

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE TEKNOLOJİ PERAKENDE
SEKTÖRÜNDE EK GARANTİ SATIŞ MODELLEMESİ

Seda ALACAN

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Doç. Dr. Ersoy ÖZ

Haziran, 2020

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE TEKNOLOJİ PERAKENDE
SEKTÖRÜNDE EK GARANTİ SATIŞ MODELLEMESİ

Seda ALACAN tarafından hazırlanan tez çalışması 17/06/2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Ersoy ÖZ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Ersoy ÖZ, Danışman

Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Öyküm Esra AŞKIN, Üye

Yıldız Teknik Üniversitesi

Doç. Dr. Seda BAĞDATLI KALKAN, Üye

İstanbul Ticaret Üniversitesi

Danışmanım Doç. Dr. Ersoy ÖZ sorumluluğunda tarafımca hazırlanan Makine Öğrenmesi ile Teknoloji Perakende Sektöründe Ek Garanti Satış Modellemesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Seda ALACAN

İmza



Sevgili aileme

TEŞEKKÜR

Bu tezin hazırlanmasında beni yönlendiren, bilgi, tecrübe ve güler yüzü ile çalışmama ışık tutan, ayrıca kendimi geliştirmeye yönelik de birkaç adım ileride olmamı sağlayan, danışman hocam Doç. Dr. Ersoy Öz'e, tezimi tamamlamamda emeği geçen ve benden maddi, manevi hiçbir desteği esirgemeyen aileme teşekkürlerimi sunuyorum.

Seda ALACAN



İÇİNDEKİLER

SİMGE LİSTESİ	vii
KISALTMA LİSTESİ	viii
ŞEKİL LİSTESİ	ix
TABLO LİSTESİ	x
ÖZET	xi
ABSTRACT	xiv
1 GİRİŞ	1
1.1 Literatür Özeti.....	3
1.2 Tezin Amacı.....	9
1.3 Hipotez.....	10
2 EK GARANTİ/GENİŞLETİLMİŞ GARANTİ	11
3 MAKİNE ÖĞRENMESİ	14
3.1 Gözetimli Öğrenme.....	15
3.1.1 Sınıflandırma Algoritmaları.....	17
3.1.2 Regresyon Algoritmaları.....	27
3.2 Gözetimsiz Öğrenme.....	27
3.3 Başarı Ölçüm Metrikleri.....	28
3.4 Çapraz Geçerlilik Yöntemleri.....	30
4 MATERYAL VE YAPILANDIRMA	31
4.1 Teknoloji Platformu.....	31
4.2 Deneysel Tasarım.....	32
4.3 Veri Seti.....	33
4.4 Veri Ön İşleme.....	34
5 UYGULAMA	38
5.1 Veri Modelleme.....	38

5.1.1	%5 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları.....	39
5.1.2	%10 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları.....	40
5.1.3	%15 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları.....	42
5.1.4	%20 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları.....	43
5.1.5	%25 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları.....	44
5.2	Tüm Tasarım Modellerinin Karşılaştırılması.....	46
6	SONUÇ VE ÖNERİLER	48
	KAYNAKÇA	51
	TEZDEN ÜRETİLMİŞ YAYINLAR	57

SİMGE LİSTESİ

w	Ağırlık Vektörü
n	Bağımsız Değişken Katsayısı
λ	Bozulma Faktörü
C	Düzenleme Parametresi
b	Eğilim Değeri
z	Fayda Fonksiyonu
x_i	Girdi Değeri
x_{max}	Girdi Seti İçerisindeki En Büyük Sayı
x_{min}	Girdi Seti İçerisindeki En Küçük Sayı
ε	Hata Terimi
x'	Normalize Edilmiş Veri
k	Örnek Sayısı
φ	Özellik Haritası
$\beta_1, \beta_2, \dots, \beta_n$	Regresyon Katsayıları
β_0	Regresyon Sabiti
y	Sınıf Etiketi
ξ_i	Yapay Değişken

KISALTMA LİSTESİ

DO	Doğruluk Oranı
DUO	Duyarlılık Oranı
DVM	Destek Vektör Makinesi
F1	F-Ölçütü
HO	Hata Oranı
KA	Karar Ağaçları
KOBİ	Küçük ve Orta Boy İşletme
KO	Kesinlik Oranı
LR	Lojistik Regresyon
MİY	Müşteri ilişkileri Yönetimi
NB	Naïve Bayes
RO	Rastgele Orman
SGİ	Stokastik Gradyan İniş
TÖ	Topluluk Öğrenmeleri
UKSB	Uzun Kısa Süreli Bellek
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 3.1 Gözetimli Öğrenme Modeli Süreci.....	16
Şekil 3.2 Destek Vektör Makinesi.....	18
Şekil 3.3 Doğrusal Ayrılabilen Veri Setleri için Hiperdüzlemin Belirlenmesi...	19
Şekil 3.4 Doğrusal Ayrılamayan Veri Setleri için Hiperdüzlemin Belirlenmesi.....	20
Şekil 3.5 İnsan Nöronunun (Sinir Hücresi) Yapısı.....	22
Şekil 3.6 Bir Nöronun (sinir hücresi) Matematiksel Modeli.....	23
Şekil 3.7 Tek ve Çok Gizli Katmanlı Yapay Sinir Ağı Modelleri.....	24
Şekil 3.8 n Kez Çapraz Geçerlilik.....	30
Şekil 4.1 Araştırmanın Akış Diyagramı.....	33
Şekil 5.1 Tasarım Modeli 1 Başarı Metriklerinin Karşılaştırılması.....	39
Şekil 5.2 Tasarım Modeli 2 Başarı Metriklerinin Karşılaştırılması.....	41
Şekil 5.3 Tasarım Modeli 3 Başarı Metriklerinin Karşılaştırılması.....	42
Şekil 5.4 Tasarım Modeli 4 Başarı Metriklerinin Karşılaştırılması.....	44
Şekil 5.5 Tasarım Modeli 5 Başarı Metriklerinin Karşılaştırılması.....	45

TABLO LİSTESİ

Tablo 3.1	Sınıflandırma Algoritmaları.....	17
Tablo 3.2	Karmaşıklık Matrisi.....	28
Tablo 4.1	Makine Öğrenmesi Kütüphaneleri ve Özellikleri.....	32
Tablo 4.2	One Hot Encoder Fonksiyonu ile Kategorik Değişken Dönüşümü...	35
Tablo 4.3	Öznitelik Sayıları ve Açıklamaları.....	36
Tablo 5.1	Tasarım Modelleri.....	38
Tablo 5.2	Tasarım Modeli 1 Çapraz Geçerlilik Başarı Ölçümlerinin Karşılaştırılması.....	39
Tablo 5.3	Tasarım Modeli 1 Test Seti Başarı Ölçümlerinin Karşılaştırılması.....	40
Tablo 5.4	Tasarım Modeli 2 Çapraz Geçerlilik Başarı Ölçümlerinin Karşılaştırılması.....	40
Tablo 5.5	Tasarım Modeli 2 Test Seti Başarı Ölçümlerinin Karşılaştırılması.....	41
Tablo 5.6	Tasarım Modeli 3 Çapraz Geçerlilik Başarı Ölçümlerinin Karşılaştırılması.....	42
Tablo 5.7	Tasarım Modeli 3 Test Seti Başarı Ölçümlerinin Karşılaştırılması.....	43
Tablo 5.8	Tasarım Modeli 4 Çapraz Geçerlilik Başarı Ölçümlerinin Karşılaştırılması.....	43
Tablo 5.9	Tasarım Modeli 4 Test Seti Başarı Ölçümlerinin Karşılaştırılması.....	44
Tablo 5.10	Tasarım Modeli 4 Çapraz Geçerlilik Başarı Ölçümlerinin Karşılaştırılması.....	45
Tablo 5.11	Tasarım Modeli 4 Test Seti Başarı Ölçümlerinin Karşılaştırılması.....	46
Tablo 5.12	Tasarım Modelleri Test Seti Başarı Ölçüm Karşılaştırılması.....	46

Makine Öğrenmesi ile Teknoloji Perakende Sektöründe Ek Garanti Satış Modellemesi

Seda ALACAN

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Ersoy ÖZ

Bu çalışmanın amacı, teknoloji perakende sektöründe satılan ürünlerin yanında ilgili ürüne ait ek garantinin satışına yönelik bir modelleme yapmaktır. Müşterilerin önceki alışverişlerindeki alışkanlıkları tespit edilerek bir sonraki alışverişinde ek garanti satın alıp almayacağına yönelik olarak tahminler yapılacaktır. Bu amaç doğrultusunda makine öğrenmesi yöntemlerinden sınıflandırma algoritmalarının performansları karşılaştırılacaktır. Algoritmalar arasında en az hata oranına sahip olan modelin tahmin sonuçları uygulamanın çıktısı olacaktır. Bu kapsamda ek garanti ürün satışlarının şirketlere ek ciro yaratması ve dönüştürme oranının artırılması hedeflenmektedir. Teknoloji perakendesinde faaliyet gösteren şirketlerden bir tanesinin son 2 yıllık ham satış verisi işlenmiş ve son iki yılda alışveriş yapan, aktif tekil müşterilerin geçmiş yıllardaki davranışları modele dahil edilerek doğruluk ve F1 skor metrikleri ile model başarıları test edilmiştir. Python programlama dili ile geliştirilen çalışmada, makine öğrenmesi yöntemlerinden yapay sinir ağları, stokastik gradyan iniş, destek vektör makineleri, karar ağaçları, naïve bayes, topluluk yöntemleri ve

lojistik regresyon yöntemleri kullanılmıştır. 5 farklı tasarım modeli ile oluşturulan çalışmada farklı başarı oranları hesaplanmıştır. Çalışma sonucunda tasarım modellerinde çoğunlukla destek vektör makineleri, topluluk yöntemleri ve yapay sinir ağları algoritmaları yüksek başarı gösterirken naïve bayes ve karar ağaçları algoritmalarının ek garanti satın alımı için yapılan sınıflandırma yöntemlerinde düşük başarı gösterdiği görülmüştür. Ayrıca, yapılan modelleme sonucunda perakende sektöründe ek ciro yaratması beklenen ve satış danışmanı prim sistemine katkısı olabilecek bir ürün ortaya çıkartılmıştır.

Anahtar Kelimeler: Makine öğrenmesi, sınıflandırma, gözetimli öğrenme, ek garanti.



Additional Warranty Sales Modeling in Technology Retail Industry with Machine Learning

Seda ALACAN

Department of Statistics

MSc. Thesis

Advisor: Assoc. Prof. Dr. Ersoy ÖZ

The purpose of this study is to create a model for the sale of additional warranty for the related product, as well as products sold in the technology retail sector. The habits of customers in their previous purchases will be determined and estimates will be made as to whether they will purchase additional guarantees in their next purchase. For this purpose, the performances of classification algorithms, one of the machine learning methods, will be compared. Estimation results of the model with the least error rate among the algorithms will be the output of the application. In this context, it is targeted that additional warranty product sales create additional turnover for companies and increase the conversion rate. For the last 2 years, raw sales data of one of the companies operating in technology retailing have been processed, and the behaviors of the active unique customer who has been shopping for the past two years are included in the model in the past years and the accuracy and F1 score metrics and model achievements have been tested. In the study developed with Python programming language, artificial neural networks, stochastic gradient landing, support vector machines, decision trees, naïve bayes, community methods and logistic regression

methods were used. In the study, which created 5 different design models, different success rates were calculated. As a result of the study, mostly in design models it was observed that support vector machines, artificial neural networks and community methods algorithms showed high success, while naïve bayes and decision trees algorithms showed low success in classification methods for additional warranty purchase. In addition, as a result of the modeling, a product, which is expected to create additional turnover in the retail sector and may contribute to the sales consultant premium system, has been created.

Keywords: Machine learning, classification, supervised learning, additional warranty.



1

GİRİŞ

Günümüzde tekil ana ürün odaklı satışlarının yanında ana ürünlerle birlikte gelen ek ürünler ve satış sonrası hizmetleri içeren paketlerin satışlarına doğru giderek artan bir trend bulunmaktadır. Firmalar son zamanlarda, ürün ile birlikte gelen ek hizmetler, bakım ve genişletilmiş ek garanti satmanın düşük riskli, yüksek marjlı iyi bir kaynak olduğunu keşfetmeye başlamıştır [1].

Satış sonrası hizmetler konusunda başarı göstermiş firmalar ürünlerini satmaya devam etmenin yanında ürünlerinin kalitesini destekleme hizmetleri de vermektedirler. Bu süreç normal olarak daha çok ürün satma isteğiyle hareket eder ve amacı satışları iyileştirmektir. Kalite ve verimliliğe yönelik güçlü bir girişimle sağlanan ek garanti hizmetleri müşteriler arasında güvenilir bir servis sağlayıcı olarak firmanın da tanınabilir olmasına olanak sağlar [2]. Bazı işletmeler müşterilerine satmış olduğu ürünlerin garanti süresinin bitiminde başlayan ek garanti/genişletilmiş garanti hizmeti satın alma fırsatı sunarlar. Müşteriler de isteğe bağlı olarak satın aldıkları bu hizmet ile ürünlerine ait ek teminat hakkı elde etmiş olurlar. Pahalı onarım maliyetlerine karşılık olarak satın aldıkları bu sigorta müşterilerde ek garanti sözleşmesi süresince huzur ve firmanın güvenilirliğine dair fikirler verir [3].

Literatürde satışları artırma yöntemleri müşteri memnuniyeti, müşteri sadakati, müşterilerin tekrar satın alma kararı gibi kavramlar ile ilişkilendirilmektedir. Artan rekabet ortamında, firmaların mevcut müşteriye koruma ve yeni müşterileri edinmesinde hizmet kalitesi ve müşteri memnuniyetinin önemi vurgulanmakta, pazarlama ve satış süreçlerinde bu konu ile ilgili dijitalleşme ile ilgili aksiyonlar alındığı vurgulanmaktadır.

Çoğu sektörde olduğu gibi teknoloji perakende sektöründe de ek garanti ürün satışlarının yapıldığı satış sonrası hizmet kalitesinin toplam şirket gelirlerini ve satılan ürün veya hizmet için ödenen birim fiyatı etkilediği düşünülmektedir. Bu

sebeple; arařtırmada ek garanti satıřlarını arttırarak řirket gelirlerine ek katkı yaratmak ve bunu yaparken g n m z teknolojisinden yararlanarak akıllı sistemler kurmak amalanmıřtır.

G n m zde kullanılan akıllı sistemlerin, fazla miktarda depolanan ve anlamlı bilgi ıkarım ihtiyacının olduėu veri k meleri; kriminoloji biliminden astronomiye, tıp bilim dalından meteoroloji alanına kadar birok dalda makine  ėrenmesi uygulamalarından yararlanması m mk nd r.

Bu alıřmada uygulama alanı olarak perakende sekt r  m řteri ve satıř veri tabanları kullanılacak ve kapsamı daraltmak amacıyla makine  ėrenmesi y ntemlerinden sınıflandırma algoritmalarına iliřkin teknikler  zerinde durulacaktır. Bu erevede, makine  ėrenmesi sınıflandırma algoritmalarına iliřkin teknikler gerek yařam verileri  zerinde uygulanacak, makine  ėrenmesi s reci t m ařamaları ile gerekleřtirilecek, uygulamadan elde edilen sonulara iliřkin farklılıklar arařtırılarak hangi tekniėin hangi kořullarda uygulanmasının daha etkin olacaėına y nelik  nerilerde bulunulacaktır. B ylece teknoloji perakende sekt r nde satılan ana  r nlerin yanında satılması planlanan ek garanti/geniřletilmiř garanti  r nlerinin satıř potansiyelinin arttırılarak karın ve cironun da artıřı hedeflenecek ve tahmin algoritmalarından sınıflandırma metotlarının bařarılarının bu noktada test edilmesi saėlanacaktır.

Bu tez kapsamında teknoloji perakendesinde faaliyet g steren řirketlerden bir tanesinin son 2 yıllık ham satıř verisi iřlenmiř ve son iki yılda alıřveriř yapan, 2 milyon tekil m řterinin gemiř yıllardaki davranıřları; yaptıkları iadeler, alıřveriř sıklıkları, prestij skorları, indirimli  r ne duyarlılıkları, nakit-kredi kullanım adetleri, m řteri segmentleri, demografik  zetleri modele dahil edilerek doėruluk oranı ve F1 skor metrikleri ile model bařarıları test edilecektir. eřitli oranlarda test ve eėitim setine ayrılan veri k mesi kullanılarak 5 farklı tasarım modeli oluřturulması hedeflenmektedir. Oluřturulan her bir tasarım modeli  zerinde 7 farklı sınıflandırma algoritması apraz geerlilik y ntemi test edilerek her tasarım modelinde 7 farklı sınıflandırıcı iin bařarı skoru hesaplanacaktır. Her bir tasarım modelinde apraz geerlilik y ntemi sonucunda en iyi 3 sınıflandırıcı belirlenerek bu 3 sınıflandırıcı iinden test veri setinde en uygun tasarım modeli ve en bařarılı

algoritmanın tespit edilmesi planlanmıştır. Python programlama dili ile geliştirilen bu çalışmada, veri ön işleme aşamasında kategorik bağımsız değişkenler nümerik değerlere çevrilerek yanlılık ortadan kaldırılmıştır. Aynı skalada bulunmayan nümerik bağımsız değişkenler normalize edilerek gerekli veri manipilasyonları yapılmıştır. Veri ön işleme aşaması tamamlandıktan sonra makine öğrenmesi yöntemlerinden yapay sinir ağları (YSA), stokastik gradyan iniş (SGİ), destek vektör makineleri (DVM), karar ağaçları (KA), naïve bayes (NB), topluluk öğrenme (TÖ) ve lojistik regresyon (LR) yöntemleri kullanılmıştır. Modellerin başarıları rasgele seçilen veri kümelerinin çapraz geçerlilik yönteminin kullanılmasıyla ölçülmüş ve modeller üzerinde parametre optimizasyonları yapılmıştır.

Çalışmanın ilk bölümünde, kavramsal ve teorik literatür çalışmaları sunulmuş olup ikinci bölümde ek garanti tanımı kavramsal olarak açıklanmıştır. Üçüncü bölümde makine öğrenmesi kavramı ve gelişimi, makine öğrenmesi yöntemleri ve sınıflandırma metotları teorik olarak incelenmiştir. Dördüncü bölümde araştırmada kullanılan veriler, araştırmanın tasarım süreci ile projeyi geliştirilen teknoloji platformu hakkında bilgilere yer verilmiş olup yapılan analizler sonucunda ulaşılan sonuçlara beşinci bölümde yer verilmiştir. Çalışmanın son bölümünde ise araştırmanın sonuçları tartışılmış ve gelecek çalışmalar için öneriler de bulunulmuştur.

1.1 Literatür Özeti

Günümüzde yaygın olarak kullanılan makine öğrenmesi yöntemlerinin birçok alanda uygulaması mevcuttur. Bu sebeple bu bölümde öncelikle makine öğrenmesi yöntemlerinden sınıflandırma algoritmalarını konu alan çeşitli sektörlerdeki tezler, kitaplar ve makaleler incelenerek temel teorik, ampirik bilgiler ve argümanlar açıklanmış ardından perakende sektöründeki uygulamalar incelenmiştir.

Makine öğrenmesi terimi ilk olarak Amerikan bilgisayar bilimcisi Arthur Samuel tarafından 1959'da bir araya getirilen ve bilgisayarın net bir şekilde programlanmadan öğrenme yeteneği olarak tanımlanmaktadır. Arthur Samuel bir

dama programı yazarak programı kendisi ile 10,000 defa oynamış ve yazdığı programı eğiterek kazanç ve kayıplara bağlı olarak hangi pozisyonların iyi ve kötü olduğunu belirlemiştir [4].

Stanford Üniversitesi Elektrik Mühendisliği ve Bilgisayar Bilimlerinde profesör olan Andrew Ng, üniversitenin yapay zeka laboratuvarında oldukça önemli çalışmalar yapmaktadır. Andrew Ng, karmaşıklığın geleneksel programlama yöntemlerinde kodlama bölümünde olduğunu ancak makine öğrenmesinde ise algoritmaların prensip olarak basit ancak asıl karmaşıklığın verilerde olduğunu vurgular. Makine öğrenmesinin merkezinde de bu karmaşıklığı otomatik olarak öğrenip çözüm üretmek olduğunu belirterek, makine öğrenmesinde sıkça duyduğumuz veri ön işleme aşamalarının önemini vurgulamaktadır [5].

Russel ve Norvig 1995 yılında yazdıkları makalede yapay zeka üzerinde modern bir yaklaşımda bulunmuş ve makine öğrenmesi algoritmalarını hedeflenen sonuca göre birkaç sınıfa ayırdığından bahsetmiştir. Bu sınıflar; girdileri hedef çıktılara eşitleyen bir algoritma üreten gözetimli öğrenme, bir girdi kümesini modelleyen gözetimsiz öğrenme ve yaşadıkça öğrenmeye dayalı bir öğrenme biçimi olarak ifade edilen pekiştirmeli öğrenme olarak tanımlanabilir [6].

Bhardwaj ve Pal eğitim sektörü ile ilgili çalışmalarında, öğrencilerin akademik performanslarını etkileyen faktörleri inceleyerek sınıflandırma algoritmalarını kullanmış ve bir tahmin modeli geliştirmişlerdir. Öğrencilerin akademik performansını; kişisel, sosyal, psikolojik ve diğer çevresel değişkenler gibi çeşitli faktörlere bakarak incelemişlerdir. Çalışmalarında, önceki yılın öğrenci bilgileri alınarak öğrencilerin akademik olarak başarılı olup/olmayacağını tahmin etmek için bizimde çalışmamızda kullandığımız yöntemlerden olan NB sınıflandırma yöntemini kullanılmışlardır. Bu çalışmayı öğrencilerin ve öğretmenlerin, öğrencinin başarı durumlarını geliştirmelerine yardımcı olacağını düşünerek ve aynı zamanda başarısızlık oranını azaltmak, doğru zamanda doğru eylemde bulunmak için özel dikkat gerektiren öğrencileri belirlemek için yapmışlardır. Mevcut çalışma sonucunda, öğrencilerin akademik performanslarının her zaman kendi çabalarına bağlı olmadığını, psikolojik ve çevresel faktörlerinde

performanslarında etkisi olduğunu göstermektedir. Bu sonuç, mevcut yöntemlere ilişkin iyileştirmelerin olmasını sağlamıştır [7].

Danso ise sigorta sektörü üzerinde direkt e-posta üzerinden ulaştıkları müşterilerin, e-postalara cevap vermeleri ile ilgili sınıflandırılma gerçekleştirmiştir. Çalışmasında sınıflandırma algoritmalarından YSA, KA, LR ve NB yöntemlerini kullanmış ve tıpkı Bhardwaj ve Pal gibi oda NB yönteminin sadece tanımladığı müşteri sayısı için değil, aynı zamanda e-postaların maliyeti açısından değerlendirilen tüm yönlerden diğer tekniklere göre üstün bir performans gösterdiğini açıkça vurgulamıştır [8].

Sağlık sektöründe uygulanan sınıflandırma algoritmaları incelendiğinde ise Shajaahan, Shanthi, ManoChitra'nın meme kanseri üzerine yaptıkları çalışma incelenmiştir. Bu çalışmayı meme kanserinin ciddi bir tehdit oluşturduğu, günümüzde kadınlarda ikinci önde gelen ölüm nedeni olması ve ayrıca gelişmiş ülkelerde en sık görülen kanser türü olması sebebiyle iyi tanı konmanın önemli olduğu gerekçesiyle araştırmışlardır. Araştırmada, en iyi sınıflandırıcıyı tahmin etmek için çeşitli gözetimli öğrenme algoritmalarını karşılaştırılmışlardır. Deneysel sonuçlar hassasiyet (precision), hatırlama (recall) ve doğruluk (accuracy) ölçüm metrikleri kullanılarak değerlendirilmiştir. Çeşitli sınıflandırma teknikleri arasında en yüksek doğruluk oranına sahip algoritma Rastgele Orman (RO) olarak çıkmıştır. Bu nedenle Shajaahan, Shanthi ve ManoChitra klinik araştırmalarında, RO algoritmasını meme kanseri gibi yaşamı tehdit eden son derece önemli olan bu hastalığın doğasını belirlemek için etkili bir sınıflandırıcı olarak tanımlamışlardır [9].

Cunningham ve Holmes, tarım sektörü için geliştirdikleri projede mantarlar için kalite sınıflandırması üzerine çalışmışlardır. Weka paket programı üzerinde çalıştıkları araştırmalarında C4.5 KA algoritmasını kullanan J4.8 sınıflandırıcısının mantarlar için kalite sınıflandırması ve pazar fiyatlandırması için daha objektif standartlar geliştirmede yararlı olabileceği sonucuna ulaşmışlardır [10].

Son dönemlerde duygu analizi de makine öğrenmesi yöntemlerinde sıkça kullanılan bir alan haline gelmiştir. Pang, Lee ve Vaithyanathan da çalışmalarında bu alana yönelerek standart makine öğrenmesi tekniklerinin insan tarafından

retilen temel hatlardan kesin olarak daha iyi performans gsterdiđini tespit etmiřlerdir. Sınıflandırıcı olarak DVM ve NB ile maksimum entropi sınıflandırıcılarını kullanmışlar ve film deđerlendirmeleri iin yaptıkları alıřmada “Movie Review” veri setini kullanmışlardır [11].

Hormozi ve Giles, makalelerinde veri madenciliđinde kullanılan makine đrenmesi yntemlerinin bir kuruluřun rakiplerine gre avantaj elde etmesini sađlayacak deđerli bir ara olduđundan bahsetmişlerdir. Pazarlama, risk ynetimi, sahtekarlık tespiti (Fraud Detection), mřteri kazanımı ve elde tutma dahil ancak bunlarla sınırlı olmamak zere veri madenciliđinin nemli kullanımlarına odaklanmışlardır. Bankacılık ve perakende sektrnn bu alanlarda veri madenciliđini nasıl etkin bir řekilde kullandıđına dair rnekler vermişlerdir [12]. Hormozi ve Giled gibi Raju, Bai ve Chaitanya’da perakende sektr iin veri madenciliđinin onlara rekabet avantajı sađlayabileceđini alıřmalarında vurgulamıştır [13]. Sudhakar ve Reddy alıřmalarında, firmaların (perakende ve bankacılık sektrlerinde) mřterilerin kredi talebini onaylamak veya reddetmek iin dođru kararı vermelerine yardımcı olan iki adımlı bir kredi gvenilirlik tahmin sistemi zerinde alıřmış ve sınıflandırma tahmin algoritmalarından KA algoritmasını kullanmışlardır [14].

Cumby, Fano, Ghani ve Krema, perakende sektrndeki bir mađazanın mřterileri iin alışveriş listelerini tahmin eden bir alıřma yapmışlardır. Kmeleme veya segmentasyon gibi geleneksel kiřiselleřtirme yaklařımlarını kullanmak yerine, her mřteri iin gemiř iřlem verilerinden đrendikleri bilgileri kullanarak, belirlenen mřterilerin, belirli bir alışveriş gezisinde hangi rnleri satın alacađı konusunda ayrıntılı ve dođru tahminler yapılabildiđini ortaya koymuşlardır. Alışveriş listesi tahminini bir sınıflandırma sorunu olarak erevelendirmiş ve yaptıkları alıřmayı perakende iř dnyasını etkilemesini amalamışlardır [15]. Online perakende sektrnde ise Ahmeda, Shehaba, Morsya ve Mekawiea’nın beraber hazırladıkları alıřmaları incelenmiştir. Satıř verilerindeki gizli iliřkiler incelenerek mřterinin bazı e-ticaret web sitelerinde aradıđı rnleri bulmasına yardımcı olan karar tablosu sınıflandırıcısına dayanan bir tavsiye sistemi (recommendation engine) geliřtirmişlerdir. Tavsiye sistemi, mřteriler ve rnler hakkındaki bilgileri đrenir

ve istenen ürünleri bulmak için müşterilere uygun kişiselleştirilmiş öneriler sunar [16]. Abor çalışmasında endüstri sınıflandırmasının Gana'daki küçük ve orta boy işletme (KOBİ)'lerin sermaye yapısı üzerindeki etkisini incelemeyi amaçlamaktadır. KOBİ'lerin endüstriler arasındaki sermaye yapısındaki farklılıkları incelemek için varyans analizi ve diğer parametrik olmayan testler de kullanılmıştır. Çalışmanın sonucunda, tarım sektöründeki KOBİ'lerin en yüksek sermaye ile varlık yapısına ve teminat değerine sahip olduğunu gösterirken, toptan ve perakende ticaret endüstrisi en düşük borç oranına ve varlık yapısına sahip olduğu ortaya çıkmıştır [17]. Ramageri ve Desai de çalışmalarında Hormozi ve Bai'nin çalışmaları gibi perakende sektöründe veri madenciliği yöntemlerinin önemini vurgulamıştır. Şirketlerin müşteri yaşam döngüsünü belirleme, müşterileri segmente etme gibi konularda bu alanın kullanılmasının kârlı müşterileri hedeflemek için veri madenciliği teknolojisini benimsemişse, bu rekabetçi pazarda kazanacağını ve bu stratejiyi sürdürürse şirketlerin giderek daha başarılı olacağını vurgulamıştır [18].

Perakende sektöründe yapılan çalışmalar özelinde incelemeler yapıldığında, müşteri odaklı çözümlerin veri madenciliği yöntemleri kullanılarak hayata geçirilirse rakiplere göre avantaj sağlayacağı yönünde olduğu açıkça gözlemlenmiştir. Yine Minami ve Dawson'ın çalışması incelendiğinde de genel olarak müşteri ilişkileri yönetimi (MİY) üzerine yapılan araştırmalar, MİY ile müşteri memnuniyetinin etkileri, müşteriyi elde tutma ve kar yönetimi ve MİY tekniğinin performans üzerindeki etkileri üzerine odaklanmış ve veri madenciliğine dikkat çekmişlerdir [19].

Perakende sektörünün, satış, müşteri alışveriş geçmişi, ürün taşımacılığı, tüketim ve hizmet hakkında büyük miktarda veri topladığı için veri madenciliği teknikleri için önemli bir uygulama alanı olduğunu söyleyen Tinabo, çalışmasında toplanan veri miktarının özellikle e-ticarette yürütülen işlerin kolaylığının, kullanılabilirliğinin ve popülerliğinin artması nedeniyle hızla genişlemeye devam ettiğini belirtmiştir. Makalesinde, perakende sektöründe müşteriyi tutundurma sorununa yönelik çeşitli potansiyel veri madenciliği tekniğini araştırmış ve KA yönteminin bu anlamda etkili teknik olduğunu önermiştir [20].

Bhasin, makalesinde veri madenciliğini, perakende sektörü için mevcut verilerden önemli bilgileri almak, daha iyi karar vermeyi sağlamak için kullanılan bir araç olarak tanımlamıştır. Veri madenciliği yapılabilmesi için veri tabanlarındaki çeşitli verileri kabul edilebilir bir biçimde birleştirmek için veri ambarlarının kullanıldığını ardından verilerin analiz edildiğinden ve yakalanan bilgilerin firma genelinde karar almayı desteklemek için kullanılmasının öneminden bahsetmiştir. Veri madenciliğinin endüstrilerde birçok kullanımı olduğunu ve bunların, kredi kartı sahtekarlık tespiti, müşteri satın alım davranış tespiti, web kullanıcılarının kişisel bilgilerinin tespiti, üretim süreçlerinin optimizasyonu vb. pratik uygulamalarda oldukça başarılı olduğunu savunmaktadır. Ayrıca, bilgisayarların geçmiş deneyimlerden otomatik olarak nasıl öğrenebilecekleri hakkında yorumlar yaparak makine öğrenmesi konusuna da değinmiştir [21].

Ling ile Li'de perakende ve bankacılık sektörlerinde giderek daha fazla kullanılan makine öğrenmesi yöntemleri sırasında çeşitli sorunlar ortaya çıktığını, örneğin, örnek dağılımlarındaki sınıfların son derece dengesiz olması durumunda doğruluk metriğinin artık öğrenme yöntemlerini değerlendirmek için uygun olmadığını ve örnek sayısının artırılması vb. gibi yöntemlere başvurulması gerektiğini tartışmışlardır. Makalelerinde, veri madenciliği kullanan doğrudan pazarlama projeleri konusundaki deneyimlerine dayanarak bu sorunlarla başa çıkma yöntemlerini incelemişlerdir [22]. Zhang ve Ruan ise çalışmalarında makine öğrenmesi yöntemlerinden birliktelik kurallarına değinmişlerdir. Perakende sektöründe çok sayıda biriken satış verisine Apriori algoritmasını kullanarak ürünler arasındaki ilişki kurallarını bulunabilir ve perakende endüstrisinin kârını arttırılabilir olduğunu kanıtlamak için çapraz satış stratejileri oluşturulabileceğinden bahsetmişler ve böylece algoritmanın uygulanabilirliğinin kanıtlanmasını açıklamışlardır [23]. Perakende sektörü için en yaygın kullanılan veri madenciliği alanlarından birinin pazarlama departmanı olduğuna değinen Hemalatha araştırmasında özellikle mağaza özelliklerini bağımsız değişkenler olarak kullanarak perakende memnuniyetini tahmin etmek için Hint perakende endüstrilerinde bir veri madenciliği aracı olan çoklu regresyon tekniğinin uygulanmasına odaklanmaktadır [24].

Gu ise perakendeciliğin dijitalleştirilmesi gerektiği ile ilgili çalışmasında, internete bağlı dijital teknolojileri perakendeciler ve tüketiciler arasındaki ara yüzlere entegre etmek için devam eden bir süreç olarak tanımlamaktadır. Araştırmasında makine öğrenmesi destekli ürünleri geliştiren tasarımcıların gerçekleştirebileceği bir tasarım yaklaşımını konu almaktadır [25]. Perakendecilere satış kaybı ve aşırı talepten fayda sağlamanın satış tahminlemesi yoluyla mümkün olacağını söyleyen Yu, Wang ve Strandhagen, makalelerinde 66 ürünün 45 haftalık satış verileri üzerinde, mevsimsellik ve promosyonların etkisini dikkate almadan, Uzun Kısa Süreli Bellek (UKSB) ile Tekrarlayan Sinir Ağları algoritmalarını test etmişler ve UKSB algoritmaları ile düşük bir tahmin hatası ile çalışmalarını sonlandırmışlardır [26].

1.2 Tezin Amacı

Günümüzde ekonominin yarattığı etkilere ek olarak hızlı bir şekilde artan rekabetçi ortamda, firmalar fark yaratabilmek adına geleceğe dair tahminlerde bulunabilmek ve stratejilerini de fark yaratabileceği alanlara göre belirlemek istemektedir. Bu bağlamda firmalar risklerini en aza indirmek için, istatistiksel analizleri ve tahmin yöntemlerini kullanmaktadırlar. Bu kapsamda bu çalışmanın amacı da, ilgili firmanın rakiplerine göre avantaj elde etmesini sağlayacak değerli bir araç olan makine öğrenmesi tekniklerinin perakende sektöründe satılan teknolojik ürünlerin yanında, satılan ürüne ait ek garantinin/genişletilmiş ek garantinin satışını arttırmak için kullanılabilir bir yöntem olduğunu ortaya koymaktır. Bu amaç doğrultusunda müşterilerin önceki alışverişlerindeki alışkanlıklarının tespit edilerek bir sonraki alışverişinde ek garanti satın alınıp alınmayacağının tahmin edilebilir olduğu ortaya koyularak satışların desteklenmesi planlanmıştır.

Tahminleme yapabilmek için makine öğrenmesi yöntemlerinden sınıflandırma algoritmalarının performanslarını karşılaştırmak, şirketin mağazalarındaki kasa entegrasyonu ile model sonucunda müşterinin aldığı ürün bazında ek garanti satın alıp almayacağının kodlanması gerekmektedir. Müşteri ve satın aldığı ürün grubu bazında kodlanan bu müşteriler satın alım yapar/yapmaz şeklinde sınıflanır.

Çalışmanın nihai amaçlarından bir diğeri ise tez çalışmasında kurulacak model ile satış danışmanının ek garanti alması tahmin edilen ancak satın alım yapmayan müşteriye ürün hakkında bilgilendirmenin yapılması ve hatta mümkünse belli bir oranda indirim uygulayarak ek garanti ürününü satın alıma dönüştürmektir. Bunun yanısıra ek garanti satın almayacağı tahmin edilen müşteriye satın alıma dönüştüren satış danışmanlarının prim sistemlerine artı puan yazılabileceği bir performans ölçüm takibi yapılarak satışların arttırılması da çalışmanın çıktılarından biridir. Tüm bu kapsamlar dahilinde dönüştürme oranını arttırılması hedeflenmektedir.

1.3 Hipotez

Perakende sektöründe ek ürünler karlılığı yüksek ürünler olması sebebiyle ana ürün ile birlikte satışlarının yapılabilmesi son derece önemlidir. Teknoloji perakendesinde ise ana ürün satış fiyatlarının oldukça yüksek olması ek ürün satışlarının yapılabilmesi diğer sektörlerle nazaran daha karmaşık ve zor bir süreci beraberinde getirmektedir. Bu süreci kolaylaştırmak için teknolojik yaklaşımlar ile akıllı sistemler kullanılmaktadır.

Ana ürün yanında satılması planlanan ek garanti ürünleri için satışların arttırılması, bu amaçla uygun modellerin oluşturulması ve makine öğrenmesi yöntemlerinden sınıflandırma algoritmalarının sektörel bir problem üzerinde kullanılabilir ve etkin yöntemler olduğu ortaya konacaktır.

EK GARANTİ/GENİŞLETİLMİŞ GARANTİ

Ürünlere yönelik sunulan garanti hizmeti, son zamanlarda giderek daha fazla önem kazanan bir pazarlama ürünü olarak öne çıkmaktadır. Garanti temelde üretici firmanın, bir ürün veya hizmetin temsil edildiği ve nasıl temsil edileceği konusunda tüketiciye verdiği güvence olarak tanımlanır. Ürünlerin satışı sırasında üretici ile tüketici arasındaki bir sözleşme olarak kabul edilir ve yeni satın alınan ürün belirlenen süreler arasında garanti altındadır. Bu sözleşme, ürün garanti süresi içinde ve garanti kapsamındaki maddeler doğrultusunda arızalandığında, ücretsiz olarak tamir ettirme garantisini verir. Üretici firmalar müşterilerine diğer firmalardan farklı garanti koşulları da sunabilmektedir. Bazı arıza durumlarında, ürün tamir edilebilir olsa dahi işletmenin ürünü yenisi ile değiştirdiği durumlar da bu farklılığa örnek olarak verilebilir [27].

Ürünle birlikte verilen ve ücretsiz olan temel garantiye ek olarak, tüketiciler genellikle garanti kapsamı süresini uzatan genişletilmiş garanti adı verilen ürünleri de satın alma seçeneğine sahiptir. İsteğe bağlı satılan uzatılmış garantiler, ürünün genel garanti ve servis paketinin önemli bir parçası haline gelmiştir. Risk almaktan kaçınan müşteriler ürüne temel garantinin yanında ek olarak uzatılmış garantiyi satın alırken, diğer müşteriler ürünlerin arızalanma risklerini alarak temel ücretsiz garanti ile devam etmektedirler [28]. Ek garanti satın alan çoğu müşterinin temel nedeni olarak sözleşme süresi boyunca verdiği iç huzur ile beraber pahalı onarım maliyetleri riskine karşı bu garantiyi bir sigorta olarak gördükleri söylenebilir [29].

Garantinin iki temel rolü bulunmaktadır. Bunlar ürünü koruma ve ürün hakkında bilgilendirilme. Ek garantinin koruma rolü, ürün talimatlara uygun bir şekilde kullanıldığında arızalanma durumu söz konusu olur ise tüketiciye maddi hakları sağlanır, bilgi rolünde ise tüketiciler; nispeten daha uzun garanti süresine sahip ürünün, diğer ürünlere göre daha güvenilir ve daha uzun ömürlü olduğunu iddia ederler [30].

Garanti hizmetleri, işletmeye ek maliyetler getirmesine karşın yeni ürünlerin pazarlanmasında da önemli bir unsur haline gelmiştir. Genel anlamda garantinin amacı; ürünün arızalanması durumunda üretici ve müşteri arasında bir sorumluluk oluşturmaktır. Ürünler genellikle bir çeşit garanti ile birlikte satılır. Günümüzde elektronik ürünlerden, dayanıklı tüketim ürünleri ve ticari ürünlere kadar pek çok ürün türü için garantiler, satış işlemlerinin önemli bir parçası haline gelmiştir. Ürünler için sunulan ek garanti şekilleri genelde iki tiptir.

- **Temel Garanti:** Üretici ile müşteri arasında sözleşmeye dayalı bir anlaşma olarak müşteriyi ürünün malzeme ve işçilik hatalarına, beklenmeyen arızalara ve düşen performanslara karşı korur. Temel garantinin maliyeti ilgili ürünün satış fiyatına dahildir.
- **Genişletilmiş / Uzatılmış Ek Garanti:** Müşteri ile üretici arasındaki bir hizmet sözleşmesidir. Müşterilerin isteğe bağlı olarak satın aldıkları garanti türüdür. Satın alınan ürünün garantisi sona erdikten sonra ürün için ek teminat sağlamayı amaçlar. Ek garanti satın alan müşterilerin temel nedeninin, sözleşme süresi boyunca verdiği iç huzur olarak ölçülmüştür. Ayrıca ek garantiyi, pahalı onarım maliyetleri riskine karşı da bir sigorta olarak görmektedirler [29].

Farklı sektörlerde, farklı ürün grupları için ek garanti satışları yapılarak karlılık amaçlanmaktadır. Bu tezde teknoloji perakendesi özelinde bir araştırma planlandığı için bu alandaki ek garanti satışları incelenmiştir.

Perakende sektöründe faaliyet göstermiş bir şirket, genişletilmiş garantilerin kâr marjlarının %50 ile %60 arasında değiştiğini, ürün satışlarındaki tipik marjın yaklaşık 18 katı olduğunu belirtmiştir. Yüksek kârlılıklarından dolayı, bugün müşteriler için yalnızca tüketici elektroniği pazarında değil, daha uzun bir garanti satışı için ek garanti satışı kampanyaları duyma olasılıkları giderek artmaktadır [31].

2004 yılında garanti yönetimi için yayınlanan bültende teknoloji perakendesi sektöründe, Dell marka ürünlerin yaklaşık %30'unun ve Apple marka ürünlerin net gelirinin yaklaşık %43'ünün ek garanti satışlarından kaynaklandığı

yayınlanmıştır [32]. Üreticilerin ürünleri için temel garanti sunmalarına ek olarak perakendeci aracılığıyla genişletilmiş garantiler sunmanın, üretici aracılığıyla sunmaya göre daha uzun garanti kapsamı ve daha yüksek sistem karı sağladığı belirtilmektedir [33].

Teknolojik ve elektronik ürün satışlarında perakendecilik yapmanın, ek garanti satışlarını da avantajlı duruma getirdiği görülmüştür. Elektronik ürün satışlarındaki ek garanti karlılığının, ürünün kendi karından daha yüksek olması ve makine öğrenmesi yöntemlerinin ek garanti satışlarının yönetiminde uygulanabileceği pek çok potansiyel alan görülmesi sebebiyle bu alan üzerinde araştırma yapılmıştır.



Teknolojinin ilerlemesi ile verilerin dijital ortamda saklanması daha basit hale gelmiş, veriler daha ulaşılabilir ve depolanabilir hale gelmekle birlikte ulaşılabilir verinin miktarı gün geçtikçe artmaktadır. Verilerin depolanabilir olması ise veri madenciliği alanının oldukça kullanılabilir ve geniş bir alan olmasına neden olmaktadır. Veri madenciliği yöntemleri üzerine yapılan ilk çalışmalar 1950'li yıllarda bazı matematikçiler tarafından bilgisayar ve mantık bilimleri alanında gerçekleştirilmiştir. Bu çalışmaların ardından veri madenciliği yöntemleri 1960'lı yıllardan günümüze kadar gelişmeye devam etmiş ve yapay zeka kavramı ile makine öğrenme yöntemleri yeni çalışma ve araştırma alanları olarak kendini göstermiştir [34].

Makine öğrenmesi yöntemleri daha önce edinilmiş bilgilere ve gözlemlere dayanarak gerçeğe en yakın tahminler yapabilmeyi öğrenebilmek amacıyla otomatik tekniklerin geliştirilmesi olarak tanımlanabilir [35]. Makine öğrenmesi yöntemlerinde yapılan çalışmaların artmasıyla beraber hesaplanan öğrenme teorileri, YSA, örüntü tanıma ve istatistik gibi bilimsel araştırma alanları arasında da bağlantı kurularak bu alanların birlikte çalışması sağlanabilmiştir. Böylece makine öğrenmesi teknikleri geleneksel problemlere ek olarak dil işleme, veritabanlarında bilgi keşfi ve robotik problemlere de uygulanmaya başlamıştır [36]. Makine öğrenmesi yöntemlerinde veri oldukça önemli bir rol oynamakla birlikte, öğrenme algoritmaları özellikle veriye ait bilgi ve özelliklerin keşfinde de önemli bir rol üstlenir [37]. Kullanılan veriler, temel olarak etiketli ve etiketsiz olmak üzere iki ayrı tipten oluşmaktadır. Etiketli veri seti bir algoritmayı eğitmek için kullanılırken, etiketsiz veri seti ise eğitilmiş algoritmayı veya model ve sistemi test etmek için kullanılmaktadır. Bu sebeple bu veri tipleri eğitim ve test seti olarak da adlandırılır. Makine öğrenmesi yöntemlerini kullanılarak oluşturulan sistemler genelde iki farklı öğrenme modeli kullanılmaktadır. Bu iki model gözetimli

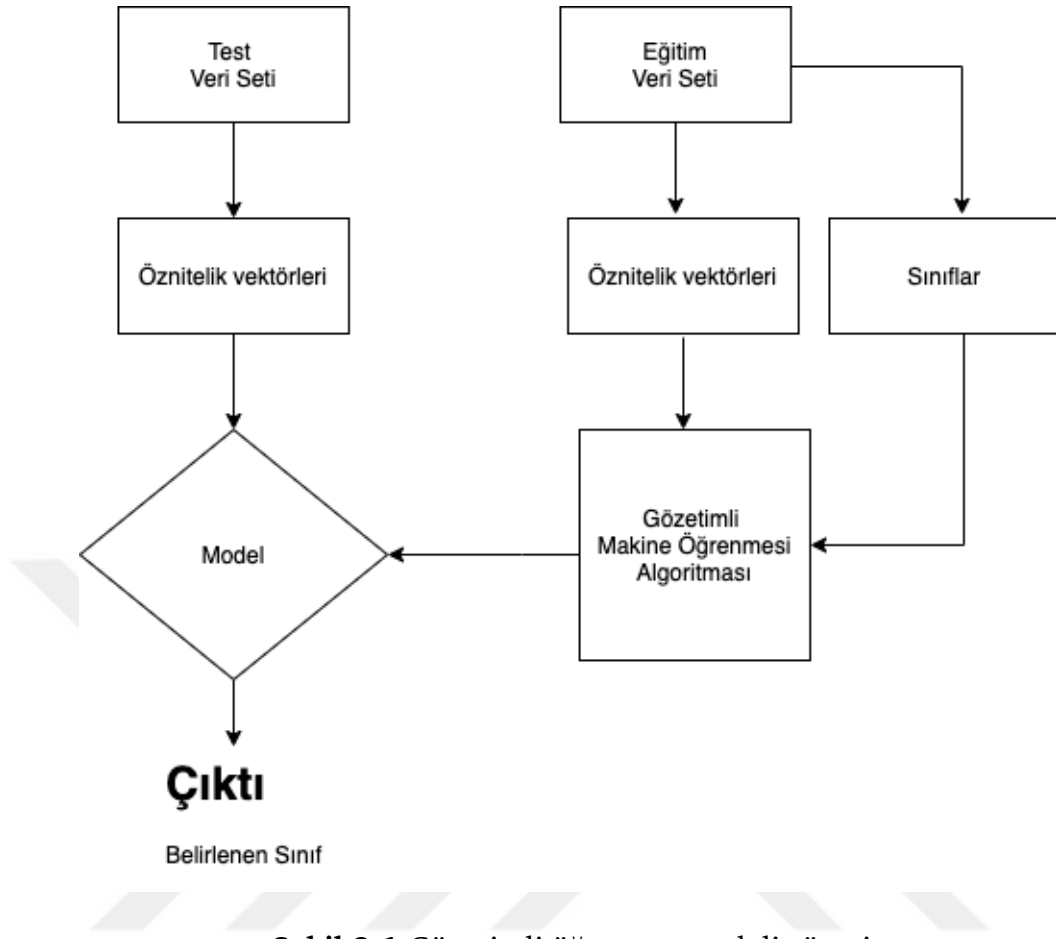
(Supervised Learning) ve gözetimsiz (Unsupervised Learning) öğrenme modeli olarak adlandırılır [38].

3.1 Gözetimli Öğrenme

Gözetimli öğrenme yöntemi, eğitim verileri üzerinden fonksiyon üreterek sonuçlar çıkartan bir makine öğrenmesi tekniğidir. Başka bir ifadeyle, bu öğrenme tekniğinde algoritma girdiler ile (etiketlenmemiş veri) çıktılar (etiketlenmiş veri) arasında eşleme yapan bir fonksiyon üretir [39]. Gözetimli öğrenme teknikleri tahmin edici modeller olarak da tanımlanırlar. Bu tahmin edici modeller tahmin sonuçları ile kararlar alarak sonuca varma süreçlerinde oldukça önemli role sahiptirler. Bu tür modellerde çıktıları bilinen etiketli verilerden yola çıkılarak model geliştirilmesi ve kurulan tahmin modelinden yararlanılarak çıktıları bilinmeyen veri setleri için çıktı değerlerinin tahmin edilmesi amaçlanmaktadır [40]. Örnek vermek gerekirse, bir bankanın daha önceki dönemlerde müşterilerine vermiş olduğu kredilere ilişkin müşteri verileri veri tabanında bulunmaktadır. Bu verilerde girdiler kredi alan banka müşterilerinin çeşitli özellikleridir ve çıktı ise kredinin geri ödenip ödenmemesidir. Oluşturulan veri setine en uygun model kurulur, test edilir ve en başarılı algoritma, daha sonraki kredi alma taleplerinde ilgili müşteri özelliklerine bakarak verilmesi planlanan kredinin geri ödenip ödenmeyeceği tahmininde bulunur [41].

Makine öğrenmesi yöntemlerinde çıktı değişkenleri (bağımlı değişkenler) nümerik veya kategorik değişkenler olabilmektedirler. Çıktı değişkenlerinin nümerik değişkenler olması durumunda, Regresyon algoritmaları uygulanırken, kategorik olması durumunda sınıflandırma algoritmaları tercih edilir. Üzerinde araştırma yapılan bu tezde, çıktılar kategorik değişken olduğundan sınıflandırma algoritmaları derinlemesine incelenmiştir.

Sınıflandırma çalışmalarında girdiler projenin içeriğini, çıktılar ise sınıfını işaret eder. Test kümesi de sistemin doğrulanması ve geliştirilmesi için kullanılır. Bu aşamada öğrenme yöntemleri sınıfı belli olmayan test değerine, eğitim kümesinde bulunan sınıflardan tahmin ettiği sonucun atamasını yapar. Sınıflandırma algoritmalarında modelleme Şekil 3.1'deki gibi gerçekleşmektedir.



Şekil 3.1 Gözetimli öğrenme modeli süreci

DVM, SGI, TÖ, YSA, LR, NB, RO ve KA algoritmaları gözetimli öğrenme modelleri için yaygın olarak kullanılan sınıflandırma yöntemlerindendir [42]. Sınıflandırma amacıyla kullanılan algoritmalar uygulanan problemin türüne göre farklılık gösterebilmektedir. Bu sebeple sınıflandırma problemlerinde sonuçların tek kategorili veya çok kategorili olarak üretilmesi de beklenen durumlar arasındadır. Tek kategorili sınıflandırmada test veri setine mevcut sınıf kümesinden sadece bir kategori atanırken, çok kategorili sınıflandırmada birden çok kategori işaretlenebilir. Ek olarak tek kategorili sınıflandırmanın özel bir durumu olan ikili sınıflandırmada (binary classification) ise mevcut iki sınıf arasından tek bir kategori atanarak sınıflandırma işlemi yapılmaktadır [43].

3.1.1 Sınıflandırma Algoritmaları

Sınıflandırma algoritmaları tahmin edici, ya da bir diğer ifadeyle kestirimci algoritmalar [44]. Bu tür algoritmalar veri kümesinde bulunan verilerin ortak özelliğine göre kategorilere ayrılmasını sağlamaktadır. Bu amaç doğrultusunda da çeşitli algoritmalar geliştirilmiştir. En yaygın kullanılan sınıflandırma algoritmaları Tablo 3.1'deki şekilde özetlenebilir [45]:

Tablo 3.1 Sınıflandırma algoritmaları

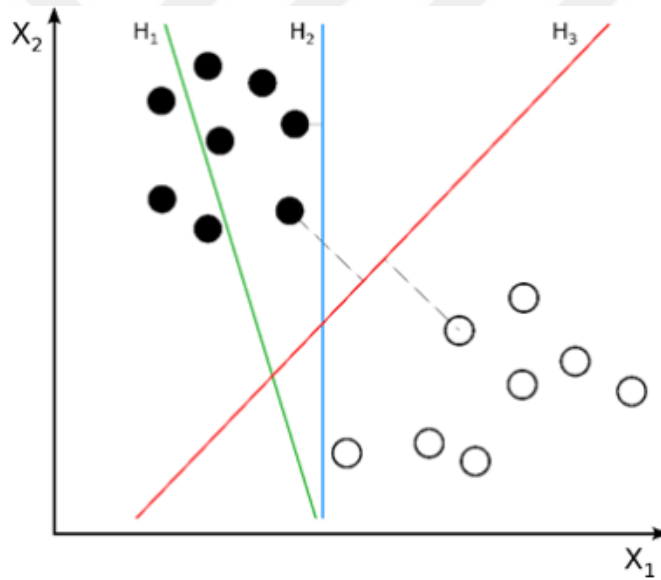
Sınıflandırma Algoritmaları	Örnek
Entropi Tabanlı Sınıflandırma	C4.5 algoritması, C5.0 algoritması
Regresyon ve Karar Ağaçları	Gini algoritması, Twoing algoritması
Bellek Tabanlı Algoritmalar	K-en yakın komşu algoritması
Bayes Sınıflandırıcılar	Naïve Bayes
Regresyon Ağaçları	
Topluluk Öğrenmeleri	RO algoritması

Sınıflandırma yöntemlerinin uygulanma aşamasında veri setinin rasgele seçilen bir kısmı eğitim veri seti olarak kabul edilir ve kalan kısmı ise test seti olarak alınır. Başarı metriğine göre karşılaştırma yapılan sınıflandırma modelleri, eğitim setinde sınıfı bilinen örnekler için öğrenme işlemi yapacaklardır. Hesaplanan veri seti üzerinde sınıflandırma algoritmaları uygulanarak çeşitli sınıflandırma modelleri oluşturulur. Ardından test setindeki örnekler için sınıf tahminleri yapılarak, model sonucunun performans başarı değeri elde edilir.

3.1.1.1 Destek Vektör Makineleri

DVM, istatistiksel öğrenme teorisine dayalı bir gözetimli öğrenme algoritmasıdır. Veriye ilişkin herhangi birleşik dağılım fonksiyonu bilgisine ihtiyaç duymadığı için de dağılımdan bağımsız öğrenme yöntemlerindendir [46]. DVM'ler günümüzde teorik araştırmacılar ve uygulamalı bilim adamları için büyük ilgi görmektedir. Yeni çekirdek (kernel) yöntemleri teknolojisi sayesinde DVM'ler yüksek derecede

doğrusal olmayan sınıflandırıcılar oluşturmada çok başarılı olmuşlardır. DVM sınıflandırıcılarının çok fazla değişkenin olduğu durumlarda ve karmaşık yapılandırılmış verilerle başa çıkmada oldukça başarılı oldukları gözlemlenmiştir. Hem doğrusal hem de doğrusal olmayan problemler için kullanılabilirler. Doğrusal olmayan problemlerde öznitelik uzayını başka bir öznitelik uzayına dönüştürmek için çekirdek (kernel) fonksiyonları kullanılır. Daha sonra dönüşümle elde edilen yeni öznitelik uzayında sınıfları birbirinden ayıran hiperdüzlem bulunur [47]. Şekil 3.2’de DVM’ler için optimum hiperdüzlemin olası konumları gösterilmektedir. H_1 ile gösterilen doğru (genel anlamda hiperdüzlem) iki sınıfı ayırma özelliğine sahip değildir. H_2 ile gösterilen doğru ve H_3 ile gösterilen doğru iki sınıfı ayırıcı özelliktedir. Bu iki doğrudan H_3 sınıflar arasındaki mesafe dikkate alındığında H_2 ’ye göre daha optimum görülmektedir.



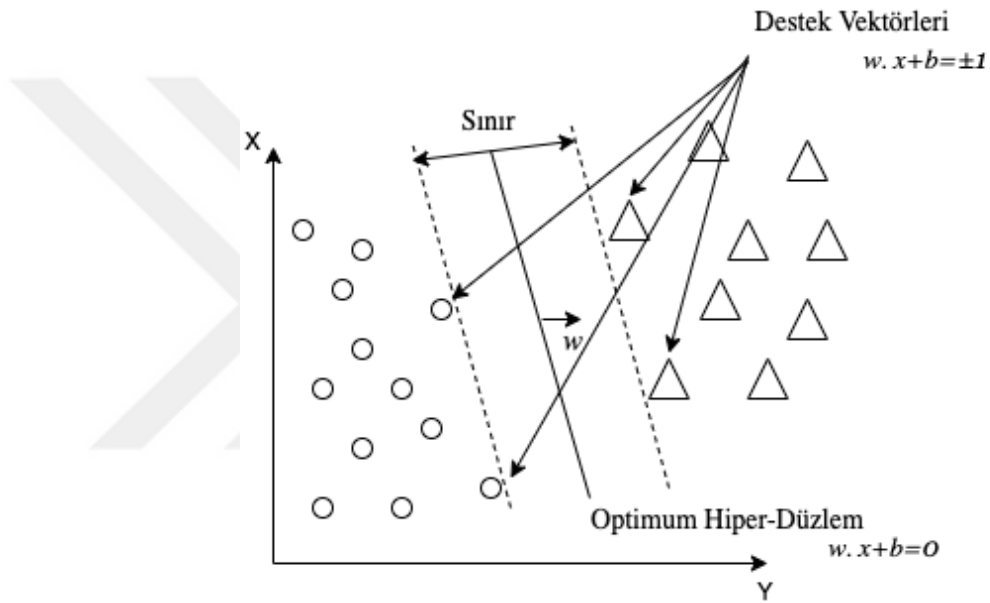
Şekil 3.2 Destek vektör makinesi [48]

DVM’ler doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan veriler için iki farklı şekilde incelenebilirler.

- **Doğrusal Olarak Ayrılabilen Veriler için DVM:** Bir sınıflandırma probleminde veriler doğrusal formda ayrılabilir ise, DVM’nin öğrenmesi için k sayıda örnekten oluşan eğitim verisinin $\{x_i, y_i\}$, $i = 1, 2, 3, \dots, k$ şeklinde olduğu kabul edilir ve optimum hiperdüzleme ait eşitsizlikler denklem 3.1’deki gibi yazılır.

$$\begin{cases} w \cdot x_i + b \geq 1, & y = 1 \\ w \cdot x_i + b \leq -1, & y = -1 \end{cases} \quad (3.1)$$

Burada $x \in R^N$ olup N-boyutlu uzayı, $y \in \{-1, +1\}$ sınıf etiketlerini, w ağırlık vektörünü bir diğer deyişle hiperdüzlemin normalini, b ise eğilim değerini göstermektedir. Optimum hiperdüzlemin belirlenebilmesi için bu düzleme paralel iki hiperdüzlemin belirlenmesi gereklidir. Bu hiperdüzlemleri oluşturan noktalarda “destek vektörleri” olarak adlandırılır ve düzlemler $w \cdot x_i + b = \pm 1$ şeklinde ifade edilirler. Şekil 3.3’te optimum hiperdüzleme paralel sınırlar ve destek vektörleri gösterilmiştir [47].



Şekil 3.3 Doğrusal ayrılabilen veri setleri için hiperdüzlemin belirlenmesi

Optimum hiperdüzleme paralel sınırın maksimuma değer alabilmesi için w ifadesi minimum hale getirilmelidir. Optimum hiperdüzlemin belirlenmesi ise denklem 3.2’deki kısıtlı optimizasyon probleminin çözümünü gerektirir.

$$\min \left[\frac{1}{2} \|w\|^2 \right] \quad (3.2)$$

Bu optimizasyon problemine bağlı sınırlamalar ise denklem 3.3’teki gibi olup optimizasyon problemi Lagrange denklemleri kullanılarak çözülebilir [47].

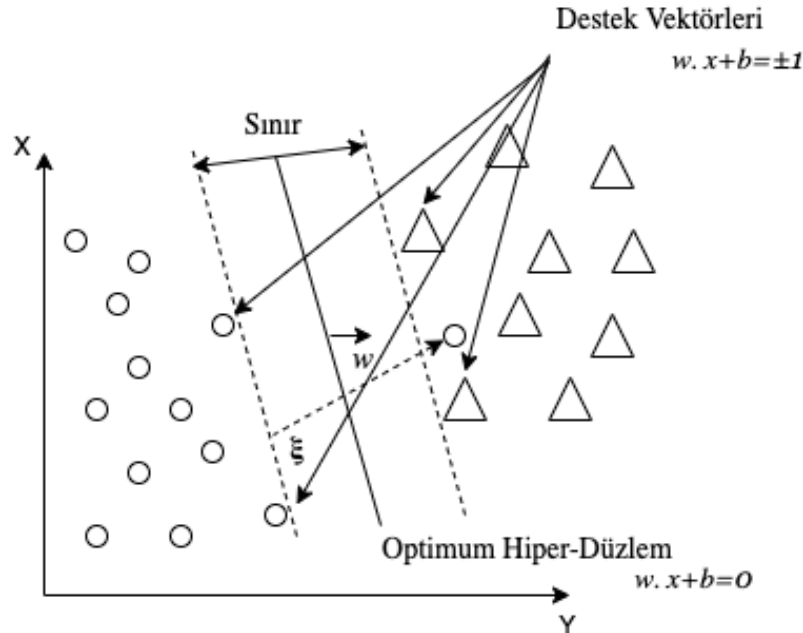
$$y_i(w \cdot x_i + b) - 1 \geq 0 \text{ ve } y_i \in \{+1, -1\} \quad (3.3)$$

Denklem çözüldüğünde, doğrusal olarak ayrılabilen iki sınıflı bir problem için karar fonksiyonu bozulma faktörü λ 'yı da içeren denklem 3.4'teki şekliyle yazılır.

$$f(x) = \text{sign}\left(\sum_{i=1}^k \lambda_i y_i (x \cdot x_i) + b\right) \quad (3.4)$$

- **Doğrusal Olarak Ayrılamayan Veriler için DVM:** Bazı problemlerin sınıflandırılması doğrusal olarak mümkün değildir. Bu durumlarda eğitim verilerinden birçoğunun optimum hiperdüzlemin diğer tarafında kalması sebebiyle yaşanan problem, pozitif bir yapay değişken olan ξ_i 'nin eklenmesiyle çözülür. Şekil 3.4'de hiperdüzlemin farklı tarafında kalan örnekler görülebilir. Hiperparametre sınırının maksimum hale getirilmesi ile yanlış sınıflandırma hatalarının minimuma dönüştürülmesi arasındaki denge; C ile ifade edilen ve değer aralığı $(0, \infty)$ arasında olan bir düzenleme parametresi yardımıyla kontrol edilebilir hale gelir [47]. ξ_i ve C parametreleri kullanılarak doğrusal olarak ayırılmaz veriler için optimizasyon problemi denklem 3.5'teki gibi olmaktadır.

$$\min \left[\frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^r \xi_i \right] \quad (3.5)$$



Şekil 3.4 Doğrusal ayrılamayan veri setleri için hiperdüzlemin belirlenmesi

DVM'ler matematiksel olarak $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ şeklinde ifade edilen bir çekirdek fonksiyonu yardımıyla doğrusal olmayan dönüşümler yapabilirler. φ değeri özellik haritası (feature map) olarak adlandırılır, doğrusal ilişki olmayan problemlerde özellik uzayını ifade eder. Bu çekirdek fonksiyonunu kullanarak doğrusal olarak ayıramayan iki sınıflı problem çözümü ile ilgili karar kuralı denklem 3.6'da gösterildiği gibi yazılır:

$$f(x) = \text{sign}\left(\sum_i \alpha_i y_i \varphi(x_i) \cdot \varphi(x_j)\right) + b \quad (3.6)$$

DVM problemlerinde düzenleme parametresi modeli kuran kişi tarafından belirlenmelidir. Bu parametrenin aldığı değer olması gerekenden çok küçük ya da çok büyük ise optimum hiperdüzlem doğru belirlenemez. Bu durumda sınıflandırma algoritmasının doğruluk değeri düşer. Diğer taraftan düzenleme parametresinin sonsuza eşit olması da DVM modelini sadece doğrusal olarak ayrılabilen veri setleri için uygun hale getirir [47].

3.1.1.2 Gradyan İniş

Gradyan iniş sınıflandırma yöntemlerinde amaç fonksiyonunu doğru şekilde minimize ederek modelin elde ettiği global minimum, veri miktarına bağlı olarak, parametre güncellemesinin doğruluğu ile güncelleme yapılması için geçen süre arasında bir denge kurulması gerekmektedir.

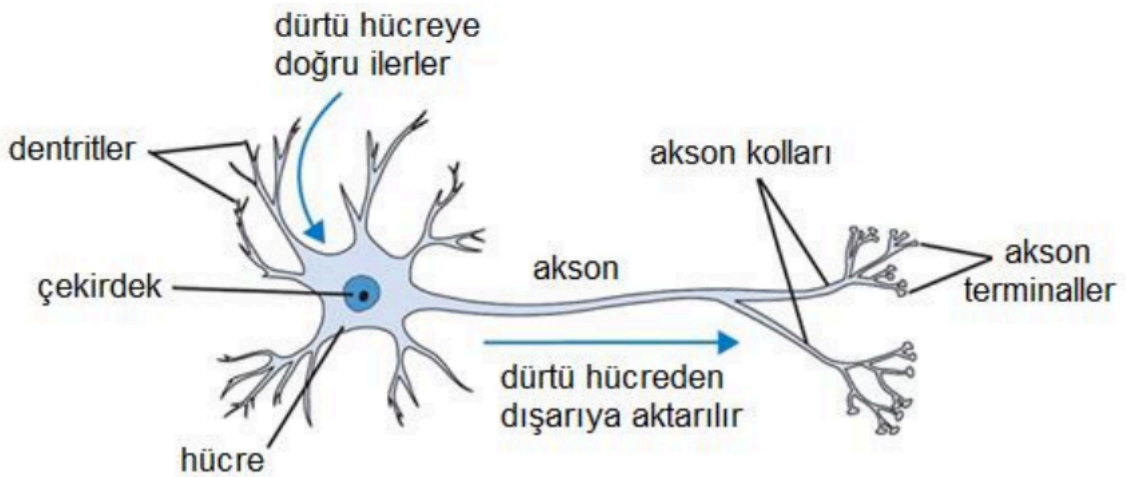
- **Batch (Toplu) Gradyan İniş:** Maliyet fonksiyonunun gradyanını hesaplar, tüm veri kümesinin yalnızca bir güncelleme yapması için gradyanlar hesaplanması gerektiğinden, batch gradyan iniş çok yavaş olabilir ve belleğe sığmayan veri kümeleri için anlaşılmasız olmaktadır.
- **Stokastik Gradyan İniş:** Batch gradyan iniş, her parametre güncellemesinden önce benzer örnekler için gradyanları yeniden hesaplar, SGİ her eğitim örneğindeki bağımsız ve bağımlı değişkenler için bir parametre güncellemesi gerçekleştirir, bu sebeple çok daha hızlıdır.
- **Mini Batch Gradyan İniş:** Mini batch gradyan iniş her iki inişinde en iyisini alır ve her mini batch, n eğitim örneği için bir güncelleme gerçekleştirir. Bu

sayede daha kararlı yakınsamaya yol açabilecek parametre güncellemelerinin varyansını azaltır. Mini batch gradyan inişi, aynı zamanda tipik olarak bir nöral ağın eğitimi sırasında tercih edilen algoritmalardandır [50].

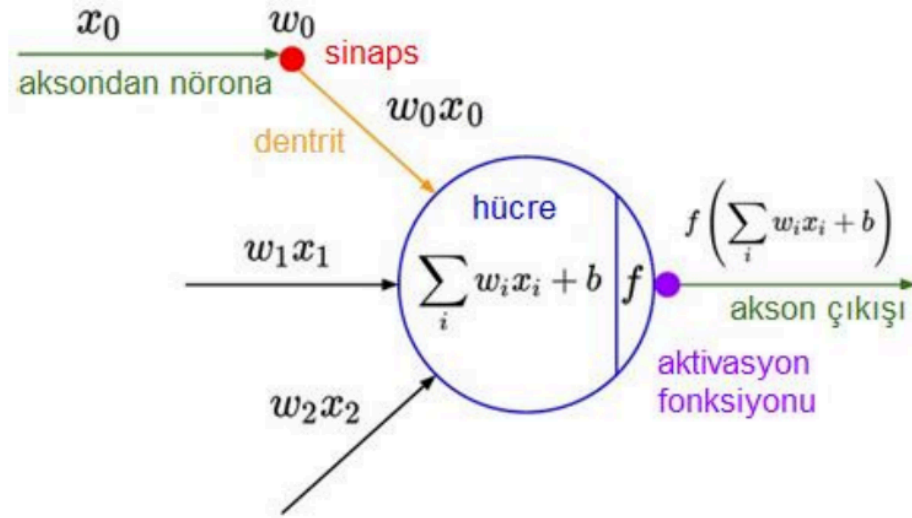
3.1.1.3 Yapay Sinir Ağları

YSA, insan beyninin en temel özelliği “öğrenme” fonksiyonunu gerçekleştiren bilgisayar sistemleridir. Öğrenmeyi örnekler yardımı ile gerçekleştiren bu ağlar birbirine bağlı yapay sinir hücrelerinden oluşur. Her bağlantının bir ağırlık değeri vardır ve YSA sahip olduğu bilgi bu ağırlık değerlerinde saklı olup ağa yayılmıştır [51]. YSA’nın temel konusu, canlıların sinir sisteminin ve karar mekanizmalarının matematiksel olarak modellenebilmesidir. YSA’da amaç, eğitilebilir, kendi kendine organize olarak uyarılabilir halde değerlendirme yapabilen YSA ile canlı beyninin öğrenme yapısının yapay olarak modellenmesi hedeflenmektedir. Makinelerin de bu yönlendirme ile insansı bir öğrenme ve karar verme mekanizmasına sahip olması beklenmektedir [52].

Şekil 3.5 ile şekil 3.6’da insan nöron sistemi ile yapay zeka nöron sistemi arasındaki benzerlikler dikkat çekmektedir.



Şekil 3.5 İnsan nöronunun (sinir hücresi) yapısı [53]

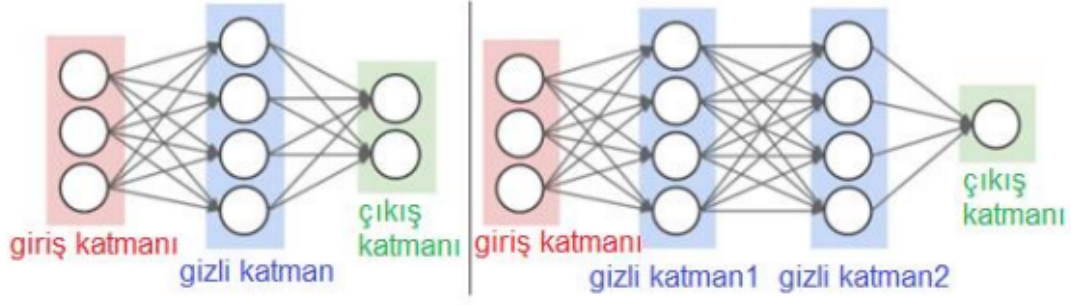


Şekil 3.6 Bir nöronun (sinir hücresi) matematiksel modeli [53]

YSA, tekli ve çoklu katmanlardan oluşabilir. Tüm katmanların geliştirilen modele katkısı olabileceği gibi, bu katmanlardaki nöronların da katmana bir katkısı vardır ve dolayısıyla da modele katkı sağlamaktadır.

Klasik bir YSA'da bir katmandaki nöronların diğer nöronlar ile aralarında bir etkileşim yoktur ve bilgiler sadece bir önceki katmandan bir sonraki katmana veya çıkışa doğru aktarılabilir. Arka arkaya gelen iki katmandaki nöronlar birbirlerini çeşitli aktivasyon değerleriyle etkilerler. Katmanlardaki nöron sayıları da modelin başarısını dolaylı yoldan da olsa etkilemektedir. Bir YSA modeli oluştururken ağırlık vektörlerinin başlangıç değerler atanması önemlidir. Ağırlıklar ilk etapta rastgele oluşturulabileceği gibi geçmişte öğrenmiş bir modelin ağırlıkları da başlangıç ağırlıkları olarak atanabilir. Ancak bu değerler tanımlanırken (+) ya da (-) değerlerden rastgele dağılım oluşturulmalıdır. Başarılı bir şekilde tasarlanan YSA modelinin başarısını ölçen metrik kayıp fonksiyonu (loss function)'dur. Bu fonksiyonun değeri optimizasyon teknikleriyle hesaplanır ve 0'a yakınsaması beklenir [54].

Şekil 3.7'de de tek ve çok katmanlı YSA modellemesinde giriş çıkış katmanları ve gizli katmanların yapıları incelenmiştir.



Şekil 3.7 Tek ve çok gizli katmanlı YSA modelleri [52]

3.1.1.4 Lojistik Regresyon

Regresyon analizi, bağımsız değişkenlerin herhangi bir veya birden çok değişkenle arasındaki ilişkiyi matematiksel olarak tanımlayan bir fonksiyondur [55]. LR ise bağımlı değişkenin ikili sınıftan oluştuğu (1 ve 0 gibi), bağımsız değişkenlerin ise kategorik veya sürekli olduğu bir çoklu regresyon çeşididir [56]. LR algoritması, bağımlı ve bağımsız değişkenler arasındaki ilişkileri, en az değişken kullanarak en iyi uyuma sahip olacak şekilde tanımlamayı amaçlar [57]. İlişkiler $y = f(x)$ gibi basit bir fonksiyon ile temsil edilebilir. Bu fonksiyonda x bağımsız değişkeni, y bağımlı değişkeni temsil eder. İstatistiksel ilişkiler göz önüne alındığında x değeri biliniyor ise y değeri belirli bir hata oranı ile tahmin edilebilir. Denklem 3.7 lojistik kümülatif dağılım fonksiyonunu tanımlamak için kullanılır.

$$y = f(x) = \frac{1}{1 + e^{-z}} \quad (3.7)$$

Burada z , “Fayda Fonksiyonu” olarak tanımlanmakta ve denklem 3.8’deki gibi gösterilmektedir.

$$z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon \quad (3.8)$$

Bu formüle göre β_0 sabit, $\beta_1, \dots, \beta_n$ ’ler ise tahmin değerleri ile çarpılan regresyon katsayılarını temsil etmektedir. n sayısı veri seti içerisindeki bağımsız değişken sayısını ifade ederken, ε değeri hata terimini ifade etmektedir. LR kesin meydana gelen bir olayın (satın alınma veya alınmama gibi) olasılığının tahminini yapabilmek için kullanıldığında, maksimum olasılırlık yaklaşımı, çoklu

iterasyonlarla çözümler deneyerek, gözlenen ve tahminlenen değerler arasındaki olası en küçük sapmayı kullanır [58].

3.1.1.5 Topluluk Öğrenmeleri

TÖ, tekli sınıflandırıcı yerine, çoklu sınıflandırıcıların sonuçlarının birleşimine dayalı, daha az hata oranına sahip tahmin modeli yaratmayı hedefler. Sınıflandırıcıların tahmin sonuçlarını birleştirmesi için istifleme (Stacking) yöntemi ilk olarak David Wolpert tarafından sunulmuştur [59]. Günümüze kadar, TÖ yöntemleri araştırmacılar tarafından farklı disiplinlerde kullanılmıştır. Bu disiplinler arasında, makine öğrenmesi ve örüntüleri tanıma öne çıkan başlıklar olmuştur [60]. Genel anlamda TÖ yöntemleri ağırlıklandırılmış ve meta-düzey olmak üzere iki grupta incelenir.

- **Ağırlıklandırılmış Yöntemler:** Sınıflandırıcılar aynı amaç için çıktı üretirlerse ve sonuçlar karşılaştırılabilir seviyede ise kullanıma uygundur [61]. Bu alandaki en popüler algoritmalarından bir tanesi oylama (Voting) algoritmasıdır.
- **Meta-Düzey Yöntemler:** Bu yaklaşımda birçok sınıflandırıcı bulunmasına karşın yalnızca bir tane ana/meta sınıflandırıcı bulunmaktadır. İlk olarak temel sınıflandırıcılar aracılığı ile tahminler yapılır, sonrasında ise bu tahminler kullanılarak öğrenme işlemi gerçekleşir ve meta seviye sınıflandırıcı oluşmuş olur. Meta sınıflandırıcının sonucu, aynı zamanda sistemin de sonucudur. Bu alandaki en bilindik algoritmalarından birisi istifleme algoritmasıdır [62].

RO yöntemi, sınıflandırma yöntemleri için kullanılan bir TÖ yöntemidir. RO, her biri birbirinden bağımsız olarak aynı dağılımı kullanan ve eğitim setinden tamamen rastgele seçilmiş bir örnekleme dayanan, karar ağaçlarından oluşan bir topluluktur. Bu yöntem, eğitim verisini işlerken birçok karar ağacı oluşturur ve daha sonra tahminleme sırasında bu karar ağaçlarının sınıflandırma sonuçlarından yararlanarak, girdinin sınıfına çoğunluk oyu aracılığıyla karar verilir. RO algoritması da dolayısı ile ağırlıklandırılmış yöntemler arasındadır [63].

3.1.1.6 Karar Ağaçları

KA yöntemleri, böl ve yönet yaklaşımından faydalanan hiyerarşik bir gözetimli öğrenme modelidir. Herhangi bir parametre gerekmeden verimli bir şekilde sınıflandırma işlemleri gerçekleştirebilen bir yöntemdir. KA, karar ve bitiş düğümlerinden oluşur. Karar düğümleri dalları etiketleyen ayrık bir test fonksiyonunu gerçekleştirir. Girdiye bu test fonksiyonu uygulanır ve sonucuna göre girdiye uygun bir dal seçilir. Bu süreç kökte başlayıp, özyinelemeli olarak bitiş düğümüne ulaşana kadar sürdürülür. Böylece girdinin ait olduğu sınıf tespit edilebilir. KA algoritmalarının diğer algoritmalara nazaran en önemli avantajı yorumlanması basit ve anlaşılır bir yapıya sahip olmasıdır.

KA, sınıflandırma amaçlı kullanıldığında dallara bölünme amacıyla farklı dallara ayırma ölçütlerinden yararlanır. Günümüzde sıklıkla denklem 3.9'daki bilgi kazanımını ölçmek için kullanılan düzensizlik (entropy) veya denklem 3.10'daki GINI katışıklığı (impurity) fonksiyonları kullanılmaktadır. Logaritmik fonksiyon hesaplaması gerektirmediği için GINI karışıklığı fonksiyonunun maliyeti, düzensizlik fonksiyonundan daha düşüktür [52].

$$H(X) = - \sum_x P(x) \log_2 P(x) \quad (3.9)$$

$$GINI(X) = 1 - \sum_x P(x)^2 \quad (3.10)$$

Hesaplama açısından düşük maliyetli olan KA algoritmaları eksik değerlerle veya ilgisiz özelliklerle çalışabilir, ancak aşırı uyum sorunu problem yaratabilir. Aşırı uyum problemini çözmek için budamadan ve TÖ yöntemlerinden RO algoritmasından yararlanılmaktadır [64].

3.1.1.7 Naïve Bayes

NB olasılıktan yararlanarak sınıflandırma amacıyla kullanılan diğer yöntemlere göre nispeten daha basit bir yöntemdir. Bayesian karar kuramına dayanır ve özetle en yüksek olasılığa sahip olan kararın seçilmesi olarak özetlenebilir [64]. NB algoritmasının en iyi performans ile çalışması için iki koşulun gerçekleşmesini

gerekir. İlk koşul, sınıfların belirlenen koşullar altında birbirlerinden bağımsız olması, ikincisi ise sonuçları etkileyebilecek olan değişkenlerin belirli olmasıdır [65]. NB yöntemi olasılık ve koşullu olasılıktan yararlanarak, girdilerin ait olduğu sınıfı bulmaya çalışır. Bu amaçla denklem 3.11'deki i . sınıf sayısını ifade etmek üzere Bayes teoreminden yararlanır [64].

$$p(c_i|x, y) = \frac{p(x, y|c_i)}{p(x, y)} \quad (3.11)$$

İki sınıflı örnek için aşağıdaki kural, tahmin edilecek sınıfa karar vermek için denklem 3.12 formunda kullanılır:

$$C(x) = \begin{cases} c_1, & p(c_1 | x, y) \geq p(c_2 | x, y) \\ c_2, & p(c_1 | x, y) < p(c_2 | x, y) \end{cases} \quad (3.12)$$

3.1.2 Regresyon Algoritmaları

Makine öğrenmesi yöntemlerinden olan regresyon algoritmaları, sınıflandırma algoritmaları gibi girdi değişkenlerini kullanarak, çıktı değişkenlerinin tahminini yapabilmek amacıyla kullanılır. Sınıflandırma algoritmalarından farkı, çıktıları kategorik olarak değil sürekli değerlerin yer aldığı tahminler ile oluşturmalarıdır. Regresyon problemleri nümerik değişken tahminlerinde kullanılır. Örneğin ev satışlarında satılacak eve verilmesi planlanan fiyat tahmini, bir ürünün satış tahmini vb. problemlerde kullanılır [66].

3.2 Gözetimsiz Öğrenme

Gözetimsiz öğrenme, etiketlenmemiş ham veri kümesindeki veriler üzerinde açıkça görülmeyen ilişkilerin açığa çıkarılması işlemidir [67]. Bu öğrenme yöntemi gözlemlere dayalı olarak öğrenir. Bir diğer deyişle, model çıktıları kullanmadan sadece girdiler üzerinde öğrenme işlemini gerçekleştirir ve gerçeklemek istediği bir hedef değişkeni bulunmadığından gözetimli öğrenmenin aksine bir hatayı minimize etmeyi hedeflemez [68]. Gözetimsiz öğrenmede amaç tahminleme yapmak değildir, genellikle öznitelikler arasındaki ilişkilerin bulunması, kümeleme, olasılık yoğunluk tahmini ve boyut indirgeme vb. amaçlar için

kullanılır. Gözetimsiz öğrenme yöntemi, gözetimli öğrenmeden farklı olarak, verileri sebep-sonuç şeklinde etiketleme yapmadan, veri içerisinde mevcut olan ilişkilerin ve yapıların öğrenilmesini esas alır. Veri seti içinde bulunan gözlemlerin birbirine olan uzaklıklarını, ilişkilerini ve yoğunluklarını kullanarak veriyle ilgili çıkarımlar yapılmasını sağlar [38].

Etiketlenmemiş verilerdeki gizli örüntüleri bulmayı amaçlayan gözetimsiz öğrenme tekniği genellikle olasılık teorisine dayanan kümeleme algoritmaları için kullanılır. Kümeleme yöntemleri, birbirine benzer olan verilerin bir araya getirilerek sınıflandırılması işlemidir. Her kümede birbirleri ile en fazla ilişkili gözlemler olmalı ve birbiriyle benzerlik göstermeyen diğer gözlemler de mümkün olduğunca farklı kümelerde olmalıdır yani kümeler arası benzerliğin en aza indirilmesi gerekmektedir. Kümeleme, makine öğrenmesinin yanısıra örüntü tanıma ve görüntü analizi gibi birçok alanda da kullanılan genel bir tekniktir [38].

3.3 Başarı Ölçüm Metrikleri

Sınıflandırma algoritmalarında model başarılarının ölçülmesinde birçok metrik kullanılır. Bu metrikler karmaşıklık matrisinden oluşturulan gerçek değerler ile tahmin değerleri arasındaki ilişkilerin formülize edilmiş halleridir. Başarı ölçüm metrikleri Tablo 3.2’de yer alan ifadeler kullanılarak oluşturulmaktadır.

Tablo 3.2 Karmaşıklık matrisi

		Mevcut Durum	
		Pozitif Durumlar	Negatif Durumlar
Tahmin	Pozitif Tahminler	TP	FP
	Negatif Tahminler	FN	TN

Sınıflandırma algoritmalarında problemin türüne göre kullanılan metrikler aşağıdaki gibidir:

- **Doğruluk Oranı (DO):** DO, doğru sınıflandırılmış (tahmin edilmiş) örnek sayısının (TP+TN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır. DO, Denklem 3.13'teki gibi gösterilir.

$$DO = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.13)$$

- **Hata Oranı (HO):** HO, Doğruluk oranının 1'e tamlayanıdır. Diğer bir ifadeyle yanlış sınıflandırılmış (tahmin edilmiş) örnek sayısının (FP+FN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır. HO, denklem 3.14'teki gibi formülüne edilir.

$$HO = \frac{FP + FN}{TP + TN + FP + FN} \quad (3.14)$$

- **Kesinlik Oranı (KO):** Var olan durumda doğru olan örneğin sınıfını doğru olarak tahmin edilmiş, True Pozitif (TP) örnek sayısının, sınıfı doğru olarak tahmin edilmiş tüm örnek sayısına (TP+FP) oranıdır. Diğer bir deyişle modelin doğru olarak tahmin ettiği değerlerin ne kadarını doğru olduğunu ifade eder. KO, denklem 3.15'teki gibi formülüne edilir.

$$KO = \frac{TP}{TP + FP} \quad (3.15)$$

- **Duyarlılık Oranı (DUO):** Var olan durumda doğru olan örneğin sınıfını doğru olarak tahmin edilmiş TP örnek sayısının, var olan durumdaki doğru olan toplam pozitif örnek sayısına (TP+FN) oranıdır. Diğer bir ifadeyle gerçekte doğru olan sınıfların modelin, ne kadarını doğru tahmin edebildiğini ifade eder. DUO, denklem 3.16'daki gibi formülüne edilir.

$$DUO = \frac{TP}{TP + FN} \quad (3.16)$$

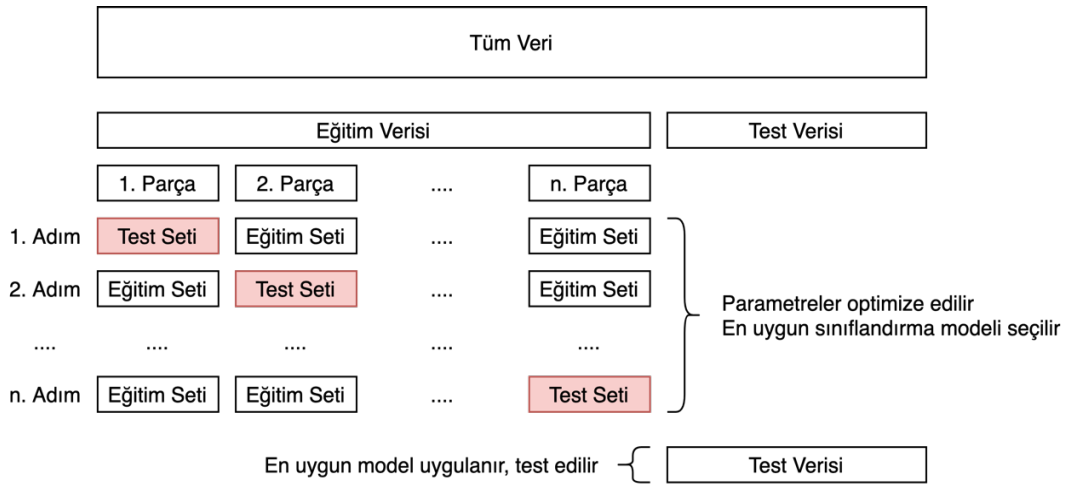
- **F-ölçütü (F1):** Kesinlik ve duyarlılık metrikleri her zaman tek başına anlamlı bir karşılaştırma sonucu çıkartılmasına yeterli olmayabilir. Her iki metriği beraber değerlendirmek genellikle daha doğru sonuçlar verir. Bunun için F1 skor tanımlanmıştır. F1 skor, kesinlik ve duyarlılık metriklerinin harmonik ortalamasıdır [67]. F1 skor, denklem 3.17'deki gibi formülüne edilir.

$$F1 = 2 * \frac{KO * DUO}{KO + DUO} \quad (3.17)$$

Başarı metriklerinin aldıkları değerler minimuma yaklaştıkça hata oranı düşeceğinden daha iyi performans gösterdikleri bilinir. Modelleme aşamasında başarı ölçümü yapılırken ölçülen kritere göre en düşük değeri alan metrik başarılı olur.

3.4 Çapraz Geçerlilik Yöntemleri

Eğitim verisi alt kümelerine ayrılarak veri setinin bir kısmı eğitim için kullanılırken kalan kısmı test/doğrulama işlemi yapmak için kullanılır. Bu işlem çapraz bir şekilde tüm alt kümeler için tekrar edilir. Modelleme yapmadan önce veri kümesi eğitim ve test seti olarak ayrılır. Ardından eğitim seti n adet eşit parçalara ayrılır. Her eşit parçada ilk grup test kümesi, diğerleri eğitim kümesi olarak atanarak sınıflandırma algoritmaları uygulanır. Bu işlem n kere her alt grup için aynı şekilde tekrar eder. Sistemin genel başarısı bu her sınıflandırma sonrası elde edilen tüm sonuçların ortalaması alınarak hesaplanmaktadır [66]. n kere çapraz geçerlilik için veri kümesi n eşit parçaya ayrılır. Her bir işlem adımında veri kümesinin $(n - 1) / n$ lik bölümü eğitim kümesi, $1 / n$ lik bölümü ise test kümesi olmak üzere kullanılır. Her bir adımda (ta ki n adım tamamlanıncaya dek) farklı bir parça test kümesi olarak alınır, kalan parçalar eğitim kümesi olacak şekilde işlem sürdürülür [69]. n kere çapraz geçerlilik Şekil 3.8'deki gibi özetlenmiştir.



Şekil 3.8 n kez çapraz geçerlilik

Tez çalışmasının uygulama kısmında kullanılacak platform, deneysel tasarım ve veri seti hakkında bilgiler bu bölümde açıklanmıştır.

4.1 Teknoloji Platformu

Çalışmada veri ön işleme aşaması, deneysel tasarım ve model kurma süreçleri Python programlama dili üzerinde gerçekleştirilmiştir. Python (C, C++ ve benzerleri gibi) programlama dili diğer programlama dilleri gibi, bilgisayarın akıllı bir şekilde kullanılmasını sağlar. Bu programlama dili Hollandalı programcı Guido Van tarafından 90'lı yılların başında geliştirilmeye başlanmıştır. Rossum bu dili, *The Monty Python* adlı bir İngiliz komedi grubunun, Monty Python's Flying Circus isimli gösterisinden esinlenerek vermiştir. Python programlarının en büyük özelliklerinden birisi, C gibi programlama dillerinin aksine, derlenmeye gerek kalmadan çalıştırılabilir olmasıdır.

Python programlama dilinde derleme işlemi olmadığı için, bu dil ile oldukça hızlı program geliştirmesi yapılabilir. Python programlama dilinin söz diziminin basit olması sayesinde hem program yazmak, hem de başkası tarafından yazılmış bir programı okumak, diğer programlama dillerine kıyasla çok daha kolaydır. Python programlama dili birçok farklı işletim sistemi ve platformlar üzerinde çalışabilir. GNU/Linux, Windows, Mac OS X, Solaris, VMS, iOS ve Android gibi pek çok ortamda Python programlama dili uygulamaları rahatlıkla geliştirilebilmektedir. Ek olarak herhangi bir ortamda yazılan Python programı, üzerinde hiçbir değişiklik yapılmadan dahi başka ortamlara taşınarak kolaylıkla çalıştırılabilir [70].

Tüm Python programlama dili sürümleri açık kaynaktır (Açık kaynak tanımı için <https://opensource.org/> adresi incelenebilir). Tarihsel olarak geliştirilen Python

programlama dili sürümlerinin hepsi olmasa da çoğu genel kamu lisansına uygundur [71]. Python programlama dili 1990 yılından bu yana geliştirilen bir dil olduğundan, bu süre içinde pek çok değişime uğramıştır ve farklı sürümlerde Python programı yazılarak programcılarının kullanımına sunulmuştur. Şu anda piyasada Python programlama dilinin 2.x serisinden bir sürümle yazılmış pek çok program bulunmakla beraber 3.x serisi de son yıllarda yaygınlık kazanmaya başlamıştır [70]. Python programlama dilinde, makine öğrenmesi metotlarının kullanılabilmesi için yazılan ve geliştirilen kütüphanelerin bulunması ve sıkça kullanılmasından dolayı sıklıkla tercih edilen programlama dilidir. Python programlama dili üzerinde matematiksel, istatistiksel işlemlerin yapılabilmesi ve makine öğrenmesi metotlarının uygulanabilmesi adına çeşitli üniversiteler ve firmalar tarafından geliştirilmiş farklı özelliklere sahip birçok kütüphane bulunmaktadır [72].

Bu çalışma kapsamında kullanılan kütüphaneler ve özellikleri tablo 4.1’de özetlendiği gibidir.

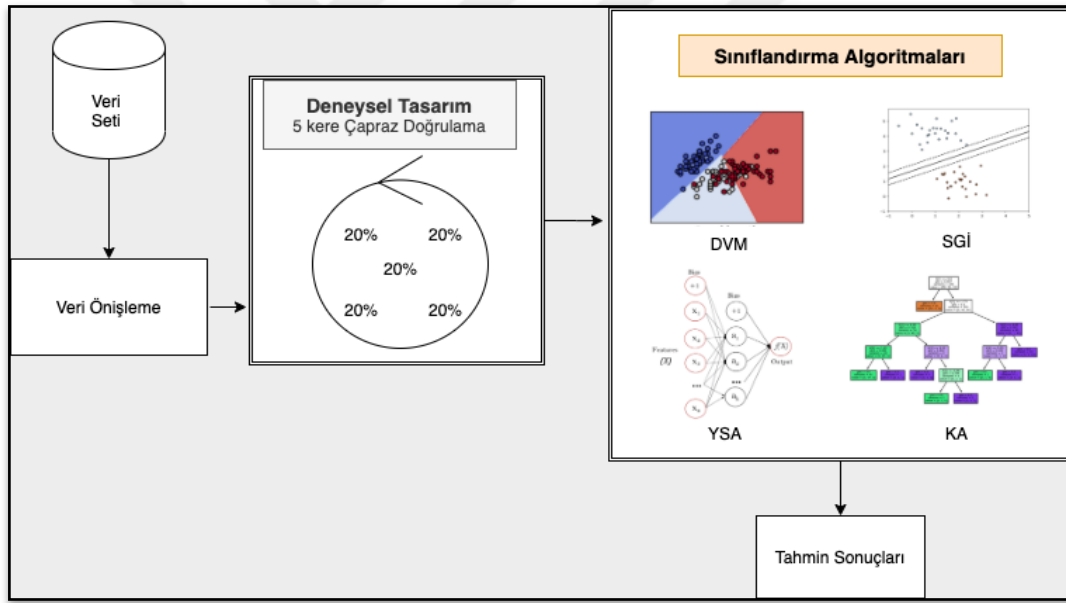
Tablo 4.1 Makine öğrenmesi kütüphaneleri ve özellikleri

Kütüphane	Özellikleri
Pandas	Veri manipülasyonu ve analizi yapmak için kullanılır. NumPy üzerine kurulmuş bir kütüphanedir.
Numpy	Bilimsel hesaplamalar yapmak için kullanılır.
Sklearn	Veri ön işleme aşamalarını gerçekleştirmek ve tahmin modelleri geliştirmek için kullanılır.
Matplotlib	Veri görselleştirme için kullanılır.

4.2 Deneysel Tasarım

Deneysel tasarım aşamasında veri setine uygulanacak adımların önceden tasarlanması ve akış diyagramının oluşturulması gerekir. Veri ön işleme aşamaları,

eğitim-test verilerinin seçimi ve veri modelleme aşamaları bu kısımda incelenir. Eğitim ve test kümelerinin seçilmesi aşamasında 5 defa çapraz geçerlilik yöntemi kullanılmıştır. Sonraki aşamada sınıflandırma işlemleri gerçekleştirilmiştir. 7 farklı sınıflandırma algoritması ile gerçekleştirilen deneylerde kullanılan algoritmalar; DVM, SGİ, YSA, LR, RO, KA ve NB algoritmalarıdır. Deneysel tasarım kısmından sonra ise önerilen modellerin etkinliklerini test etmek için istatistiksel değerlendirme kriterlerinden yararlanılmıştır. Bu çalışma kapsamında sınıflandırma algoritmalarının başarıları F1 skor ve DO baz alınarak hesaplanmış ve en başarılı algoritma bu iki metrikte de en yüksek skoru alan model seçilmiştir. Çalışmanın akış diyagramı aşağıdaki gibi olup bir sonraki bölümde veri ön işleme aşamasında uygulanan yöntemler ve sınıflandırma algoritmalarının eğitim setinde öğrendikleri modelleri test ettikleri veri modellemesi aşamasına geçilmiştir.



Şekil 4.1 Araştırmanın akış diyagramı

4.3 Veri Seti

Bu çalışmada teknoloji perakende sektöründe faaliyet gösteren şirketlerden bir tanesinin son 2 yıllık (2018-2019) ham satış verisi ile satın alımı yapan müşterilere ait bazı demografik ve müşteri satın alım hareketlerini içeren bilgiler kullanılmıştır. Maskelenmiş müşteri id'lerini içeren ve üzerinde öğrenme

gerçekleştirilen veri seti 2,190,325 örnek ve bir gerçek değerli cevabı (ek garanti satın alınıp/alınmayacağı) tahmin etmeyi amaçlayan 23 özellikten/değişkenden oluşmaktadır. Her bir örnek ayrı bir sipariştir. Veri kümesi, oldukça büyük olduğundan çalışma 2 milyon gözlemi içeren ana ürün gruplarından akıllı telefon, televizyon ve dizüstü bilgisayarları içeren siparişler üzerinde incelenmiştir. İncelenen veri seti farklı müşteriler tarafından yapılan sipariş detayları ile ilgili müşterilerin alışveriş davranışlarını içermektedir.

Bu davranışlar müşterilerin;

- Müşteri veritabanı üzerinden hesaplanmış değer ve davranış segment bilgilerini,
- Daha önceki dönemlerde yaptıkları ürün iade adetlerini,
- Son 5 yıl içerisindeki toplam harcama tutarlarını,
- Kaç farklı kategoriden alışveriş yaptıklarını,
- Yeni lanse edilen ürünü hemen satın alma adetlerini,
- Üst segment ürünleri satın alma adetlerini,
- Müşterilerin firmadaki yaşı, gelme sıklığı ve firmaya bıraktığı maddi kazanç bilgilerini,
- Kredi/nakit/kart kullanım adetlerini içermektedir.

Tüm bu müşteri ve satış verisi üzerinden oluşturulan veri kümesi üzerinde bir öğrenme yapılarak, müşterinin bir sonraki ana ürün alışverişinde ek garanti satın alıp almayacağını tahmini yapılacaktır. Veri setinin oldukça büyük olması 8GB ön belleğe sahip olan yerel bilgisayar üzerinde performans sorunlarına yol açtığından, algoritmalar mevcut makinenin 4 katı daha hızlı olan 32GB'lık ön belleğe sahip sanal sunucu üzerinde çalıştırılmıştır.

4.4 Veri Ön İşleme

İlk aşamada sınıflandırma tahminlerinin daha verimli yapılması için sayısal değer içeren verilere 0-1 arasında normalizasyon işlemi uygulanmıştır. Bu işlem her değişkeni aynı ölçeğe getirmiştir. Normalizasyon yöntemi olarak denklem 5.1'de verilen Min-Max yöntemi kullanılmıştır.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (5.1)$$

Min-max yöntemi ile sayısal değer alan değişkenler gerçek değerlerini temsil eden sayılar ile normalize edilerek x' formunda güncellenmiştir.

Veri kümesinde bulunan sipariş tarihlerini ek garanti satın alım modeli hakkında daha çok bilgi sahibi olmamız için eğitirken kullanabilmek adına siparişlerin yapıldığı hafta, ay, yıl ve hafta içi/hafta sonu(dönem) bilgileri de yeni alanlar olarak veri kümesine eklenmiştir.

Veri kümesinde, müşterilerin yaşadıkları il bilgisi yüksek oranda (%86) boş olduğu için bu alanın modeli yanıltmaması adına veri kümesinden çıkartılmıştır.

Kategorik değerler makine öğrenmesi yöntemleri ile tahmin algoritmalarına dahil edilebilmesi için sayısal formata getirilmelidir. Python programlama dili üzerinde Sklearn kütüphanesinde yer alan One Hot Encoding fonksiyonu, bu işlemi yapabilmeyi sağlar. Bu fonksiyon kategorik değişkenleri ikili olarak temsil eder. Bu işlem öncelikle kategorik değerlerin tamsayı değerleriyle eşlenmesini esas alır. Daha sonra, her tamsayı değeri 1 ile işaretlenmiş tamsayı indeksi dışında ki tüm değerleri sıfır olan ikili vektör olarak temsil eder.

Tablo 4.2’de seçilen müşteri id’leri için müşterilerin yaptıkları alışverişlerde satın alınan ana ürün grupları üzerinde bu One Hot Encoding fonksiyonunun nasıl çalıştığı örneklenmiştir.

Tablo 4.2 One Hot Encoder fonksiyonu ile kategorik değişken dönüşümü

Müşteri ID	Segment		Müşteri ID	S_Akıllı_Telefon	S_Dizüstü_Bilgisayar	S_TV
1	Akıllı Telefon		1	1	0	0
5	Dizüstü Bilgisayar	→	5	0	1	0
6	Akıllı Telefon		6	1	0	0
10	TV		10	0	0	1

Tablo 4.2'ye göre Müşteri ID değeri 1 iken akıllı telefon satın alımı yapıldığı, Müşteri ID değeri 5 iken dizüstü bilgisayar alımı yapıldığı, Müşteri ID değeri 6 iken akıllı telefon alımı yapıldığı ve Müşteri ID değeri 10 iken TV alımı yapıldığı görülmektedir. Burada üç kategori ürün olduğundan dönüşüm aşamasında üç sütun oluşturulup satın alınan ürüne 1, diğer ürünlere 0 değeri atanmaktadır.

Gözlem değerleri incelendiğinde %60 ve üzeri gözlemin boş olduğu alanlar modele dahil edilmemiştir. Bu duruma örnek olarak maskelenen müşterilerin sipariş yaptıkları il bilgileri verilebilir. İl bilgilerinin %78'i boş değer içerdiğinden öznitelik uzayına dahil edilmemiştir. Boş gözlem içeren alanlar, modelleme algoritmalarına girdi olarak dahil edilemezler ancak bu alanlar modeli eğitirken yok sayılabileceği gibi bazı yöntemler ile veri manipilasyonu yapılarak modele dahil edilebilir. Ancak bu işlem, gözlem değerlerinin yüksek miktarda boş olması durumunda sentetik veri üretilerek model başarısını düşüreceği varsayımı yapılarak müşterilerin sipariş yaptıkları il bilgileri veri setinden çıkartılmıştır.

Kategorik değişkenlerin sayısallaştırılması/nümerikleştirilmesi ve veri ön işleme aşamasının tamamlanması ile veri kümesi modellemeye hazır hale gelmiş ve müşteri id hariç tutulduğunda toplam 36 öznitelik ve bir hedef değişkeni oluşmuştur. Tablo 4.3'de modele girdi olarak verilen öznitelikler özetlenmiştir. Tablo 4.3'de Müşteri Ürün Satın Alım Bilgileri özniteliği, iade adedi, harcama tutarı, sipariş tarihi/ay-gün-yıl, nakit/kredi kartı kullanım, müşteri yaşı vb. özniteliklerden oluşmaktadır.

Tablo 4.3 Öznitelik sayıları ve açıklamaları

Öznitelik Tanımı	Öznitelik Adedi	One Hot Encoder Sonrası Öznitelik Adedi
Müşteri Segment Bilgileri	4	15
Müşteri Ürün Satın Alım Bilgileri	18	18
Alışveriş Yapılan Kategori	1	3
Toplam	23	36

Hedef deęiřkeni yalnızca 1 ve 0 deęerleri alabilir ve satın alınan ana ürünün yanında ek garanti satın alınıp-alınmadığını belirtir. Ek garanti hedef deęiřkeni 1'e eřit olduęunda o sipariřte ek garanti satın alındığı, 0 olarak etiketlendiğinde ise o sipariřte ek garanti satışının yapılmadığı anlaşılmaktadır.

Uygulama bölümünde farklı oranlar ile test edilen veri setinde hedef deęiřkeninin alacağı deęeri tahmin eden algoritmalar uygulanmıştır.



Bu bölümde çalışmada kullanılan veri seti üzerinde kullanılan modeller ve modellerin gösterdikleri başarıların karşılaştırması yapılarak en iyi sonucu veren modelin tercih edilmesi ve elde edilen bulgular üzerinde durulmuştur.

5.1 Veri Modelleme

Veri kümesini modelleme aşamasına geçmeden önce başarının ölçülebilmesi için model eğitim – test kümesi olarak farklı oranlarda iki gruba ayrılmıştır. Veri seti sırasıyla %5'i, %10'u, %15'i, %20'si ve %25'i test seti olacak şekilde ayrılmış ve buna karşılık geriye kalan %95, %90, %85, %80 ve %75'lik gözlemler de eğitim verisi olarak kabul edilmiştir. Her bir ayırım yeni bir tasarım modeli olarak konu alınmış ve incelenerek sonuçlar paylaşılmıştır.

5 tasarım modelindeki tüm eğitim setlerine 5 kere çapraz geçerleme yapılarak sınıflandırma algoritmalarının model başarıları test edilmiştir. Model başarısı, çapraz geçerleme sonucunda F1 skor ve DO metrikleri ile ölçülmüş ve not edilmiştir. Buna göre her tasarım modelinde en başarılı seçilen algoritmalar test veri setlerinde uygulamaya alınmıştır. Oluşturulacak tasarım modelleri ve veri seti ayrımları Tablo 5.1'de gösterilmiştir.

Tablo 5.1 Tasarım modelleri

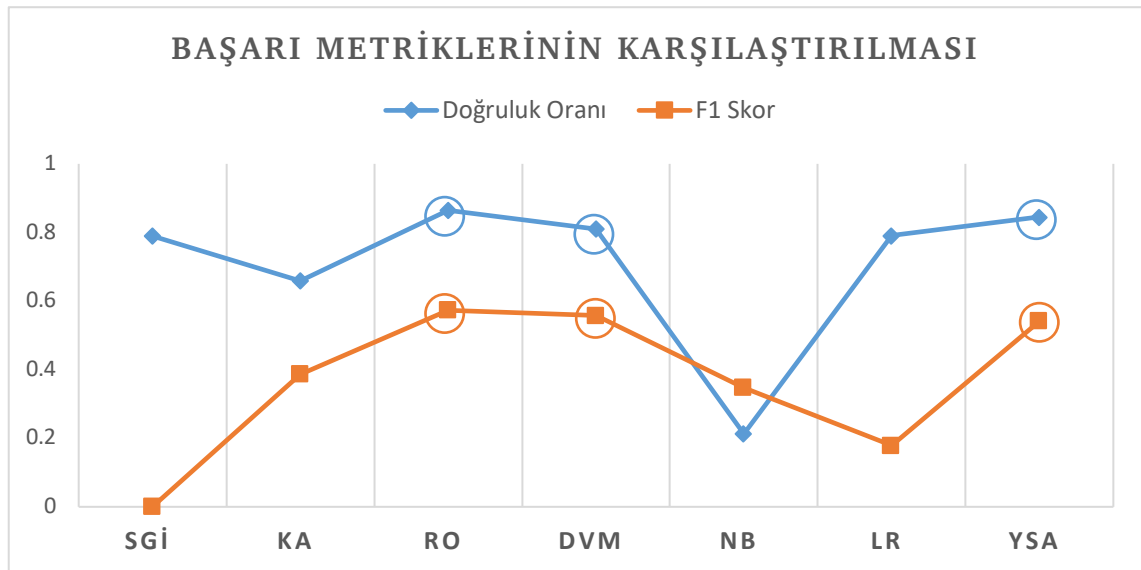
	Tasarım Modeli 1	Tasarım Modeli 2	Tasarım Modeli 3	Tasarım Modeli 4	Tasarım Modeli 5
Eğitim Seti	%95	%90	%85	%80	%75
Test Seti	%5	%10	%15	%20	%25

5.1.1 %5 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları

Tasarım modeli 1 olarak planlanan bu modelde eğitim ve test olmak üzere 2 parçaya ayrılan veri setinin %95'i eğitim için ayrılmıştır. 1,900,000 gözlem içeren eğitim veri setinde sırasıyla her biri 5 kere çapraz geçerlemeye giren algoritmaların F1 skor ve DO değerleri Tablo 5.2'de paylaşılmıştır.

Tablo 5.2 Tasarım modeli 1 çapraz geçerlilik başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
SGİ Sınıflandırıcısı	0.7896	0.0005
KA Sınıflandırıcısı	0.6590	0.3871
RO Sınıflandırıcısı	0.8645	0.5730
DVM Sınıflandırıcısı	0.8102	0.5576
NB Sınıflandırıcısı	0.2128	0.3474
LR Sınıflandırıcısı	0.7905	0.1784
YSA (MLP) Sınıflandırıcısı	0.8451	0.5416



Şekil 5.1 Tasarım modeli 1 başarı metriklerinin karşılaştırılması

F1 skor ve doğruluk oranı metriklerine göre en başarılı seçilen algoritmalar Şekil 5.1’de de görüldüğü gibi sırasıyla topluluk öğrenmeleri algoritmalarından RO, YSA ve DVM algoritmalarıdır. En başarılı ilk üç algoritma test veri setinde aldığı sonuçlar tablo 5.3’teki gibidir.

Tablo 5.3 Tasarım modeli 1 test seti başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
RO Sınıflandırıcısı	0.8766	0.5705
YSA (MLP) Sınıflandırıcısı	0.8405	0.5130
DVM Sınıflandırıcısı	0.8007	0.5117

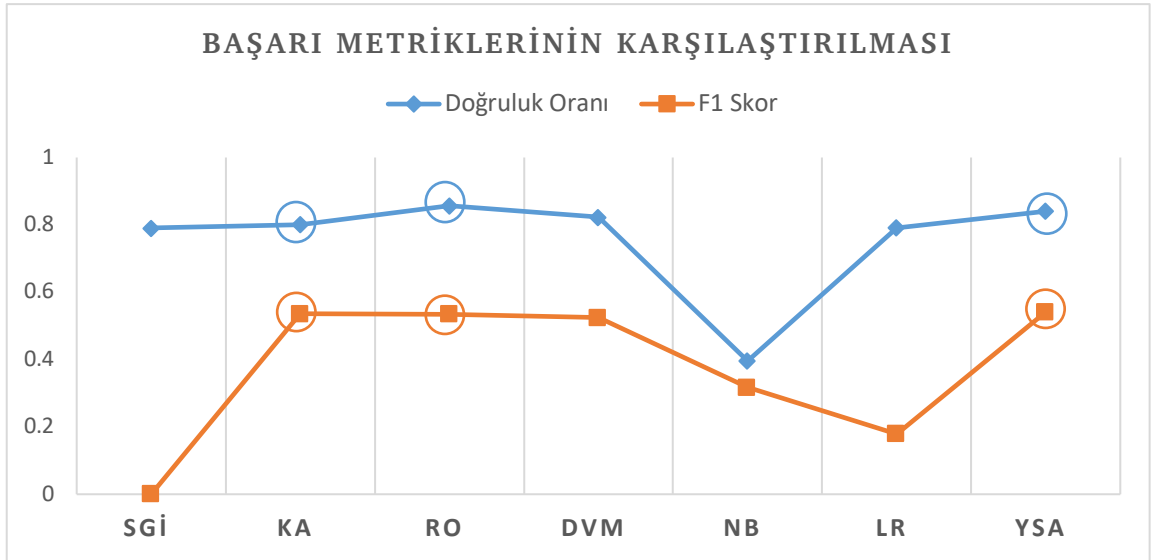
En yüksek DO ve F1 skor değerine sahip olan algoritma RO sınıflandırıcısıdır.

5.1.2 %10 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları

Tasarım modeli 2 olarak planlanan bu modelde eğitim ve test olmak üzere 2 parçaya ayrılan veri setinin %90’ı eğitim için ayrılmıştır. 1,800,000 gözlem içeren eğitim veri setinde sırasıyla her biri 5 kere çapraz geçerlemeye giren algoritmaların F1 skor ve DO değerleri Tablo 5.4’te paylaşılmıştır.

Tablo 5.4 Tasarım modeli 2 çapraz geçerlilik başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
SGİ Sınıflandırıcısı	0.7905	0.0001
KA Sınıflandırıcısı	0.8035	0.5408
RO Sınıflandırıcısı	0.8581	0.5407
DVM Sınıflandırıcısı	0.8031	0.5305
NB Sınıflandırıcısı	0.5660	0.3238
LR Sınıflandırıcısı	0.7915	0.1776
YSA (MLP) Sınıflandırıcısı	0.8405	0.4961



Şekil 5.2 Tasarım modeli 2 başarı metriklerinin karşılaştırılması

F1 skor ve doğruluk oranı metriklerine göre en başarılı seçilen algoritmalar sırasıyla topluluk öğrenmeleri algoritmalarından RO, YSA ve KA algoritmalarıdır. En başarılı ilk üç algoritma test veri setinde aldığı sonuçlar tablo 5.5'te özetlenmiştir.

Tablo 5.5 Tasarım modeli 2 test seti başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
RO Sınıflandırıcısı	0.8632	0.5672
YSA (MLP) Sınıflandırıcısı	0.8336	0.4832
KA Sınıflandırıcısı	0.8096	0.5556

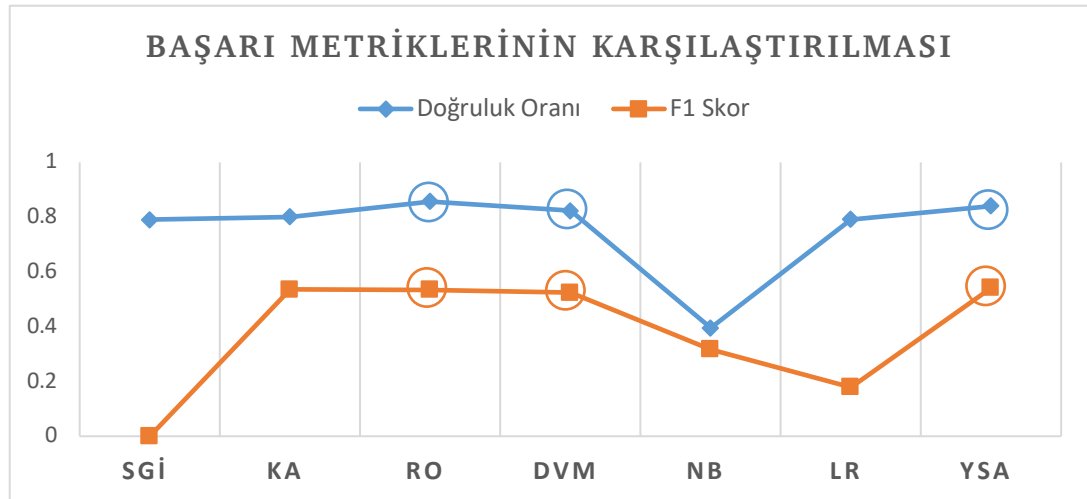
Tablo 5.5'te görüldüğü gibi en yüksek DO ve F1 skor değerine sahip olan algoritma RO sınıflandırıcısıdır.

5.1.3 %15 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları

Tasarım modeli 3 olarak planlanan bu modelde eğitim ve test olmak üzere 2 parçaya ayrılan veri setinin %85'i eğitim için ayrılmıştır. 1,700,000 gözlem içeren eğitim veri setinde sırasıyla her biri 5 kere çapraz geçerlemeye giren algoritmaların F1 skor ve DO değerleri Tablo 5.6'da paylaşılmıştır.

Tablo 5.6 Tasarım modeli 3 çapraz geçerlilik başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
SGİ Sınıflandırıcısı	0.7907	0.0004
KA Sınıflandırıcısı	0.8006	0.5333
RO Sınıflandırıcısı	0.8582	0.5401
DVM Sınıflandırıcısı	0.8480	0.5362
NB Sınıflandırıcısı	0.3948	0.3200
LR Sınıflandırıcısı	0.7920	0.1782
YSA (MLP) Sınıflandırıcısı	0.8408	0.5112



Şekil 5.3 Tasarım modeli 3 başarı metriklerinin karşılaştırılması

F1 skor ve doğruluk oranı metriklerine göre en başarılı seçilen algoritmalar sırasıyla topluluk öğrenmeleri algoritmalarından RO, DVM ve YSA algoritmalarıdır. En başarılı ilk üç algoritma test veri setinde aldığı sonuçlar tablo 5.7’de özetlenmiştir.

Tablo 5.7 Tasarım modeli 3 test seti başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
RO Sınıflandırıcısı	0.8630	0.5522
DVM Sınıflandırıcısı	0.8501	0.4670
YSA (MLP) Sınıflandırıcısı	0.8330	0.4664

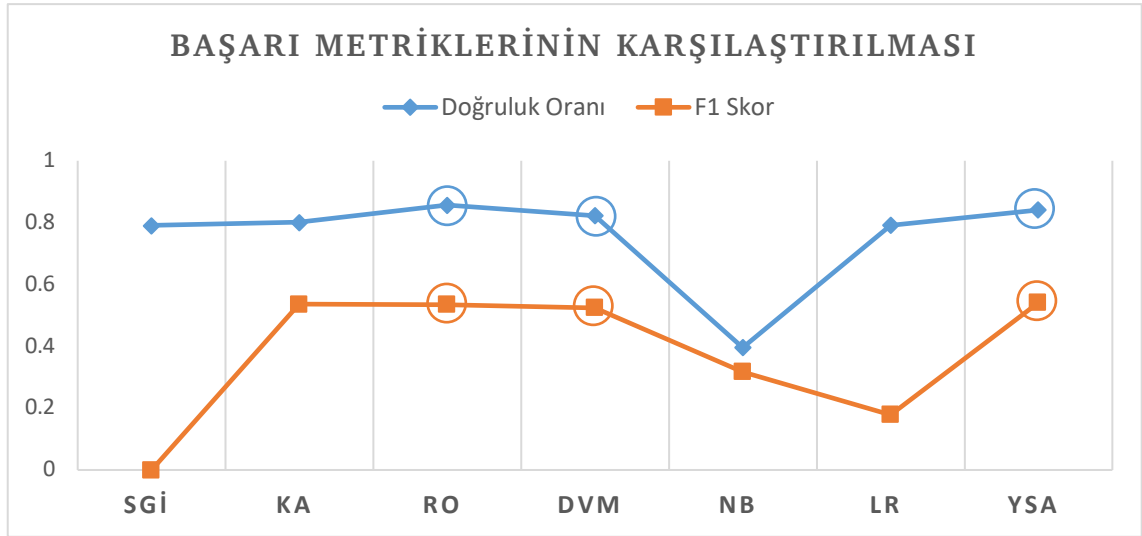
Tablo 5.7’de görüldüğü gibi en yüksek DO ve F1 skor değerine sahip olan algoritma RO sınıflandırıcısıdır.

5.1.4 %20 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları

Tasarım modeli 4 olarak planlanan bu modelde eğitim ve test olmak üzere 2 parçaya ayrılan veri setinin %80’i eğitim için ayrılmıştır. 1,600,000 gözlem içeren eğitim veri setinde sırasıyla her biri 5 kere çapraz geçerlemeye giren algoritmaların F1 skor ve DO değerleri Tablo 5.8’de paylaşılmıştır.

Tablo 5.8 Tasarım modeli 4 çapraz geçerlilik başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
SGİ Sınıflandırıcısı	0.7906	0.0001
KA Sınıflandırıcısı	0.8011	0.5356
RO Sınıflandırıcısı	0.8568	0.5352
DVM Sınıflandırıcısı	0.8231	0.5243
NB Sınıflandırıcısı	0.3958	0.3171
LR Sınıflandırıcısı	0.7917	0.1788
YSA (MLP) Sınıflandırıcısı	0.8387	0.5460



Şekil 5.4 Tasarım modeli 4 başarı metriklerinin karşılaştırılması

F1 skor ve doğruluk oranı metriklerine göre en başarılı seçilen algoritmalar sırasıyla topluluk öğrenmeleri algoritmalarından RO, YSA ve DVM algoritmalarıdır. En başarılı ilk üç algoritma test veri setinde aldığı sonuçlar tablo 5.9'da özetlenmiştir.

Tablo 5.9 Tasarım modeli 4 test seti başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
RO Sınıflandırıcısı	0.8530	0.5500
YSA (MLP) Sınıflandırıcısı	0.8234	0.5430
DVM Sınıflandırıcısı	0.8202	0.5234

Tablo 5.9'da görüldüğü gibi en yüksek DO ve F1 skor değerine sahip olan algoritma RO sınıflandırıcısıdır.

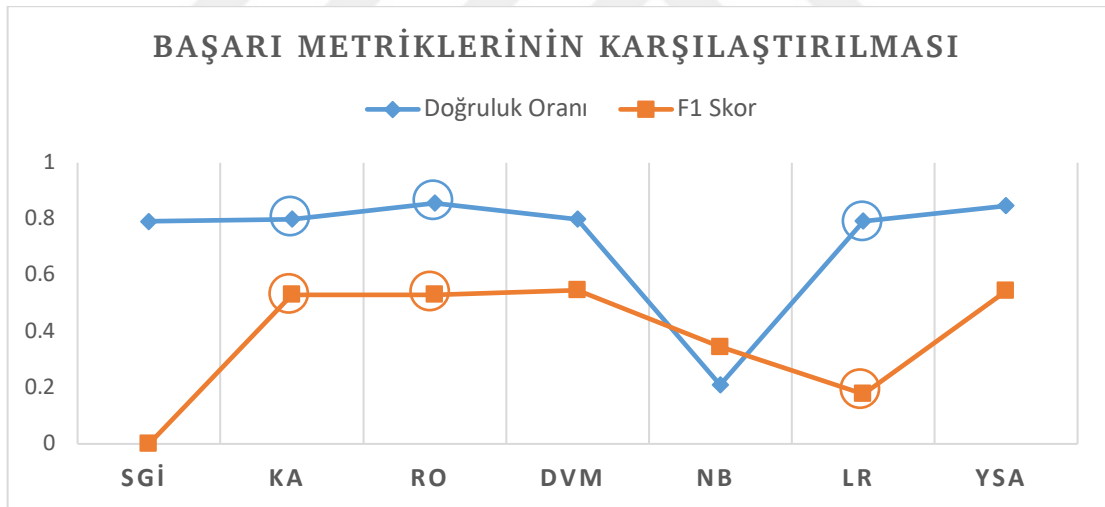
5.1.5 %25 Test Setinin Kullanılması ile Gerçekleştirilen Sınıflandırma Performansları

Tasarım modeli 5 olarak planlanan bu modelde eğitim ve test olmak üzere 2 parçaya ayrılan veri setinin %75'i eğitim için ayrılmıştır. 1,500,000 gözlem içeren

eğitim veri setinde sırasıyla her biri 5 kere çapraz geçerlemeye giren algoritmaların F1 skor ve DO değerleri Tablo 5.10'da paylaşılmıştır.

Tablo 5.10 Tasarım modeli 5 çapraz geçerlilik başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
SGİ Sınıflandırıcısı	0.7912	0.0000
KA Sınıflandırıcısı	0.7992	0.5303
RO Sınıflandırıcısı	0.8564	0.5306
DVM Sınıflandırıcısı	0.8001	0.5467
NB Sınıflandırıcısı	0.2099	0.3456
LR Sınıflandırıcısı	0.7923	0.1780
YSA (MLP) Sınıflandırıcısı	0.8475	0.5449



Şekil 5.5 Tasarım modeli 5 başarı metriklerinin karşılaştırılması

F1 skor ve doğruluk oranı metriklerine göre en başarılı seçilen algoritmalar sırasıyla topluluk öğrenmeleri algoritmalarından RO, YSA ve DVM algoritmalarıdır. En başarılı ilk üç algoritma test veri setinde aldığı sonuçlar tablo 5.11'de özetlenmiştir.

Tablo 5.11 Tasarım modeli 5 test seti başarı ölçümlerinin karşılaştırılması

Sınıflandırma Algoritması	Doğruluk Oranı	F1 Skor
RO Sınıflandırıcısı	0.8502	0.5588
YSA (MLP) Sınıflandırıcısı	0.8395	0.4532
DVM Sınıflandırıcısı	0.7878	0.4530

Tablo 5.11’de görüldüğü gibi en yüksek DO ve F1 skor değerine sahip olan algoritma RO sınıflandırıcısıdır.

5.2 Tüm Tasarım Modellerinin Karşılaştırılması

Yapılan çalışma sonucunda teknoloji perakendesinde satılan veriler üzerinde makine öğrenmesi yöntemleri farklı tasarım modelleri üzerinde test edilmiştir. Gerçekleştirilen çalışmada çapraz geçerlilik yöntemi ile sistemin eğitim ve test performansı ölçülmüştür. Çapraz geçereleme için $k=5$ olarak alınmıştır. Gerçekleştirilen çalışma sonuçlarına göre RO algoritması diğer makine öğrenmesi algoritmalarından daha yüksek F1 ve DO değeri ile sınıflandırmayı tamamlamıştır.

Tasarım modellerinde çapraz geçerlilik yöntemi sonuçlarında algoritmaların başarı sıralamaları değişse de aldıkları puanlar arasındaki farkın çok fazla olmadığı görülmüştür. Tablo 5.12’de tasarım modellerindeki test sonuçlarının farklılaştığı noktalar özetlenmiştir.

Tablo 5.12 Tasarım modelleri test seti başarı ölçümlerinin karşılaştırılması

Tasarım Modeli	En Başarılı Sınıflandırıcı	Doğruluk Oranı	F1 Skor
Tasarım Modeli 1	RO Sınıflandırıcısı	0.8766	0.5705
Tasarım Modeli 2	RO Sınıflandırıcısı	0.8632	0.5672
Tasarım Modeli 3	RO Sınıflandırıcısı	0.8630	0.5522
Tasarım Modeli 4	RO Sınıflandırıcısı	0.8530	0.5500
Tasarım Modeli 5	RO Sınıflandırıcısı	0.8502	0.5588

Tablo 5.12’de de görüldüğü gibi, test setindeki veriler başarı metrikleri ile ölçümlendiğinde en başarılı sınıflandırıcı her tasarım modelinde RO sınıflandırıcısı olarak görülmektedir. Veri setinin %5’ini test setine ayrıldığında eğitim verisi diğer tasarım modellerindeki eğitim setlerine göre en fazla gözlem değerine sahiptir. Bu durum veri setindeki gözlem değerleri arttıkça, modelin gözlem değerlerini daha iyi ilişkilendirdiğini ve öğrenmeyi daha başarılı yaptığını göstermektedir.

Diğer tasarım modellerinde, test veri setleri üzerinde yapılan karşılaştırmalarda elde edilen sonuçların birbirlerine oldukça yakın olduğu görülmektedir.



Bu çalışmada, makine öğrenmesi yöntemleri, ek garanti bilgisi ve satış tarihçesi günümüze kadarki gelişimi ve elde edilen başarıları aktarmanın yanında temel olarak perakende sektöründe tercih edilen makine öğrenmesi mimarileri teorik ve detaylı olarak anlatılmıştır.

Çalışmada, makine öğrenmesi algoritmalarından sınıflandırma yöntemleri ile modelleme çalışmaları gerçekleştirilmiştir. Önerilen yöntemde makine öğrenmesinin önemli aşamalarından olan veri ön işleme, modelleme ve performans değerlendirme adımları bulunmaktadır.

Çalışma sonucunda gerçek bir veri kümesi üzerindeki makine öğrenmesi yöntemlerinin performansı karşılaştırılmıştır. Bu tez çalışması, bir deneyimin aktarılmasıdır. Bu çalışma, bundan sonraki benzer çalışmalara bir temel oluşturmakla beraber üzerinde çalışılan makine öğrenmesi algoritması için karşılaştırma yapılabilmesini sağlayacak çıktılar sunmaktadır. Gerçekleştirilen çalışma sonucunda RO algoritmasının yüksek başarısı, bundan sonraki perakende sektörü üzerinde ek garanti yönetimini veriye dayandırarak satış artışı sağlayacak ve satın alım tahminlemesi yapılması için bu tür çalışmalarda temel oluşturacaktır.

DO ve F1 skor metriklerine göre değerlendirilen sınıflandırma algoritmaları çeşitli oranlarda eğitim ve test setine ayrılan tasarım modelleri ile ölçümlenmiştir. Oluşturulan 5 farklı tasarım modeli için çapraz geçerlilik yöntemi sonucunda en yüksek skoru alan ilk üç algoritma test verisi üzerinde değerlendirmeye alınmıştır. Test verisi üzerinde ölçülen metriklere göre, DO ve F1 skorlarında topluluk öğrenmesi yöntemlerinden RO sınıflandırma algoritmaları tüm tasarım modellerinde en iyi sonucu vermekle beraber, DVM ve YSA algoritmaları da yüksek başarı göstermiştir. DVM ve YSA algoritmalarından bir kaç puan fark ile

öne çıkan RO sınıflandırıcıları perakende sektöründe satın alım yaparken ek ürün tahminlemede kullanılabilecek, en yüksek DO ve F1 skor değerine sahip olan algoritma olmaktadır.

Çalışmanın öne çıkardığı bir diğer bulgu ise veri setinin önemidir. Tasarım modeli 1'in en başarılı sonuçları vermesinden de anlaşılacağı gibi veri setindeki gözlem miktarı ne kadar artarsa algoritma başarılarının o ölçüde artacağı gözlemlenmiştir. Eğitim verisinde en yüksek gözlem değerine sahip olan tasarım modeli 1 (%95 eğitim seti) için RO sınıflandırıcısının yüzdelik skalada değer alan DO değeri %87.66 iken, eğitim verisinde en düşük gözlem miktarına sahip olan tasarım modeli 5'de (%75 eğitim seti) bu oran %85.02 olmaktadır. %85 ve üzerinde olan DO değerlerini 1 birim arttırmanın oldukça zor olduğu bilindiğinden bu artışın ciddi bir farkı belirttiği gözardı edilmemelidir.

Tasarım modeli 2'de ise DVM sınıflandırıcısı yerini KA sınıflandırıcısına bırakmıştır. KA sınıflandırıcısı, test veri setinin %10'unu oluşturduğu modelde %80.35 DO değerine sahipken; DVM sınıflandırıcısı KA sınıflandırıcısına çok yakın olan %80.31 değerini almıştır. Bu kapsamda tasarım modeli 2 için KA sınıflandırıcısı DVM sınıflandırıcısına alternatif olarak düşünülebilmektedir.

LR sınıflandırıcısı tüm tasarım modellerinde DO değeri ortalama %79 olmaktadır. Hiç bir model için test setini ölçümlemeye girecek kadar başarılı olamasa da DVM sınıflandırıcısına alternatif olarak düşünülebilmektedir.

F1 skor değerleri genel anlamda DO değerleri ile pozitif bir ilişki içinde olduğu görülse de LR sınıflandırıcısı için DO değeri %79'larda iken F1 skor'un %17'lerde kalması bu durumun her zaman doğru olmadığını göstermektedir. Bu anlamda en başarılı modeli seçerken yalnızca DO veya sadece F1 metriklerine bakmanın araştırmacıyı yanıltacağını, en azından iki farklı başarı metriği ile beraber incelenmesinin daha sağlıklı sonuçlar vereceği de çalışmanın bir çıktısı olmaktadır.

Tüm tasarım modelleri incelendiğinde NB sınıflandırıcısı tüm sınıflandırıcılar içerisinde en düşük başarı oranlarını veren algoritma olmaktadır. NB

sınıflandırıcısı zaman kaybı ve olası maliyetleri engellemek adına, teknoloji perakendesinde ek ürün satın alımı yapılırken değerlendirmeye alınmaması kararı bu çalışma sonucunda ortaya çıkan önemli bir diğer bulgu olmaktadır.

Bu bulgular doğrultusunda literatürdeki çalışmalar ile geliştirilen tez konusu, makine öğrenmesi yöntemlerinin farklı sektör uygulamalarında modellerin başarı ölçümleri ile karşılaştırmalarını aktarılmıştır. Ek garanti satışlarını veriye dayandırarak yönetmek isteyen firmalar için de yol gösterici bir içerik oluşturulmuştur. Son yılların büyük problemlerinden biri olan karlılığı yüksek ürün gruplarının satış yönetimini analitik yöntemlerle gerçekleştirilmesi ve insan gücünü en aza indirerek tüm süreçleri otomatize etmesi için makine öğrenmesi üzerinde çalışması gereken başarılı değerlendirmeler, alınan hem akademik hem pratik bir çalışma alanıdır.

İlerleyen çalışmalarda ek garanti yönetimi satış tahmininde giriş parametrelerinin tespiti için öznitelik seçme algoritmalarından yararlanılacaktır. Ana ürün grupları ile veri setini kısıtlamadan tüm ürün gruplarının ek garanti satışları ile ilişkisi incelenmeli yani daha kapsamlı bir veri seti üzerinde de deneylerin tekrarlanması ve elde edilen sonuçların bu çalışmada elde edilen sonuçlarla da karşılaştırılması hedeflenmektedir.

Araştırma teknoloji perakendesi şirketlerinden yalnızca biri baz alınarak yapılmıştır. Sektördeki diğer şirketlerin de araştırmaya dahil edilmesiyle sektörel analizin daha detaylı ve anlamlı sonuçlar vereceği düşünülmektedir. Çalışma aynı zamanda sadece akıllı telefon, dizüstü bilgisayar ve tv satışları özelinde yapılmıştır, sektördeki tüm ürün gruplarının çalışmaya dahil edilmesi ek fayda açısından yararlı olacağı düşünülmektedir.

- [1] Crothers, K., 5 Advantages to selling after-sales services, 2014. <https://impact.cisco.com/c/2014/05/>, [Eriřim Tarihi: 10 Ocak 2020].
- [2] Olivia, R., Kallenberg ve R., “Managing the transition from products to services”, International Journal of Service Industry Management. 14 (2), ss. 160-172, 2003.
- [3] Aaker D.A., “Measuring brand equity across products and markets”, California Management Review. 38, ss. 102– 120, 1996.
- [4] Samuel A., “Some Studies in Machine Learning Using the Game of Checkers”, IBM Journal of Research and Development, Temmuz 1959.
- [5] Pandey M., “Machine Learning in Formal Verification”, FMCAD 2016.
- [6] Russel S. ve Norvig P., “Artificial Intelligence: A Modern Approach (2.Bas.)”, Prentice Hall, 1995-2003.
- [7] Bhardwaj B. K. ve Pal S., “Data Mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security, Vol. 9, No. 4, Nisan 2011.
- [8] Danso S. O., “An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain”, Eylül 2006.
- [9] Shajahaan S.S., Shanthi S. ve ManoChitra V., “Application of Data Mining Techniques to Model Breast Cancer Data, International Journal of Emerging Technology and Advanced Engineering”, Volume 3, Issue 11, Kasım 2013.
- [10] Cunningham S. J. ve Holmes G., “Developing innovative applications in agriculture using data mining”, 1999.
- [11] Pang B., Lee L. ve Vaithyanathan S., “Thumbs up?: sentiment classification using machine learning techniques”, Proceedings of the ACL 02 conference on Empirical methods in natural language processing, ss. 79-86, 2002.

- [12] Hormozi A. M., ve Giles S., “Data Mining: A Competitive Weapon for Banking and Retail Industries,” *Information Systems Management*, 21:2, ss. 62-71, 2004.
- [13] Raju P.S., Bai V. R. ve Chaitanya G. K., “Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries”, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2, Issue 1, Ocak 2014.
- [14] Sudhakar M. ve Reddy C. V. K, “Two Step Credit Risk Assesment Model For Retail Bank Loan Applications Using Decision Tree Data Mining Technique”, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Volume 5 Issue 3, Mart 2016.
- [15] Cumby C., Fano A. E. ve Ghani R., Krema M., “Predicting customer shopping lists from point-of-sale purchase data”, *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* August, ss. 402-409, 2004.
- [16] Ahmeda R. A. E., Shehaba M. E., Morsya S. ve Mekawiea N., “Performance Study of Classification Algorithms for Consumer Online Shopping Attitudes and Behavior Using Data Mining”, *Fifth International Conference on Communication Systems and Network Technologies*, 2015.
- [17] Abor J., “Industry classification and the capital structure of Ghanaian SMEs”, *Studies in Economics and Finance*, Ağustos 2007.
- [18] Ramageri B. M. ve Desai B. D., “Role Of Data Mining In Retail Sector”, *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 5, Ocak 2013.
- [19] Minamia C. ve Dawson J., “The CRM process in retail and service sector firms in Japan: Loyalty development and financial return”, *Journal of Retailing and Consumer Services*, Volume 15, Issue 5, ss. 375-385, Eylül 2008.
- [20] Tinabo R., “Decision Tree Technique for Customer Retention in Retail Sector”, *International Conference on Integrated Computing Technology*, *Integrated Computing Technology* ss. 123-131, 2011.

- [21] Dr. Bhasin M. L, “Data Mining: A Competitive Tool in the Banking and Retail Industries”, The Chartered Accountant, ss. 588-594, Ekim 2006.
- [22] Ling C. X. ve Li C., “Data Mining for Direct Marketing: Problems and Solutions”, KDD, 1998.
- [23] Zhang C. ve Ruan J., “A Modified Apriori Algorithm with Its Application in Instituting Cross-Selling Strategies of the Retail Industry”, International Conference on Electronic Commerce and Business Intelligence, Temmuz 2019.
- [24] Hemalatha M., “A Predictive Modeling of Retail Satisfaction: A Data Mining Approach to Retail Service Industry”, 2013.
- [25] Gu N., “Digitalization of Retailing with Machine Learning: A Service-Product Design for Digital Merchandisers to Implement Machine Learning”, Ağustos 2019.
- [26] Yu Q., Wang K., Strandhagen J. O. ve Wang Y., “Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study”, International Workshop of Advanced Manufacturing and Automation, Advanced Manufacturing and Automation VII ss. 11-17, 2017.
- [27] Rinsaka, K. ve Sandoh, H., “A stochastic model on an additional warranty service contract. Computers and Mathematics with Applications”. 51, ss. 179-188, 2006.
- [28] Desai, P.S. ve Padmanabhan, P., Durable good, extended warranty and channel coordination”, Review of Marketing Science. 2, pp.1-23, 2004.
- [29] Jack, N. ve Murthy, D.N.P., “A flexible extended warranty and related optimal strategies. Journal of the Operational Research Society”, ss. 1612-1620, 2007.
- [30] Murthy, D.N.P., Solem, O. ve Roren, T., “Product warranty logistics: Issues and challenges”, European Journal of Operational Research. 156, ss. 110-126, 2004.
- [31] Berner, R., “The warranty windfall: Service contracts are cash cows—but retailers are mum about their importance”, Business Week, ss. 84–86, Aralık 2004

- [32] Warranty Week, “Extended Warranty Profits”, <http://www.warrantyweek.com/archive/ww20041026.html>, [Eriřim Tarihi: 08 Mart 2020].
- [33] Li, K., Chhajed, D., ve Mallik, S., “Design of extended warranties in supply chains. Working paper”, Urbana, IL: University of Illinois, 2005.
- [34] Kaya, H., ve Köymen, K, “Veri madencilięi kavramı ve uygulama alanları, Doęu Anadolu Bölgesi Arařtırmaları”, ss. 159-164, 2008.
- [35] Schapire, R. E., “The boosting approach to machine learning: An overview. In Nonlinear estimation and classification”, ss. 149-171, 2003.
- [36] Dietterich, T. G., “Machine-learning research, AI magazine”, 18(4), 97, 1997.
- [37] Chao W. L., “Machine Learning Tutorial”, 2011.
- [38] Gentleman R., Huber W ve Carey V. J., S”upervised machine learning-In Bioconductor Case Studies”, ss. 121-136, 2008.
- [39] Uzun E., “İnternet Tabanlı Bilgi Eriřimi Destekli Bir Otomatik Öğrenme Sistemi” Doktora Tezi, Trakya Üniversitesi, Edirne, 2007.
- [40] Arslan H., “Sakarya üniversitesi web sitesi erişim kayıtlarının web madencilięi ile analizi”. Sakarya Üniversitesi Fen Bilimler Enstitüsü, Sakarya, 2008.
- [41] Onat A., “Veri madencilięinin web tabanlı uygulamalarda insan uyumluluklarının tespiti üzerine bir çalıřma”, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 2008.
- [42] Caruana, R. ve Niculescu-Mizil, A., “Anempirical comparison of supervised learning algorithms”. In Proceedings of the 23rd international conference on Machine learning, ACM, ss. 161-168, 2006.
- [43] Sebastiani, F., “Machine learning in automated text categorization”, ACM computing surveys (CSUR)", 34(1), 1-47, 2002.
- [44] Han, J. ve Kamber, M., “Data Mining: Concepts and Techniques”, Elsevier Inc., Third Edition, USA, 2012.
- [45] Özkan Y. ve Erol Ç.S., “Biyoenformatik DNA Mikrodizi Veri Madencilięi”, Papatya Yayıncılık, İstanbul, 2015.

- [46] Soman, K.P., Loganathan, R. ve Ajay V., “Machine learning with SVM and other kernel methods”. PHI Learning Pvt. Ltd.2011.
- [47] Bao H. T., “Kernel Methods and Support Vector Machines”, Japan Advance Institute of Science and Technology John von Neumann Institute, VNU-HCM.
- [48] ZackWeinberg, Wikimedia Commons, File: Svm separating hyperplanes [https://commons.wikimedia.org/w/index.php?title=File:Svm_separating_hyperplanes_\(SVG\).svg&oldid=217578095](https://commons.wikimedia.org/w/index.php?title=File:Svm_separating_hyperplanes_(SVG).svg&oldid=217578095), [Eriřim Tarihi: 01 Mart 2020].
- [49] Karpathy A., “Stanford CS class CS231n: Convolutional Neural Networks for Visual Recognition”, Stanford University, Course Notes, 2018.
- [50] DevHunter, “Gradyan İniř Optimizasyon Algoritmalarına Genel Bakıř”, <https://devhunteryz.wordpress.com/2019/06/04/gradyan-inis-optimizasyon-algoritmalarina-genel-bakis/> [Eriřim Tarihi: 09 Mart 2020]
- [51] Öztemel E., “Yapay Sinir Ağları”, Papatya Yayıncılık Eğitim, İstanbul, 2003.
- [52] Alpaydın E., “Yapay Öğrenme”, Boğaziçi Üniversitesi Yayınevi, Türkiye, 2011.
- [53] A. Karpathy, Stanford University, Stanford CS class CS231n: Convolutional Neural Networks for Visual Recognition, Course Notes, [Eriřim Tarihi: 09.03.2020].
- [54] D. E. Rumelhart, G. E. Hinton ve R. J. Williams, “Learning Representations by Back-Propagating Errors”, Nature, 323, ss. 533-536, 1986.
- [55] Hosmer, David W. ve Lemeshow, Stanley ve Sturdivant, Rodney X., “Applied Logistic Regression”, Third edition, Wiley, 2013
- [56] Field ve Andy, “Discovering Statistics Using SPSS”, Sage Publications, 2005.
- [57] Çokluk Ö., “Lojistik regresyon Analizi: Kavram ve Uygulama,” Kuram ve Uygulamada Eğitim Bilimleri, cilt 10, sayı 3, ss. 1357-1407, 2010.
- [58] McCullagh P. ve Nelder J. A., “Generalized linear models”, Vol. 37 CRC press, 1989.
- [59] Wolpert David H., ”Stacked generalization.” Neural networks 5.2, ss. 241-259, 1992.

- [60] Sewell ve Martin. “Ensemble learning”, RN 11.02, 2008.
- [61] Sigletos G., “Combining information extraction systems using voting and stacked generalization.” The Journal of Machine Learning Research 6, ss. 1751-1782, 2005.
- [62] Dzeroski S. ve Zenko B., “Is combining classifiers with stacking better than selecting the best one?”, Machine learning 54.3, ss. 255-273, 2004.
- [63] Breiman L., “Random Forests”, Machine Learning, 45(1), ss. 5-32, 2001.
- [64] Harrington P., “Machine Learning in Action”, New York, USA, Manning Publications, 2012.
- [65] G. H. John ve P. Langley, “Estimating continuous distributions in Bayesian classifiers,” in: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, ss. 338-345, 1995.
- [66] Kutlugün M. A., “Gözetimli Makine Öğrenmesi Yoluyla Türe Göre Metinden Ses Sentezleme”, İstanbul Zaim Üniversitesi, İstanbul, 2017.
- [67] Nizam H., “Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması”, XIX. Türkiye'de İnternet Konferansı, 2014.
- [68] Hinton G. ve Sejnowski T.J., “Unsupervised Learning and Map Formation: Foundations of Neural Computation”, MIT Press, ISBN 0-262-58168-X. MIT Press, ISBN 0-262- 58168-X,1999.
- [69] Biricik, G., “Metin Sınıflama İçin Yeni Bir Özellik Çıkarım Yöntemi.” İstanbul, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, 2011.
- [70] Özgül F., “Python 3 için Türkçe Klavuz”, Sürüm 3, 2016.
- [71] Rossum G. V., “Python Tutorial Realise 3.8.1”, Python Software Foundation, 2020.
- [72] Y. Poyraz ve S. Sevgen, “GPU Programlama Tekniği ile Yüksek Performanslı Araç Takibi”, Bilişim Teknolojileri Dergisi, 10(3),ss. 255-261, 2017.
- [73] Scikit-learn yazılım geliştiricileri, Python Kütüphanesi, https://scikitlearn.org/stable/supervised_learning.html#supervised-learning, 2007-2019, [Erişim Tarihi: 05 Mart 2020].

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgisi: sedalacan@gmail.com

Konferans Bildirileri

1. Alacan S. ve Öz E., “Makine Öğrenmesi Yöntemleri ile Ek Garanti Satın Alım Tahminlemesi”, 20. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik (Eyi2020) Sempozyumu, 2020.

