: 499-500).

Yapay sinir ağları, sınıflandırma, sayısal tahmin ve hatta denetimsiz örüntü (pattern) tanıma gibi hemen hemen tüm öğrenme görevlerinde uygulanabilmektedir (Lantz, 2013: 207).

Yapay sinir ağları, girdi ve çıktı verilerinin anlaşılır olduğu ya da yeterli düzeyde basit olduğu problemlerde oldukça iyi bir şekilde uygulanabilmekte, diğer taraftan girdileri çıktılara bağlayan süreç oldukça karmaşık bir yapıya da sahip olabilmektedir (Lantz, 2013: 207).

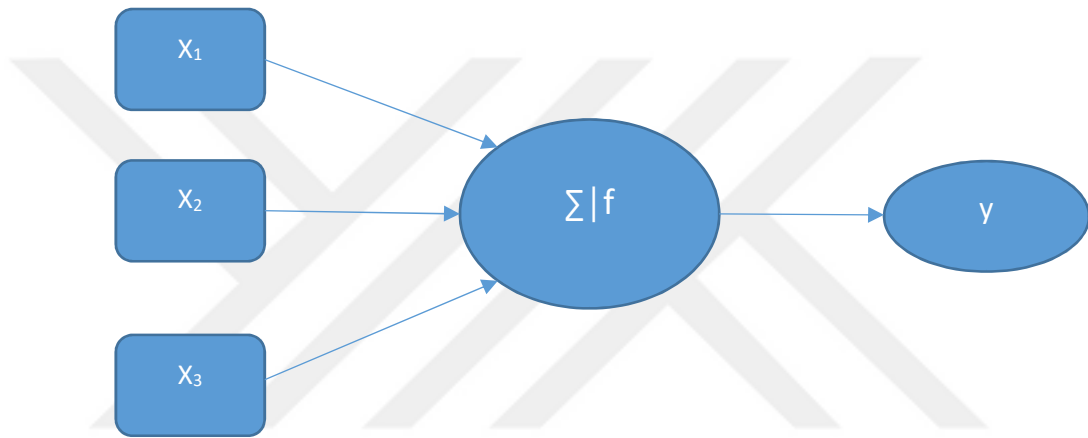
Yapay sinir ağları, insan beyninin çalışma prensibi baz alınarak tasarlandığı için öncelikle biyolojik nöron fonksiyonlarının anlaşılması gerekmektedir. Şekil 5’de gösterildiği üzere (Biyoloji Dersleri, 21.01.2018), gelen sinyaller dendrit olarak adlandırılan hücreler tarafından alınmakta, buradaki biyokimyasal süreç, hücreye gelen uyarıya, uyarının önemi (şiddeti) ya da sıklığına göre ağırlık atamaktadır (Eagleman, 2018: 100). Hücre gövdesi, gelen sinyalleri toplamakta, eşik değerine ulaşıldığında sinir hücresi ateşlenmekte, akabinde çıktı sinyali aksondaki elektrokimyasal süreç vasıtasıyla iletilmektedir. Akson terminalinde (ucunda) ise elektrik sinyali, tekrar kimyasal bir sinyal olarak işlenmekte, sonrasında komşu nörona, aradaki sinaps denilen boşluk vasıtasıyla iletilmektedir (Yıldız, 2009: 10-11).

Şekil 5: Sinir Hücresi Yapısı



Yapay nöronun çalışma modeli, biyolojik versiyonuna oldukça benzemektedir. Şekil 6’de, yönlendirilmiş ağ diyagramının, dendritler tarafından alınan girdi sinyalleri (x değişkenleri) ile çıktı sinyali (y değişkeni) arasındaki ilişkiyi tanımlaması tasvir edilmiştir (Schmidt, 30.01.2018). Biyolojik nöronda olduğu gibi, her bir dendritin sinyaline, sinyalin önemine göre ağırlık (w_i) atanmaktadır (Kuhnert ve Verables, 2005: 251). Girdi sinyalleri hücre gövdesinde toplanmakta, aktivasyon fonksiyonuna (f) göre sinyalin iletimi sağlanmaktadır (Tuffery, 2011: 218).

Şekil 6: Yapay Nöron



Yapay nöronun, n adet girdi dendriti ile matematiksel olarak aşağıdaki formülle ifade edilmesi mümkündür (Torres, 2016: 83). N adet girdinin (x), girdi sinyallerine etkisi, ağırlık ölçütü (w) ile belirlenmektedir. Toplamda elde edilen değer, aktivasyon fonksiyonu (f(x)) tarafından kullanılmakta, buna göre sonuç sinyali (y(x)) çıktı aksonu olarak ifade edilmektedir (Zaccone, Karim ve Menshaw, 2017: 14).

$$y(x) = f\left(\sum_{i=1}^n w_i x_i\right)$$

Sinir ağları, bu şekilde tanımlanan nöronları kullanarak veri setleri için kompleks modeller geliştirmektedir (Russell ve Norvig, 2003: 748). Çok sayıda sinir ağı varyasyonu bulunmakla birlikte, üç temel karakteristiğin söz konusu varyasyonlarda bulunduğu ifade edilebilir (Zocca, Spacagna, Slater ve Roelants, 2017: 39):

- Bir nöronun net girdi sinyalini, çıktı sinyaline dönüştürerek bu sinyalin ağ içerisindeki yayılmasını sağlayan aktivasyon fonksiyonu,

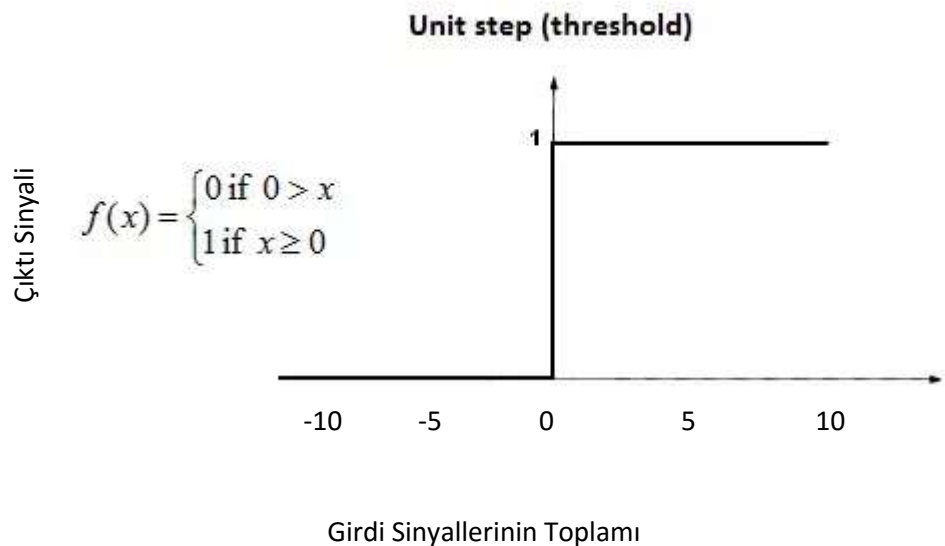
- Modeldeki nöronların sayısını, katmanların sayısını ve bunların birbirleriyle olan bağımlı şekillendiren yolu içeren ağ mimarisi,
- Girdi sinyalinde, nöronları engelleyen ya da harekete geçiren ağırlıkların belirlendiği öğrenme algoritması.

Aktivasyon fonksiyonu, yapay nöronun bilgiyi işlediği ve ağ boyunca iletimini sağladığı mekanizma olarak tanımlanabilir (McClure, 2017: 14). Aktivasyon fonksiyonunun modellenmesinde, biyolojik nöronların çalışma prensibinden etkilenilmiştir.

Biyolojik sinir hücresinde, aktivasyon fonksiyonu toplam girdi sinyallerini toplayan ve söz konusu sinyal toplamının ateşleme için aşılması gereken eşik değerine ulaşmış ulaşmadığı kararını veren bir süreç olarak düşünülebilmektedir (Zaccone, 2016: 96). Eşik değerine ulaşıldığında sinyal iletilmekte, aksi takdirde hiçbir aktivite gerçekleşmemektedir (Zocca, Spacagna, Slater ve Roelants, 2017: 41). Yapay sinir ağları terminolojisinde, bu durum eşik aktivasyon fonksiyonu olarak adlandırılmakta, ancak belirli düzeyde girdi sinyali geldiğinde çıktı sinyali üretilmektedir (Nabiyev, 2003: 581).

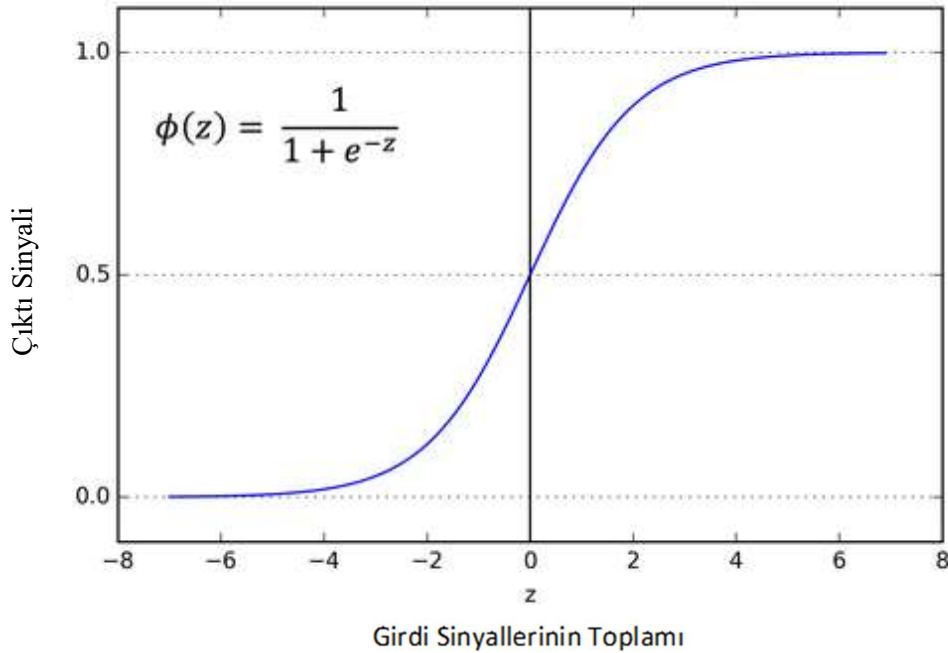
Şekil 7’de tipik eşik fonksiyonu tasvir edilmiş, girdi sinyallerinin toplamı en az sıfır olduğunda nöronun ateşlendiği ifade edilmiştir (Cox, 22.01.2018). Şekil 7, biçiminden ötürü birim aktivasyon fonksiyonu olarak da isimlendirilmektedir (Zaccone, Karim ve Menshawy, 2017: 14-15).

Şekil 7: Birim Aktivasyon Fonksiyonu



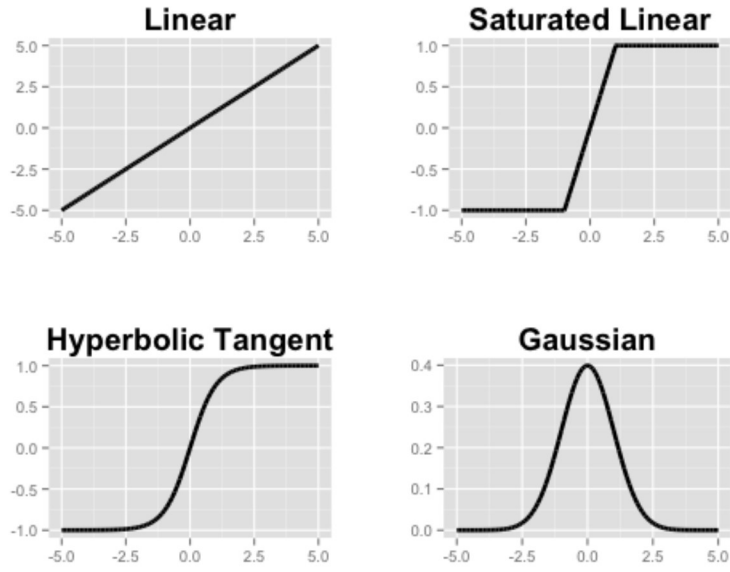
Biyolojik versiyonuna oldukça benzer bir yapıya sahip olmasına rağmen, eşik (birim) aktivasyon fonksiyonun, yapay sinir ağlarında nadiren kullanıldığı söylenebilir (Hastie, Tibshirani ve Friedman, 2001: 352-353). Biyokimyanın kısıtlarından sıyrılmış olarak, yapay sinir ağlarında aktivasyon fonksiyonlarının seçimi, söz konusu fonksiyonların istenen matematiksel karakteristikleri sergileyebilmesine ve veri içerisindeki ilişkileri modelleyebilmesine göre yapılmaktadır (Alpaydın, 2004: 233-236). Bu durumdan ötürü, en çok kullanılan alternatif fonksiyonun sigmoid aktivasyon fonksiyonu olduğu ifade edilebilir (Zocca, Spacagna, Slater ve Roelants, 2017: 46). Şekil 8’de betimlenen sigmoid aktivasyon fonksiyonu, eşik aktivasyon fonksiyonuna benzer bir “S” şekline sahip görünse de çıktı sinyali ikili bir sonuç olmaktan ziyade 0 ile 1 arasında değerler alabilmektedir (Sharma, 22.01.2018). Ayrıca, sigmoid fonksiyonunun türevi alınabilmekte, yani girdi aralığındaki tüm değerler için türev hesaplanabilmektedir (Bishop, 2006: 227-229).

Şekil 8: Sigmoid Aktivasyon Fonksiyonu



Sigmoid en yaygın kullanılan aktivasyon fonksiyonu olmakla birlikte bazı sinir ağları algoritmaları farklı alternatifleri tercih edebilmektedir. Söz konusu alternatifler arasında yer alan fonksiyonlar Şekil 9’da görsel olarak ifade edilmiştir (Lantz, 2013: 210).

Şekil 9: Alternatif Aktivasyon Fonksiyonları



Belirtilen alternatif aktivasyon fonksiyonları arasındaki temel farklılık, çıktı sinyal aralığı olarak karşımıza çıkmaktadır. Çıktı sinyal aralığı (0, 1), (-1, +1) ya da $(-\infty, +\infty)$ olarak görülebilmektedir (Lantz, 2013: 210). Aktivasyon fonksiyonun seçimi, sinir ağının uygulanacağı veri tiplerine bağlı olmakla birlikte, uygun fonksiyonun seçimi veriye göre özelleştirilmiş sinir ağlarının inşasına da imkan sağlamaktadır (Bishop, 2006: 227-228).

Aktivasyon fonksiyonlarının büyük çoğunluğunda, girdi sinyallerinin aralığının göreceli olarak dar olması gözden kaçırılmaması gereken bir ayrıntı olarak karşımıza çıkmaktadır (Hastie, Tibshirani ve Friedman, 2001: 358). Örneğin, sigmoid fonksiyonunda, girdi sinyalinin -5'in altında ya da +5'in üstünde olması durumunda çıktı sinyali her zaman 0 ya da 1 olarak hesaplanacaktır. Sinyalin bu şekilde sıkıştırılması üretilen sinyalin kalitesini olumsuz yönde etkileyebilecektir (Tuffer, 2011: 499). Bu tarz problemlerin önüne geçmek adına tüm sinir ağındaki girdilerin sifıra yaklaşık değerler aralığına indirgenmesi gerekebilmektedir (Tuffery, 2011: 219). İndirgeme işlemi, veri setindeki özellikler için normalizasyon ya da standardizasyon uygulanarak yapılabilmektedir (Lantz, 2013: 210). Girdi değerlerinin belirli bir aralığa indirgenmesi ile aktivasyon fonksiyonu tüm aralık için aktif bir şekilde çalışabilecek, geniş aralıkta değerler alan özelliklerin küçük aralığa sahip özellikleri bastırmasının önüne geçilebilecektir (Tuffery, 2011: 219). Bu yaklaşımın bir diğer faydası ise modelin eğitim safhasında daha hızlı olmasının sağlanmasıdır. Çünkü algoritma girdi

değerlerinin aralığında daha hızlı yineleme (deneme) yapabilir hale gelmektedir (Lantz, 2013: 210).

Sinir ağlarının öğrenme hususundaki kapasitesini belirleyen unsur, söz konusu ağın topolojisi yani ağın modeli, bir diğer deyişle yapısıdır (Tuffery, 2011: 220). Ağın yapısı birçok farklı şekilde dizayn edilebilmekle beraber bu yapıyı etkileyen üç temel etmeni:

- Katmanların sayısı,
- Ağ içerisindeki bilginin, geriye doğru iletilmesine imkân olup olmadığı,
- Her bir katmandaki düğüm sayısı,

Şeklinde sıralamak mümkündür (Duda, Hart ve Stork, 2001: 283).

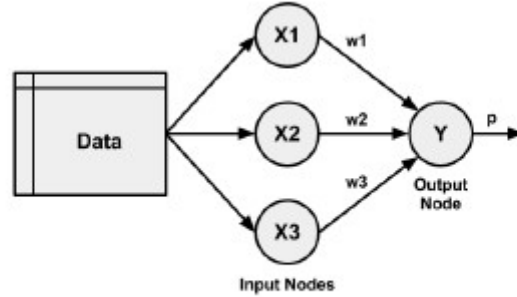
Ağın topolojisi, öğrenilmesi amaçlanan görevlerin kompleksliğine karar vermekte, genellikle geniş ve kompleks yapıya sahip ağlar, veri setindeki keşfi zor detayları ve karışık karar sınırlarını tanımlamada başarılı olabilmektedir (Russell ve Norvig, 2003: 748). Ancak ağın gücü sadece ağın genişliği ile değil ağ içerisindeki birimlerin nasıl tasarlandığı ile ilgilidir (Alpaydın, 2004: 259).

Topolojinin tanımlanması için, yapay nöronların ağ içerisindeki pozisyonlarına göre ayırt edilmesini sağlayan bir terminolojiye ihtiyaç duyulmaktadır (Hastie, Tibshirani ve Friedman, 2001: 353). Şekil 10'da oldukça basit bir ağın topolojisi sunulmuştur (Lantz, 2013: 212). Bir dizi nörondan oluşan ve işlenmemiş sinyalleri doğrudan girdi verisinden alan bölümler girdi düğümleri (input nodes) olarak isimlendirilmektedir. Her bir girdi düğümü, veri setindeki bir özelliğin işlenmesinden sorumlu olmakta, özelliğin aldığı değer, düğümün aktivasyon fonksiyonu tarafından dönüştürülmektedir. Girdi düğümlerinden çıkan sinyaller de çıkış düğümüne (output node) iletilmekte, çıkış düğümü kendi aktivasyon fonksiyonunu kullanarak nihai öngörüde (p) bulunmaktadır.

Girdi ve çıktı düğümleri, katman (layer) olarak isimlendirilen gruplar altında toplanmaktadır (Berry ve Linoff, 2004: 221). Girdi düğümleri, gelen veriyi olduğu gibi işlediğinden, ağın sadece bir set bağlantı ağırlıkları bulunmaktadır. Şekil 10'da söz konusu ağırlıklar w_1 , w_2 , w_3 şeklinde gösterilmiştir. Bu durumdan ötürü Şekil 10'daki gösterim tek katmanlı ağ (single-layer network) olarak adlandırılmaktadır (Russell ve Norvig, 2003: 740). Tek katmanlı ağlar basit sınıflandırma modellemeleri için kullanılmakta, özellikle veri seti için lineer ayrımların yapılmasının mümkün olduğu

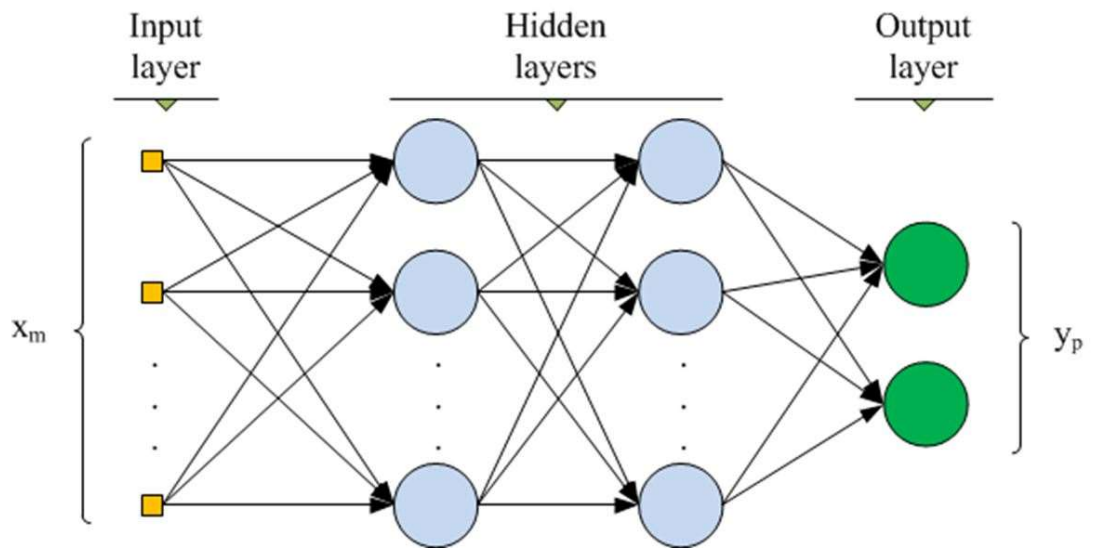
durumlarda başarılı neticeler alınabilmektedir (Torres, 2016: 83). Ancak daha karmaşık öğrenme modelleri için daha sofistike ağlara ihtiyaç duyulabilmektedir.

Şekil 10: Tek Katmanlı Ağ (Single-Layer Network)



Kompleks ağların inşası için daha fazla katman eklenmesi gerekmektedir (Torres, 2016: 104). Şekil 11’de, bir ya da daha fazla gizli katmanın (hidden layer) bulunduğu çok katmanlı ağda, sinyaller gizli katmanda işlenerek çıktı düğümüne iletilmektedir (Caparrini, 25.01.2018). Çok katmanlı ağların büyük çoğunluğunda, ağdaki birimlerin tamamı birbirine bağlantılıdır. Ancak çok katmanlı ağlarda, her bir nöron sonraki katmanda yer alan tüm nöronlarla bağlantılı olmalı, aynı katmandaki nöronlar arasında bağ olmamalı, birbirini takip etmeyen katmanların nöronları arasında bağ olmamalı, katman sayısı ve her bir katmandaki nöron sayısı probleme göre belirlenmelidir (Zaccone, Karim ve Menshawy, 2017: 19).

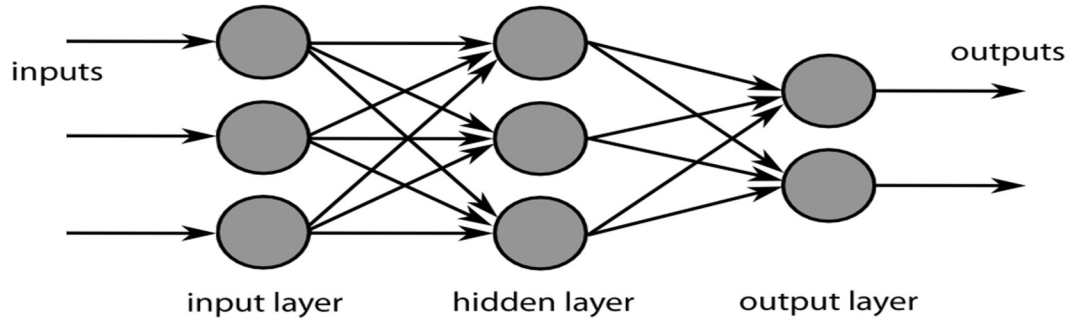
Şekil 11: Çok Katmanlı Ağ (Multi-Layer Network)



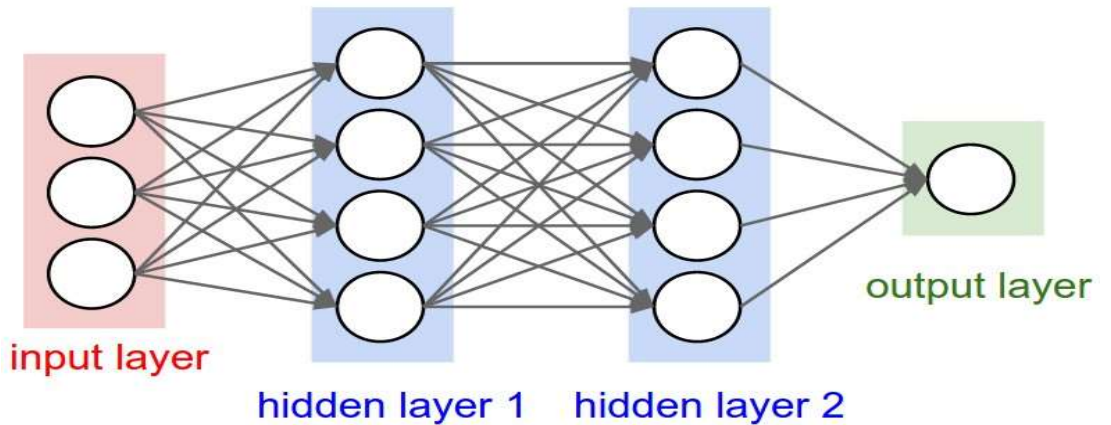
Gerek tek katmanlı gerekse çok katmanlı ağlar için sıralanan görsellerde, bağlantıları sağlayan okların uçları, ağ içerisinde iletilen sinyallerin tek yönlü olduğunu ifade etmektedir. Girdi sinyallerinin sürekli olarak sadece bir yönde, bağlantıdan bağlantıya iletiildiği ve en nihayetinde çıktı katmanına ulaşıldığı ağlar, ileri beslemeli ağlar (feedforward networks) olarak adlandırılmaktadır (Zaccone, Karim ve Menshawy, 2017: 74).

Bilgi akışında bir sınırlama bulunsa da ileri beslemeli ağlar şaşırtıcı derecede esneklik sağlamaktadırlar (Lantz, 2013: 213). Örneğin, katman sayısı ve her bir katmandaki düğüm sayısı değişebilmekte, çok sayıda çıktı aynı anda modellenebilmekte (Şekil 12) ya da çok sayıda gizli katman (Şekil 13) eklenebilmektedir (Feng, 25.01.2018; Rosebrock, 25.01.2018).

Şekil 12: Çoklu Çıktı Düğümleri (Multiple Output Nodes)



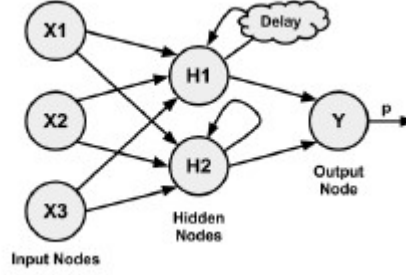
Şekil 13: Çoklu Gizli Katmanlar (Multiple Hidden Layers)



İleri beslemeli ağlardan farklı olarak, tekrarlı ya da geri beslemeli ağlarda (feedback network) sinyallerin döngüler sayesinde hem ileri hem de geri taşınmasına imkân sağlanmaktadır (Hackeling, 2014: 194). Söz konusu ağın bu özelliği, biyolojik

sinir ağlarının çalışma prensibine ışık tutmakla birlikte, oldukça kompleks modellerin öğrenilmesini mümkün kılmaktadır (Russell ve Norvig, 2003: 739). Kısa dönemli hafızanın eklenmesiyle birlikte geri beslemeli ağların gücü oldukça artırılmış olmaktadır. Basit bir geri beslemeli ağ modeli Şekil 14’de gösterilmiştir (Lantz, 2013: 214).

Şekil 14: Geri Beslemeli Ağ (Feedback Network)



Yüksek bir potansiyele sahip olmasına rağmen geri beslemeli ağların oldukça teorik olduğunu ve uygulamada nadiren kullanıldığını, buna karşın ileri beslemeli ağların geniş bir uygulama alanına sahip olduğunu söylemek mümkündür (Lantz, 2013: 214). Aslında, çok katmanlı ileri beslemeli ağların (Multilayer Perceptron) yapay sinir ağları topolojisinde fiili standart olduğu ifade edilebilir (Murphy, 2012: 563).

Katman sayısındaki çeşitlilik ve bilginin akış yönündeki farklılıkların yanı sıra, sinir ağlarının kompleks yapısı, her bir katmandaki düğüm sayısına göre de değişmektedir (Lantz, 2013: 214). Girdi düğümlerinin sayısı, girdi verilerindeki özellik sayısına göre önceden belirlenmektedir (Duda, Hart ve Stork, 2001: 283). Benzer şekilde, çıktı düğüm sayısı da modellenmek istenen çıktı sayısına, bir diğer deyişle çıktındaki sınıf sayısına göre önceden şekillendirilmektedir (Berry ve Linoff, 2004: 227). Ancak gizli düğüm noktalarının sayısı, modelin eğitilmesinden önce kullanıcının kararına bırakılmaktadır (Bell, 2015: 97). Lakin gizli katmandaki nöron sayısının belirlenmesinde güvenilir bir kural bulunmamaktadır (Duda, Hart ve Stork, 2001: 283). Uygun nöron sayısı, girdi düğüm sayısına, eğitim verisinin büyüklüğüne, anlamsız verilerin çokluğuna ve modelin öğrenme konusunun karmaşıklığına dayanmaktadır (Raschka, 2016: 345).

Genellikle, daha kompleks yapıdaki ağ topolojileri, daha karmaşık problemlerin öğrenilmesine imkan sağlamaktadır (Alpaydın, 2004: 241-242). Daha

fazla nöron, eğitim verisinin daha iyi yansıtılmasına olanak sunmakla birlikte, modelin aşırı uyumu (overfitting) söz konusu olabilecektir (Tuffery, 2011: 220-221). Geniş sinir ağları ayrıca yüksek işlem maliyetine sahip olmakla birlikte, eğitimi uzun zaman alan yapılardır (Shwartz ve David, 2014: 276).

En ideal uygulama, yeterli düzeyde performans gösteren az sayıda düğümün kullanılması ile oluşturulan modeldir (Alpaydın, 2004: 259). Birçok uygulama örneğinde, az sayıda gizli düğüme sahip sinir ağının yüksek derecede öğrenme kabiliyeti sergilediği görülebilmektedir (Lantz, 2013: 215).

Ağ topolojisinin kendi kendine öğrenmesi mümkün olmamakta, ağın deneyimler vasıtasıyla öğrenmesine imkan sağlanması gerekmektedir (Nabiyev, 2003: 593). Girdi verilerin sinir ağı tarafından işlenmesiyle birlikte nöronlar arasındaki bağlantılar güçlendirilmekte ya da zayıflatılmakta, böylelikle ağdaki bağlantıların ağırlıkları modeli yansıtır hale getirilmektedir (Russell ve Norvig, 2003: 737).

Bağlantı ağırlıklarının ayarlanmasıyla sinir ağının eğitilmesi oldukça yoğun bir hesaplama içermektedir (Duda, Hart ve Stork, 2001: 224-227). Bu yönde geliştirilen algoritma geri yayılım (backpropagation) yöntemi olarak isimlendirilmekte, söz konusu stratejide geri yayılım hata oranları kullanılmaktadır (Zocca, Spacagna, Slater ve Roelants, 2017: 51). Birçok makine öğrenmesi algoritmasına kıyasla yavaş olmasına rağmen geri yayılım metodu yapay sinir ağlarına olan ilginin yeniden canlanmasını sağlamıştır (Raschka, 2016: 372). Bu durumun sonucu olarak çok katmanlı ileri beslemeli ağlar geri yayılım algoritmasını kullanır hale gelmiştir (Witten, Frank ve Hall, 2011: 469). Bu tarz modellerin sahip olduğu avantajlar ve dezavantajlar Tablo 9’da sıralanmıştır (Lantz, 2013: 215).

Tablo 9: Geri Yayılım Algoritmasının Avantaj ve Dezavantajları

Avantajları	Dezavantajları
- Sınıflandırma ya da sayısal öngörü problemlerinde kullanılabilirlik.	- Yoğun hesaplama gerektirmekle birlikte eğitimi uzun zaman almaktadır.
- En isabetli modelleme yaklaşımlar arasındadır.	- Eğitim verisine aşırı uyum ya da zayıf uyum kolaylıkla gerçekleşebilmektedir.

- Veri setindeki ilişkilere dair az sayıda varsayımda bulunmaktadır.	- Sonuçların anlaşılır kılınması oldukça zor bir hal alabilmektedir.
--	--

En genel yapısıyla geri yayılım algoritması, iki sürecin çok sayıdaki döngülerini tekrarlamaktadır (Bell, 2015: 99). Her bir yineleme dönem (epoch) olarak adlandırılmaktadır (Witten, Frank ve Hall, 2011: 471). Başlangıçta ağ, öncül bilgilere sahip olmadığından, ağ içerisindeki ağırlıklar rastgele belirlenmektedir (Tuffery, 2011: 225). Sonrasında algoritma, durma kriterlerine ulaşana dek, süreç boyunca döngüleri çalıştırmaktadır (Zaccone, Karim ve Menshawy, 2017: 75). Döngü ileri aşama (forward phase) ve geri aşama (backward phase) şeklinde 2 temel unsuru içermektedir (Hastie, Tibshirani ve Friedman, 2001: 353). İleri aşamada, nöronlar giriş katmanından çıkış katmanına sırasıyla aktive edilmekte, her bir nöron için ağırlık ve aktivasyon fonksiyonu uygulanmaktadır (Berry ve Linoff, 2004: 229). Bitiş katmanına ulaşılmasının ardından çıktı sinyali üretilmektedir. Geri aşamada ise, ağın çıktı sinyali eğitim verisindeki gerçek değer ile karşılaştırılmaktadır (Tuffery, 2011: 218). Ağın ürettiği çıktı sinyali ile gerçek sonuçlar arasındaki hata, geri yayılım ile ağdaki ağırlıkların yeniden gözden geçirilmesinde kullanılmakta, böylelikle hata oranlarının ortadan kaldırılması amaçlanmaktadır (Zaccone, Karim ve Menshawy, 2017: 15).

Zaman içerisinde ağ, geriye doğru gönderdiği bilgiyi toplam hata oranının azaltılmasında kullanılmaktadır (Tuffery, 2011: 226). Söz konusu işlemler yapılırken, algoritmanın ağdaki ağırlıkları ne ölçüde değiştireceği sorusu karşımıza çıkmaktadır (Duda, Hart ve Stork, 2001: 282). Ağdaki ağırlıkların değiştirilmesinde gradyan iniş (gradient descent) olarak isimlendirilen teknik kullanılmaktadır (Zaccone, Karim ve Menshawy, 2017: 16). Bu tekniğin çalışması, ormanda su arayan gezginin izleyeceği yola benzetilmektedir. Öncelikle arazinin yapısını anlamaya çalışan gezgin, aşağı doğru en yüksek eğime sahip bölgeyi bularak, su yatağı olma ihtimali en yüksek olan en alçak bölgeye doğru bu yoldan hızla ilerlemeye çalışacaktır. Benzer şekilde, geri yayılım algoritması her bir nöronun aktivasyon fonksiyonunun türevini kullanarak her bir girdi ağırlığındaki değişimin eğimini tanımlamaktadır (Zocca, Spacagna, Slater ve Roelants, 2017: 57-58). Eğitim sayesinde, ağırlıklardaki her bir değişimin hata oranları üzerindeki etkisi araştırılmaktadır (Zaccone, Karim ve Menshawy, 2017: 17). Algoritma hata oranında en fazla azalmayı sağlayan ölçüde ağırlıklarda değişiklik

yapmakta, söz konusu ölçü, öğrenme oranı (learning rate) olarak adlandırılmaktadır (Witten, Frank ve Hall, 2011: 471). Öğrenme oranı ne denli büyük olursa algoritmanın o denli hızda hedefine ulaşması mümkün olmaktadır.

2.4.7. ZeroR Algoritması

Makine öğrenmesi algoritmaları arasında yer almakla birlikte en temel ve en basit algoritmanın ZeroR algoritması olduğu ifade edilebilir (Şeker, 2016: 55). ZeroR algoritması, veri setindeki baskın sınıfa göre seçim yapmaktadır (Arvanitakis, Karydis, Kermanidis ve Avlonitis, 2017: 3). Algoritmanın çalışma prensibi basit bir şekilde örneklendirilecek olursa; kredi için başvuru yapmış 100 kişiden, 65'inin medeni halinin "evli" ve aylık net gelirinin "5.000 TL üzerinde" olduğu, geri kalan kişilerin ise medeni halinin "bekar" ve aylık net gelirlerinin "5.000 TL'den az" olduğu kabul edilsin. İlk gruptaki yani evli olup geliri yüksek olanların, kullanmış oldukları kredileri vadesinde ödedikleri, diğer grubun ise kredilerini vadesinde ödemedikleri sonucu elde edilmiş sayılsın. Kredi için başvuracak yeni bir kişinin medeni halinin bekâr, aylık net gelirinin ise 2.000 TL olduğu durumda, bu şahsın ikinci gruba, yani kredisini vadesinde ödemeyen gruba benzer özelliklere haiz olduğu, dolayısıyla kredisini vadesinde ödemeyecek şeklinde sınıflandırılması gerektiği söylenebilir. Ancak, ZeroR algoritmasının değerlendirmesi farklı olmaktadır. 100 kişilik veri setinde baskın olan sınıfın, kredisini vadesinde ödeyen ve toplam eleman sayısı 65 olan sınıf olduğu dikkate alındığında, ZeroR algoritması yeni gelen bireyi, medeni hali ve aylık net geliri her ne olursa olsun, kredisini vadesinde ödeyecek kişi olarak sınıflandıracaktır. Bu durum daha önce de ifade edildiği üzere, algoritmanın veri setinde çoğunluğa sahip sınıfı esas almasından ve bu baskın sınıfa göre yeni eleman için sınıflandırma yapmasından kaynaklanmaktadır (Chen, Chang, Su ve Shieh, 2010: 2504-2505; Şeker ve Ocak, 2017: 5).

Algoritmanın yukarıda açıklanan en temel özelliği, ilk bakışta basit ve anlamsız gelebilmektedir. Ancak, ZeroR algoritması ile elde edilecek doğru sınıflandırma başarısı, diğer algoritmaların performansının değerlendirilmesinde "referans değer (baseline)" olarak kabul edilecektir (Cotha ve Sokolova, 2015: 42). ZeroR algoritması, verilerin sahip oldukları özelliklere, her bir özelliğe göre alınan değerlere ve özellikler arasındaki ilişkiler gibi kompleks çıkarımlara bakmaksızın yalnızca veri setinde

çoğunluğu oluşturan sınıfı temel almakta, sınıflandırma kararını da çoğunluğun oluşturduğu sınıftan yana kullanmakta, özünde tahmin gücü barındırmayan bir algoritma olarak karşımıza çıkmaktadır (Lavesson, Boldt, Davidsson ve Jacobsson, 2011: 296). Bu durumda, doğru sınıflandırmada başarı iddiası bulunan bir diğer algoritmanın, en azından, böylesine basit bir prensiple karar veren algoritmadan daha yüksek düzeyde doğru sınıflandırma oranı elde etmesi gerekmektedir (Hussain, Calvo ve Chen, 2014: 263). Aksi takdirde, daha kompleks yapıya sahip olan diğer algoritmalar için daha fazla kaynak kullanımı manasız olacak, çok daha basit ve hızlı bir şekilde çalışan ZeroR algoritması kullanıcının tercihi olacaktır. Netice itibariyle, ZeroR algoritması, herhangi bir algoritmanın (modelin) ne denli kullanışlı olduğunu ya da sıradan bir doğru sınıflandırma oranı sunabildiğini ortaya koyan temel ölçüt olarak kullanılmaktadır (Murgia, Ortu, Tourani, Adams ve Demeyer, 2018: 545).

2.5 KREDİ ANALİZİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMINA İLİŞKİN LİTERATÜR İNCELEMESİ

Makine öğrenmesi algoritmalarının, farklı birçok alanda kullanılmakta olduğu görülmekle birlikte kredi analizinde, müşteri kredi skorlarının belirlenmesinde, kredi kararlarının verilmesinde söz konusu algoritmaların kullanımına ilişkin çalışmalar yapıldığı ve modeller geliştirildiği görülmektedir. Kredi analizi başlığı altında değerlendirilebilecek literatürdeki bazı çalışmalar ve bu çalışmalarda öne çıkan hususlar aşağıda açıklanmıştır.

Ghodselahe (2011), kredi skorlamasında kullanılmak üzere destek vektör makineleri algoritması temelli bir hibrit model geliştirmiş, söz konusu hibrit modelde kümeleme ve sınıflandırma teknikleri kombine edilmiştir. Geliştirilen hibrit modelin, tekli yaklaşımlara kıyasla daha yüksek düzeyde doğru sınıflandırma başarısı sunabildiği gösterilmiştir.

Chen vd. (2012), iki aşamalı hibrit bir veri madenciliği tekniği geliştirmişlerdir. Kümeleme aşamasında homojen gruplar oluşturulmuş, tutarsız örnekler yeniden etiketlenmiştir. Sınıflandırma aşamasında, destek vektör makineleri, yeni etiketlenmiş verileri kullanarak bir skorlama modeli oluşturmuştur. Diğer kredi skorlama modellerinden farklı olarak, veri seti iyi ve kötü kredi sınıfı gibi ikili ayırım yapmak

yerine üç ya da dört farklı sınıfa bölünmüştür. Yerel bir Çin bankası verileri üzerinde yapılan bu çalışma, veri seti içerisinde uygun bölme noktaları belirlendiği takdirde, iyi ve kötü kredilerin sınıflandırılmasında çok iyi bir düzeye ulaşılabileceğini ortaya koymuştur.

Ünlüer ve Güneş (2013), Eskişehir ilinde faaliyet gösteren tarımsal işletmelere ait veriler üzerinden, lojistik regresyon analiziyle, kredilerin takibe düşmesine neden olabilecek faktörlerin belirlenmesini amaçlamışlar, araştırma bulgularına göre kredinin zamanında ödenebilirliğini etkileyen en önemli değişkenin ekonomik rantabilite olduğunu ortaya koymuşlardır. Kredinin ödenebilirliğini etkileyen diğer değişkenlerin kaldıraç oranı ve kredinin vadesi olarak ifade edildiği görülmekle birlikte, kaldıraç oranındaki ve kredinin vadesindeki artışın, ödenebilirliğe etkisinin olumsuz, ekonomik rantabilitedeki artışın ise ödenebilirliğe etkisinin olumlu olduğu belirtilmiştir.

Sönmez (2015), kredi talep eden müşterilerin kredi taleplerinin değerlendirilmesi için yapay sinir ağları metodolojisini baz alan bir yazılım modeli önermeyi amaçlamış, gerçek verilere dayalı uygulama sonuçlarını ayrıca karar ağaçları algoritmasıyla geliştirilen model ile kıyaslamıştır. Uygulama sonuçlarına göre yapay sinir ağları metodolojisiyle geliştirilen modelin karar ağaçlarına dayalı modele kıyasla daha yüksek düzeyde doğru sınıflandırma yapabildiği ortaya konulmuştur. Ancak her iki yöntemde de Tip II hatanın, yani düşük performansla sahip kredi isteklisinin tespitine dair verilen sınıflandırma kararındaki hatanın yüksek olduğu görülmüştür.

Soydaner ve Kocadağlı (2015), Alman kredi veri seti üzerinde yapay sinir ağları ve lojistik regresyon algoritmalarını kullanarak kredi isteklisinin kredibilitesini tahmin edebilecek bir model geliştirmeyi amaçlamışlar, elde edilen bulgular ışığında yapay sinir ağları algoritmasıyla geliştirilen modelin, lojistik regresyon algoritmasına kıyasla daha yüksek düzeyde performans sergilediğini ortaya koymuşlardır.

Dima ve Vasilache (2016), Romanya'da faaliyet gösteren bankaya kredi için başvuran uluslararası sermayeli firmaların iflas riskinin tahmin edilmesi amacıyla, başvuran firmaları Standard&Poors'un belirlediği puan kategorilerine göre sekiz farklı sınıfa ayırmışlar, sonrasında puanlar arasında olumsuz yönde geçişin olasılığını tahmin etmeyi hedefleyen bir model geliştirmişlerdir. Bu bağlamda lojistik regresyon

ve yapay sinir ağıları algoritmaları kullanılmış ve yapay sinir ağıları algoritmasınca geliştirilen modelin daha iyi neticeler sunduğu belirtilmiştir.

Punniyamorthy ve Sridevi (2016), kredi riskinin değerlendirilmesinde, başvuru sahiplerini sınıflandırmak adına yapay sinir ağıları ve bulanık destek vektör makineleri algoritmalarını kullanmışlardır. Bulanık destek vektör makinelerinin kullanılmasındaki amacı, veri seti içerisindeki aykırı değerlerin etkisinin giderilmesi olarak ifade etmişler, bulanık destek vektör makineleri algoritmasıyla ortaya konulan modelin, yapay sinir ağıları algoritmasınca belirlenen modelden daha iyi sonuçlar sağladığını ifade etmişlerdir.

Tunç (2016), kredi başvurusunda bulunanlara ilişkin geliştirilecek sınıflandırma modelinde bayes algoritması ve gri kurt optimizasyon algoritmalarını kullanmış, bayes algoritmasıyla daha yüksek düzeyde doğru sınıflandırma yapabilen bir model geliştirmiştir. Gri kurt optimizasyon algoritmasının, doğru sınıflandırma başarısının, bayes algoritmasına yakın olduğu ifade edilmekle birlikte, bayes algoritmasına kıyasla söz konusu algoritmanın işlem süresinin daha uzun olduğu belirtilmiştir.

Addo vd. (2018), kredi temerrüt olasılığının tahmin edilmesi için makine öğrenmesi ve derin öğrenme modellerine dayalı ikili sınıflandırıcıları kullanarak veri setindeki önem arz eden değişkenleri tespit etmişler, daha sonra bu özellikleri kullanarak ikili sınıflandırıcıların farklı veriler üzerindeki tutarlılığını test etmişlerdir. Çok katmanlı yapay sinir ağlarına kıyasla ağaç temelli modellerin daha tutarlı olduğunu ortaya koymuşlardır.

Kalaycı (2018), KOBİ'lere kullanılan kredilerin belirli bir zaman diliminde tahsil edilebilirliğini öngörmeyi hedefleyen bir çalışma yürütmüş, bu çalışmada lojistik regresyon, karar ağaçları, destek vektör makineleri, sinir ağıları, rastgele orman algoritması ve meylli hızlandırma yöntemleri kullanılmıştır. En yüksek düzeyde başarının, meylli hızlandırma yöntemiyle elde edildiği ifade edilmiştir. Çalışmada ayrıca, kredi müşterilerinin statik özniteliklerinin kullanılması eleştirilerek, kredi müşterilerinin ödeme davranışlarına dair özniteliklere dayalı öznitelik vektörünün oluşturulduğu belirtilmiş, ödeme davranışlarının eklenmesiyle modelin daha dengeli bir hale geldiği ifade edilmiştir. Ayrıca, makine öğrenmesi yöntemlerinin hibrid bir

yapıda kullanılması durumunda daha yüksek düzeyde başarı elde edilebileceği ortaya konulmuştur.

Çizer (2018), kredi talebinde bulunan müşterilerin kredi risklerinin belirlenmesi için karar ağacı analizi, destek vektör makineleri analizi, bulanık mantık ve genetik algoritması analizi ve yapay sinir ağları analizi yöntemlerini kullanmış, söz konusu makine öğrenmesi algoritmalarından en yüksek düzeyde doğru sınıflandırma başarısı gösteren algoritmanın yapay sinir ağları algoritması olduğunu ifade etmiştir.

Kredi analizlerine ilişkin olarak literatürde yer alan makine öğrenmesi algoritmaları arasında destek vektör makineleri, yapay sinir ağları, lojistik regresyon, bayes, karar ağaçları algoritmalarının yaygın olarak kullanılmakta olduğu görülmektedir. Yapılan çalışmalarda yapay sinir ağları, öne çıkan başarılı algoritmalar arasında yer almakla birlikte, farklı algoritmaların kombinasyonu ile hibrit modeller oluşturulduğu, söz konusu hibrit modeller ile daha iyi neticeler elde edildiği gözlenmektedir.

Tarımsal kredilerin analiz edilmesine ilişkin çalışmamızda, literatürde yer alan yapay sinir ağları, lojistik regresyon, bayes, karar ağaçları algoritmaları kullanılacak olmakla birlikte, ikili sınıflandırma kararlarında başarılı bulunan doğrusal ayırma analizi algoritması ile efektif ve öğrenme safhası hızlı olan k en yakın komşu algoritmasına yer verilecektir. Böylelikle farklı algoritmaların, kredi analizi hususunda ortaya koyduğu performansın değerlendirilmesi amaçlanmaktadır.

3. BÖLÜM

UYGULAMA ÖRNEĞİ

3.1 ARAŞTIRMANIN KONUSU

Uygulama örneğinde yurt genelinde bir kredi kuruluşu olarak faaliyet gösteren Türkiye Tarım Kredi Kooperatiflerinden, farklı coğrafi bölgelerde yer alan kooperatiflerin, kullandırmış oldukları tarımsal kredilere esas olan değerlendirme ölçütleri baz alınarak, kredilerin vadesinde ödenip ödenmeyeceğini tahmin eden bir model geliştirilmeye çalışılmıştır. Söz konusu modelin geliştirilmesi için tezimizin 2. bölümünde sıralanan algoritmalar kullanılmış, her bir algoritmanın geliştirdiği model ve elde edilen sınıflandırma başarısı ilerleyen bölümlerde açıklanmıştır.

3.2 ARAŞTIRMANIN AMACI

Çalışmanın amacı, tarımsal kredilerin vadesinde ödenmesini/ödenmemesini etkileyen unsurları saptamak, söz konusu tespitleri yapabilmek adına makine öğrenmesinde kullanılan bazı algoritmaların uygulanabilirliğini ortaya koymak, kredilerin vadesinde tahsil edilip edilemeyeceğini önceden tespit edebilen bir model geliştirerek kurum kaynaklarının karşılaşılabileceği riskin minimize edilmesine katkı sağlamaktır.

Araştırmanın amacı doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- Tarımsal kredi kullanan üreticilere kullanılacak kredi tutarını etkileyen unsurlar nelerdir?
- Kredi tutarının belirlenmesinde etkili olan unsurlardan hangileri kredilerin vadesinde ödenip ödenmeyeceğini etkilemektedir?
- Tarımsal kredilerin vadesinde ödenip ödenmeyeceğinin tahmininde, makine öğrenmesi algoritmaları etkin bir modelleme performansı göstermekte midir?

3.3 ARAŞTIRMANIN KAPSAMI

Araştırmanın genel evreni, Türkiye çapında faaliyet gösteren 1625 tarım kredi kooperatifinden oluşmaktadır. Ancak ülkenin tarımsal deseninin birkaç temel grupta toplanabilmesi ve genel evrendeki birim sayısının oldukça fazla olması nedeniyle, çalışma evreni olarak, farklı bölgelerden 19 tarım kredi kooperatifi belirlenmiştir (Bkz. Ek-1). Tarımsal üretim, ani iklimsel değişikliklere göre ciddi dalgalanmalar gösterebildiğinden, salt bir yıllık veriler yerine 2012-2016 dönemindeki veriler kullanılmıştır. Şahıslara kullandırılan kredi tutarının belirlenmesine esas olan ve her bir şahıs için ayrıca düzenlenen ortak beyannamesinde yer alan 47 etmen, çalışmamızda değişken olarak belirlenmiş ve bu değişkenler ilerleyen bölümlerde detaylıca açıklanmıştır.

3.4 ARAŞTIRMANIN YÖNTEMİ

Çalışmamızda kredilerin vadesinde ödenmesine ilişkin yapılacak öngöründe, k en yakın komşu, naive bayes, karar ağacı, doğrusal ayırma analizi, lojistik regresyon ve yapay sinir ağları algoritmaları kullanılmış, ZeroR algoritması ise ileride değinileceği üzere, öngöründe bulunan algoritmaların performansının değerlendirilmesinde eşik değeri belirleyici algoritma olarak kullanılmıştır. Söz konusu algoritmalar, Yeni Zelanda'da yer alan Waikato Üniversitesi tarafından geliştirilen WEKA programı ile uygulanmıştır. Bu program Java temelli olup, açık erişime sahiptir. WEKA, verinin ön işleme, sınıflandırma, regresyon, kümeleme, ilişkilendirme kuralları ve görselleştirme gibi araçlar içermektedir (University Of Waikato, 25.03.2018).

3.5 HİPOTEZ GELİŞTİRME

Çalışmanın temel amacının, tarımsal kredilerin vadesinde ödenip ödenmeyeceğine dair makine öğrenmesi algoritmalarının etkin bir tahmin modellemesi yapabilmesi olduğu göz önüne alındığında, her bir algoritmanın eşik değer üzerinde tahmin modellemesi yapıp yapamayacağına dair hipotezler Tablo 10'da sunulmuştur.

Tablo 10: Araştırmanın Hipotezleri

Hipotez-1	K-En Yakın Komşu Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.
Hipotez-2	Naive Bayes Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.
Hipotez-3	Karar Ağacı Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.
Hipotez-4	Doğrusal Ayırma Analizi Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.
Hipotez-5	Lojistik Regresyon Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.
Hipotez-6	Yapay Sinir Ağları Algoritması, kredilerin vadesinde ödenip ödenmeyeceğine dair, eşik değeri üzerinde tahmin modellemesi geliştirmiştir.

3.6 VERİ SETİ

Farklı yörelerde faaliyet gösteren 19 tarım kredi kooperatifinin 2012-2016 dönemine ait veriler, söz konusu kurumların işlemlerini gerçekleştirdikleri E-Koop isimli merkezi otomasyon sisteminden temin edilmiş, verilerin tez çalışmamızda kullanılabilmesi için gerekli yazılı izin, Türkiye Tarım Kredi Kooperatifleri Merkez Birliği Genel Müdürlüğü Rehberlik ve Teftiş Kurulu Başkanlığı'ndan alınmıştır.

Çalışma evrenini oluşturan kooperatiflerin çok sayıda ortağı bulunması sebebiyle, bu kooperatiflerden, kredi verileri kullanılacak 437 farklı şahıs basit tesadüfi örnekleme yoluyla seçilmiştir.

İlgili Kurum tarafından, şahıslara kullanılacak kredi tutarının belirlenmesinde ortak beyannamesi esas alınmakta, söz konusu beyannamede şahısların tarımsal faaliyetlerine dair tüm veriler yer almakta, bu veriler ışığında kredi limiti belirlenerek limit kapsamında kredi kullanılmaktadır. Çalışmamızdaki 47 farklı bağımsız değişkeni ortak beyannamesinde (Bkz. Ek-2) yer alan unsurlar ile yine E-Koop otomasyon sisteminde yer alan ve şahısların demografik verilerini içeren ortak kartındaki (Bkz. Ek-3) veriler oluşturmaktadır. Bağımlı değişken ise kredinin vadesinde ödenmesi ya da ödenmemesi şeklinde belirlenmiştir. Çalışmamıza konu olan 437 kişiden, 154'ü kredisini vadesinde ödememiş, geriye kalan 283'ü kredisini vadesinde ödemiş kişilerdir.

Her bir şahsın beyannamesinde yer alan veriler öncelikle Excel ortamına tek tek aktarılmıştır. Söz konusu verilerin WEKA'ya aktarılabilmesi için dosya biçimi ve içeriği CSV (comma-separated values) formatına dönüştürülmüş, akabinde veriler WEKA'ya aktarılabilir ve bu program tarafından işlenebilecek hale getirilmiştir.

3.7 DEĞİŞKENLERİN TANIMLANMASI

Tarım Kredi Kooperatiflerinde, şahıslara tespit edilecek kredi limitini, ortak beyannamesinde yer alan ve ilgilinin tarımsal faaliyetlerinin hacmini, elde edebileceği geliri yaklaşık olarak ortaya koyan değişkenler belirlemektedir. Bunun yanı sıra, bazı demografik unsurların, kredilerin tam ve zamanında ödenmesine etkisi olabileceği düşünülerek, söz konusu demografik unsurlar çalışmamızdaki değişkenler arasına dahil edilmiştir. Üreticilerin beyannamelerinde yer alan değişkenlerle birlikte demografik değişkenlerin, kredilerin vadesinde ödenmesinde ne denli öneme sahip olduğu ya da bir öneme sahip olup olmadığı hususları araştırılmış, hali hazırda kullanılmakta olan değişkenlerden öneme sahip olanlarla kredilerin vadesinde ödenip ödenmeyeceğini öngören bir model geliştirilmeye çalışılmıştır.

Uygulama örneğimizde kullanılan bağımlı ve bağımsız değişkenler, sırasıyla Tablo 11 ve Tablo 12'de izah edilmiştir.

Tablo 11: Uygulama Örneğindeki Bağımlı Değişken

<i>Gösterim</i>	<i>Değişken Adı</i>	<i>Açıklama</i>
Y	Durum	Kredinin vadesinde ödenip ödenmediğini gösteren bağımsız değişkendir. Uygulama örneğimizde “YES-NO” şeklinde değerlendirilmiştir. “NO” şeklindeki gösterimler vadesinde ödenmeyen kredileri, “YES” şeklindeki gösterimler ise vadesinde ödenen kredileri ifade etmektedir.

Tablo 12: Uygulama Örneğinde Kullanılan Bağımsız Değişkenler

<i>Gösterim</i>	<i>Değişken Adı</i>	<i>Açıklama</i>
X ₁	cinsiyet	Kredi kullanan kişinin cinsiyeti kadınlar için “0”, erkekler için “1” olarak belirlenmiştir.
X ₂	nufuskayıt_yeri	Kredi kullanan kişinin, nüfusa kayıtlı olduğu yerin kooperatifin faaliyet alanı içerisinde olması durumunda bu değer “0”, aksi durumlarda bu değer “1” olarak gösterilmiştir.
X ₃	egitim_durumu	Kredi kullanan kişinin eğitim durumunun ilkökul, ortaokul, lise ve üniversite olmasına göre sırasıyla 0-4 aralığında değerler eğitim durumunu göstermek amacıyla kullanılmıştır.
X ₄	medeni_hali	Kredi kullanan kişinin bekâr olması durumunda bu değer “0”, evli olması durumunda bu değer “1” olarak değerlendirilmiştir.
X ₅	yas<23	Kredi kullanan kişinin yaşının 23’ten küçük olması durumunda söz konusu

		değişken “1” değerini almakta, aksi durumlarda bu değişkenin değeri “0” olmaktadır.
X ₆	yas>65	Kredi kullanan kişinin yaşının 65’ten büyük olması durumunda söz konusu değişken “1” değerini almakta, aksi durumlarda bu değişkenin değeri “0” olmaktadır.
X ₇	kendi_arazisi	Üreticinin, mülkiyeti kendisine ait olan arazisi var ise söz konusu arazinin dekar cinsinden büyüklüğüdür.
X ₈	dekar_degeri	Kredi kullanan kişinin mülkiyeti kendisine ait olan arazisi bulunduğu takdirde, bu değişkenin değeri arazinin 1 dekarının değeri olarak gösterilmektedir.
X ₉	icar	Kredi kullanan kişinin tarımsal faaliyetlerini kiralık arazi üzerinde sürdürmesi durumunda, kiralama işlemine konu olan arazinin dekar cinsinden büyüklüğüdür.
X ₁₀	uretim_yapilan	Üreticinin gerek kendisine ait olan gerekse kiralama yoluyla işlediğini beyan ettiği arazilerin dekar cinsinden toplamıdır.
X ₁₁₋₁₅	arazi_urun1-5	Kredi kullanan kişinin, tarımsal faaliyetlerini sürdürdüğü arazilerde ekip biçtiğini beyan ettiği 5 farklı ürün için X ₁₁₋₁₅ aralığında 5 farklı ürün belirtilmektedir. Değişkenlerin alacağı değer metin cinsinden ifade edilmiş olup, 5’ten daha az çeşitte üretim yapanlar için bulunmayan

		ürün çeşitleri “0” olarak değerlendirilmiştir.
X ₁₆₋₂₀	dekar_urun1-5	Her bir ürün çeşidi için m ² 'den alındığı beyan edilen ürün miktarının kg cinsinden gösterimidir.
X ₂₁₋₂₅	arazi_urun_fiyat	M ² 'den alınan ürünün 1 kg'ının parasal değeridir.
X ₂₆	buyukbas	Üreticinin sahip olduğunu beyan ettiği büyükbaş hayvan sayısıdır.
X ₂₇	buyuk_deger	Sahip olunan büyükbaş hayvanın 1 adedinin parasal değeridir.
X ₂₈	buyukbas_uretim_et	Sahip olunan büyükbaş hayvan besi amacıyla kullanılıyor ise bu hayvandan elde edilebilecek etin kg cinsinden ifadesidir.
X ₂₉	buyukbas_uretim_fiyat_et	Besi niteliğindeki büyükbaş hayvandan elde edilebilecek etin kg fiyatıdır.
X ₃₀	buyukbas_uretim_sut	Sahip olunan büyükbaş hayvan sütçülük amacıyla kullanılıyor ise bu hayvandan elde edilebilecek sütün kg cinsinden ifadesidir.
X ₃₁	buyukbas_uretim_fiyat_sut	Sütçü büyükbaş hayvandan elde edilebilecek sütün kg fiyatıdır.
X ₃₂	buzagi_ikinciurun	Sahip olunan büyükbaş hayvanlardan elde edilebilecek toplam buzağının, parasal değeridir.
X ₃₃	kucuk_bas	Üreticinin sahip olduğu küçükbaş hayvan sayısını ifade etmektedir.
X ₃₄	kucuk_deger	Sahip olunan her bir küçükbaş hayvanın parasal tutarıdır.

X35	kucukbas_uretim_et	Üreticinin besi amacıyla kullandığı küçükbaş hayvanın bir adedinden elde edilebilecek etin kg cinsinden ifadesidir.
X36	kucukbas_uretim_fiyat_et	Besi amacıyla her bir küçükbaştan elde edilebilecek etin 1 kg'ının fiyatıdır.
X37	kucukbas_uretim_sut	Üreticinin sütçülük amacıyla kullandığı küçükbaş hayvanın bir adedinden elde edilebilecek sütün kg cinsinden ifadesidir.
X38	kucukbas_uretim_fiyat_sut	Sütçülük amacıyla kullanılan her bir hayvandan elde edilen sütün 1 kg'ının fiyatıdır.
X39	kuzu_ikinciurun	Üreticinin sahip olduğu küçükbaş hayvanlardan elde edilebilecek toplam kuzuların parasal değeridir.
X40	kovan_sayi	Arıcılık faaliyetiyle uğraşan üreticilerin sahip oldukları kovan sayısıdır.
X41	kovan_deger	Arıcılık yapan üreticinin sahip olduğu her bir kovanın parasal değeridir.
X42	kovan_uretim	Her bir kovandan elde edilecek balın kg cinsinden ifadesidir.
X43	kovan_uretim_fiyat	Üreticinin elde ettiği balın 1 kg'ının fiyatıdır.
X44	alet_ekipman	Üreticinin sahip olduğu zirai alet ve ekipmanın toplam parasal değeridir.
X45	teminat	Kredi kullanan kişinin, krediye karşılık vermiş olduğu maddi teminatın parasal tutarıdır.
X46	disari_borçlar	Kredi kullanmak isteyen şahsın, tarım kredi kooperatifleri haricinde diğer kredi kuruluşlarına olan borç tutarıdır.

X ₄₇	kredi	Şahıslara tespit edilen kredi limitidir.
-----------------	-------	--

3.8 UYGULAMA ÖRNEĞİNDE BAZ ALINACAK PERFORMANS ÖLÇÜTLERİ

Çalışmamızda kullanmış olduğumuz bilgisayar programındaki çıktı sonuçlarında bir takım istatistiki oranlar yer almakla birlikte, algoritma bazında doğru sınıflandırma işlemine dair bazı ölçütler bulunmaktadır. Aşağıda sırasıyla söz konusu oranlar/ölçütler açıklanmıştır.

3.8.1. Kappa Katsayısı (Kappa Statistics)

Söz konusu katsayı, iki ya da daha fazla gözlemcinin kararları arasındaki uyumun güvenilirliğinin ölçülmesinde kullanılmaktadır (Kılıç, 2015: 142). Kappa katsayısı hesaplanırken iki temel olasılık kullanılmaktadır. Aşağıdaki formülde kappa katsayısının nasıl hesaplandığı gösterilmiştir. Söz konusu formülde Pr(a), iki gözlemcinin uyumlu karar verdikleri gözlemlerin, tüm gözlemlere oranını ifade etmekte, Pr(e) ise iki gözlemci arasında uyumun şansa bağlı olarak gerçekleşme olasılığını göstermektedir (Kanik, Erdoğan ve Temel, 2012: 154).

$$K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

Kappa katsayısı -1 ile +1 arasında değerler almakla birlikte, söz konusu değerler aşağıdaki şekilde yorumlanmaktadır (Keskin, 2004: 169):

- K=+1 olarak hesaplandığında iki gözlemcinin sonuçlarının birbiriyle tamamen uyumlu olduğu,
- K=0 olarak hesaplandığında iki gözlemci arasındaki uyumun şansa bağlı olduğu,
- K=-1 olarak hesaplandığında iki gözlemci arasında tam bir uyumsuzluk bulunduğu,

Kabul edilmektedir.

Söz konusu katsayı için hesaplanabilecek değerlere dair geliştirilen daha detaylı yaklaşım Tablo 13’de gösterilmiştir (Vehid ve Eral, 2014: 61).

Tablo 13: Kappa Katsayısı Uyum Düzeyleri

K Değeri	Yorum
<0	Şans eseri gerçekleşebilecek uyumdan daha kötü düzeyde uyum
0,01-0,20	Önemsiz düzeyde uyum
0,21-0,40	Zayıf düzeyde uyum
0,41-0,60	Orta düzeyde uyum
0,61-0,80	İyi düzeyde uyum
0,81-1,00	Çok iyi düzeyde uyum

3.8.2. Hata Oranları

Hata oranları genel itibariyle, modelin öngördüğü değer ile gerçek değer arasındaki farkı ortaya koyan oranlardır. Aşağıdaki bölümlerde, çalışmamızda kullanılan program çıktısında yer alan bazı hata ölçütleri sırasıyla açıklanacaktır. Her bir hata oranına ilişkin formülde yer alan p tahmin edilen değeri (predicted value), a gerçekleşen değeri (actual value) ve n ise tahmin yapılan gözlem sayısını ifade etmektedir.

3.8.2.1. Ortalama Mutlak Hata (Mean Absolute Error-MAE)

Söz konusu hata oranı, tahmini değerler ile gerçekleşen değerler arasındaki farkların mutlak değerlerinin toplamının, tahmin işlemine konu olan gözlem sayısına bölünmesiyle hesaplanmaktadır (Willmott ve Matsuura, 2005: 80).

$$\text{Mean Absolute Error} = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Farkların mutlak değeri alındığından, toplam hatanın ortalama büyüklüğü ölçülmektedir. Birim bazında farklılıklar ortalama eşit ağırlığa sahip olmaktadır.

3.8.2.2. Ortalama Hata Karelerinin Karekökü (Root Mean Squared Error-RMSE)

Tahmini değerler ile gerçekleşen değerler arasındaki hata ortalama büyüklüğünü ölçen bu oranda, örnek bazındaki hataların kareleri toplamı alınarak örnek sayısına bölünmekte, akabinde bulunan değerın karekökü alınmaktadır (Wang, Fang, Chai ve Zheng, 2012: 93).

$$\text{Root Mean Squared Error} = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_1)^2}{n}}$$

Ortalama hesaplamadan önce hataların karesi alındığından, bu yaklaşımda büyük hataların ağırlığı fazla olmaktadır (Chai ve Draxler, 2014: 1247). Geliştirilmek istenilen modelde, büyük hatalar (sapmalar) istenilmediğinde söz konusu oran kullanışlı olmaktadır.

MAE ve RMSE birlikte kullanılabilen, RMSE, MAE'ye eşit ya da ondan daha büyük olarak gerçekleşmektedir. Her iki oran arasındaki farkın büyümesi, birim bazında hata varyansının arttığıın göstergesidir. Her iki oran 0 ile ∞ arasında değişmekle birlikte bu oranların mümkün olduğunca sıfıra yakın olması arzulanan durumdur (Veri Bilimcisi, 04.04.2018).

3.8.2.3. Nispi Mutlak Hata (Relative Absolute Error-RAE)

Bu yaklaşımda, tahmini değerlerin gerçekleşen değerlerden sapmasıyla, gerçekleşen değerlerin kendi ortalamalarından sapması karşılaştırılmaktadır. Aşağıdaki formülde de gösterildiği üzere hem pay hem de paydada yer alan farkların mutlak değeri alınmaktadır (Saedsayad, 04.04.2018).

$$\text{Relative Absolute Error} = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{|a_1 - \bar{a}| + \dots + |a_n - \bar{a}|}$$

3.8.2.4. Nispi Hata Karelerinin Karekökü (Root Relative Squared Error-RRSE)

Nispi mutlak hata oranına benzer bir yaklaşım benimsenmekle birlikte, RMSE'de olduğu gibi büyük sapma gösteren örneklere bu oranda önem verilmektedir. RAE'den farklı olarak, bu oranda, tahmini değerler ile gerçekleşen değerler arasındaki

farkların kareköklerinin toplamı, gerçekleşen değerlerin kendi ortalamalarından sapmalarının karekök toplamıyla karşılaştırılmaktadır (Gepsoft, 05.04.2018).

$$\text{Root Relative Squared Error} = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}}$$

RAE ve RRSE’de, tahmini değerlerin gerçekleşen değerlerden sapması ile gerçekleşen değerlerin kendi ortalamalarından sapmaları karşılaştırılarak bir performans ölçümü yapılmaktadır. Her iki oranın örtüştüğü nokta, formülün payında yer alan sapma değerlerinin, paydada yer alan sapma değerlerini aşmaması gerektirir. Özetle, modelin başarılı sayılabilmesi için, tahmini değerlerin, gerçekleşen değerlerden daha az sapma büyüklüğüne sahip olması gerekmektedir.

3.8.3. Doğru Sınıflandırmaya İlişkin Detaylı Oranlar

Bu bölümde, doğru sınıflandırma başarısını ilgilendiren oranlar, program çıktısında yer alan sıralarına göre izah edilecektir.

Program çıktısında, TP (True Positive) ve FP (False Positive) şeklinde iki temel oran bulunmakta ve bu oranlara bağlı olarak doğru sınıflandırmaya ilişkin diğer oranlar hesaplanmaktadır. TP ve FP değerlerinin hesaplanması Tablo 14’de gösterilmiştir (Witten, Frank ve Hall, 2011: 164). Söz konusu tablo oluşturulurken, çalışmamızda elde edilen çıktıların sunuluş biçimi esas alınmıştır. “NO” şeklindeki gösterim kredi borçlarını vadesinde ödemeyen örnekleri, “YES” şeklindeki gösterim ise kredi borcunu vadesinde ödeyen örnekleri ifade etmektedir.

Tablo 14: Sınıflandırmaya İlişkin Örnek Çıktı Tablosu

		<i>Tahmin Edilen Değerler</i>	
		a	b
<i>Gerçekleşen Değerler</i>	a=NO	True Positive (TP)	False Negative (FN)
	b=YES	False Positive (FP)	True Negative (TN)

True Positive: Gerçekleşen değeri “NO” olup, tahmin edilen değeri “NO” olan örnek sayısıdır.

False Positive: Gerçekleşen değeri “YES” olup, tahmin edilen değeri “NO” olan örnek sayısıdır.

True Negative: Gerçekleşen değeri “YES” olup, tahmin edilen değeri “YES” olan örnek sayısıdır.

False Negative: Gerçekleşen değeri “NO” olup, tahmin edilen değeri “YES” olan örnek sayısıdır.

Program çıktısında yer alan ve hesaplanırken yukarıda izah edilen TP, FP ve FN değerleri kullanılan oranlar aşağıda açıklanmıştır.

Precision (Kesinlik): Pozitif olarak tahmin edilen değerlerin aslında kaçının pozitif olduğunu ortaya koyan orandır.

$$Precision = \frac{TP}{TP + FP}$$

Recall (Hassasiyet): Aslında pozitif olarak gerçekleşen değerlerden kaçının pozitif olarak tahmin edildiğini ifade eden orandır.

$$Recall = \frac{TP}{TP + FN}$$

F-Measure (F-Skoru): Precision ve Recall oranlarının armonik ortalaması kombine edilmek suretiyle hesaplanan ölçüdür. Armonik ortalama, toplam ölçüm sayısının, ölçüm karşılıkları toplamına bölünmesiyle hesaplanmaktadır (IBM Knowledge Center, 08.04.2018).

$$F = 2 \frac{precision * recall}{precision + recall}$$

İkili sınıflandırma kararlarında, veri setinde sınıfların dağılımında bir denge var ise yani veri seti yaklaşık yarı yarıya iki ayrı sınıfa bölünmüş ise, performans ölçütü olarak doğru sınıflandırma oranı büyük ölçüde yeterli olabilecektir. Ancak, veri setinde, sınıfların dengesiz bir şekilde dağılmış olması durumunda, modelin performansının tam manasıyla değerlendirilebilmesi için precision, recall ve bu iki oranın armonik ortalamasının alındığı f-measure gibi ölçütlere ihtiyaç duyulmaktadır (Patil ve Sherekar, 2013: 257). Yukarıda bu oranlar için sunulmuş formüller göz önüne alındığında, bu oranların 1'e yakın olması istenilen bir durumdur. Ancak bazı sınıflandırma kararlarında, sınıflardan birisine ilişkin verilen kararların önemi daha

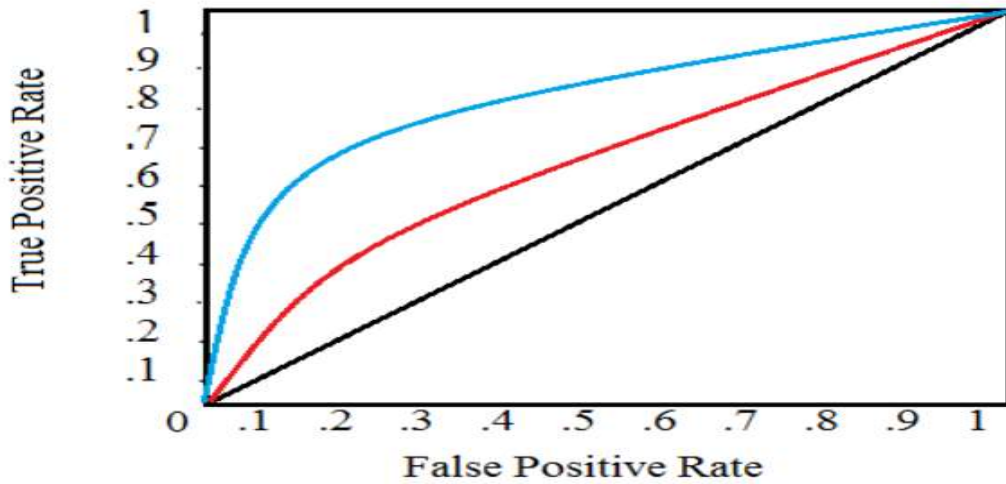
fazla ise bu sınıfa ait oranlar öncelikli olarak değerlendirilmek suretiyle performans sıralaması yapılabilmektedir.

MCC (Matthews Correlation Coefficient): Söz konusu oran, ikili sınıflandırma kararlarının verildiği makine öğrenmesi uygulamalarında, korelasyon katsayısının özünü yansıtır şekilde -1 ile +1 arasında değerler almaktadır (Scikit Learn, 08.04.2018). +1 mükemmel tahmini ifade ederken, -1 tam tersi tahmin kararı verildiğini göstermektedir. 0 ise ortalama rastgele bir tahmin yapıldığı manasına gelmektedir. MCC, aşağıdaki formülde gösterilen şekilde hesaplanmaktadır (Boughorbel, Jarray ve El-Anbari, 2017: 4).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}}$$

ROC (Receiver Operating Characteristics Curve): Program çıktısında söz konusu eğrinin altında kalan alanın büyüklüğü belirtilmekte, bu eğri TP ve FP değerleri karşılaştırılarak, Şekil 15’de gösterilen şekilde çizilmektedir (Statistics How To, 09.04.2018).

Şekil 15: ROC- Karar Vericinin Etkinliği Eğrisi



Tüm tahminlerin, gerçekleşen değerlerle örtüştüğü durumda eğrinin şekilde (0,1) noktasından geçmesi gerekmektedir. Rastgele nitelikte olan bir tahmin modeli ise (0,0) noktasından (1,1) noktasına çizilecek köşegen üzerinde hareket edecektir. Makul seviyede tahmin başarısı gösteren modellere ait eğriler ise köşegen ile (0,1) noktası arasında çizilebilecek eğrilerden oluşmaktadır. Tahmin modeline ait eğrinin köşegenin altında kalması durumunda ise yapılan tahminlerin doğru sonuçlardan çok yanlış sonuçlar içerdiği söylenebilecektir. (Witten, Frank ve Hall, 2011: 172-173).

Program çıktısında, eğri altında kalan alanın bir performans ölçütü olarak değerlendirildiği dikkate alındığında, az önce ifade edilen rastgele tahmin yapan modellere ait köşegenin altında kalan alan 0,5, mükemmel tahmin yapan model için çizilecek eğrinin altında kalan alan ise 1 olacaktır. Tahmin için geliştirilen modele ait eğrinin altında kalan alanın bu iki değer arasında yer alması beklenmektedir.

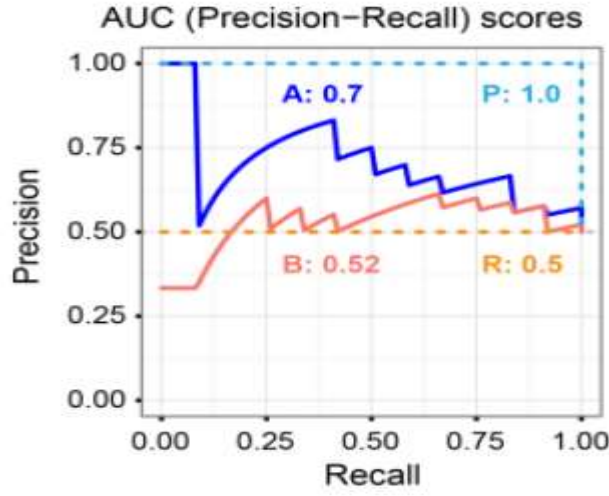
PRC (Precision-Recall Curve): Kesinlik-Hassasiyet eğrisi, özellikle dengesiz sınıf dağılımlarının olduğu veri setlerindeki sınıflandırma tahminlerinin değerlendirilmesinde önem arz etmektedir (Saito ve Rehmsmeier, 2015: 7). Kesinlik (precision) ve hassasiyet (recall) arasındaki dengenin gösterildiği eğrinin altında kalan alanın büyük olması, bu iki oranın da yüksek olduğu anlamına gelmektedir. Yüksek kesinlik oranı düşük FP'nin, yüksek hassasiyet oranı ise düşük FN'nin göstergesi olmaktadır. Böylelikle yüksek kesinlik ve hassasiyet oranları neticesinde sınıflandırma tahminlerinde daha yüksek başarı elde edilmektedir (Scikit Learn, 09.04.2018).

Kesinlik-hassasiyet eğrisi için eşik değer hesaplanırken, pozitif ve negatif değerler kullanılarak aşağıdaki formülde gösterilen hesaplama yapılarak $y=a$ doğrusu altında kalan alan belirlenmeye çalışılır (Ekelund, 09.04.2018).

$$a = \frac{P}{P + N}$$

Eşik doğrusu olarak da nitelendirilebilecek değer belirlendikten sonra, kesinlik-hassasiyet eğrisi için aşağıdaki örnek Şekil 16 çizilebilecektir (Classeval, 09.04.2018). Söz konusu gösterimde, farklı iki (A ve B) sınıflandırıcı modelin kesinlik-hassasiyet eğrileri karşılaştırılmış, bu iki sınıflandırıcının birbirlerine ve eşik değer doğrusuna göre konumları gösterilmiştir. İdeal olan durumun, eğrinin altında kalan alanın 1'e eşit olduğu konumda gerçekleştiği dikkate alınır, A sınıflandırıcısının, B'ye kıyasla daha iyi bir performans sergilediği, her iki modelin nerdeyse eşik değer doğrusu üzerinde olduğu, ancak B modelinin eşik değere oldukça yakın seyrettiği, nitekim B modeline ait eğrinin altında kalan alanın A modeline kıyasla daha küçük olduğu görülmektedir. Netice itibarıyla, uygulama örneğimizdeki çıktılarda, PRC Area (Kesinlik-Hassasiyet Eğrisi Altındaki Alan) şeklindeki gösterimin mümkün olduğunca 1'e yakın olması istenilen durumdur.

Şekil 16: Örnek Kesinlik-Hassasiyet Eğrileri



3.9 UYGULAMA SONUÇLARI

Çalışmamızda kullandığımız farklı algoritmalara ilişkin teorik açıklamalar, algoritmaların çalışma prensipleri önceki bölümlerde değerlendirilmiştir. İlerleyen bölümlerde algoritma bazında elde edilen çıktılar sırasıyla değerlendirilecektir. Ancak algoritma sonuçlarına geçmeden önce, her bir çıktıda karşılaşılabilecek ve programın çalışma prensiplerini etkileyen bazı hususlar aşağıda açıklanmıştır.

Test Mode: Bu bölümde programın öğrenme ve test için kullanacağı verilerin nasıl belirlendiğine ilişkin yöntem belirtilmektedir. Çalışmamızda ağırlıklı olarak “split 66.0% train, remainder test” yöntemi benimsenmiştir. Söz konusu yöntemde, veri setinin %66’sı makinenin (algoritmanın) öğrenmesi için, geri kalan %34’ü ise makinenin öğrenmiş olduğu modelin test edilmesinde kullanılmaktadır. Çalışmamızda kullandığımız bir diğer yöntem ise “10-fold cross-validation” olarak isimlendirilmektedir. Bu yöntemde, orijinal veri seti 10 ayrı alt kümeye bölünmekte, bu alt kümelerden 9 tanesi eğitim için 1 tanesi ise test için kullanılmakta, söz konusu işlem 10 kez tekrarlanarak her bir alt kümenin test aşamasında kullanılması sağlanmaktadır (OpenML, 14.04.2018).

Algoritma çıktıları, önceki bölümlerde yapılan açıklama sırasına göre aşağıda yer almakla birlikte, ZeroR algoritmasının, diğer algoritma sonuçlarının değerlendirilmesinde eşik değeri (baseline) belirlemesi sebebiyle, söz konusu algoritmaya ait çıktı öncelikli olarak ele alınacaktır.

3.9.1. ZeroR Algoritması Uygulama Çıktısı

Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-3). Algoritmanın çalışma prensibinin, veri setinde hâkim olan sınıfa göre, tüm birimleri bu sınıfa aitmiş gibi sınıflandırmak olduğu belirtilmişti. Aşağıdaki tablolardan da görüleceği üzere, veri setinde hâkim olan sınıfın, kredisini vadesinde ödeyen “YES” etiketli sınıf olması nedeniyle makine testteki tüm verileri “YES” olarak sınıflandırmıştır.

Tablo 15’de test için seçilen 149 örnekten, 104’ünün doğru sınıflandırıldığı, bunun da %69,7987 doğru sınıflandırma oranına denk geldiği görülmektedir. Kappa katsayısı 0 olarak hesaplandığından sonucun tamamen şans eseri gerçekleştiği anlaşılmakta, nitekim algoritmanın karar verme prensibinin tamamen hâkim olan sınıfa bağlı olarak şekillendiği görülmektedir. Kredisini vadesinde ödemeyen örneklerin tamamı yanlış sınıflandırıldığından ortalama hata oranları 0’dan uzak olmakla birlikte, nispi hata oranları %100 olarak gerçekleşmiştir.

Tablo 15: ZeroR Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	104	69.7987 %
Kappa katsayısı	0	
Ortalama Mutlak Hata	0.4522	
Ortalama Hata Karelerinin Karekökü	0.4656	
Nispi Mutlak Hata	100	%
Nispi Hata Karelerinin Karekökü	100	%
Toplam Örnek Sayısı	149	

Sınıf bazındaki detaylı doğru sınıflandırma oranlarına bakıldığında, Tablo 16’nın üst satırında yer alan ve kredisini vadesinde ödemeyen örneklere ait oranlarda, TP ve FP oranları 0 olduğundan, precision, recall, f-measure oranları da 0 olarak hesaplanmıştır. MCC’nin 0 olmasının rastgele bir tahmin yapıldığı manasına geldiği daha önce de ifade edilmiş olmakla birlikte, kappa katsayısının 0, ROC Area’nın 0,50 olması bu durumu doğrulamaktadır. PRC Area’nın 0,302 olarak hesaplandığı, söz

konusu değerin bu oran için hesaplanabilecek eşik değeri olduğu, netice itibariyle “NO” için yapılan tahminlerin anlam ifade etmediği anlaşılmaktadır.

“YES” olarak etiketlenen detaylı sınıflandırma oranları da söz konusu algoritmanın rastgele bir tahmin modeli yaptığını ortaya koymaktadır. Tablo 16’nın altındaki ağırlıklı ortalamalar ise her iki sınıfın, test veri seti içerisindeki ağırlıklarına göre hesaplanmaktadır.

Tablo 16: ZeroR Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.302	NO
	1.000	1.000	0.698	1.000	0.822	0.000	0.500	0.698	YES
Ağ. Ort.	0.698	0.698	0.487	0.698	0.574	0.000	0.500	0.578	

Algoritmaların genel sınıflandırma başarısı, hata matrisi olarak adlandırılan ve ZeroR algoritması için örneği Tablo 17’de gösterilen tabloya göre değerlendirilmektedir. Daha önceki tablolarda yer alan verilerin esas alındığı, TP, FP gibi esas oranların hesaplandığı rakamlar söz konusu tablodan çekilmektedir. Hata matrisinde, tüm tahminlerin, örneklerin kredilerini vadesinde ödeyeceğine yönelik yapıldığı görülmektedir.

Tablo 17: ZeroR Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
0	45	a=NO	Gerçekleşen Değerler
0	104	b=YES	

Netice itibariyle, ZeroR uygulaması, test veri setinde hâkim olan yani daha fazla sayıda örnek barındıran sınıfa göre tahmin yapmakta, bu şekilde geliştirdiği modelin % 69,7987 doğru sınıflandırma başarısı elde ettiği görülmektedir. Rastgele

olarak nitelendirilebilecek bu yaklaşımın ortaya koyduğu doğru sınıflandırma oranı, diğer algoritmalar için eşik değer kabul edilmekte, sonraki bölümlerde farklı algoritmaların başarısı bu değere göre incelenmektedir.

3.9.2. K En Yakın Komşu Algoritması Uygulama Çıktısı

Algoritmanın, sınıflandırılmak istenilen birime en yakın k tane birimden, daha fazla sayıda birimin dahil olduğu sınıfa göre, yeni birim için sınıflandırma tahmini yaptığı açıklanmıştı. Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-5). Eğitim ve test veri setlerinin seçiminde yine “split 66.0% train, remainder test” metodu kullanılmış olup, algoritmanın tahmin kararında etkili olan k sayısı 11 olarak belirlenmiştir. K sayısının, tek bir sayı olarak belirlenmesinin sağlayacağı avantajlar algoritmaya ilişkin teorik açıklamalarda ifade edilmişti. Ancak, 11 sayısı belirlenirken, doğru sınıflandırma oranının k sayısının artışıyla birlikte aldığı değerler takip edilmiştir. Sırasıyla 1’den başlayarak 20’ye kadar olan k değerleri tek tek denenmiş, 20’den sonraki k değerlerinden rastgele birkaç tek sayı seçilerek doğru sınıflandırma oranının artış ve azalışları izlenmiştir. Netice itibariyle, doğru sınıflandırma oranının maksimum değere k=11 iken ulaştığı, diğer k değerlerinde erişilen doğru sınıflandırma oranlarının dalgalanmalar göstermekle birlikte program çıktısında elde edilen değer kadar olmadığı gözlenmiştir.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, Tablo 18’de test için seçilen 149 örnekten, 111’inin doğru sınıflandırıldığı, bunun da %74,4966 doğru sınıflandırma oranına denk geldiği, söz konusu oranın ZeroR algoritmasının eşik değerinin üzerinde olduğu görülmektedir. Kappa katsayısı, teorik aralığa göre zayıf düzeyde uyumu işaret etse de söz konusu değer orta düzey uyum sınırına yakın olduğu, yani gerçek durum ile tahmin modeli arasında, şans eseri benzerlikten ziyade bir uyumdan söz edilebileceği anlaşılmıştır. Hata oranlarına bakıldığında, ZeroR algoritmasına kıyasla daha düşük düzeyde hata oranları elde edilmişse de büyük hataların ağırlığının daha fazla olduğu ortalama hata karelerinin karekökü ve nispi hata karelerinin karekökü oranlarında belirgin bir farklılık bulunduğu görülmüş, bu durumdan tahmin modeli içerisinde sapma büyüklüğü fazla olan birimlerin bulunduğu anlaşılmıştır.

Tablo 18: K En Yakın Komşu Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	111	74.4966 %
Kappa katsayısı	0.3951	
Ortalama Mutlak Hata	0.382	
Ortalama Hata Karelerinin Karekökü	0.4385	
Nispi Mutlak Hata	84.4773 %	
Nispi Hata Karelerinin Karekökü	94.1718 %	
Toplam Örnek Sayısı	149	

Sınıf bazında doğru sınıflandırma oranları Tablo 20’deki hata matrisi baz alınarak Tablo 19’da sunulmuştur. ZeroR algoritmasının aksine, burada ve bundan sonra değerlendireceğimiz tüm algoritmalarda, TP ve FP oranları performansın değerlendirilmesinde önem arz edecektir.

TP oranı, gerçekleşen değeri “NO” olup, tahmin edilen değeri “NO” olan, yani gerçekleşen durumda kredisini vadesinde ödemeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örneklerin oranını göstermekte ve aşağıdaki şekilde hesaplanmaktadır.

$$\text{“NO” sınıfı için: } TP \text{ Rate} = \frac{26}{26+19} = 0,578$$

Tablo 19’da TP oranının altında belirtilmesine rağmen, literatürde ve yukarıda yapmış olduğumuz açıklamalarda TN (True Negative) olarak gösterilen, gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örneklerin oranı aşağıda gösterildiği gibi hesaplanmaktadır.

$$\text{“YES” sınıfı için: } TP \text{ Rate} = \frac{85}{85+19} = 0,817$$

Her iki sınıf için TP oranlarına bakıldığında, kredisini vadesinde ödeyenler için yapılan tahminlerin daha isabetli olduğu, kredisini vadesinde ödemeyenlerin %57,8 oranında tespit edilebildiği anlaşılmıştır.

FP oranı, gerçekleşen değeri “YES” olup, tahmin edilen değeri “NO” olan, yani gerçekleşen durumda kredisini vadesinde ödeyenler içerisinde, algoritma tarafından

kredisini vadesinde ödemeyecek olarak sınıflandırılan örneklerin oranını göstermekte ve aşağıdaki şekilde hesaplanmaktadır.

$$\text{“NO” sınıfı için: } FP \text{ Rate} = \frac{19}{19+8} = 0,183$$

Tablo 19’da FP oranının altında belirtilmesine rağmen, literatürde ve yukarıda yapmış olduğumuz açıklamalarda FN (False Negative) olarak gösterilen, gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örneklerin oranı aşağıda gösterildiği gibi hesaplanmaktadır.

$$\text{“YES” sınıfı için: } FP \text{ Rate} = \frac{19}{19+2} = 0,422$$

Her iki sınıf için FP oranları değerlendirildiğinde, kredisini vadesinde ödemesine rağmen, model tarafından kredisini vadesinde ödemeyecek şeklinde sınıflandırılanların oranının %18,3 gibi düşük sayılabilecek bir seviyede bulunduğu, ancak kredisini vadesinde ödememesine rağmen, ödeyecekmış gibi sınıflandırılanların oranının %42,2 olduğu, bu durumda kredisini vadesinde ödemeyenlerin yarıya yakınının model tarafından tespit edilemediği anlaşılmıştır.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemeyecek olduğu aşağıdaki gibi hesaplanmaktadır.

$$\text{“NO” sınıfı için: } Precision = \frac{26}{26+1} = 0,578$$

Kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödeyecek olduğu şu şekilde hesaplanmaktadır.

$$\text{“YES” sınıfı için: } Precision = \frac{85}{85+} = 0,817$$

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden kaçının kredisini vadesinde ödemeyecek olarak etiklendiği, aşağıdaki şekilde hesaplanmaktadır.

$$\text{“NO” sınıfı için: } Recall = \frac{26}{26+19} = 0,578$$

Gerçek durumda kredisini vadesinde ödeyenlerden kaçının kredisini vadesinde ödeyecek olarak etiklendiği, şu şekilde hesaplanmaktadır.

$$\text{“YES” sınıfı için: } Recall = \frac{85}{85+1} = 0,817$$

FP ve FN değerlerinin ikisinin de 19 olması, yukarıdaki hesaplamalarda benzer oranların bulunmasına neden olmuştur. Precision ve Recall oranlarının armonik ortalaması olması ve bu iki oranın her iki sınıf için aynı olması sebebiyle f-skoru bu oranlarla aynı değerlere sahip olmuştur. Söz konusu skor için arzulanan durumun 1'e yakınlık olduğu dikkate alındığında, modelin kredisini vadesinde ödeyenler için daha iyi bir sonuç ortaya koyduğunu söylemek mümkün olacaktır.

MCC'nin pozitif değer alması, model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değerın 0'a daha yakın olması, model ile gerçekleşen durum arasında çok güçlü bir korelasyon olmadığını ifade etmektedir.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği manasına gelmektedir.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir modelleme yapıldığı görülmekte, PRC eğrileri için eşik değer doğrusu aşağıdaki şekilde hesaplandığında, her iki sınıf için yapılan tahminlerin bu eşik değer doğrusu üzerinde yer aldığı, sonuç itibarıyla her iki sınıf için geliştirilen tahmin modelinin başarılı olarak değerlendirilebileceği anlaşılmıştır.

$$\begin{aligned} \text{Eşik Değer Doğrusunun Alanı} &= \frac{\text{Positives}}{\text{Positives} + \text{Negatives}} = \frac{26 + 19}{26 + 19 + 19 + 85} \\ &= 0,3020 \end{aligned}$$

Tablo 19: K En Yakın Komşu Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.578	0.183	0.578	0.578	0.578	0.395	0.707	0.544	NO
	0.817	0.422	0.817	0.817	0.817	0.395	0.707	0.809	YES
Ağ. Ort.	0.745	0.350	0.745	0.745	0.745	0.395	0.707	0.729	

Tablo 20’de sunulan hata matrisi incelendiğinde, yanlış tahmin edilen örnek sayısının her iki sınıf için de 19 olduğu görülmektedir. Ancak yukarıda diğer oranlara ilişkin yapmış olduğumuz değerlendirmelerden de anlaşılacağı üzere, algoritmanın geliştirdiği tahmin modelinin, kredisini vadesinde ödeyen örnekler için daha iyi bir sonuç ortaya koyduğu, kredisini vadesinde ödemeyen örneklerin yaklaşık yarısının tespit edilebildiği anlaşılmıştır.

Tablo 20: K En Yakın Komşu Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
26	19	a=NO	Gerçekleşen Değerler
19	85	b=YES	

3.9.3. Naive Bayes Algoritması Uygulama Çıktısı

Algoritmanın çalışma prensibi Bayes Teorisi’ni esas almakla birlikte, Naive Bayes yaklaşımında, veri setindeki tüm özelliklerin eşit derecede öneme sahip olduğu ve özelliklerin birbirinden bağımsız olduğu kabul edilmektedir. Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-6). Eğitim ve test veri setlerinin seçiminde “split 66.0% train, remainder test” metodu kullanılmıştır.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, Tablo 21’de test için seçilen 149 örnekten, 111’inin doğru sınıflandırıldığı, bunun da %74,4966 doğru sınıflandırma oranı olarak hesaplandığı, söz konusu oranın ZeroR algoritmasının eşik değerinin üzerinde olduğu görülmektedir. Kappa katsayısı, gerçekleşen durum ile tahmin edilen durum arasında zayıf bir uyum bulunduğunu göstermektedir. Ortalama hata oranları, daha önce değerlendirilen algoritmalara kıyasla daha düşük düzeyde seyretmekle birlikte, nispi hata oranları göstermektedir ki bu algoritmadaki tahminlerin gerçekleşen değerlerden sapması, gerçekleşen değerlerin kendi ortalamalarından saptığı düzeyin daha da altına inmiştir.

Tablo 21: Naive Bayes Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	111	74.4966 %
Kappa katsayısı	0.3078	
Ortalama Mutlak Hata	0.3464	
Ortalama Hata Karelerinin Karekökü	0.4108	
Nispi Mutlak Hata	76.6117 %	
Nispi Hata Karelerinin Karekökü	88.222 %	
Toplam Örnek Sayısı	149	

Tablo 22’de sunulan doğru sınıflandırmaya ilişkin detaylı oranlar incelendiğinde aşağıdaki sonuçların elde edildiği görülmüştür.

TP oranı, kredisini vadesinde ödemeyenler içerisinde, kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,356; gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,913 olarak hesaplanmıştır.

Her iki sınıf için TP oranlarına bakıldığında, kredisini vadesinde ödeyenler için yapılan tahminlerin daha isabetli olduğu, kredisini vadesinde ödemeyenlerin %35,6 oranında tespit edilebildiği anlaşılmıştır.

FP oranı, kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,087; gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,644 olarak hesaplanmıştır.

Her iki sınıf için FP oranları değerlendirildiğinde, kredisini vadesinde ödemesine rağmen, model tarafından kredisini vadesinde ödemeyecek şeklinde sınıflandırılanların oranının %8,7 olduğu, ancak kredisini vadesinde ödememesine rağmen, ödeyeceymiş gibi sınıflandırılanların oranının %64,4 olarak hesaplandığı, bu durumda kredisini vadesinde ödemeyenlerin yarısından fazlasının model tarafından tespit edilemediği anlaşılmıştır.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,640; kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,766 olduğu görülmektedir.

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanların oranının 0,356; gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanların oranının 0,913 olarak hesaplandığı görülmüştür.

TP ve FP oranlarının yanı sıra precision ve recall oranlarından da anlaşılacağı üzere, algoritmanın, kredisini vadesinde ödeyecek örneklere ilişkin elde ettiği başarı ile kredisini vadesinde ödemeyecek örneklere dair elde ettiği başarı arasında fark bulunmakta, algoritmanın geliştirdiği model, vadesinde ödeme yapan örnekleri tespit etmede daha başarılı olmaktadır. F-skoru için arzulanan durumun 1'e yakınlık olduğu dikkate alındığında, modelin kredisini vadesinde ödeyenler için daha iyi bir sonuç ortaya koyduğunu söylemek mümkün olacaktır.

MCC'nin pozitif değer alması, model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değer, model ile gerçekleşen durum arasında çok güçlü bir korelasyon olmadığını ifade etmektedir.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği manasına gelmektedir.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir modelleme yapıldığı görülmekte, bu sınıf için çizilen eğrinin altında kalan alanın 0,849; "NO" sınıfı için çizilen eğrinin alanının 0,647 olduğu gözlenmektedir. PRC eğrileri için eşik değer doğrusu hesaplandığında;

$$\text{Eşik Değer Doğrusunun Alanı} = \frac{\text{Positives}}{\text{Positives} + \text{Negative}} = \frac{16+9}{16+9+29+9} =$$

0,1678 değerinin elde edileceği, her iki sınıf için PRC eğrileri altında kalan alanın bu değerden yüksek olduğu, netice itibariyle her iki sınıf için geliştirilen tahmin modelinin rastgele tahmin yapan bir modelden daha iyi netice sundukları anlaşılmıştır.

Tablo 22: Naive Bayes Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.356	0.087	0.640	0.356	0.457	0.331	0.753	0.647	NO
	0.913	0.644	0.766	0.913	0.833	0.331	0.752	0.849	YES
Ağ. Ort.	0.745	0.476	0.728	0.745	0.720	0.331	0.752	0.788	

Tablo 23’de sunulan hata matrisi incelendiğinde, kredisini vadesinde ödemeyen örneklere ilişkin tahminlerde yapılan hata sayısının ve oranının daha fazla olduğu görülmektedir. “NO” sınıfı için yapılan hatalı tahmin sayısı fazla olmakla birlikte, aslında “NO” sınıfına dahil olan yani kredisini vadesinde ödemeyenlerin çoğunluğunun tespit edilemediği anlaşılmıştır. Kredisini vadesinde ödeyen örnekler içinse, daha başarılı bir sınıflandırma kararı verilmiş olduğu görülmektedir.

Tablo 23: Naive Bayes Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
16	29	a=NO	Gerçekleşen Değerler
9	95	b=YES	

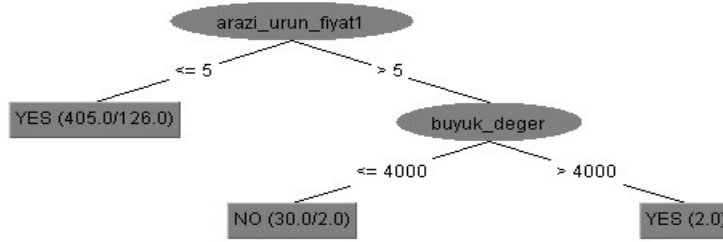
3.9.4. Karar Ağacı Algoritması Uygulama Çıktısı

Algoritma tarafından oluşturulan modelin, akış şemalarındaki benzer şekilde mantıksal kararlardan oluştuğu, model içerisindeki akışın, herhangi bir özelliğe ilişkin belirli kararların verildiği düğüm noktalarından geçerek nihai sonuca ulaştığı, modelin nihai şeklinin ağaç yapısına benzemesi nedeniyle algoritmanın karar ağacı olarak isimlendirildiği önceki bölümlerde açıklanmıştır.

Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-7). Eğitim ve test veri setlerinin seçiminde, daha önceki algoritmalarda olduğu gibi,

“split 66.0% train, remainder test” metodu kullanılmıştır. Program çıktısında, “J48 pruned tree” ifadesinin yer aldığı, J48’in karar ağacı algoritmasının WEKA programında ifade edilmiş yöntemi olduğu, “pruned tree” ile, teorik olarak önceki bölümlerde açıkladığımız gibi, ağaçta budama işlemi yapıldığının belirtildiği görülmektedir. Ağaç budama işlemine tabi tutularak, algoritma tarafından önemsiz olduğu ya da değerlendirilmesi gereksiz maliyet içerecek özellikler işleme dahil edilmemekte, en kısa şekilde ağacın dallarının sınıflandırma kararına ulaşması amaçlanmaktadır. Budanmış ağaçta, 2 özellik kullanılarak sınıflandırma kararı verildiği görülmektedir. Bu özellikler sırasıyla “arazi_urun_fiyat1” ve “buyuk_deger” olarak belirlenmiştir. Üreticinin ekim-dikimini yaptığını beyan ettiği ilk ürünün kg fiyatına göre ağaç şekil almaya başlamakta, sonrasında üreticinin sahip olduğu büyükbaş hayvanların 1 adedinin değerine göre sınıflandırma kararı verilmekte, nihai durumda karar ağacı Şekil 17’deki gibi oluşmaktadır.

Şekil 17: Uygulama Örneğindeki Karar Ağacı



Uygulama çıktısındaki karar ağacı yakından incelendiğinde, ilk ürünün kg fiyatı 5,00 TL’ye küçük eşit olduğunda, bu üreticinin kredisini vadesinde ödeyeceği tahmini yapılmaktadır. İlk ürünün kg fiyatının 5,00 TL’den büyük olması durumunda, üreticinin sahip olduğu büyükbaş hayvanın değerine bakılmakta, söz konusu değer 4.000,00 TL’ye küçük eşit olması durumunda kredinin vadesinde ödenmeyeceği, 4.000,00 TL’den yüksek olması durumunda ise kredinin vadesinde ödeneceği öngörüsünde bulunduğu görülmektedir.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, Tablo 24’de test için seçilen 149 örnekten, 115’inin doğru sınıflandırıldığı, bunun da % 77,1812 doğru sınıflandırma oranına karşılık geldiği, söz konusu oranın ZeroR algoritmasının eşik değerinin üzerinde olduğu görülmektedir. Kappa katsayısı, gerçekleşen durum ile

tahmin edilen durum arasında zayıf bir uyum bulunduğunu göstermektedir. Ortalama hata oranları, şu ana dek hesaplanan hata oranlarına yakın seviyede gerçekleşmiştir. Ancak nispi hata oranları, tahmin değerlerinin sapmasının, gerçekleşen değerlerinin sapmasına yaklaştığını göstermektedir.

Tablo 24: Karar Ağacı Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	115	77.1812%
Kappa katsayısı	0.3428	
Ortalama Mutlak Hata	0.3773	
Ortalama Hata Karelerinin Karekökü	0.4267	
Nispi Mutlak Hata	83.4407 %	
Nispi Hata Karelerinin Karekökü	91.6402 %	
Toplam Örnek Sayısı	149	

Tablo 25’de sunulan doğru sınıflandırmaya ilişkin detaylı oranlar incelendiğinde aşağıdaki sonuçların elde edildiği görülmüştür.

TP oranı, kredisini vadesinde ödemeyenler içerisinde, kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,311; gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,971 olarak hesaplanmıştır.

Her iki sınıf için TP oranlarına bakıldığında, kredisini vadesinde ödeyenler için yapılan tahminlerin daha isabetli olduğu ve büyük çoğunluğunun tespit edildiği, kredisini vadesinde ödemeyenlerin %31,1 oranında tespit edilebildiği anlaşılmıştır.

FP oranı, kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,029; gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,689 olarak hesaplanmıştır.

Her iki sınıf için FP oranları değerlendirildiğinde, kredisini vadesinde ödemesine rağmen, model tarafından kredisini vadesinde ödemeyecek şeklinde sınıflandırılanların oranının %2,9 olduğu, ancak kredisini vadesinde ödememesine

rağmen, ödeyecekmiş gibi sınıflandırılanların oranının %68,9 olarak hesaplandığı görülmüştür. Bu durumda kredisini vadesinde ödemeyenlerin yarısından fazlasının model tarafından tespit edilemediği ve Naive Bayes algoritmasına kıyasla bu durum için daha kötü bir netice elde edildiği anlaşılmıştır.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,824; kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödediğine ilişkin oranın 0,765 olduğu görülmektedir. Bu durumda algoritmanın, vadesinde ödenmeyen kredilere ilişkin tahminlerinin kendi içerisinde tutarlı olduğu, öyle ki bu tutarlılığın vadesinde ödenen kredi örneklerine kıyasla daha yüksek düzeyde bulunduğu, daha önceki algoritmalara kıyasla bu konuda daha başarılı bir sonuç elde edildiği anlaşılmaktadır.

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanların oranının 0,311; gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanların oranının 0,971 olarak hesaplandığı görülmüştür. Söz konusu oranın TP oranında elde edilen sonuçlarla aynı şekilde yorumlanabileceği anlaşılmıştır.

F-skorunun, daha önceki algoritmalarda olduğu gibi, vadesinde ödenen kredi sınıfı için daha olumlu bir sonuca işaret ettiği gözlenmiştir.

MCC'nin pozitif değer alması (0,408), model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değer çok güçlü bir korelasyon için yeterli olmadığı, ancak daha önceki algoritmalara kıyasla bu oranın yüksek olduğu görülmüştür.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği göstermekle birlikte, şu ana dek eğri altında kalan alanın diğer algoritmalarda (ZeroR hariç) daha yüksek olduğu görülmüştür.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir

modelleme yapıldığı görülmekte, bu sınıf için çizilen eğrinin altında kalan alanın 0,770; “NO” sınıfı için çizilen eğrinin alanının 0,444 olduğu gözlenmektedir. PRC eğrileri için eşik değeri doğrusu hesaplandığında;

$$\text{Eşik Değer Doğrusunun Alanı} = \frac{\text{Positives}}{\text{Positives+Neg}} = \frac{14+3}{14+3+31+10} =$$

0,1141 değerinin elde edileceği, her iki sınıf için PRC eğrileri altında kalan alanın bu değerden yüksek olduğu, ancak ideal olan durumda eğri altında kalan alanların mümkün mertebe 1'e yakın olması beklenirken, kredinin vadesinde ödenmediği sınıf için ideal durumdan uzaklaşmış olduğu görülmektedir.

Tablo 25: Karar Ağacı Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.311	0.029	0.824	0.311	0.452	0.408	0.626	0.444	NO
	0.971	0.689	0.765	0.971	0.856	0.408	0.626	0.770	YES
Ağ. Ort.	0.772	0.490	0.783	0.772	0.734	0.408	0.626	0.671	

Tablo 26'da sunulan hata matrisi incelendiğinde, gerçekleşen durumda kredisini vadesinde ödemeyen örneklere ilişkin tahminlerde hata sayısının ve oranının daha fazla olduğu, gerçek durumda kredisini vadesinde ödeyen örnekler için daha başarılı bir sınıflandırma kararı verilmiş olduğu gözlenmektedir. Ancak precision oranını dikkate alındığında, vadesinde ödeme yapmayacak şekilde tahmin edilen örneklerin, büyük çoğunluğunun aslında vadesinde kredisini ödemeyen örneklerden oluştuğu, dolayısıyla modelin “NO” sınıfı olarak etiketleme işleminin kendi içerisinde başarılı olduğu söylenebilmektedir.

Tablo 26: Karar Ağacı Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
14	31	a=NO	

3	101	b=YES	Gerçekleşen Değerler
---	-----	-------	----------------------

3.9.5. Doğrusal Ayırma Analizi Algoritması Uygulama Çıktısı

Algoritma, veri setindeki özelliklerin istatistiksel dağılım özelliklerini baz alarak bir ayırma yapma yoluna gitmekte, dağılımların sergiledikleri eğitim farklılıklarına göre sınıflandırma yapılmakta, istatistiksel dağılımların belirlenmesi için de kovaryans matrislerinin oluşturulması gerekmektedir.

Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-8). Eğitim ve test veri setlerinin seçiminde, “split 66.0% train, remainder test” metodu kullanılmıştır. Program çıktısı incelendiğinde, özellik sayısının 43 olarak belirtildiği görülmektedir. Doğrusal ayırma analizi algoritması sadece sayısal değerleri işleyebildiğinden, üreticilerin arazilere ekip biçtikleri ürünleri simgeleyen ve metin niteliğinde olan “arazi_urun1-5” özellikleri veri setinden çıkarılmış, geri kalan verilerin tamamının sayısal olması sağlanmıştır.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, Tablo 27’de test için seçilen 149 örnekten, 99’unun doğru sınıflandırıldığı, bunun da % 66,443 doğru sınıflandırma oranına karşılık geldiği, söz konusu oranın ZeroR algoritmasının eşik değerinin altında olduğu görülmektedir. Kappa katsayısı, gerçekleşen durum ile tahmin edilen durum arasında zayıf bir uyum bulunduğunu göstermekle birlikte, katsayının önemsiz düzeyde uyum sınırına yakın olduğu gözlenmektedir. Ortalama hata oranlarının, şu ana kadar değerlendirilen tüm algoritmalar içerisinde en yüksek düzeye ulaştığı görülmekle birlikte, rastgele tahminde bulunan ZeroR algoritmasının hata oranlarının dahi doğrusal ayırma analizi algoritmasının hata oranlarından daha düşük olduğu görülmüştür. Nispi hata oranları değerlendirildiğinde, gerçek değerlerin kendi ortalamalarından saptığının üzerinde, tahmin modelinde sapma bulunduğu da anlaşılmıştır. Hataların karelerinin alındığı oranlar ile sadece farkların alındığı hata oranları arasında ciddi bir fark bulunmadığı, bu durumdan sapmaların büyüklüklerinin sınırlı olduğu anlaşılmaktadır.

Tablo 27: Doğrusal Ayırma Analizi Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	99	66.443 %
Kappa katsayısı	0.2686	
Ortalama Mutlak Hata	0.4705	
Ortalama Hata Karelerinin Karekökü	0.4773	
Nispi Mutlak Hata	104.0463 %	
Nispi Hata Karelerinin Karekökü	102.5244 %	
Toplam Örnek Sayısı	149	

Tablo 28’de sunulan doğru sınıflandırmaya ilişkin detaylı oranlar incelendiğinde aşağıdaki sonuçların elde edildiği görülmüştür.

TP oranı, kredisini vadesinde ödemeyenler içerisinde, kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,60; gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,692 olarak hesaplanmıştır.

Her iki sınıf için TP oranları değerlendirildiğinde, sınıflar için yapılan isabetli tahminlerin birbirine yakın olduğu, ancak göreceli olarak daha iyi tahmin sonucu alınan kredisini vadesinde ödeyenlerin %69,2 oranında tespit edilebildiği anlaşılmıştır.

FP oranı, kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,308; gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,40 olarak hesaplanmıştır.

FP oranları değerlendirildiğinde, kredisini vadesinde ödeyen örneklerin, kredisini vadesinde ödemeyecek şeklinde en yüksek oranda sınıflandırılması, bu algoritma tarafından geliştirilen modelce yapılmıştır. Diğer taraftan, gerçek durumda kredisini vadesinde ödemeyenlerin yarısından azı kredisini vadesinde ödeyecekmiş gibi sınıflandırılmıştır. Bu konuda karar ağacı algoritmasından daha iyi bir netice elde edilmiş olduğu söylenebilmektedir.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,458; kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,80 olduğu görülmektedir. Bu durumda algoritmanın, vadesinde ödenmeyen kredilere ilişkin tahminlerinin yarısından daha azının başarılı olduğu, diğer taraftan kredisini vadesinde ödeyecek şeklinde yapılan tahminlerin isabetli olarak değerlendirilebileceği gözlenmiştir.

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanların oranının 0,60; gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanların oranının 0,692 olarak hesaplandığı görülmüştür. Bu durumda her iki sınıf için istenilen isabetli tespit seviyesine yaklaşamadığı anlaşılmaktadır.

F-skorunun, kredinin vadesinde ödendiği örnekler için istenilen düzeye (1'e) daha yakın olduğu gözlenmiştir.

MCC'nin pozitif değer alması (0,274), model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değer zayıf bir ilişkiye işaret ettiği anlaşılmaktadır.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği anlamına gelmekle birlikte, eğri altında kalan alanın 0,682 olduğu, ancak gelişmiş tahmin yaptığı bilinen ZeroR algoritmasının, doğrusal ayırma analizi algoritmasından daha yüksek düzeyde sınıflandırma başarısı sağladığı gözlenmiştir. Bu durum, veri setinin dengesiz dağılımından, yani kredinin vadesinde ödendiği örnek sayısının, kredinin vadesinde ödenmediği örnek sayısına kıyasla bariz şekilde fazla olmasından kaynaklanmaktadır.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir modelleme yapıldığı görülmekte, bu sınıf için çizilen eğrinin altında kalan alanın 0,798; "NO" sınıfı için çizilen eğrinin alanının 0,558 olduğu gözlenmektedir. PRC eğrileri için eşik değer doğrusu hesaplandığında;

$$Eşik Değer Doğrusunun Alanı = \frac{Positives}{Positives+Negatives} = \frac{27+32}{27+32+18+72} =$$

0,3960 değerinin elde edileceği, her iki sınıf için PRC eğrileri altında kalan alanın bu değerden yüksek olduğu, ancak ideal olan durumda eğri altında kalan alanların mümkün olduğunca 1'e yakın olması beklendiğinden, kredinin vadesinde ödendiği sınıf için ideal duruma daha yakın olduğu görülmektedir.

Tablo 28: Doğrusal Ayırma Analizi Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,600	0,308	0,458	0,600	0,519	0,274	0,682	0,558	NO
	0,692	0,400	0,800	0,692	0,742	0,274	0,682	0,798	YES
Ağ. Ort.	0,664	0,372	0,697	0,664	0,675	0,274	0,682	0,726	

Tablo 29'da sunulan hata matrisi incelendiğinde, kredinin vadesinde ödendiği örnekler için daha fazla sayıda hatalı tahmin yapıldığı ancak oran olarak vadesinde ödenmeyen örnekler için daha yüksek düzeyde hata yapıldığı görülmektedir. Yapılan tahminler açısından değerlendirildiğinde, vadesinde ödenmeyecek şekilde etiketlenen örneklerin çoğunluğunun hatalı olduğu, diğer yandan vadesinde ödenecek şekilde etiketlenen örneklerde daha yüksek düzeyde başarı elde edildiği anlaşılmaktadır. Genel sınıflandırma başarısı açısından doğrusal ayırma analizi algoritmasının, ZeroR algoritmasından daha düşük seviyede sınıflandırma başarısı sunması nedeniyle, başarısız olarak değerlendirilebilecek bir model geliştirdiği söylenebilmektedir.

Tablo 29: Doğrusal Ayırma Analizi Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
27	18	a=NO	Gerçekleşen Değerler
32	72	b=YES	

3.9.6. Lojistik Regresyon Algoritması Uygulama Çıktısı

Değişkenler arasındaki ilişkinin, doğrusal bir denklem ile modellenebilecek kadar basit olmadığı durumlarda, logaritmik bir yaklaşım benimsenerek olasılık hesaplamasına dayalı bir sınıflandırma yöntemi kullanılmakta, söz konusu yöntem de lojistik regresyon olarak adlandırılmaktadır.

Algoritmaya ilişkin program çıktısı detaylı olarak ekte sunulmuştur (Bkz. Ek-9). Eğitim ve test veri setlerinin seçiminde, diğer algoritmalarından farklı olarak, “10-fold cross-validation” metodu kullanılmıştır. Veri setlerinin kullanılmasında bu kez farklı bir yöntem kullanılmasının sebebi, tercih edilen yöntemde daha yüksek düzeyde doğru sınıflandırma oranının elde edilmiş olmasıdır. Söz konusu algoritmada 48 özelliğin tamamının kullanıldığı program çıktısından anlaşılmaktadır.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, örneklerin tamamının test için kullanılmış olduğu, Tablo 30’dan test için seçilen 437 örnekten, 305’inin doğru sınıflandırıldığı, bunun da % 69,7941 doğru sınıflandırma oranına karşılık geldiği, söz konusu oranın ZeroR algoritmasının eşik değerinin çok az bir farkla altında olduğu görülmektedir. Kappa katsayısı, gerçekleşen durum ile tahmin edilen durum arasında zayıf bir uyum bulunduğunu göstermektedir. Gerek ortalama hata oranlarında gerekse nispi hata oranlarında, sapmaların karesinin alındığı oranların belirgin bir şekilde diğer oranlardan yüksek olduğu görülmüş, bu durumdan tahminlerdeki sapma büyüklüklerinin yüksek olduğu anlaşılmıştır. Öyle ki nispi hata karelerinin karekökünde, tahmin edilen değerlerin, gerçek değerlerin kendi ortalamalarından saptığının üzerinde bir sapma olduğu gözlenmiştir.

Tablo 30: Lojistik Regresyon Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	305	69.7941 %
Kappa katsayısı	0.2988	
Ortalama Mutlak Hata	0.3658	
Ortalama Hata Karelerinin Karekökü	0.4854	
Nispi Mutlak Hata	80.1005 %	
Nispi Hata Karelerinin Karekökü	101.6055 %	
Toplam Örnek Sayısı	437	

Tablo 31’de sunulan doğru sınıflandırmaya ilişkin detaylı oranlar incelendiğinde aşağıdaki sonuçların elde edildiği görülmüştür.

TP oranı, kredisini vadesinde ödemeyenler içerisinde, kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,448; gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,834 olarak hesaplanmıştır.

TP oranları değerlendirildiğinde, kredisini vadesinde ödeyenler için daha yüksek düzeyde başarı sağlandığı, ancak kredinin vadesinde ödenmediği örneklerin yarısından daha azının tespit edilebildiği anlaşılmıştır.

FP oranı, kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,166; gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,552 olarak hesaplanmıştır.

Her iki sınıf için FP oranları incelendiğinde, gerçek durumda kredisini vadesinde ödeyecek olanlar için daha sınırlı düzeyde hata yapılmış olduğu, kredisini vadesinde ödemeyen örneklerin ise yarısından fazlası için hatalı tahmin yapıldığı gözlenmiştir.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,595; kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,735 olduğu görülmektedir. Bu durumda algoritmanın, her iki sınıf için yapmış olduğu tahminlerin başarısı arasında ciddi bir farklılık bulunmadığı söylenebilmekle beraber, kredinin vadesinde ödeneceğine ilişkin tahminlerin daha isabetli olduğu görülmektedir.

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanların oranının 0,448; gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanların oranının 0,834 olarak hesaplandığı görülmüştür. Bu durumda

kredinin vadesinde ödendiği örnekler için başarılı sayılabilecek bir netice elde edildiği tespit edilmiştir.

F-skorunun, kredinin vadesinde ödendiği örnekler için istenilen düzeye (1'e) daha yakın olduğu gözlenmiştir.

MCC'nin pozitif değer alması (0,305), model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değer zayıf bir ilişkiye işaret ettiği anlaşılmaktadır. Ancak bu ilişkinin doğrusal ayırma analizi algoritmasına kıyasla daha güçlü olduğu görülmüştür.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği anlamına gelmekle birlikte, eğri altında kalan alanın 0,664 olduğu, ancak geliş güzel tahmin yaptığı bilinen ZeroR algoritmasının, doğrusal ayırma analizi algoritmasından daha yüksek düzeyde sınıflandırma başarısı sağladığı gözlenmiştir. Bu durum, veri setinin dengesiz dağılmış olmasından, yani kredisinin vadesinde ödendiği örnek sayısının, kredinin vadesinde ödenmediği örnek sayısına kıyasla bariz şekilde fazla olmasından kaynaklanmaktadır.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir modelleme yapıldığı görülmekte, bu sınıf için çizilen eğrinin altında kalan alanın 0,753; "NO" sınıfı için çizilen eğrinin alanının 0,526 olduğu gözlenmektedir. PRC eğrileri için eşik değeri doğrusu hesaplandığında;

$$\text{Eşik Değer Doğrusunun Alanı} = \frac{\text{Positives}}{\text{Positives} + \text{Negatives}} = \frac{69+4}{69+47+85+2} =$$

0,2654 değerinin elde edileceği, her iki sınıf için PRC eğrileri altında kalan alanın bu değerden yüksek olduğu, ancak ideal olan durumda eğri altında kalan alanların mümkün olduğunca 1'e yakın olması beklendiğinden, kredinin vadesinde ödendiği sınıf için ideal duruma daha yakın olduğu görülmektedir.

Tablo 31: Lojistik Regresyon Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,448	0,166	0,595	0,448	0,511	0,305	0,664	0,526	NO

	0,834	0,552	0,735	0,834	0,781	0,305	0,664	0,753	YES
Ağ. Ort.	0,698	0,416	0,686	0,698	0,686	0,305	0,664	0,673	

Tablo 32’de sunulan hata matrisi incelendiğinde, kredinin vadesinde ödenmediği örnekler için daha fazla sayıda hatalı tahmin yapıldığı, hatalı tahmin oranının da diğer sınıfa kıyasla daha yüksek düzeyde bulunduğu görülmektedir. Yapılan tahminler açısından değerlendirildiğinde, vadesinde ödenecek şekilde etiketlenen örneklerde daha yüksek düzeyde başarı elde edildiği anlaşılmaktadır. Genel sınıflandırma başarısı açısından lojistik regresyon algoritmasının, ZeroR algoritmasından daha düşük seviyede sınıflandırma başarısı sunması nedeniyle, başarısız sayılabilecek bir model sunduğu ifade edilebilecektir.

Tablo 32: Lojistik Regresyon Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
69	85	a=NO	Gerçekleşen Değerler
47	236	b=YES	

3.9.7. Yapay Sinir Ağları Algoritması Uygulama Çıktısı

Algoritmanın, bir takım girdi sinyalleri ve çıktı sinyalleri arasındaki ilişkiyi, biyolojik beyinin algısal girdilere verdiği tepkilere benzer şekilde modelleyerek sınıflandırma yaptığı açıklanmıştır.

Algoritmaya ilişkin program çıktısı çok uzun olduğundan, aşağıda yapacağımız değerlendirmeye esas olacak bölümlere ekte yer verilmiştir (Bkz. Ek-10). Eğitim ve test veri setlerinin seçiminde, “split 66.0% train, remainder test” metodu kullanılmıştır.

Algoritma çıktılarına ait özet tablolar sırasıyla incelendiğinde, Tablo 33’de test için seçilen 149 örnekten, 111’inin doğru sınıflandırıldığı, bunun da %74,4966 doğru sınıflandırma oranına karşılık geldiği, söz konusu oranın ZeroR algoritmasının eşik değerinin üzerinde olduğu görülmektedir. Kappa katsayısı, gerçekleşen durum ile

tahmin edilen durum arasında zayıf bir uyum bulunduğunu göstermektedir. Ortalama hata oranları arasındaki farkın, diğer algoritmalara kıyasla oldukça yüksek olduğu göz önüne alındığında, tahminlerin gerçekleşen değerlerden sapma büyüklüklerinin oldukça fazla olduğunu söylemek mümkün olmaktadır. Bu durum nispi hata oranlarından da anlaşılmaktadır. Nispi hata karelerinin karekökü, gerçekleşen değerlerin sahip olduğu hata düzeyine nerdeyse ulaşılmış olduğunu göstermektedir.

Tablo 33: Yapay Sinir Ağları Algoritması Program Çıktı Özeti

Doğru Sınıflandırılmış Örnek Sayısı	111	74.4966 %
Kappa katsayısı	0.2765	
Ortalama Mutlak Hata	0.2513	
Ortalama Hata Karelerinin Karekökü	0.4647	
Nispi Mutlak Hata	55.5729 %	
Nispi Hata Karelerinin Karekökü	99.7997 %	
Toplam Örnek Sayısı	149	

Tablo 34’de sunulan doğru sınıflandırmaya ilişkin detaylı oranlar incelendiğinde aşağıdaki sonuçların elde edildiği görülmüştür.

TP oranı, kredisini vadesinde ödemeyenler içerisinde, kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,289; gerçekte kredisini vadesinde ödeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,942 olarak hesaplanmıştır.

Her iki sınıf için TP oranları değerlendirildiğinde, kredisini vadesinde ödeyenler için yapılan tahminlerin oldukça isabetli olduğu, kredisini vadesinde ödemeyenlerin ancak %28,9 oranında tespit edilebildiği anlaşılmıştır.

FP oranı, kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için 0,058; gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için 0,711 olarak hesaplanmıştır.

FP oranlarından anlaşıldığı üzere, vadesinde ödenen örnekler için hatalı tahmin düzeyi oldukça sınırlı olmasına rağmen, vadesinde ödenmeyen örnekler için hatalı tahmin düzeyinin oldukça yüksek olduğu görülmektedir.

Precision (kesinlik) oranları her iki sınıf için incelendiğinde;

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,684; kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin oranın 0,754 olduğu görülmektedir.

Recall (hassasiyet) oranları her iki sınıf için incelendiğinde;

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanların oranının 0,289; gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanların oranının 0,942 olarak hesaplandığı görülmüştür.

TP ve FP oranlarının yanı sıra precision ve recall oranlarından da anlaşılacağı üzere, algoritmanın, kredisini vadesinde ödeyecek örneklere ilişkin daha başarılı netice sağladığı gözlenmiştir.

F-skoru için arzulanan durumun 1'e yakınlık olduğu dikkate alındığında, modelin kredisini vadesinde ödeyenler için daha iyi bir sonuç ortaya koyduğunu söylemek mümkün olacaktır.

MCC'nin pozitif değer alması, model ile gerçek durum arasında pozitif bir korelasyon katsayısı bulunduğunu göstermekle birlikte, söz konusu değer, model ile gerçekleşen durum arasında çok güçlü bir korelasyon bulunmadığını göstermektedir.

ROC eğrisi altında kalan alanın 0,50'den yüksek olması, modelin rastgele yapılacak bir tahminden daha iyi sonuç verdiği manasına gelmekle birlikte, söz konusu alan algoritmalar arasında elde edilen en yüksek değerler arasında yer almaktadır.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, "YES" sınıfı yani kredisini vadesinde ödeyenlerin dahil olduğu sınıf için daha iyi bir modelleme yapıldığı görülmekte, bu sınıf için çizilen eğrinin altında kalan alanın 0,830; "NO" sınıfı için çizilen eğrinin alanının 0,591 olduğu gözlenmektedir. PRC eğrileri için eşik değer doğrusu hesaplandığında;

$$\text{Eşik Değer Doğrusunun Alanı} = \frac{\text{Positives}}{\text{Positives} + \text{Ne}} = \frac{13+6}{13+6+32+98} =$$

0,1275 değerinin elde edileceği, her iki sınıf için PRC eğrileri altında kalan alanın bu değerden yüksek olduğu, netice itibariyle her iki sınıf için geliştirilen tahmin modelinin bu oran baz alındığında başarılı olarak değerlendirilebileceği anlaşılmıştır.

Tablo 34: Yapay Sinir Ağları Algoritması Sınıf Bazında Detaylı Doğru Sınıflandırma Oranları

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,289	0,058	0,684	0,289	0,406	0,318	0,731	0,591	NO
	0,942	0,711	0,754	0,942	0,838	0,318	0,731	0,830	YES
Ağ. Ort.	0,745	0,514	0,733	0,745	0,707	0,318	0,731	0,758	

Tablo 35’de sunulan hata matrisi incelendiğinde, kredisini vadesinde ödemeyen örneklere ilişkin tahminlerde yapılan hata sayısının düşük olduğu görülmektedir. “NO” sınıfı için yapılan hatalı tahmin sayısı düşük olmakla birlikte, aslında “NO” sınıfına dahil olan yani kredisini vadesinde ödemeyenlerin çoğunluğunun tespit edilemediği anlaşılmıştır. Gerçekleşen durumda kredisini vadesinde ödeyen örneklerin büyük çoğunluğunun model tarafından tespit edildiği, ancak vadesinde ödenecek şekilde tahmin edilen örneklerde hata oranının yükseldiği gözlenmiştir.

Tablo 35: Yapay Sinir Ağları Algoritması Hata Matrisi (Confusion Matrix)

Tahmin Değerleri			
a	b		
13	32	a=NO	Gerçekleşen Değerler
6	98	b=YES	

3.10 UYGULAMA SONUÇLARININ KARŞILAŞTIRILMASI

Çalışmamızda temel olarak 6 farklı algoritma kullanılmış olmakla birlikte, ZeroR algoritması, diğer 6 algoritmanın performansının değerlendirilmesinde eşik değer sağlayıcı olarak kullanılmıştır. Tablo 36'dan algoritmaların genel doğru sınıflandırma oranlarına bakıldığında, en yüksek doğru sınıflandırmanın %77,1812 oranı ile karar ağacı algoritması tarafından ortaya konulduğu, doğrusal ayırma analizi ve lojistik regresyon algoritmalarının, ZeroR algoritmasıyla ortaya konulan eşik değerinin altında kaldığı görülmektedir. K en yakın komşu, naive bayes ve yapay sinir ağları algoritmalarının aynı düzeyde, %74,4966 doğru sınıflandırma oranı elde ederek, eşik değer üzerinde doğru sınıflandırma modellemesi yaptıkları gözlenmiştir.

TP oranı, "NO" sınıfı için incelendiğinde, gerçekleşen durumda kredinin vadesinde ödenmediği örneklerin, algoritmalar tarafından vadesinde ödenmeyecek şeklinde tahmin edilmesinde, en yüksek oranın doğrusal ayırma analizi algoritmasıyla oluşturulan model tarafından elde edildiği, her ne kadar bu algoritma genel sınıflandırma başarısında, eşik değer altında kalmış olsa da kredi verenlerce öncelikli olarak önem arz eden vadesinde ödenmeyecek krediler için daha yüksek oranda başarılı tahmin yapıldığı anlaşılmıştır.

TP oranı, "YES" sınıfı için incelendiğinde, doğrusal ayırma analizi algoritması haricinde, genel olarak tüm algoritmaların yüksek sayılabilecek düzeyde, vadesinde ödenecek krediler için doğru tahminde bulunduğu görülmüştür. Gerçekte vadesinde ödenen örneklerin, vadesinde ödenecek olarak tahmin edilmesinde, en başarılı algoritmanın karar ağacı algoritması olduğu görülmektedir.

Kredisini vadesinde ödeyenler içerisinde, algoritma tarafından kredisini vadesinde ödemeyecek olarak sınıflandırılan örnekler için FP oranları incelendiğinde, en yüksek hatanın 0,308 ile doğrusal ayırma analizine ait olduğu gözlenmiştir. Bu konuda en az hatalı model geliştiren algoritmanın 0,029 oranı ile karar ağacı algoritması olduğu anlaşılmıştır.

Gerçekte kredisini vadesinde ödemeyenler arasında, modelin kredisini vadesinde ödeyecek olarak sınıflandırdığı örnekler için FP oranları değerlendirildiğinde, en yüksek hatanın 0,711 oranı ile yapay sinir ağları algoritması, en düşük hatanın 0,40 oranı ile doğrusal ayırma analizi algoritması tarafından yapıldığı gözlenmiştir. Kredi verenlerce önem arz eden hususun, vadesinde ödenmeyecek olan

kredilerin tespiti olduğu dikkate alındığında, bu konuda en az düzeyde hata yapan doğrusal ayırma analizi algoritmasının, diğer algoritmalara kıyasla daha başarılı olduğu söylenebilmektedir.

Kredisini vadesinde ödemeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin precision oranından, karar ağacı algoritmasının en yüksek düzeyde isabet oranına (0,824) sahip olduğu görülmektedir. Kredisini vadesinde ödeyecek olarak tahmin edilen birimlerin, aslında kaçının kredisini vadesinde ödemediğine ilişkin precision oranında, k en yakın komşu algoritmasının, diğer algoritmalara kıyasla daha başarılı tahminde (0,817) bulunduğu anlaşılmıştır.

Gerçek durumda kredisini vadesinde ödemeyen örneklerden, kredisini vadesinde ödemeyecek olarak sınıflandırılanlara ilişkin recall oranları incelendiğinde, k en yakın komşu ve doğrusal ayırma analizi algoritmalarının diğerlerine kıyasla daha başarılı olduğu, ancak en yüksek recall oranının doğrusal ayırma analizi algoritmasında elde edildiği görülmektedir. Gerçek durumda kredisini vadesinde ödeyen örneklerden, kredisini vadesinde ödeyecek olarak sınıflandırılanlara dair recall oranları değerlendirildiğinde, karar ağacı algoritmasının 0,971 oranıyla en yüksek recall oranına sahip olduğu gözlenmiştir.

Algoritmaların her iki sınıf için f skorları incelendiğinde, TP ve FP oranlarına paralel olarak, kredinin vadesinde ödendiği sınıf için f skorlarının daha yüksek seyrettiği, karar ağacı algoritmasının söz konusu sınıf için en yüksek orana (0,856) sahip olduğu görülmüştür.

Tahminlere esas olan modeller ile gerçek durum arasındaki korelasyon düzeylerine bakıldığında, en yüksek MCC oranının (0,408) karar ağacı algoritmasına ait olduğu anlaşılmıştır.

Modelin rastgele yapılacak bir tahminden daha iyi sonuç verip vermediğini değerlendiren ROC eğrisi altında kalan alanlar incelendiğinde, naive bayes algoritmasının ROC eğrisi altında kalan alanının (0,752), diğer algoritmalarından daha yüksek olduğu görülmüştür.

Her iki sınıf için PRC eğrileri altında kalan alanlar karşılaştırıldığında, genel olarak kredinin vadesinde ödendiği “YES” sınıfı örnekleri için bu alanların daha

yüksek hesaplandığı, en yüksek alana sahip algoritmanın 0,849 ile naive bayes algoritması olduğu gözlenmiştir.

Tablo 36: Algoritma Sonuçlarının Karşılaştırılması

Oranlar	ZeroR	K-NN	Naive Bayes	Karar Ağacı	Doğrusal Ayırma Analizi	Lojistik Regresyon	Yapay Sinir Ağları
Doğru Sınıflandırma Oranı	%69,7987	%74,4966	%74,4966	%77,1812	%66,443	%69.7941	%74,4966
TP Oranı ("NO")	0	0.578	0.356	0.311	0.60	0.448	0.289
TP Oranı ("YES")	1	0.817	0.913	0.971	0.692	0.834	0.942
FP Oranı ("NO")	0	0.183	0.087	0.029	0.308	0.166	0.058
FP Oranı ("YES")	1	0.422	0.644	0.689	0.40	0.552	0.711
Precision ("NO")	0	0.578	0.640	0.824	0.458	0.595	0.684
Precision ("YES")	0.698	0.817	0.766	0.765	0.80	0.735	0.754
Recall ("NO")	0	0.578	0.356	0.311	0.60	0.448	0.289
Recall ("YES")	1	0.817	0.913	0.971	0.692	0.834	0.942
F-Skoru ("NO")	0	0.578	0.457	0.452	0.519	0.511	0.406
F-Skoru ("YES")	0.822	0.817	0.833	0.856	0.742	0.781	0.838
MCC	0	0.395	0.331	0.408	0.274	0.305	0.318
ROC Alanı	0.50	0.707	0.752	0.626	0.682	0.664	0.731
PRC Alanı ("NO")	0.302	0.544	0.647	0.444	0.558	0.526	0.591
PRC Alanı ("YES")	0.698	0.809	0.849	0.770	0.798	0.753	0.830

Tablo 37'den görülebileceği üzere, program çıktısındaki tüm kappa istatistiği değerleri tahmin modeli ile gerçekleşen durum arasında zayıf düzeyde uyuma işaret etse de k en yakın komşu algoritmasının orta düzeyde uyum sınırına yakın olduğu görülmüştür. Ayrıca, doğrusal ayırma analizi algoritması, vadesinde ödenmeyecek

krediler için daha yüksek düzeyde başarı elde etmiş gibi görünse de en düşük kappa istatistiğine bu algoritmanın sahip olduğu, sonuç itibariyle gerçek durum ile tahmin modeli arasında en zayıf uyum derecesinin bu algoritma çıktısında görüldüğü tespit edilmiştir.

Tablo 37: Kappa İstatistik Değerlerinin Karşılaştırılması

<i>Oran</i>	<i>ZeroR</i>	<i>K-NN</i>	<i>Naive Bayes</i>	<i>Karar Ağacı</i>	<i>Doğrusal Ayırma Analizi</i>	<i>Lojistik Regresyon</i>	<i>Yapay Sinir Ağları</i>
Kappa İstatistiği	0	0,3951	0,3078	0,3428	0,2686	0,2988	0,2765

Çalışmamızın temel amacı üzerine kurulan hipotezler değerlendirildiğinde, ZeroR algoritmasının belirlediği eşik değerinin altında doğru sınıflandırma modeli geliştiren doğrusal ayırma analizi ve lojistik regresyon algoritmalarına ait Hipotez-4 ve Hipotez-5 ret edilmiş, diğer 4 hipotez ise kabul edilmiştir.

SONUÇ

Kredi kuruluşlarının, tasarruf sahiplerinin fonlarını ekonomik sisteme kazandırması, fonların ihtiyaç sahiplerince amacına uygun olarak etkin bir şekilde kullanımının sağlanması ve sonrasında fonları kullanan tarafların kredi kurumlarına geri ödemelerini tam ve zamanında yapmaları, ekonomik sistemlerin sürdürülebilirliğinin önemli bir unsurudur.

Dünya genelinde fonların büyük çoğunluğunun kredi kurumlarının kontrolü altında bulunduğu dikkate alındığında, kredi kullandıran yapıların, söz konusu fonları talep eden tarafları ve talep gerekçelerini oldukça iyi analiz etmeleri beklenmektedir. Yapılan bu analiz çalışmalarında gözden kaçırılan ya da yanlış değerlendirilen unsurlar kredi sisteminde temerrüt durumlarının oluşmasına, fonların kaybedilmesine, nihai olarak kaynakların israf edilmesine sebebiyet vermektedir. Söz konusu analiz faaliyetleri için gerek global ölçekte gerekse ülke bazında yaşanan krizlerden sonra, daha iyi analiz yapılabilmesi adına araştırmalar yapılmış, hali hazırda farklı branşlar için kullanılan metotların yanı sıra teknolojik gelişmelerle birlikte yeni metotların kullanılmasının önü de açılmıştır. Çalışmamızda söz konusu iki farklı yaklaşımın uyumlaştırılması olarak değerlendirilebilecek makine öğrenmesi algoritmalarından yararlanılarak, tarımsal kredilerin vadesinde ödenip ödenmeyeceğini tahmin edebilecek bir model oluşturulması amaçlanmıştır. Tarımsal kredi kullanan gerçek kişilerin, kredi tespit işlemlerinde kullanılan unsurlar modelin değişkenleri olarak belirlenmiş, yurdun farklı bölgelerinde faaliyet gösteren üreticilerin kredilerine esas olan veriler, WEKA yazılımı bünyesinde 7 farklı algoritmanın modelleme yapması için kullanılmıştır. Bu algoritmalarından ZeroR algoritması, diğer algoritmaların performansının değerlendirilebilmesi için eşik değer modeli ortaya koymuştur. Daha önce de açıklandığı üzere ZeroR algoritması veri setinde hakim olan sınıfı baz alarak tahmin modeli oluşturmakta, yani sınıflandırma kararı için kendisine sunulan yeni bir örneği, veri setinde hakim olan sınıfa dahilmiş gibi sınıflandırmaktadır. Bunun manası, eğer veri setinde bir sınıf çoğunluğu oluşturuyorsa, sınıflandırılmak istenen yeni örnek de bu sınıfa ait olmalıdır. Söz konusu yaklaşım oldukça basit ve rastgele olarak değerlendirilebilecek bir çalışma prensibi benimsediğinden, diğer algoritmaların başarılı sayılabilmesi için, rastgele tahmin yapan ZeroR algoritmasının elde etmiş olduğu doğru sınıflandırma oranından daha üst düzeyde bir doğru sınıflandırma oranı ortaya koymaları beklenmektedir. ZeroR algoritmasının ortaya koyduğu model,

%69,7987 oranında doğru sınıflandırma yapabilmıştır. Tezimizde kullandığımız doğrusal ayırma analizi ve lojistik regresyon algoritmaları, söz konusu oranın altında doğru sınıflandırma yapabilen modeller ortaya koymuşlardır. Çalışmamızda kullanılan k-en yakın komşu, naive bayes, karar ağacı ve yapay sinir ağı algoritmalarınca ortaya konulan modeller söz konusu eşik değerin üzerinde doğru sınıflandırma başarısı elde etmişlerdir. En yüksek doğru sınıflandırma oranı, %77,1812 ile karar ağacı algoritmasınınca ortaya konulan model tarafından elde edilmiştir. Ancak kredi kurumlarınca, daha riskli olarak nitelendirilen, vadesinde ödenmeyecek kredilerin doğru sınıflandırılması noktasında, doğrusal ayırma analizi algoritması daha yüksek düzeyde doğru sınıflandırma yapabilmıştır. Doğrusal ayırma analizi algoritmasının genel doğru sınıflandırma oranı %66,443 olsa da “TP NO”, “FP YES” ve “Recall” oranları göstermektedir ki gerçekte vadesinde ödenmeyen kredileri, vadesinde ödenmeyecek olarak sınıflandırmada bu algoritma tarafından geliştirilen model daha yüksek oranda başarı sağlamıştır. Diğer yandan modelin rastgele yapılacak bir tahminden ne denli farklı olduğunu belirleyen kappa istatistiği ve MCC değerlerinin doğrusal ayırma analizi için diğer algoritmalara kıyasla daha düşük olduğu görülmektedir.

Netice itibariyle, tarımsal kredi tespitinde kullanılan bazı değişkenler değerlendirilerek, farklı ve daha gelişmiş bir yaklaşım ortaya konulması amaçlanmıştır. Bu bağlamda %77,1812 oranında doğru sınıflandırma yapabilen bir model ortaya konulmuştur. Ancak, gerçekte vadesinde ödenmeyen kredileri daha yüksek düzeyde doğru sınıflandırılacak bir model geliştirilmesi için, veri setinin genişletilmesi ve daha fazla değişkenin kullanılması ile daha yüksek bir genel başarı oranı da elde edilebileceği öngörülmektedir. Tezimizde kullanılan algoritmaların, söz konusu ekleme ve değişimlerle birlikte yeniden çalıştırılması durumunda daha başarılı neticeler sunabilecek modeller geliştirilmesi mümkün olacaktır.

KAYNAKÇA

Adanacıođlu, H., Artukođlu, M. ve Guneş, E. (2017). “Türkiye’de Tarımsal Kredi Performansının Çok Boyutlu Ölçekleme Yaklaşımıyla Analizi”, Tarım Ekonomisi Dergisi, C:3, No:2, ss. 195-204.

Addo, P. M., Guegan, D. ve Hassani, B. (2018). “Credit Risk Analysis Using Machine and Deep Learning Models”, Risks, C:6, No:38, ss.1-20.

Akbay, C., Bilgiç, A. ve Miran, B. (2008). “Türkiye’de Önemli Gıda Ürünlerinin Talep Esneklikleri”, Tarım Ekonomisi Dergisi, C: 14, No: 2, ss. 55-65.

Akbulak, Y. (2012). “Kredi Derecelendirmesi veya Rating: Kavram ve Ölçütler”, Mali Çözüm Dergisi, Mayıs-Haziran 2012, ss. 171-184.

Akgüç, Ö. (2014). *Kredi Taleplerinin Deđerlendirilmesi*, 10. Baskı, Arayış Basım ve Yayıncılık, İstanbul.

Akman, G., Özkan, C. ve Eriş, H. (2008). “Strateji Odaklılık ve Firma Stratejilerinin Firma Performansına Etkisinin Analizi”, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, C: 7, No: 13, ss. 93-115.

Albright, J., (Çevrimiçi), <https://www.methodsconsultants.com/tutorial/what-is-the-difference-between-logit-and-probit-models/>, 16.07.2018.

Alpaydın, E. (2004). *Introduction to Machine Learning*, The MIT Press, England.

Anderson, D.R., Sweeney, D.J., Williams, T.A., Freeman, J. ve Shoemith, E. (2009). *Statistics for Business and Economics*, Seng Lee Press, Singapore.

Arvanitakis, K., Karydis, I., Kermanidis, K. L. ve Avlonitis, M. (2017). “A Machine Learning Approach for Asperitis’ Location”, *Evolving Systems*, ss. 1-10.

Balkaş, K. (2004). “Kredi Kavramı ve Sektör Kredilerine Göre Türkiye’deki Belli Başlı Sektörlerin Analizi”, Basılmamış Yüksek Lisans Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*, Cambridge University Press, Cambridge.

Barnes, J. (2015). *Azure Machine Learning: Microsoft Azure Essentials*, Microsoft Press, Washington.

Bankacılık Düzenleme ve Denetleme Kurulu, (2016). Bankaların Kredi Yönetimine İlişkin Rehber, Sayı:6827.

Bayar, Y. (2015). “Kredi Derecelendirme Kuruluşları ve Yunanistan Borç Krizi”, *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, C:8, No:15, ss. 51-57.

Bayramoğlu, Z., (Çevrimiçi), http://www.zmo.org.tr/genel/bizden_detay.php?kod=19245&tipi=24&sube=15, 30.07.2018.

Bektöre, S., Çömlekçi, F. ve Sözbilir H. (2013). *Mali Tablolar Analizi: Tekdüzen Hesap Planına Göre*, Nisan Kitapevi, Eskişehir.

Bell, J. (2015). *Machine Learning: Hands-on for Developers and Technical Professionals*, John Wiley & Sons Press, Indiana.

Berry, M. J. A. ve Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*, 2. Baskı, Wiley Publishing, USA.

Berry, M.W. ve Kogan, J. (2010). *Text Mining Applications and Theory*, Wiley Publishing, UK.

Billari, F.C., Fürnkranz, J. ve Prskawetz, A. (2006). “Timing, Sequencing and Quantum of Life Course Events: A Machine Learning Approach”, *European Journal of Population*, C:22, No:1, ss. 37-65.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer, Singapore.

Blangiardo, M. ve Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*, Wiley Publishing, UK.

Boschetti, A. ve Massaron, L. (2016). *Python Data Science Essentials*, 2.baskı, Packt Publishing, Birmingham.

Boughorbel, S., Jarray, F. ve El-Anbari, M., *Optimal Classifier for Imbalanced Data Using Matthews Correlation Coefficient Metric*, PLOS ONE 12(6), 2017, (Çevrimiçi), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5456046/>, 08.04.2018.

Buglear, J. (2014). *Practical Statistics: A Handbook for Business Projects*, Kogan Page, The UK.

Cardoso, D.O., Carvalho, D.S., Alves, D.S.F., Souza, D.F.P., Carneiro, H.C.C., Pedreira, C.E., Lima, P.M.V. ve França, F.M.G. (2016). “Financial Credit Analysis Via A Clustering Weightless Neural Classifier”, *Neurocomputing*, No:183, ss. 70-78.

Canepa, G. A. (2016). *What You Need To Know About Machine Learning*, Packt Publishing, Birmingham.

Caparrini, F.S., (Çevrimiçi), <http://www.cs.us.es/~fsancho/?e=135>, 25.01.2018.

Chai, T. ve Draxler, R. R. (2014). “Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) – Arguments Against Avoiding RMSE in the Literature”, *Geoscientific Model Development*, No:7, ss. 1247-1250.

Chen, P. C., Chang, J. B., Su, Y. L. ve Shieh, C. K. (2010). “On-Demand Data Co-Allocation with User-Level Cache for Grids”, *Concurrency and Computation: Practice and Experience*, C:22, ss. 2488-2513.

Chen, W., Xiang, G., Liu, Y. ve Wang, K. (2012). “Credit Risk Evaluation by Hybrid Data Mining Technique”, *Systems Engineering Procedia*, No:3, ss.194-200.

Cheng, C.H., Chen, Y.H. ve Liu, J.W. (2009). “Classifying Cinnamomums Using Rough Sets Classifier Based On Interval Discretization”, *Plant Systematics and Evolution*, C:280, No: 1/2, ss. 89-97.

Cichosz, P. (2015). *Data Mining Algorithms Explained R*, John Wiley & Sons Press, UK.

Coşkun, M. (2016). “Türkiye ile Aynı Kredi Notuna Sahip Ülkelerin Makro Ekonomik Göstergelerinin Karşılaştırılması”, *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 2016 Kasım Özel Sayısı, ss. 259-276.

Cotha, N. K. P. ve Sokolova, M. (2015). “Multi-Label Learning in Classification of Patients’ Quasi-Identifiers”, *Progress in Artificial Intelligence*, C:4, ss. 37-48.

Cox, J.A., (Çevrimiçi), <https://www.quora.com/What-is-the-unit-step-function-in-Artificial-Neural-Network>, 22.01.2018.

Cui, G., Wong, M. L. ve Lui, H. K. (2006). "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming", *Management Science*, C:52, No:4, ss. 597-612.

Çabuk, A. ve Lazol, İ. (2014). *Mali Tablolar Analizi*, 14. Baskı, Ekin Basın Yayın Dağıtım, Bursa.

Çizer, E. B. (2018). "Makine Öğrenmesi Teknikleriyle Kredi Risk Analizi", Basılmamış Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü.

Dayan, V. ve Karğın, S. (2013). "Basel II Düzenlemeleri Çerçevesinde Kullanılan Kredi Riski Modelleri: Karşılaştırmalı Bir Çalışma", *Journal of Yaşar University*, C:8, No:32, ss. 5433-5464.

Demir, M. (2014). "Kredi Derecelendirme Kuruluşları Derecelendirme Süreçleri ve Türkiye", Basılmamış Yüksek Lisans Tezi, Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü.

Demirci, S. ve Astar, M. (2011). "Türkiye'de Özel Sigortayı Etkileyen Faktörler: Logit Modeli", *Trakya Üniversitesi Sosyal Bilimler Dergisi*, C:13, No:2, ss. 119-130.

Derelioğlu, G. ve Gürgen, F. (2011). "Knowledge Discovering Using Neural Approach for SME's Credit Risk Analysis Problem in Turkey", *Expert Systems with Applications*, No:38, ss. 9313-9318.

Desterci, P. (2009). "Türkiye Bankacılık Sistemi Açısından Krediler ve Kredi Değerlendirmede Kullanılan Mali Analiz Yöntemleri", Basılmamış Yüksek Lisans Tezi, Namık Kemal Üniversitesi Fen Bilimleri Enstitüsü.

Dima, A. M. ve Vasilache, S. (2016). "Credit Risk Modeling For Companies Default Prediction Using Neural Networks", *Romanian Journal of Economic Forecasting*, C:19, No:3, ss. 127-143.

Duda, R.O., Hart, P.E. ve Stork, D.G. (2001). *Pattern Classification*, 2. Baskı, John Wiley & Sons Press, USA.

Eagleman, D. (2018). *Beyin: Senin Hikâyen*, 11. Baskı, çev: Zeynep Arık Tozar, Domingo-Bkz Yayıncılık, İstanbul.

Ekelund, S., *Precision-Recall Curves – What Are They and How Are They Used?*, *Acute Care Testing*, 2017, (Çevrimiçi), <https://acutecaretesting.org/en/articles/precision-recall-curves-what-are-they-and-how-are-they-used>, 09.04.2018.

Ertaş, F. C., Kaban, İ. ve Sobacı, F. (2016). “Bireysel Kredi Kullanan Finansal Tüketicilerce Üstlenilecek Masraflar; BDDK Düzenlemesi Çerçevesinde Karşılaştırmalı Bir Değerlendirme”, *Finansal Araştırmalar ve Çalışmalar Dergisi*, C:8, No:14, ss. 125-149.

Eşsiz, F. P. ve Karabulut, K. (2018). “Finansal İstikrarın Sağlanması Adına Makro İhtiyati Politikaların Kredi Büyümesi Üzerindeki Etkinliğinin Dinamik Panel Veri Yöntemiyle Ölçülmesi: 2000-2013”, *Atatürk Üniversitesi İİBF Dergisi*, C:32, No: 3, ss. 661-681.

Fan, G.Z., Ong, S.E. ve Koh, H.C. (2006). “Determinants of House Price: A Decision Tree Approach”, *Urban Studies*, C:43, No:12, ss. 2301-2315.

Feng, W., (Çevrimiçi), <http://web.utk.edu/~wfeng1/spark/fnn.html>, 25.01.2018.

Filiz, E. ve Öz, E. (2017). “Classification of BIST-100 Index’ Changes Via Machine Learning Methods”, *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Cilt:39, No:1, ss. 117-129.

Gaganis, C., Pasiouras, F., Spathis, C. ve Zopounidis, C. (2007). “A Comparison of Nearest Neighbors, Discriminant and Logit Models For Auditing Decisions”, *Intelligent Systems in Accounting, Finance and Management*, No:15, ss. 23-40.

Geçer, T. (2014). “Kredi İstihbaratı”, *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, Yıl:13, No:25, ss. 21-35.

Gerlein, E. A., McGinnity, M., Belatreche, A. ve Coleman, S. (2016). “Evaluating Machine Learning Classification for Financial Trading: An Empirical Approach”, *Expert Systems with Application*, No:54, ss. 193-207.

Ghodselahe, A. (2011). “A Hybrid Support Vector Machine Ensemble Model For Credit Scoring”, *International Journal of Computer Applications*, C:17, No: 5, ss. 1-5.

Girginer, N. (2008). "Ticari Kredi Taleplerinin Değerlendirilmesinde Çok Kriterli Yaklaşım: Özel ve Devlet Bankası Karşılaştırması", Muhasebe ve Finansman Dergisi, No:37, ss. 132-142.

Gorakala, S. K. ve Usulli, M. (2015). *Building a Recommendation System with R*, Packt Publishing, Birmingham.

Gurny, P. ve Gurny, M. (2013). "Comparing of Credit Scoring Models on Probability of Default Estimation for US Banks", No:2, ss. 163-181.

Gümrah, Ü. (2009). "Kredi Türevleri ve Gelişmekte Olan Ülkelerde Kredi Temerrüt Swapları Üzerine Bir Araştırma", Yayınlanmış Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İktisadi Araştırmalar Vakfı, İstanbul.

Günden, C. ve Miran, B. (2008). "Çiftçilerin Temel İşletmecilik Kararlarının Öncelik ve Destek Alma Açısından Analizi", Tekirdağ Ziraat Fakültesi Dergisi, C:5, No:2, ss. 67-80.

Gürsakar, N. (2014). *Büyük Veri*, 2. Baskı, Dora Yayıncılık, Bursa.

Hackeling, G. (2014). *Mastering Machine Learning with Scikit-Learn*, Packt Publishing, Birmingham.

Han, J. ve Kamber, M. (2006). *Data Mining Concepts And Techniques*, 2.baskı, Morgan Kaufmann Publishers, San Francisco.

Haspolat, F. B. (2015). "Ülke Kredi Notlarının Belirleyicileri: Türkiye'nin Kredi Notunun Ülke Karşılaştırmalı Analizi", Yayınlanmış Uzmanlık Tezi, T.C. Kalkınma Bakanlığı Ekonomik Modeller ve Stratejik Araştırmalar Genel Müdürlüğü, Yayın No:2923.

Hastie, T., Tibshirani, R. ve Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, Kanada.

Hayran, S. ve Gül, A. (2018). "Mersin İlinde Çiftçilerin Tarımsal Kredi Kullanım Kararlarını Etkileyen Faktörler", İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, C:8, No:1, ss. 271-277.

Horasan, M. ve Horasan, İ. (2012). "Kredilendirme Süreci ve Basel II Kriterleri İle Karşılaştırılması", Marmara Üniversitesi İİBF Dergisi, C:32, No:1, ss. 201-230.

Hussain, M. S., Calvo, R. A. ve Chen, F. (2014). "Automatic Cognitive Load Detection from Face, Physiology, Task Performance and Fusion During Affective Interference", *Interacting with Computers*, C:26, No: 3, ss. 256-268.

Idris, I. (2014). *Python Data Analysis*, Packt Publishing, Birmingham.

İskender, E. S. (2014). *Kredi Riski Dayanıklılığının Analizi: Türk Bankacılık Sektörü Üzerine Politika Önerileri*, Türkiye Bankalar Birliği Yayın No: 306, İstanbul.

James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer, USA.

Kaelbling, L. P., Littman, M. L. ve Moore, A. W. (1996). "Reinforcement Learning: A Survey", *Journal of Artificial Intelligence Research*, No: 4, ss. 237-285.

Kalaycı, S. (2018). "Makine Öğrenmesi Yöntemleri İle Kredi Risk Analizi", Basılmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Kanik, E. A., Erdoğan, S. ve Temel, G. O. (2012). "İki Sonuçlu Tanı Testlerinde İki Hekim Arasındaki Uyum İstatistiklerinin Prevalanstan Etkilenme Durumları", *İnönü Üniversitesi Tıp Fakültesi Dergisi*, C:19, No:3, ss. 153-158.

Kaptan, Ö. B. (2011). "Bireysel Kredilerin Risk ve Tüketici Davranışı Açısından Analizi", Basılmamış Doktora Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü.

Karacaoğlu, K. (2009). "Rekabet Üstünlüğünü Etkileyen Unsurların Yapısal Eşitlik Modeli İle Belirlenmesi: İSO 500 Büyük Sanayi İşletmesi Örneği", *Erciyes Üniversitesi İİBF Dergisi*, No: 34, ss. 165-187.

Kasapoğlu, B. (2009). "Kredi Riskinin Hesaplanmasında Skorlama Yaklaşımı", Basılmamış Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Kayalı, N. ve Çümen, A. (2011). "Basel II Sürecinde Kredi Derecelendirme ve Bir Tekstil İşletmesinde Uygulaması", *Yönetim ve Ekonomi Dergisi*, C:18, No:2, ss. 115-126.

Keskin, S. (2004). “2X2 Tablolarında Bazı Örnek Genişlikleri ve 1. Tip Hata Seviyeleri (α) İçin Kappa (K) İstatistiğine Ait Ampirik Olarak Gerçekleşen Kritik Değerler, Tarım Bilimleri Dergisi, C:10, No:2, ss. 169-173.

Kılıç, S. (2015). “Kappa Testi”, Journal of Mood Disorders, C:5, No:3, ss. 142-144.

Kliestik, T., Kocisova, K. ve Misankova, M. (2015). “Logit and Probit Model Used For Prediction of Financial Health of Company”, Procedia Economics and Finance, No:23, ss. 850-855.

Koçyiğit, S. Ç. ve Demir, A. (2014). “Türk Bankacılık Sektöründe Kredi Riski ve Yönetimine İlişkin Bir Uygulama: Türkiye Garanti Bankası Örneği”, İşletme Araştırmaları Dergisi, C:6, No:3, ss. 222-246.

Kokol, P., Pohorec, S., Stiglic, G. ve Podgorelec, V. (2012). “Evolutionary Design of Decision Trees for Medical Application”, WIREs Data Mining Knowledge Discovery, C:2, ss. 237-254.

Kök, D. ve Aksu, G. (2013). “Müşteri Kredi Değerliliğinin Belirlenmesinde Analitik Ağ Süreci Kullanımı: Bir Model Önerisi”, Muhasebe ve Finansman Dergisi, Temmuz 2013, ss. 167-185.

Kuhnert, P. ve Venables, B. (2005). *An Introduction to R: Software for Statistical Modelling & Computing*, CSIRO, Australia.

Lantz, B. (2013). *Machine Learning With R*, Packt Publishing, Birmingham.

Lavesson, N., Boldt, M., Davidsson, P. ve Jacobsson, A. (2011). “Learning to Detect Spyware Using End User Licence Agreements”, Knowledge Information Systems, C:26, ss. 285-307.

Ledolter, J. (2013). *Data Mining And Business Analytics With R*, John Wiley & Sons Press, New Jersey.

Liu, Y., Jiang, Z.L., Yao, L., Wang, X., Yiu, S.M. ve Huang, Z. (2017). “Outsourced Privacy-Preserving C4.5 Decision Tree Algorithm Over Horizontally and Vertically Partitioned Dataset Among Multiple Parties”, Cluster Computing, ss. 1-13.

Maindonald, J. H. (2008). Using R for Data Analysis and Graphics: Introduction, Code and Commentary, (Çevrimiçi), <https://cran.r-project.org/doc/contrib/usingR.pdf>, 24.02.2018.

Maslove, D.M., Podchiyska, T. Ve Lowe, H.J. (2013). “Discretization of Continuous Features in Clinical Datasets”, Journal of the American Medical Informatics Association, C:20, No:3, ss. 544-553.

McClure, N. (2017). *TensorFlow Machine Learning Cook Book*, Packt Publishing, Birmingham.

Meenakshi, S. ve Venkatachalam, V. (2015). “FUDDT: A Fuzzy Uncertain Decision Tree Algorithm for Classification of Uncertain Data”, Arabian Journal For Science and Engineering, C:40, No:11, ss. 3187-3196.

Mesaric, J. ve Sebalj, D. (2016). “Decision Trees for Predicting The Academic Success of Students”, Croatian Operational Research Review, C:7, ss.367-388.

Murgia, A., Ortu, M., Tourani, P., Adams, B. ve Demeyer, S. (2018). “An Exploratory Qualitative and Quantitative Analysis of Emotions in Issue Report Comments of Open Source Systems”, Empirical Software Engineering, C:23, ss. 521-564.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*, MIT Press, London.

Nabiyev, V.V. (2003). *Yapay Zekâ: Problemler-Yöntemler-Algoritmalar*, Seçkin Yayıncılık, Ankara.

Nahayo, Y. ve Arı, S. (2016). “Performance of SVM, K-NN and NBC Classifiers for Text-Independent Speaker Identification With and Without Modelling Through Merging Models”, Sakarya Üniversitesi Fen Bilimleri Dergisi, C:20, No:1, ss. 1-6.

Newbold, P. (2008). *İşletme ve İktisat İçin İstatistik*, çev: Ümit Şenesen, Literatür Yayınları, İstanbul.

Nguyen, K., Tran, D., Ma, W. ve Sharma, D. (2017). “Decision Tree Algorithms for Image Data Type Identification”, Logic Journal of the IGPL, C:25, No:1, ss. 67-82.

Oktay, S. ve Temel, H. (2007). “Basel II Kriterleri Ekseninde Ticari Bankalarda Kredi Riski Yönetiminin Karşılaştırılmasına Yönelik Bir Saha Çalışması”, ZKÜ Sosyal Bilimler Dergisi, C:3, No:6, ss. 163-185.

Onan, A. (2015). “Şirket İflaslarının Tahmin Edilmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi”, *Bilişim Teknolojileri Dergisi*, C:8, No:1, ss. 9-19.

Oral, N. (2015). “Ticari Kredilerde Çok Kriterli Risk Analizi: Bir Kamu Bankasında Uygulama”, *Basılmamış Yüksek Lisans Tezi*, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Orloff, J. ve Bloom, J., (Çevrimiçi), https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading7b.pdf, 21.02.2018.

Ovalı, S. (2014). “Ülke Kredi Notu Değerlendirme Kriterleri Açısından Türkiye: AB İle Karşılaştırmalı Analiz”, *Uluslararası Yönetim İktisat ve İşletme Dergisi*, C:10, No: 23, ss. 53-80.

Özden, K. Ö. (2010). “Ticari Bankalarda Kredilendirme Süreci ve Kredi Riski Yönetimi”, *Basılmamış Yüksek Lisans Tezi*, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Patil, T. R. ve Sherekar, S. S. (2013). “Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification”, *International Journal of Computer Science and Applications*, C:6, No:2, ss. 256-261.

Peprah, W.K., Agyei, A. ve Oteng, E. (2017). “Ranking the 5C’s of Credit Analysis: Evidence From Ghana Banking Industry”, *International Journal of Innovative Research and Advanced Studies*, C:4, No:9, ss. 78-80.

Poyraz, O. (2010). “Kredi Talebinde Bulunan İşletmelerin Finansal Tablolarında Yapılan Düzeltme ve Arındırma İşlemleri Üzerine Bir İnceleme”, *Basılmamış Yüksek Lisans Tezi*, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Prajapati, V. (2013). *Big Data Analysis with R and Hadoop*, Packt Publishing, Birmingham.

Punniyamorthy, M. ve Sridevi, P. (2016). “Identification of a Standard AI Based Technique for Credit Risk Analysis”, *Benchmarking: An International Journal*, C:23, No: 5, ss. 1381-1390.

Raschka, S. (2016). *Python Machine Learning*, Packt Publishing, Birmingham.

Raschka, S., (Çevrimiçi), http://sebastianraschka.com/Articles/2014_python_lda.html, 21.02.2018.

Reis, G. ve Kötüoğlu, R. (2016). “Türk Bankacılık Sektörünün Sermaye Yeterliliği Davranışı”, *Yönetim ve Ekonomi Araştırmaları Dergisi*, C: 14, No: 3, ss. 101-110.

Rojas, R., (Çevrimiçi), http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/secretcovariance.pdf, 21.02.2018.

Rosebrock, A., (Çevrimiçi), <https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras/>, 25.01.2018.

Russell, S. ve Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, 2. Baskı, Prentice Hall, USA.

Saito, T. ve Rehmsmeier, M., *The Precision-Recall Plot is More Informative Than The ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets*, PLOS ONE 10(3), 2015, (Çevrimiçi), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4349800/>, 09.04.2018.

Schmidt, A., (Çevrimiçi), <https://www.teco.edu/~albrecht/neuro/html/node16.html>, 30.01.2018.

Selimler, H. (2015). “Sorunlu Kredilerin Analizi, Banka Finansal Tablo ve Oranları Etkisinin Değerlendirilmesi”, *Finansal Araştırmalar ve Çalışmalar Dergisi*, C:7, No:12, ss. 131-172.

Seyitoğlu, F., (Çevrimiçi), https://prezi.com/al_xpwb8mipl/probit-regresyon-analizi/, 17.07.2018.

Sharma, S., (Çevrimiçi), <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>, 22.01.2018.

Shwartz, S. S. ve David, S. B. (2014). *Understanding Machine Learning From Theory To Algorithms*, Cambridge University Press, New York.

Soydaner, D. ve Kocadağlı, O. (2015). “Artificial Neural Network with Gradient Learning Algorithm for Credit Scoring”, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, C:44, No:2, ss. 3-12.

Sönmez, F. (2015). “Kredi Skorunun Belirlenmesinde Yapay Sinir Ağları ve Karar Ağaçlarının Kullanımı: Bir Model Önerisi”, ABMYO Dergisi, No:40, ss. 1-22.

Şeker, Ş. E. (2013). *İş Zekası Ve Veri Madenciliği: Weka İle*, Cinius Yayınları, İstanbul.

Şeker, Ş. E. (2016). *Weka ile Veri Madenciliği*, Bilgisayar Kavramları Yayınları, İstanbul.

Şeker, Ş. E. ve Ocak, İ. (2017). “Performance Prediction of Roadheaders Using Ensemble Machine Learning Techniques”, Neural Computing and Applications, ss. 1-14.

Şimşek, K. Ç. (2007). “Bankacılıkta Risk ve Risk Ölçüm Yöntemleri”, Basılmamış Yüksek Lisans Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü.

Ştefea, P. ve Pelin, A. (2016). “Contribution to Improvement of Credit Analysis Techniques”, *Lucrări Ştiinţifice*, C:18, No:2, ss. 247-252.

Tanç, Ş. G. ve Altun, Ş. (2016). “Bankacılıkta Risk ve HSBC Bank Uygulama Örneği”, Selçuk Üniversitesi Sosyal Bilimler MYO Dergisi, C:19, 41. Yıl Özel Sayısı, ss. 225-242.

Tattar, P. N. (2013). *R Statistical Application Development by Example*, Packt Publishing, Birmingham.

TBB Çalışma Grubu (2006). “Kredi Riski Modelleri”, Bankacılar Dergisi, No: 57, ss. 33-66.

Tokel, Ö. E. (2004). “Kredi Risk Modelleri Kullanılarak Kredi Taleplerinin Değerlendirilmesi”, Basılmamış Yüksek Lisans Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü.

Toraman, C. ve Yürük, M. F. (2014). “Kredi Derecelendirme Kuruluşları ve Finansal Krizlere Etkileri”, Bitlis Eren Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, C:3, No:1, ss. 127-154.

Torres, J. (2016). *First Contact with TensorFlow*, Jordi Torres BSC-CNS, Barcelona.

Tuffery, S. (2011). *Data Mining and Statistics for Decision Making*, John Wiley & Sons Press, UK.

Tuna, K., (Çevrimiçi), <http://www.halkbankkobi.com.tr/NewsDetail/Kredili-Satislarda-Musteri-Riskinin-Belirlenmesi/5388>, 18.06.2018.

Tunay, K. B. (2016). “Makro Finansal Bağlantılar ve Kredi Portföyü Kalitesine Etkileri”, *Yönetim ve Ekonomi Araştırmaları Dergisi*, C: 14, No: 4, ss. 25-44.

Tunç, A. (2016). “Finans Sektörü İçin Yapay Öğrenme Teknikleri Kullanarak Kredi Kullanabilirliğin Tespiti”, *Basılmamış Yüksek Lisans Tezi*, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Tunç, B. (2012). “Kredi Hacmini Etkileyen Faktörler: Banka Kredileri Eğilim Anketi Analizi”, *Uzmanlık Yeterlilik Tezi*, Türkiye Cumhuriyeti Merkez Bankası İstatistik Genel Müdürlüğü.

Türe, E. (2015). “Bankalarda Kredilerin Yönetimi ve Sorunlu Kredilerin Yeniden Yapılandırılması Süreci”, *Basılmamış Yüksek Lisans Tezi*, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü.

Ünlüer, M. ve Güneş, E. (2013). “Tarımsal Kredilerin Geri Ödenmesinde Etkili Faktörlerin Analizi”, *Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi*, C:30, No:2, ss.86-93.

Vaidya, D., (Çevrimiçi), <https://www.wallstreetmojo.com/credit-analysis/>, 14.06.2018.

Vehid, H. E. ve Eral, G. (2014). “Nitel Verilerin Değerlendirilmesinde Uygulanan İstatistiksel Yöntemler”, *Çocuk Dergisi*, C:14, No:2, ss.60-61.

Vukovic, S., Delibasic, B., Uzelac, A. ve Suknovic, M. (2012). “A Case-Based Reasoning Model That Uses Preference Theory Functions for Credit Scoring”, *Expert Systems with Applications*, No:39, ss. 8389-8395.

Vurucu, M. ve Arı, M. U. (2017). *Güncel Gelişmeler Işığında Bankacılıkta Kredi Bilgisi*, Seçkin Yayıncılık, Ankara.

Wang, H., Wang, Y., Fang, J., Chai, H. ve Zheng, H. (2012). "Simulation Research on a Minimum RMSE Rotation-Fitting Algorithm for Gravity Matching Navigation", *Science China Earth Sciences*, C:55, No:1, ss. 90-97.

Webb, A. (2002). *Statistical Pattern Recognition*, 2. Baskı, John Wiley & Sons Press, England.

Williams, G. (2011). *Data Mining with Rattle and R*, Springer Science+Business Media, New York, USA.

Willmott, C. J. ve Matsuura, K. (2005). "Advantages of the Mean Absolute Error (MAE) Over The Root Mean Square Error (RMSE) in Assessing Average Model Performance", *Climate Research*, C:30, No:1, ss. 79-82.

Witten, I.H., Frank, E. ve Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, 3. Baskı, Morgan Kaufman Publications, USA.

Wong, M. L. ve Leung, K. S. (2002). *Data Mining Using Grammar Based Genetic Programming and Applications*, Kluwer Academic Publishers, USA.

Yang, Y. ve Webb, G.I. (2001). "Proportional k-Interval Discretization for Naive-Bayes Classifiers", *European Conference on Machine Learning, Lecture Notes on Computer Science*, No: 2167, ss. 564-575.

Yıldırım, E. (2007). "Bankalarda Ticari Kredilendirme Süreci: Karşılaşılan Sorunlar ve Çözüm Yolları İle İlgili Ampirik Bir Çalışma", *Basılmamış Yüksek Lisans Tezi*, Selçuk Üniversitesi Sosyal Bilimler Enstitüsü.

Yıldız, B. (2009). *Finansal Analizde Yapay Zeka*, Beta Yayınevi, Eskişehir.

Yiğitbaş, Ş. B. (2012). *Bankaların Kredi Verme Davranışı Üzerine Asimetrik Bilginin Etkisi ve Reel Sektör Yansıması (Türkiye Analizi 2002-2010)*, Türkiye Bankalar Birliği Yayınları, Yayın No:288, İstanbul.

Yurtoğlu, B. İ. (2015). "Ürün Senedinin Tarım Sektörünün Finansmanındaki Rolü", *Basılmamış Yüksek Lisans Tezi*, Başkent Üniversitesi Sosyal Bilimler Enstitüsü.

Yüksel, Ö. (2011). "Makroekonomik Değişkenlere Dayalı Kredi Riski Modellemesi ve Stres Testi Analizi", *Basılmamış Yüksek Lisans Tezi*, TOBB Ekonomi ve Teknoloji Üniversitesi Sosyal Bilimler Enstitüsü.

Zaccone, G. (2016). *Getting Started With TensorFlow*, Packt Publishing, Birmingham.

Zaccone, G., Karim, M.R. ve Menshawy, A. (2017). *Deep Learning with TensorFlow*, Packt Publishing, Birmingham.

Zafarani, R., Abbasi, M. A. ve Liu, H. (2014). *Social Media Mining: An Introduction*, Cambridge University Press, New York.

Zhao, Y. (2013). *R and Data Mining: Examples and Case Studies*, Elsevier, USA.

Zocca, V., Spacagna, G., Slater, D. ve Roelants, P. (2017). *Python Deep Learning*, Packt Publishing, Birmingham.

Biyoloji Dersleri, (Çevrimiçi), <http://www.biyolojidersnotlari.com/sinir-hucresi-yapisi>, 21.01.2018.

Classeval, (Çevrimiçi), <https://classeval.wordpress.com/introduction/introduction-to-the-precision-recall-plot/>, 09.04.2018.

IBM Knowledge Center, (Çevrimiçi), https://www.ibm.com/support/knowledgecenter/tr/SSEP7J_10.2.0/com.ibm.swg.ba.cognos.ug_cr_rptstd.10.2.0.doc/c_id_obj_desc_tables.html, 08.04.2018.

Gepsoft, (Çevrimiçi), <https://www.gepsoft.com/gxpt4kb/Chapter10/Section1/SS07.htm>, 05.04.2018.

Kredi Kayıt Bürosu, (Çevrimiçi), <https://www.kkb.com.tr/urunler/bireysel-kredi-notu>, 31.07.2018.

OpenML, (Çevrimiçi), <https://www.openml.org/a/estimation-procedures/1>, 14.04.2018.

Saedsayad, (Çevrimiçi), http://www.saedsayad.com/model_evaluation_r.htm, 04.04.2018.

Scikit Learn, (Çevrimiçi), http://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html, 08.04.2018.

Scikit Learn, (Çevrimiçi), http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html, 09.04.2018.

Statistics How To, (Çevrimiçi), <http://www.statisticshowto.com/receiver-operating-characteristic-roc-curve/>, 09.04.2018.

University of Waikato, (Çevrimiçi), <https://www.cs.waikato.ac.nz/ml/weka/index.html>, 25.03.2018.

Veri Bilimcisi, (Çevrimiçi), <https://veribilimcisi.com/2017/07/14/mse-rmse-mae-mape-metrikleri-nedir/>, 04.04.2018.



EKLER

Ek-1: Çalışma Evrenindeki Tarım Kredi Kooperatifleri

<i>Sıra No</i>	<i>Kooperatifin Numarası</i>	<i>Kooperatifin Adı</i>
1	15	Bor T.K.K.
2	268	Simav T.K.K.
3	501	Karacabey T.K.K.
4	590	Yeniçağa T.K.K.
5	625	Harmanlı T.K.K.
6	763	Yenikaraağaç T.K.K.
7	846	Taşagıl T.K.K.
8	903	Balıköy T.K.K.
9	1088	Yeniçiftlik T.K.K.
10	1362	İğdir T.K.K.
11	1367	Hisarcık T.K.K.
12	1400	Karahüyük T.K.K.
13	184	Bayburt T.K.K.
14	2667	Kelkit T.K.K.
15	485	Perşembe T.K.K.
16	2052	Refahiye T.K.K.
17	1032	Bafra T.K.K.
18	683	Atatürk Tekirçiftliği T.K.K.
19	2436	Bismil T.K.K.

Ek-2: Örnek Ortak Beyannamesi

ORTAK BEYANNAMESİ

K120184 sayılı BAYBURT Tarım Kredi Kooperatifi
Yönetim Kuruluna,

Aşağıda gösterilen tarımsal varlığıma göre ihtiyacım olan
Yıllık 6000 Toplam 6.000,00 TL kredi açılmasını arz ederim

2037 Nolu Ortak

Beyanname No :418
Ortak No :2037
TC./Vergi Kimlik :00000000000
Adı/Unvanı :XXXX
Soyadı :YYYYY

Vekil Ad Soyadı :
Baba Adı :VVVV
Köy Adı :MMMM KÖYÜ

TARIMSAL ÜRETİM YAPTIĞIM ARASILAR (Tablo I)								
(1) Mevkii	(2) Nevi	(3) Belge Türü	(4) Belge No	(5) Belge Tarihi	(6) Arasının Metrekaresi	(7) Dekar Değeri (TL)	(8) Arasının Tüm Değeri	(9) Arasının Sahibi
KIRATLI	TARLA KURU	EMLAK KA	0	16.10.2016	32.055,00	0,40	12.822,00	KENDİSİ
Toplam:							12.822,00	

HAYVANLARIM VE DEĞERLERİ (Tablo VI)				
(1) Mevcut/ Kredili Talep Edilen	(2) Cinsi	(3) Adet veya Kovan Sayısı	(4) Bir Adedinin Değeri TL	(5) Tümüünün Değeri TL
Toplam:				

ZİRAAT ALET VE MAKİNALARIM (Tablo IV)			
(1) Mevcut/Kredili Talep Edilen	(2) Nevi	(3) Sayısı	(4) Değeri
Toplam:			

BANKA TEMİNAT MEKTUBU (Tablo III)	
(1) Nevi	(2) Rayic/Hasko Değeri
Toplam :	

YETİŞTİRDİĞİM ÜRÜNLER VE DEĞERLERİ (Tablo II)					
(1) Yetiştireceğim Ürünlerin Cinsi	(2) Üretim Yaptığım Arasının	(3) Metrekaresin den Alınacak Ürün Miktarı	(4) Alınacak Tüm Ürün Miktarı	(5) Bir Kg'nın Fiyatı	(6) Ürünlerin Tümüünün Değeri
KORUNGA	32055	0,55	17.630,25	0,65	11.459,66
Toplam:					11.459,66

BORÇLARIM (Tablo V)	
(1) Alacaklı	(2) Borç Miktarı
Toplam: 0,00	

HAYVANSAL ÜRÜNLERİM VE DEĞERLERİ (Tablo VII)					
(1) Ürün Cinsi	(2) Adet veya Kovan Sayısı	(3) Bir Adedinde Alınan Ürün Miktarı Kg	(4) Tümüünün Miktarı Kg/Adet	(5) Bir Kg'nın veya Adedinin Değeri TL	(6) Ürünler'in Tüm Değeri TL
Toplam:					

Bu beyannamede adı ve soyadı yazılı üreticinin Köyünde/Mahallesinde oturduğunu ve beyannamede gösterdiği bilgilerin doğruluğunu onaylarım. 28.3.2016

Köy/Mahallesi İhtiyar Kurulu

KOOPERATİF YÖNETİM KURULUNCA TESBİT EDİLECEK KREDİ AÇIKLAMA		Tablo VIII TUTAR
1	Arası, sırası alet-makina ve hayvanların değerlerinin toplamı (II/8.kolon+IV/4.kolon+VI/8.kolon)	12.822,00
2	Tarımsal varlık harici ipotek ve rehin değerinin toplamının %50'si+ (Banka Teminat Mektubunun %75'i) (III/2.KOLON)	0,00
3	Kurum dışı borçları (V/2.kolon)	0,00
4	Kalan (1.ve 2.satırdaki toplam ile 3.satırdaki toplamın farkı)	12.822,00
5	Tarımsal ürünleriyle hayvan gelirlerinin toplamı (II/6.kolon+VII/6.kolon)	11.459,66
6	Tarımsal ürünleriyle hayvan gelirleri toplamının %60 si	6.875,80
7	Müdürlükçe teklif edilen ve Yönetim Kurulu Kararı ile tesbit olunan Toplam kredi limiti	6.000,00
8	Müdürlükçe teklif edilen ve Yönetim Kurulu Kararı ile tespit olunan Yıllık Kredi limiti	6.000,00

Ek-3: Örnek Ortak Kartı



K120184 SAYILI BAYBURT KOOPERATİFİ
2037 NO'LU XXXX YYYYYY ORTAK KARTI

Tarih : GG.A.YY
Saat : 00:00

ORTAK BİLGİLERİ	
Adı Soyadı : XXXX YYYYYY	Bakmakla Yükümlü Olduğu Kiş
Kimlik No : 00000000000	Sayısı :
Baba Adı : -----	Ev Telefonu : 0000000000
Doğum Yeri/Tarihi : BAYBURT GG/AA/YYYY	Cep Telefonu : 0000000000
Medeni Hali : EVLİ	Ortaklık Giriş Tarihi : 19/10/2016
Cinsiyeti : ERKEK	Ortaklık Durumu : KAYITLI
Çocuk Sayısı :	Adresi : Bayburt

TESPİT EDİLEN KREDİLER		SERMAYE BİLGİLERİ		BEYANNAME BİLGİLERİ	
İşlem Tarihi	Kredi	Sermaye Taahhütü		Beyanname No	
19/10/2016	6.000,00	Sermaye Tahsilatı	300,00	Beyanname Tarihi	418
		Vadesi Geçen Borcu	268,00	Kredi Hacmi	19/10/2016
		Kalan Kredi Limiti	0,00	Kredi Limiti	6.876,00
			3.675,00		6.000,00

SÖZLEŞME BİLGİLERİ		KEFİL OLUNAN SÖZLEŞMELER				
Sözleşme No	1298	Ort.No	Söz.No	Ad Soyad	Tc. Kimlik	Söz.Tutarı
Sözleşme Durumu	ONAYLI	2075	1393	HHHH SSSSSS	0000000000	30.000,00
Sözleşme Tarihi	19/10/2016					
Sözleşme Tutarı	12.000,00					

SÖZLEŞME KEFİLLERİ		
Ort.No	Ad Soyad	Tc. Kimlik
1500	NNN CCCCC	0000000000
1682	BBB MMMMM	0000000000

YILLARA GÖRE KULLANILAN KREDİ DAĞILIM TABLOSU			
	Senet	2016	2017
MEKANİZASYON KREDİLERİ	Sayısı Tutarı	0 0,00	2 2.000,00
TÜKETİM KREDİSİ	Sayısı Tutarı	1 1.890,00	1 1.425,00

SENET DURUMLARI		
Son Durum	Adet	Bakiye
NORMAL TAKİP	3	5.976,15
ÖDENDİ	1	2.187,79
İPTAL	0	0,00
TOPLAM TAHSİLAT	1	2.187,79
SON TAHSİLAT TARİHİ :		GG/AA/YY

AÇIK SENETLER							
Sıra No	Eski Senet No	Yeni Senet No	Düzenleme Tarihi	Vade Tarihi	Senet Tutarı	Bakiye	Son Durum
1		18040-1/1	18/07/2017	05/07/2018	900,00	2.199,14	NORMAL TAKİP
2		18040-1/2	18/07/2017	05/07/2019	1.100,00	2.200,00	NORMAL TAKİP
3		18448-1	11/10/2017	01/10/2018	1.425,00	1.577,01	NORMAL TAKİP
Toplam :					3.425,00	5.976,15	

Ek-4: ZeroR Algoritması Detaylı Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.Remove-R1,49-
weka.filters.unsupervised.attribute.StringToNominal-R12-15

Instances: 437

Attributes: 48

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut

buzagi_ikinciurun
 kukukbas
 kukuk_deger
 kukukbas_uretim_et
 kukukbas_uretim_fiyat_et
 kukukbas_uretim_sut
 kukukbas_uretim_fiyat_sut
 kuzu_ikinciurun
 kovan_sayi
 kovan_deger
 kovan_uretim
 kovan_uretim_fiyat
 alet_ekipman
 teminat
 disari_borclar
 kredi
 durum2

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

ZeroR predicts class value: YES

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	104	69.7987 %
Kappa statistic	0	
Mean absolute error	0.4522	
Root mean squared error	0.4656	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	149	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.302	NO
	1.000	1.000	0.698	1.000	0.822	0.000	0.500	0.698	YES
W. Avg.	0.698	0.698	0.487	0.698	0.574	0.000	0.500	0.578	

=== Confusion Matrix ===

a b <-- classified as

0 45 | a = NO

0 104 | b = YES



Ek-5: K En Yakın Komşu Algoritması Detaylı Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 11 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
\"weka.core.EuclideanDistance -R first-last\""

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.StringToNominal-R12-15-
weka.filters.unsupervised.attribute.Remove-R48

Instances: 437

Attributes: 48

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut

buyukbas_uretim_fiyat_sut
 buzagi_ikinciurun
 kucukbas
 kucuk_deger
 kucukbas_uretim_et
 kucukbas_uretim_fiyat_et
 kucukbas_uretim_sut
 kucukbas_uretim_fiyat_sut
 kuzu_ikinciurun
 kovan_sayi
 kovan_deger
 kovan_uretim
 kovan_uretim_fiyat
 alet_ekipman
 teminat
 disari_borclar
 kredi
 durum2

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier

using 11 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.02 seconds

=== Summary ===

Correctly Classified Instances	111	74.4966 %
Kappa statistic	0.3951	
Mean absolute error	0.382	
Root mean squared error	0.4385	
Relative absolute error	84.4773 %	
Root relative squared error	94.1718 %	
Total Number of Instances	149	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.578	0.183	0.578	0.578	0.578	0.395	0.707	0.544	NO
	0.817	0.422	0.817	0.817	0.817	0.395	0.707	0.809	YES
W. Avg.	0.745	0.350	0.745	0.745	0.745	0.395	0.707	0.729	

=== Confusion Matrix ===

a b <-- classified as

26 19 | a = NO

19 85 | b = YES



Ek-6: Naive Bayes Algoritması Detaylı Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes -D

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.StringToNominal-Rlast-
weka.filters.unsupervised.attribute.StringToNominal-Rlast-
weka.filters.unsupervised.attribute.StringToNominal-R13-16-
weka.filters.unsupervised.attribute.Remove-R49-weka.filters.unsupervised.attribute.Normalize-S1.0-
T0.0-weka.filters.unsupervised.attribute.Remove-R1

Instances: 437

Attributes: 48

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et

buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut
buzagi_ikinciurun
kucukbas
kucuk_deger
kucukbas_uretim_et
kucukbas_uretim_fiyat_et
kucukbas_uretim_sut
kucukbas_uretim_fiyat_sut
kuzu_ikinciurun
kovan_sayi
kovan_deger
kovan_uretim
kovan_uretim_fiyat
alet_ekipman
teminat
disari_borclar
kredi
durum2

Test mode: split 66.0% train, remainder test
=== Classifier model (full training set) ===

Naive Bayes Classifier

	Class	
Attribute	NO	YES
	(0.35)	(0.65)

=====

cinsiyet

'All'	155.0	284.0
[total]	155.0	284.0

nufuskayit_yeri

'All'	155.0	284.0
[total]	155.0	284.0

egitim_durumu

'All'	155.0	284.0
[total]	155.0	284.0

medeni_hali

'All'	155.0	284.0
[total]	155.0	284.0

yas<23		
'All'	155.0	284.0
[total]	155.0	284.0
yas>65		
'All'	155.0	284.0
[total]	155.0	284.0
kendi_arazisi		
'All'	155.0	284.0
[total]	155.0	284.0
dekar_degeri		
'All'	155.0	284.0
[total]	155.0	284.0
icar		
'All'	155.0	284.0
[total]	155.0	284.0
uretim_yapilan		
'(-inf-0.001324]'	25.0	107.0
'(0.001324-inf)'	131.0	178.0
[total]	156.0	285.0
arazi_urun1		
pancar	2.0	2.0
domates	35.0	47.0
0	25.0	104.0
patates	4.0	4.0
yonca	1.0	5.0
bugdaysulu	14.0	33.0
bugdaykuru	14.0	20.0
sekerpancari	1.0	4.0
domatessera	2.0	1.0
kiraz	3.0	3.0
arpa	2.0	1.0
figkuru	2.0	1.0
sebze	7.0	5.0
misir	6.0	20.0
zeytin	1.0	5.0
misirsilaji	3.0	4.0
marul	1.0	2.0
karpuz	1.0	3.0

bezelye	1.0	2.0
kirmizibiber	3.0	2.0
incir	1.0	2.0
misir	1.0	2.0
seftali	1.0	2.0
biber	2.0	2.0
aycicegi	8.0	7.0
kurusogan	2.0	1.0
arpasulu	2.0	2.0
sarimsak	1.0	2.0
celtik	4.0	8.0
yesilsogan	1.0	2.0
kayisi	2.0	2.0
korunga	1.0	2.0
arpakuru	1.0	2.0
kurufasulye	2.0	1.0
findik	28.0	5.0
kavun	2.0	1.0
tutun	3.0	1.0
figsulu	2.0	2.0
limon	1.0	2.0
pamuk	1.0	7.0
[total]	194.0	323.0

arazi_urun2

0	101.0	183.0
misirsilaji	6.0	7.0
sekerpancari	1.0	4.0
domates	5.0	19.0
bugdaysulu	10.0	16.0
kiraz	3.0	2.0
arpakuru	2.0	1.0
cavdar	2.0	1.0
misir	18.0	24.0
arpa	2.0	1.0
bugday	2.0	1.0
sebze	7.0	2.0
bezelye	1.0	2.0
kirmizibiber	1.0	2.0

yonca	2.0	6.0
karpuz	1.0	4.0
zeytin	1.0	6.0
ceviz	1.0	2.0
bugdaykuru	3.0	3.0
aycicegi	1.0	6.0
kavun	1.0	2.0
kayisi	1.0	4.0
figsulu	1.0	4.0
lahana	2.0	1.0
karnabahar	2.0	2.0
figkuru	2.0	1.0
celtik	2.0	3.0
mercimek	1.0	2.0
[total]	182.0	311.0
arazi_urun3		
0	131.0	245.0
kavun	1.0	2.0
sekerpancari	1.0	2.0
arpasulu	1.0	2.0
cavdar	2.0	1.0
mercimek	2.0	2.0
arpakuru	2.0	1.0
sebze	3.0	1.0
visne	2.0	1.0
figsulu	2.0	1.0
domates	6.0	6.0
biber	2.0	2.0
misir	4.0	7.0
bugdaysulu	3.0	6.0
bezelye	1.0	3.0
zeytin	1.0	3.0
aycicegi	3.0	10.0
misirsilaji	4.0	2.0
figkuru	1.0	2.0
korunga	1.0	2.0
lahana	2.0	2.0
nar	1.0	2.0

[total]	176.0	305.0
arazi_urun4		
0	151.0	272.0
yonca	1.0	2.0
domates	2.0	3.0
bezelye	2.0	1.0
bugdaysulu	1.0	3.0
misir	1.0	4.0
aycicegi	2.0	1.0
sekerpancari	1.0	2.0
lahanabeyaz	2.0	1.0
figkuru	1.0	2.0
cilek	1.0	2.0
pamuk	1.0	2.0
[total]	166.0	295.0
arazi_urun5		
0	155.0	283.0
bezelye	1.0	2.0
[total]	156.0	285.0
dekar_urun1		
'(-inf-0.003]'	26.0	104.0
'(0.003-0.0145]'	39.0	8.0
'(0.0145-inf)'	92.0	174.0
[total]	157.0	286.0
dekar_urun2		
'All'	155.0	284.0
[total]	155.0	284.0
dekar_urun3		
'All'	155.0	284.0
[total]	155.0	284.0
dekar_urun4		
'All'	155.0	284.0
[total]	155.0	284.0
dekar_urun5		
'All'	155.0	284.0
[total]	155.0	284.0
arazi_urun_fiyat1		

'(-inf-0.403846]' 127.0 280.0
'(0.403846-inf)' 29.0 5.0
[total] 156.0 285.0
arazi_urun_fiyat2
'All' 155.0 284.0
[total] 155.0 284.0
arazi_urun_fiyat3
'All' 155.0 284.0
[total] 155.0 284.0
arazi_urun_fiyat4
'All' 155.0 284.0
[total] 155.0 284.0
arazi_urun_fiyat5
'All' 155.0 284.0
[total] 155.0 284.0
buyukbas
'All' 155.0 284.0
[total] 155.0 284.0
buyuk_deger
'All' 155.0 284.0
[total] 155.0 284.0
buyukbas_uretim_et
'All' 155.0 284.0
[total] 155.0 284.0
buyukbas_uretim_fiyat_et
'(-inf-0.022894]' 152.0 242.0
'(0.022894-inf)' 4.0 43.0
[total] 156.0 285.0

buyukbas_uretim_sut
'All' 155.0 284.0
[total] 155.0 284.0
buyukbas_uretim_fiyat_sut
'All' 155.0 284.0
[total] 155.0 284.0
buzagi_ikinciurun
'All' 155.0 284.0
[total] 155.0 284.0

kucukbas
'All' 155.0 284.0
[total] 155.0 284.0
kucuk_deger
'All' 155.0 284.0
[total] 155.0 284.0
kucukbas_uretim_et
'All' 155.0 284.0
[total] 155.0 284.0
kucukbas_uretim_fiyat_et
'All' 155.0 284.0
[total] 155.0 284.0
kucukbas_uretim_sut
'All' 155.0 284.0
[total] 155.0 284.0
kucukbas_uretim_fiyat_sut
'All' 155.0 284.0
[total] 155.0 284.0
kuzu_ikinciurun
'All' 155.0 284.0
[total] 155.0 284.0
kovan_sayi
'All' 155.0 284.0
[total] 155.0 284.0

kovan_deger
'All' 155.0 284.0
[total] 155.0 284.0
kovan_uretim
'All' 155.0 284.0
[total] 155.0 284.0
kovan_uretim_fiyat
'All' 155.0 284.0
[total] 155.0 284.0
alet_ekipman
'All' 155.0 284.0
[total] 155.0 284.0
teminat

```

'All'          155.0 284.0
[total]       155.0 284.0
disari_borclar
'All'          155.0 284.0
[total]       155.0 284.0
kredi
'All'          155.0 284.0
[total]       155.0 284.0

```

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

```

Correctly Classified Instances   111      74.4966 %
Kappa statistic                  0.3078
Mean absolute error              0.3464
Root mean squared error          0.4108
Relative absolute error          76.6117 %
Root relative squared error      88.222 %
Total Number of Instances       149

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.356	0.087	0.640	0.356	0.457	0.331	0.753	0.647	NO
	0.913	0.644	0.766	0.913	0.833	0.331	0.752	0.849	YES
W. Avg.	0.745	0.476	0.728	0.745	0.720	0.331	0.752	0.788	

=== Confusion Matrix ===

a b <-- classified as

16 29 | a = NO

9 95 | b = YES

Ek-7: Karar Ağacı Algoritması Detaylı Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.StringToNominal-R13-16-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R48

Instances: 437

Attributes: 48

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut

buzagi_ikinciurun
kucukbas
kucuk_deger
kucukbas_uretim_et
kucukbas_uretim_fiyat_et
kucukbas_uretim_sut
kucukbas_uretim_fiyat_sut
kuzu_ikinciurun
kovan_sayi
kovan_deger
kovan_uretim
kovan_uretim_fiyat
alet_ekipman
teminat
disari_borclar
kredi
durum2

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

arazi_urun_fiyat1 <= 5: YES (405.0/126.0)

arazi_urun_fiyat1 > 5

| buyuk_deger <= 4000: NO (30.0/2.0)

| buyuk_deger > 4000: YES (2.0)

Number of Leaves : 3

Size of the tree : 5

Time taken to build model: 0.09 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances 115 77.1812 %

Kappa statistic 0.3428

Mean absolute error 0.3773

Root mean squared error 0.4267

Relative absolute error 83.4407 %

Root relative squared error 91.6402 %

Total Number of Instances 149

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.311	0.029	0.824	0.311	0.452	0.408	0.626	0.444	NO
	0.971	0.689	0.765	0.971	0.856	0.408	0.626	0.770	YES
Weighted Avg.	0.772	0.490	0.783	0.772	0.734	0.408	0.626	0.671	

=== Confusion Matrix ===

a b <-- classified as

14 31 | a = NO

3 101 | b = YES



Ek-8: Doğrusal Ayırma Analizi Algoritması Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.functions.FLDA -R 1.0E-6

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.StringToNominal-R13-16-
weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R48-
weka.filters.unsupervised.attribute.Remove-R11-15-weka.filters.unsupervised.attribute.Standardize

Instances: 437

Attributes: 43

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut
buzagi_ikinciurun
kucukbas
kucuk_deger
kucukbas_uretim_et

kucukbas_uretim_fiyat_et
kucukbas_uretim_sut
kucukbas_uretim_fiyat_sut
kuzu_ikinciurun
kovan_sayi
kovan_deger
kovan_uretim
kovan_uretim_fiyat
alet_ekipman
teminat
disari_borclar
kredi
durum2

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Fisher's Linear Discriminant Analysis

Threshold: 0.09389904646062094

Weights:

cinsiyet: -0.05104303889306374
nufuskayit_yeri: 0.12894073145093968
egitim_durumu: 0.021596691251922148
medeni_hali: -0.09362843023573678
yas<23: -0.0400445096609753
yas>65: 0.02052868355499005
kendi_arazisi: -0.09517206079656132
dekar_degeri: 0.13552651823789658
icar: -0.27025311448646283
uretim_yapilan: 0.2068359282585822
dekar_urun1: 0.14439970670600671
dekar_urun2: -0.07882684212039684
dekar_urun3: 0.06467070523006593
dekar_urun4: -0.13210404708731363
dekar_urun5: -0.02275636672495555
arazi_urun_fiyat1: 0.407485272954619
arazi_urun_fiyat2: -0.04607108784088622
arazi_urun_fiyat3: -0.038746138119315
arazi_urun_fiyat4: 0.11260865795005968
arazi_urun_fiyat5: -0.022756365255959345

buyukbas: -0.048215208056245754
 buyuk_deger: -0.06986411273966181
 buyukbas_uretim_et: 0.007104490715122379
 buyukbas_uretim_fiyat_et: -0.03809173756056598
 buyukbas_uretim_sut: -0.28281875190246025
 buyukbas_uretim_fiyat_sut: 0.42837655781833256
 buzagi_ikinciurun: 0.060091405245886916
 kucukbas: -0.04832615671461133
 kucuk_deger: -0.04139251223791845
 kucukbas_uretim_et: -0.26771045989225795
 kucukbas_uretim_fiyat_et: 0.20841491957375927
 kucukbas_uretim_sut: 0.18636540798744372
 kucukbas_uretim_fiyat_sut: -0.19500377258908108
 kuzu_ikinciurun: 0.025725388525231995
 kovan_sayi: 0.020115750594429688
 kovan_deger: 0.01994587972928535
 kovan_uretim: 0.1860854181887085
 kovan_uretim_fiyat: -0.26571169418780466
 alet_ekipman: -0.01905592167532431
 teminat: 0.1135110210847627
 disari_borclar: -0.06405934020882426
 kredi: 0.04914005226434799

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	99	66.443 %
Kappa statistic	0.2686	
Mean absolute error	0.4705	
Root mean squared error	0.4773	
Relative absolute error	104.0463 %	
Root relative squared error	102.5244 %	
Total Number of Instances	149	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.600	0.308	0.458	0.600	0.519	0.274	0.682	0.558	NO
	0.692	0.400	0.800	0.692	0.742	0.274	0.682	0.798	YES
W. Avg.	0.664	0.372	0.697	0.664	0.675	0.274	0.682	0.726	

=== Confusion Matrix ===

a b <-- classified as

27 18 | a = NO

32 72 | b = YES



Ek-9: Lojistik Regresyon Algoritması Program Çıktısı

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -C -R 1.0E-8 -M 8 -num-decimal-places 4

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.Remove-R1,49-
weka.filters.unsupervised.attribute.StringToNominal-R12-15

Instances: 437

Attributes: 48

cinsiyet
nufuskayit_yeri
egitim_durumu
medeni_hali
yas<23
yas>65
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4
arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut

buzagi_ikinciurun
 kukukbas
 kukuk_deger
 kukukbas_uretim_et
 kukukbas_uretim_fiyat_et
 kukukbas_uretim_sut
 kukukbas_uretim_fiyat_sut
 kuzu_ikinciurun
 kovan_sayi
 kovan_deger
 kovan_uretim
 kovan_uretim_fiyat
 alet_ekipman
 teminat
 disari_borclar
 kredi
 durum2

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Variable	Class	
	NO	
=====		
cinsiyet	-0.5307	
nufuskayit_yeri	0.8547	
egitim_durumu	0.147	
medeni_hali	-0.1853	
yas<23	-3.9739	
yas>65	-0.0865	
kendi_arazisi	-0.0005	
dekar_degeri	0	
icar	-0.0028	
uretim_yapilan	0.0012	
arazi_urun1=pancar	0.8445	
arazi_urun1=domates	0.0131	
arazi_urun1=0	-0.458	
arazi_urun1=patates	0.1149	

arazi_urun1=yonca	-5.3075
arazi_urun1=bugdaysulu	0.3319
arazi_urun1=bugdaykuru	0.2027
arazi_urun1=sekerpancari	-3.7635
arazi_urun1=domatessera	5.386
arazi_urun1=kiraz	0.4329
arazi_urun1=arpa	5.2786
arazi_urun1=figkuru	4.198
arazi_urun1=sebze	1.3996
arazi_urun1=misir	-1.1239
arazi_urun1=zeytin	-3.8903
arazi_urun1=misirsilaji	-0.5371
arazi_urun1=marul	-3.6108
arazi_urun1=karpuz	-3.6946
arazi_urun1=bezelye	-3.4005
arazi_urun1=kirmizibiber	6.4286
arazi_urun1=incir	-5.8747
arazi_urun1=misir	-1.2391
arazi_urun1=seftali	-5.7022
arazi_urun1=biber	4.2706
arazi_urun1=aycicegi	0.5259
arazi_urun1=kurusogan	4.8286
arazi_urun1=arpasulu	4.936
arazi_urun1=sarimsak	-3.9323
arazi_urun1=celtik	-0.0158
arazi_urun1=yesilsogan	-5.6537
arazi_urun1=kayisi	0.3405
arazi_urun1=korunga	-4.7266
arazi_urun1=arpakuru	-5.0419
arazi_urun1=kurufasulye	4.9142
arazi_urun1=findik	2.5828
arazi_urun1=kavun	4.5329
arazi_urun1=tutun	4.5758
arazi_urun1=figsulu	2.2519
arazi_urun1=limon	-1.3365
arazi_urun1=pamuk	-5.8897
arazi_urun2=0	0.0784
arazi_urun2=misirsilaji	1.3808

arazi_urun2=sekerpancari	-3.5715
arazi_urun2=domates	1.239
arazi_urun2=bugdaysulu	-0.5666
arazi_urun2=kiraz	4.7468
arazi_urun2=arpakuru	5.2619
arazi_urun2=cavdar	5.2786
arazi_urun2=misir	0.3258
arazi_urun2=arpa	4.198
arazi_urun2=bugday	5.6188
arazi_urun2=sebze	4.1732
arazi_urun2=bezelye	-4.5522
arazi_urun2=kirmizibiber	-3.0238
arazi_urun2=yonca	-0.3603
arazi_urun2=karpuz	-3.689
arazi_urun2=zeytin	-5.8547
arazi_urun2=ceviz	-1.2391
arazi_urun2=bugdaykuru	2.3524
arazi_urun2=aycicegi	-6.0394
arazi_urun2=kavun	-3.9323
arazi_urun2=kayisi	-4.118
arazi_urun2=figsulu	-2.16
arazi_urun2=lahana	4.9403
arazi_urun2=karnabahar	-1.5729
arazi_urun2=figkuru	3.9437
arazi_urun2=celtik	-2.602
arazi_urun2=mercimek	-4.0534
arazi_urun3=0	-0.7025
arazi_urun3=kavun	-6.2163
arazi_urun3=sekerpancari	-5.4408
arazi_urun3=arpasulu	-1.7423
arazi_urun3=cavdar	5.2619
arazi_urun3=mercimek	0.3831
arazi_urun3=arpakuru	6.0001
arazi_urun3=sebze	8.0544
arazi_urun3=visne	4.4716
arazi_urun3=figsulu	4.7638
arazi_urun3=domates	0.6111
arazi_urun3=biber	4.9509

arazi_urun3=misir	-0.3532
arazi_urun3=bugdaysulu	0.3859
arazi_urun3=bezelye	-2.134
arazi_urun3=zeytin	-4.7225
arazi_urun3=aycicegi	-0.0184
arazi_urun3=misirsilaji	4.2669
arazi_urun3=figkuru	-3.383
arazi_urun3=korunga	-1.4436
arazi_urun3=lahana	0.8584
arazi_urun3=nar	-1.3365
arazi_urun4=0	1.4838
arazi_urun4=yonca	-1.7423
arazi_urun4=domates	-2.1074
arazi_urun4=bezelye	4.7638
arazi_urun4=bugdaysulu	-2.3251
arazi_urun4=misir	-2.6649
arazi_urun4=aycicegi	4.8286
arazi_urun4=sekerpancari	-1.4436
arazi_urun4=lahanabeyaz	4.5329
arazi_urun4=figkuru	-9.1215
arazi_urun4=cilek	-1.3365
arazi_urun4=pamuk	-4.5143
arazi_urun5=bezelye	-1.2391
dekar_urun1	0.0201
dekar_urun2	-0.1262
dekar_urun3	0.0035
dekar_urun4	-0.2992
dekar_urun5	-0.6884
arazi_urun_fiyat1	-0.001
arazi_urun_fiyat2	-0.6789
arazi_urun_fiyat3	-0.3853
arazi_urun_fiyat4	-0.1023
arazi_urun_fiyat5	-0.826
buyukbas	-0.0028
buyuk_deger	-0.0001
buyukbas_uretim_et	0.0003
buyukbas_uretim_fiyat_et	-0.0093
buyukbas_uretim_sut	-0.0002

buyukbas_uretim_fiyat_sut	1.5816
buzagi_ikinciurun	0.0004
kucukbas	-0.0029
kucuk_deger	0.0003
kucukbas_uretim_et	-0.0248
kucukbas_uretim_fiyat_et	0.0008
kucukbas_uretim_sut	-0.0014
kucukbas_uretim_fiyat_sut	-1.1735
kuzu_ikinciurun	-0
kovan_sayi	0.001
kovan_deger	-0.0027
kovan_uretim	0.0093
kovan_uretim_fiyat	-0.1016
alet_ekipman	0
teminat	0
disari_borclar	-0
kredi	0
Intercept	-1.1519
Odds Ratios...	
Class	
Variable	NO

=====

cinsiyet	0.5882
nufuskayit_yeri	2.3507
egitim_durumu	1.1584
medeni_hali	0.8308
yas<23	0.0188
yas>65	0.9172
kendi_arazisi	0.9995
dekar_degeri	1
icar	0.9972
uretim_yapilan	1.0012
arazi_urun1=pancar	2.3268
arazi_urun1=domates	1.0132
arazi_urun1=0	0.6326
arazi_urun1=patates	1.1218
arazi_urun1=yonca	0.005
arazi_urun1=bugdaysulu	1.3936

arazi_urun1=bugdaykuru	1.2246
arazi_urun1=sekerpancari	0.0232
arazi_urun1=domatessera	218.3346
arazi_urun1=kiraz	1.5417
arazi_urun1=arpa	196.0992
arazi_urun1=figkuru	66.5552
arazi_urun1=sebze	4.0538
arazi_urun1=misir	0.325
arazi_urun1=zeytin	0.0204
arazi_urun1=misirsilaji	0.5844
arazi_urun1=marul	0.027
arazi_urun1=karpuz	0.0249
arazi_urun1=bezelye	0.0334
arazi_urun1=kirmizibiber	619.2957
arazi_urun1=incir	0.0028
arazi_urun1=misir	0.2897
arazi_urun1=seftali	0.0033
arazi_urun1=biber	71.5678
arazi_urun1=aycicegi	1.692
arazi_urun1=curusogan	125.0319
arazi_urun1=arpasulu	139.2102
arazi_urun1=sarimsak	0.0196
arazi_urun1=celtik	0.9843
arazi_urun1=yesilsogan	0.0035
arazi_urun1=kayisi	1.4057
arazi_urun1=korunga	0.0089
arazi_urun1=arpakuru	0.0065
arazi_urun1=kurufasulye	136.217
arazi_urun1=findik	13.2335
arazi_urun1=kavun	93.0241
arazi_urun1=tutun	97.1091
arazi_urun1=figsulu	9.5055
arazi_urun1=limon	0.2628
arazi_urun1=pamuk	0.0028
arazi_urun2=0	1.0815
arazi_urun2=misirsilaji	3.9781
arazi_urun2=sekerpancari	0.0281
arazi_urun2=domates	3.4522

arazi_urun2=bugdaysulu	0.5675
arazi_urun2=kiraz	115.2185
arazi_urun2=arpakuru	192.8439
arazi_urun2=cavdar	196.0992
arazi_urun2=misir	1.3851
arazi_urun2=arpa	66.5552
arazi_urun2=bugday	275.5625
arazi_urun2=sebze	64.9231
arazi_urun2=bezelye	0.0105
arazi_urun2=kirmizibiber	0.0486
arazi_urun2=yonca	0.6975
arazi_urun2=karpuz	0.025
arazi_urun2=zeytin	0.0029
arazi_urun2=ceviz	0.2897
arazi_urun2=bugdaykuru	10.5104
arazi_urun2=aycicegi	0.0024
arazi_urun2=kavun	0.0196
arazi_urun2=kayisi	0.0163
arazi_urun2=figsulu	0.1153
arazi_urun2=lahana	139.8146
arazi_urun2=karnabahar	0.2075
arazi_urun2=figkuru	51.6084
arazi_urun2=celtik	0.0741
arazi_urun2=mercimek	0.0174
arazi_urun3=0	0.4954
arazi_urun3=kavun	0.002
arazi_urun3=sekerpancari	0.0043
arazi_urun3=arpasulu	0.1751
arazi_urun3=cavdar	192.8439
arazi_urun3=mercimek	1.4668
arazi_urun3=arpakuru	403.4795
arazi_urun3=sebze	3147.7369
arazi_urun3=visne	87.4958
arazi_urun3=figsulu	117.1934
arazi_urun3=domates	1.8425
arazi_urun3=biber	141.2998
arazi_urun3=misir	0.7024
arazi_urun3=bugdaysulu	1.471

arazi_urun3=bezelye	0.1184
arazi_urun3=zeytin	0.0089
arazi_urun3=aycicegi	0.9817
arazi_urun3=misirsilaji	71.2996
arazi_urun3=figkuru	0.0339
arazi_urun3=korunga	0.2361
arazi_urun3=lahana	2.3593
arazi_urun3=nar	0.2628
arazi_urun4=0	4.4096
arazi_urun4=yonca	0.1751
arazi_urun4=domates	0.1216
arazi_urun4=bezelye	117.1934
arazi_urun4=bugdaysulu	0.0978
arazi_urun4=misir	0.0696
arazi_urun4=aycicegi	125.0319
arazi_urun4=sekerpancari	0.2361
arazi_urun4=lahanabeyaz	93.0241
arazi_urun4=figkuru	0.0001
arazi_urun4=cilek	0.2628
arazi_urun4=pamuk	0.011
arazi_urun5=bezelye	0.2897
dekar_urun1	1.0203
dekar_urun2	0.8814
dekar_urun3	1.0035
dekar_urun4	0.7414
dekar_urun5	0.5024
arazi_urun_fiyat1	0.999
arazi_urun_fiyat2	0.5072
arazi_urun_fiyat3	0.6802
arazi_urun_fiyat4	0.9028
arazi_urun_fiyat5	0.4378
buyukbas	0.9972
buyuk_deger	0.9999
buyukbas_uretim_et	1.0003
buyukbas_uretim_fiyat_et	0.9907
buyukbas_uretim_sut	0.9998
buyukbas_uretim_fiyat_sut	4.8625
buzagi_ikinciurun	1.0004

```

kucukbas          0.9971
kucuk_deger       1.0003
kucukbas_uretim_et  0.9755
kucukbas_uretim_fiyat_et  1.0008
kucukbas_uretim_sut  0.9986
kucukbas_uretim_fiyat_sut  0.3093
kuzu_ikinciurun   1
kovan_sayi        1.001
kovan_deger       0.9974
kovan_uretim      1.0093
kovan_uretim_fiyat  0.9034
alet_ekipman      1
teminat           1
disari_borclar    1
kredi             1

```

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

```

Correctly Classified Instances  305      69.7941 %
Kappa statistic                0.2988
Mean absolute error            0.3658
Root mean squared error        0.4854
Relative absolute error        80.1005 %
Root relative squared error    101.6055 %
Total Number of Instances      437

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.448	0.166	0.595	0.448	0.511	0.305	0.664	0.526	NO
	0.834	0.552	0.735	0.834	0.781	0.305	0.664	0.753	YES
W. Avg.	0.698	0.416	0.686	0.698	0.686	0.305	0.664	0.673	

=== Confusion Matrix ===

a b <-- classified as

69 85 | a = NO

47 236 | b = YES

Ek-10: Yapay Sinir Ağları Algoritması Detaylı Program Çıktısı

(Söz konusu algoritma çıktısı, oldukça uzun olduğu için, aşağıda değerlendirmeye esas olan bölümlere yer verilmiştir.)

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Relation: tezverileriENSON-weka.filters.unsupervised.attribute.StringToNominal-Rlast-
weka.filters.unsupervised.attribute.StringToNominal-Rlast-
weka.filters.unsupervised.attribute.StringToNominal-R13-16-
weka.filters.unsupervised.attribute.Remove-R49-weka.filters.unsupervised.attribute.Normalize-S1.0-
T0.0-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.NumericToBinary-R1-2-
weka.filters.unsupervised.attribute.NumericToBinary-R4-6

Instances: 437

Attributes: 48

cinsiyet_binarized
nufuskayit_yeri_binarized
egitim_durumu
medeni_hali_binarized
yas<23_binarized
yas>65_binarized
kendi_arazisi
dekar_degeri
icar
uretim_yapilan
arazi_urun1
arazi_urun2
arazi_urun3
arazi_urun4
arazi_urun5
dekar_urun1
dekar_urun2
dekar_urun3
dekar_urun4
dekar_urun5
arazi_urun_fiyat1
arazi_urun_fiyat2
arazi_urun_fiyat3
arazi_urun_fiyat4

arazi_urun_fiyat5
buyukbas
buyuk_deger
buyukbas_uretim_et
buyukbas_uretim_fiyat_et
buyukbas_uretim_sut
buyukbas_uretim_fiyat_sut
buzagi_ikinciurun
kucukbas
kucuk_deger
kucukbas_uretim_et
kucukbas_uretim_fiyat_et
kucukbas_uretim_sut
kucukbas_uretim_fiyat_sut
kuzu_ikinciurun
kovan_sayi
kovan_deger
kovan_uretim
kovan_uretim_fiyat
alet_ekipman
teminat
disari_borclar
kredi
durum2

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Time taken to build model: 33.86 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	111	74.4966 %
Kappa statistic	0.2765	
Mean absolute error	0.2513	
Root mean squared error	0.4647	
Relative absolute error	55.5729 %	

Root relative squared error 99.7997 %
Total Number of Instances 149

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.289	0.058	0.684	0.289	0.406	0.318	0.731	0.591	NO
	0.942	0.711	0.754	0.942	0.838	0.318	0.731	0.830	YES
Weighted Avg.	0.745	0.514	0.733	0.745	0.707	0.318	0.731	0.758	

=== Confusion Matrix ===

a b <-- classified as
13 32 | a = NO
6 98 | b = YES