

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BANKA PAZARLAMA
TAHMİNİ

YÜKSEK LİSANS TEZİ
Egemen Türkmen
1101020018

Anabilim Dalı: Bilgisayar Mühendisliği
Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Prof. Dr. Özgür Koray ŞAHİNGÖZ

ŞUBAT 2021

T.C.
İSTANBUL KÜLTÜR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BANKA PAZARLAMA
TAHMİNİ

YÜKSEK LİSANS TEZİ
Egemen Türkmen
1101020018

Anabilim Dalı: Bilgisayar Mühendisliği
Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Prof. Dr. Özgür Koray ŞAHİNGÖZ
Diğer Jüri Üyeleri: Doç. Dr. Akhan AKBULUT
Dr. Öğretim Üyesi Hakan AYDIN

ŞUBAT 2021

ÖNSÖZ

“MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BANKA PAZARLAMA TAHMİNİ” adlı yüksek lisans tez çalışmam süresince bilgi ve deneyimi ile çalışmalarımı yönlendiren ve desteğini esirgemeyen değerli tez danışmanım Prof. Dr. Özgür Koray Şahingöz’e, her durumda şartsız ve koşulsuz desteklerini ve sevgilerini benden esirgemeyen aileme, katkıda bulunan tüm hocalarıma ve arkadaşlarıma teşekkürlerimi sunarım.



İÇİNDEKİLER

ÖNSÖZ	iii
İÇİNDEKİLER	iv
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ	viii
KISALTMALAR	ix
ÖZET	x
ABSTRACT	xii
1. GİRİŞ	1
1.1. Problem Tanımı	3
1.2. Amaç ve Kapsam.....	3
1.3. Literatüre Katkıları	3
1.4. Tezin Organizasyonu	4
2. ÖN BİLGİLER ve LİTERATÜR ARAŞTIRMASI	5
2.1. Bankacılıkta Pazarlama ve Pazarlama Türleri	5
2.1.1. Bankacılıkta Pazarlama.....	5
2.1.2. Pazarlama Türleri	5
2.1.2.1. Dijital Pazarlama Türleri	5
2.1.2.2. Doğrudan Pazarlama Türleri.....	6
2.2. Makine Öğrenmesi.....	7
2.2.1. Makine Öğrenmesi Tanımı ve Amacı.....	7
2.2.2. Makine Öğrenmesi Yöntemleri	7
2.2.3. Yararlanılan Makine Öğrenmesi Algoritmaları.....	11
2.2.3.1. Destek Vektör Makinesi Algoritması (SVM).....	11
2.2.3.2. Karar Ağacı Algoritması (DT).....	12
2.2.3.3. Naif Bayes Algoritması (NB).....	13

2.2.3.4.	K-En Yakın Komşu Algoritması (KNN)	13
2.2.3.5.	Rastsal Orman Algoritması (RF)	14
2.2.3.6.	Ekstra Ağaç Algoritması (ET)	15
2.2.3.7.	Torbalama Algoritması (BA).....	15
2.2.3.8.	Gradyan Artırma Algoritması (GBA)	15
2.2.3.9.	Yapay Sinir Ağı (ANN).....	16
2.3.	Derin Öğrenme	16
2.3.1.	Derin Öğrenme Tanımı ve Amacı	16
2.3.2.	Derin Öğrenme Mimarileri	16
2.3.2.1.	Derin İnanç Ağları (DBN)	16
2.3.2.2.	Konvolüsyonel Sinir Ağları (CNN).....	17
2.3.2.3.	Tekrarlayan Sinir Ağları (RNN).....	17
2.3.3.	Yararlanılan Derin Öğrenme Yöntemleri	18
2.3.3.1.	Uzun Kısa-Süreli Bellek (LSTM)	18
2.3.3.2.	Geçitli Tekrarlayan Birim (GRU)	19
2.3.3.3.	Basit Tekrarlayan Sinir Ağı (SRNN).....	19
2.4.	Literatür Araştırması	19
2.5.	Veri Seti	23
2.5.1.	Yararlanılan Veri Seti	23
2.5.2.	Yararlanılan Veri Setindeki Özellikler	24
3.	YÖNTEM.....	27
3.1.	Normalizasyon İşlemi	27
3.2.	SMOTE Uygulaması	28
3.3.	K-Katlamalı Çapraz Doğrulama (CV).....	28
3.4.	Karışıklık Matrisi (CM)	29
4.	ÖNERİLEN MODEL	31
4.1.	Modelin Akış Diyagramı	31
4.2.	Modelde Kullanılan Teknoloji	32

5. ALINAN SONUÇLAR VE DEĞERLENDİRMELERİ.....	33
6. SONUÇLAR.....	38
KAYNAKÇA.....	39



ŞEKİL LİSTESİ

Şekil 1.1-Veri Bilimi Venn Diagramı	1
Şekil 1.2-Veri Bilimi ve Alt Kırılımları [23].....	2
Şekil 2.1 - Makine Öğrenmesi Yöntemleri.....	8
Şekil 2.2 - Sınıflandırma ve Regresyon.....	8
Şekil 2.3 - Kümeleme.....	9
Şekil 2.4 - İlişkilendirme	9
Şekil 2.5 - Yarı Denetimli Öğrenme [24].....	10
Şekil 2.6 - Pekiştirmeli Öğrenme.....	11
Şekil 2.7 - Destek Vektör Makinesi Çalışma Biçimi	12
Şekil 2.8 - Karar Ağacı Yapısı.....	12
Şekil 2.9- KNN Algoritması.....	14
Şekil 2.10 - Rastsal Orman Çalışma Yöntemi	14
Şekil 2.11 - Torbalama Algoritması.....	15
Şekil 2.12 - Derin İnanç Ağları Mimarisi.....	17
Şekil 2.13 - Konvolüsyonel Sinir Ağları Mimarisi [27].....	17
Şekil 2.14 - Tekrarlayan Sinir Ağları Mimarisi [33].....	18
Şekil 3.1 - SMOTE Tekniği.....	28
Şekil 3.2 - K-Katlamalı Çapraz Doğrulama Modeli.....	29
Şekil 3.3 - Karışıklık Matrisi	29
Şekil 4.1 - Kullanılan Model	31
Şekil 5.1 - SMOTE Kullanılmadan Alınan Sonuçlar – Makine Öğrenmesi.....	33
Şekil 5.2 – SMOTE Kullanılarak Alınan Sonuçlar – Makine Öğrenmesi.....	34
Şekil 5.3 - SMOTE Kullanılmadan Alınan Sonuçlar- Derin Öğrenme.....	35
Şekil 5.4 - SMOTE Kullanılarak Alınan Sonuçlar- Derin Öğrenme	36

TABLO LİSTESİ

Tablo 2-1-Veri Seti Bilgisi	23
Tablo 2-2 Veri Seti Özellik Bilgileri.....	24
Tablo 2-3 Özelliklerin Değerleri ve Sayısal Karşılıkları.....	26
Tablo 5-1 Literatürdeki Çalışmalar ile Alınan Sonuçların Karşılaştırılması	37



KISALTMALAR

SVM	: Destek Vektör Makinesi Algoritması
KNN	: K-En Yakın Komşu Algoritması
RF	: Rastsal Orman Algoritması
NB	: Naif Bayes Algoritması
DT	: Karar Ağacı Algoritması
ANN	: Yapay Sinir Ağları
GBA	: Gradyan Arttırma Algoritması
ET	: Ekstra Ağaç Algoritması
BA	: Bagging (Torbalama) Algoritması
CV	: Çapraz Doğrulama
CM	: Karışıklık Matrisi
LSTM	: Uzun Kısa-Süreli Bellek
GRU	: Geçitli Tekrarlayan Birim
SMOTE	: Sentetik Azınlık Aşırı Örnekleme Tekniği

Üniversite	:	T.C. İstanbul Kültür Üniversitesi
Enstitüsü	:	Lisansüstü Eğitim Enstitüsü
Anabilim Dalı	:	Bilgisayar Mühendisliği
Program	:	Bilgisayar Mühendisliği
Tez Danışmanı	:	Prof. Dr. Özgür Koray ŞAHİNGÖZ
Tez Türü ve Tarihi	:	Yüksek Lisans – Şubat 2021

ÖZET

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BANKA PAZARLAMA TAHMİNİ

Günümüzde aktif olarak hayatımızın büyük bir bölümünde yer alan mobil aletler ve internet ile, insanlar ve onlara çeşitli alanlarda hizmet eden şirketler birbirleri ile etkileşim yöntemlerini çoğunlukla bu iletişim organları üzerinden yapar hale gelmiştir. Bu süreçler genellikle firmaların kendi ürettikleri programlar üzerinden ya da var olan paket programlar kullanılarak sürdürülmektedir. Kullanılan bu programlar süreç mükemmelliği açısından önemli olmakla birlikte, bu programlar üzerinden elde edilen çeşitli işlenmiş ve işlenmemiş veriler son zamanlarda çok daha önem arz eder hale gelmiştir. Elde edilen bu verilerden özellikle müşteri verileri firmalar için altın niteliği kazanmıştır. Bu veriler işlenerek ve işlenen veriler üzerinden detaylı analizler çıkartılarak firmalar müşterilerini daha çok tanımaya başlamış ve bunun sonucu olarak firmaların daha iyi pazarlama yöntemleri geliştirilmesine ve bu doğrultuda pazarlama sonuçlarında alınan verimin artmasına sebebiyet vermektedir. Bu nedenden dolayı firmalar son birkaç senedir verilerin önemini anlamış ve veri işleme çalışmalarını hızlandırmıştır. Veri işleme çalışmalarına hız veren firmalar arasında ülkelerin ekonomik gücü olan bankalarda bulunmaktadır. Bankalar müşterileri ile sayısız iletişime girerek, müşterilerinin finansal süreçlerini yönetmelerinde onlara kolaylık sağlayarak onlara destek olma amacındadırlar.

Bu tez çalışmasında bir bankanın telefon görüşmeleri ile aradıkları kişiye bankalarında vadeli bir hesap açtırmaya çalıştığı ve bu çalışmalarda başarı oranlarının sonuçlarının olduğu bir veri seti incelenecek olup, bu incelemeler doğrultusunda var olan sonuçların iyileştirilmesi için Karar Ağacı, Naif Bayes, K-En

Yakın Komşu, Destek Vektör Makinesi, Rastsal Orman, Ekstra Ağaçlar, Adaboost, Gradient Boosting gibi makine öğrenmesi algoritmaları ile Uzun Kısa-Sürelî Bellek, Geçitli Tekrarlayan Birim, Basit Tekrarlayan Sinir Ağları gibi derin öğrenme yöntemleri ilgili veri seti ile kullanılarak ayrıntılı ve karşılaştırmalı bir çalışma yapılmıştır.

Anahtar Kelimeler: Pazarlama Tahmini, Makine Öğrenmesi, Derin Öğrenme, Sınıflandırma, Phyton.



University : T.C. İstanbul Kültür University
Institute : Institute of Graduate Studies
Department : Computer Engineering
Program : Computer Engineering
Thesis Advisor : Prof. Dr. Özgür Koray ŞAHİNGÖZ
Degree Awarded And Date : MA – February 2021

ABSTRACT

BANK MARKETING ESTIMATION WITH MACHINE LEARNING METHODS

With mobile devices and the internet, which are actively involved in a large part of our lives today, people and companies serving them become on these communication organs by using their interactions with each other. These processes are usually carried out through the programs produced by the companies or by using existing package programs. Although these programs are important in terms of process excellence, various processed and unprocessed data obtained through these programs have become much more important recently. Especially customer data has gained importance for companies from these data obtained. By processing these data and making detailed analyzes on the processed data, companies become more familiar with their customers and as a result, companies develop better marketing methods and increase the efficiency of marketing results in this direction. For this reason, companies have understood the importance of data in the last few years and accelerated data processing. Among the companies that accelerate the data processing work are the banks with the economic power of the countries. Banks aim to support their customers by facilitating them in managing their financial processes by making numerous contacts with their customers.

In this thesis, a data set in which a bank tries to open a time deposit account in their banks through phone calls and the results of the success rates in these studies will be examined. In order to improve the existing results with these examinations, some machine learning and deep learning algorithms such as; Decision Tree, Naive Bayes,

K-Nearest Neighbor, Support Vector Machines, Random Forest, Extra Trees, Gradient Boosting, Bagging, Artificial Neural Network, LSTM, GRU and SimpleRNN will be use.

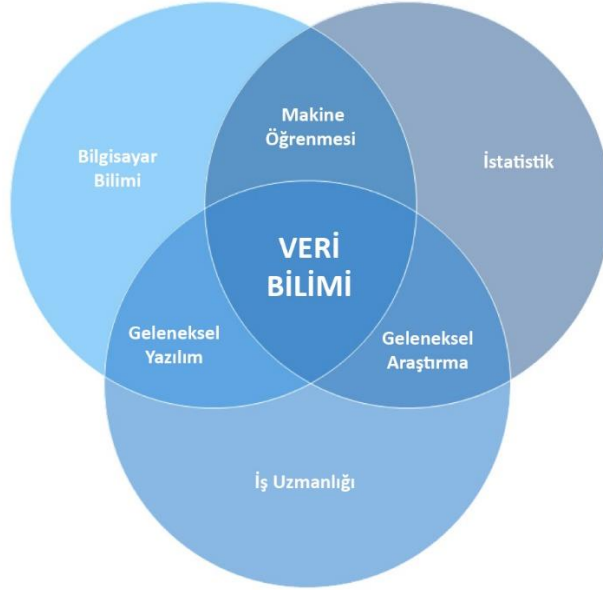
Keywords: Marketing Prediction, Machine Learning, Deep Learning, Classification, Python.



1. GİRİŞ

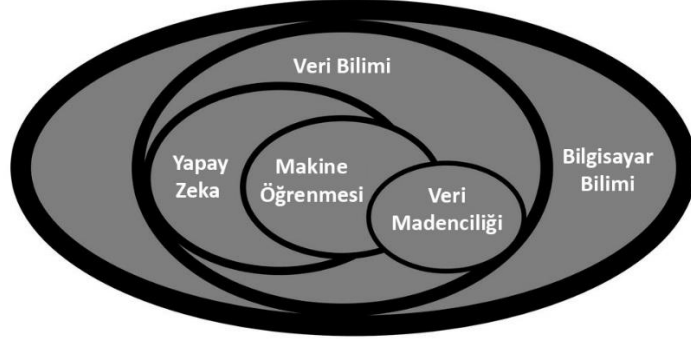
Günümüzde insanların hayatlarını sürdürmek için ve hayat kalitelerini arttırmaları için var olan her alanda yazılımlar kullanılmaktadır. Bu yazılımlar kullanıldıkları alana göre dallanmaktadır. Örneğin bir araba fabrikası daha çok robotik anlamdaki yazılımlara yönelerek imalat hızını arttırmaya yönelirken, finansal alandaki şirketler hem kendi iç işleyişlerini hem müşteriler ile olan akışlarını hemde diğer firmalar ya da devlet kurumları ile haberleşmelerini yönetebilmek için algoritmalara dayalı yazılımlar kullanırlar.

Son on senedir veri bilimi alanları, yazılımları desteklemek, kullanıcı veya müşteriye tanımak açısından önem kazanmaktadır. Veri bilimi, makine öğrenmesi, yapay zekâ gibi alt kırımları sayesinde firmalar müşterilerini ve kullanıcılarını daha iyi analiz ederek doğru kişiye doğru ürün prensibi ile müşteri ve kullanıcı memnuniyetlerini arttırmaktadırlar.[1] Veri biliminin yazılım ve iş süreçleri ile etkileşimi Şekil 1.1 üzerinden görülebilir.



Şekil 1.1-Veri Bilimi Venn Diagramı

Veri biliminin yaygınlaşmasının en önemli sebeplerinden biri de her sistemin yavaş yavaş yazılımlar üzerinden ilerlemesi ve bu yazılımlar aracılığı ile yapılan işlemlerin sıklığı, detayları, tarihleri ve benzerleri gibi sayısız bilgilerin ulaşılabilir şekilde kaydedilmesidir. Bu veriler işlenmedikleri sürece herhangi bir anlam ifade etmezler fakat bir araya getirilip anlamlandırılmaya çalışıldıklarında birçok alanda işe yarar tahminler üretilebilmektedir. Bu doğrultuda veri biliminin kapsadığı çeşitli yöntemler ile veriler anlamlandırılıp işlenebilmektedir. Veri biliminin diğer alt kırımları ile ilişkisi Şekil 1.2 üzerinden görülebilmektedir.



Şekil 1.2-Veri Bilimi ve Alt Kırılımları [23]

Bilgisayar Bilimi: Bilgisayarlar üzerine ve bilgisayarlar ile yapılan hesaplama kavram ve işlemlerinin tümünü kapsayan, içerisinde yazılım, donanım ve ağ ile ilgili konuları da barındıran geniş çaplı bir bilim dalıdır.

Donanım anlamında bilgisayarların tasarımlarını ve nasıl çalıştıkları ele alınır ve fonksiyonel bir bilgisayarın hesaplamaları nasıl yaptığı gibi konuları detaylıca inceler.

Yazılım anlamında ise programlama konseptleri ve yazılım dilleri incelenir. Programlama konseptleri altında fonksiyonlar, algoritmalar, işletim sistemleri, derleyiciler gibi temel ve detay konuları kapsar ve bunları inceler.

Veri Bilimi: Tek başına anlam ifade etmeyen ve yapılandırılmamış verilerin çeşitli algoritmalar yardımı ile anlamlandırılmasının hedeflendiği bir çalışma dalıdır. Verilerin düzgün anlamlandırılması ve daha net sonuçlar için büyük miktarda veri bilgisine ihtiyaç duyulur. Bu veriler anlamlandırılırken, bilgisayar bilimi, matematik, istatistik gibi birçok alandan ve bu alanların tekniklerinden yararlanır.

Veri bilimi günümüz dünyasında çok önemli bir hale gelmiştir. Çünkü, endüstrilere pazarlama ve müşteri kazanımında yardım eder, hızlı karar alımlarında destekleyici olur, yenilik yapma ve kullanıcı eğilimlerini tespit etmekte kullanılır.

Veri bilimi finans, imalat, sağlık hizmetleri, e-ticaret gibi günümüzde kritik yerlere sahip olan alanlarda aktif olarak kullanılmaktadır. Örneğin finans alanında müşterilere kişiselleştirilmiş deneyimler sağlarken, üretim alanında optimizasyon ve kar artırımı için kullanılmaktadır.

Yapay Zekâ: bir sistemin dışarıdan gelen bilgileri ve verileri doğru bir şekilde yorumlayarak, yorumladığı bu verilerden çıkarımlar yapıp sonrasında onları belirli hedeflere ve görevlere ulaşmak için kullanma durumu olarak açıklanabilir.[2]

Veri Madenciliği: Büyük veri depolarını otomatik olarak arayarak, basit analizlere kıyasla daha detaylı modelleri ve eğilimlerin keşfedilmesini sağlayan uygulamalar bütünüdür. Veri madenciliğinin temel özellikleri arasında, otomatik kalıp keşifleri, olası sonuçların tahmin

edilmesi, eyleme dönüştürülebilir yeni bilgi çıkarımları yapılması gibi özellikler bulunmaktadır.[3]

Önceki bölümlerde, veri biliminin genel yapısından, öneminden, kullanım alanlarından ve alt kırınımlarından bahsedilmiştir. Bu bilgiler doğrultusunda yazının ilerleyen kısımlarında alt kırınımlardan biri olan makine öğrenmesi yöntemleri ve bu yöntemler aracılığıyla alınan sonuçların incelenmesinden bahsedilecek ve bu sonuçlar doğrultusunda çıkarımlar yapılacaktır.

1.1. Problem Tanımı

Her geçen gün büyüyen müşteri ve kullanıcı kitlesi ve onlar tarafından sağlanan çok sayıda veri, kullanılmadığı ve işlenmediği sürece bir anlam ifade etmemektedir. Müşterilerin ve kullanıcıların sağladığı veriler firmalara onları daha iyi tanımaları ve onlara özel ürünler pazarlama yolunda çok faydalı olabileceği gibi, bu veriler değerlendirilmez ise potansiyel müşteri kayıplarına da yol açabilir.

Pazar tahmini, düşük maliyetli bir işlem olup düzgün bir şekilde uygulandığı zaman firmalara müşteriye iyi tanıma fırsatı vererek ilgili ürünü ilgili müşteriye sunarak daha tutarlı işlemler yapmaya olanak sağlar. Bu doğrultuda düzgün bir tahmin etme ile müşteri memnuniyeti gözle görülür bir şekilde artacak olup aynı zamanda firmalarında hem maliyet hem de zaman açısından kâra geçmesine olanak sağlayacaktır.

1.2. Amaç ve Kapsam

Yapılan tez çalışması, günümüzde verilerin aktif kullanılma ihtiyacının arttığı bu dönemde, bankacılık alanında yapılan tele-pazarlama çalışmalarında var olan başarı oranlarını arttırmayı hedef almaktadır. Bu hedef doğrultusunda başarı oranlarını arttırmak için sınıflandırma yöntemlerini kullanarak olası başarılı sonuç alınacak müşterilerin aranma öncesinde tespit edilebilmesi amaçlanmaktadır. Bu amaç kapsamında, eldeki veriler adım adım işlenerek sınıflandırmaya uygun hale getirilecek. Veriler sınıflandırmaya uygun hale geldikten sonra çeşitli makine öğrenmesi ve derin öğrenme yöntemleri ile başarı oranları elde edilecektir. Uygulanan teknikler ve işlem adımları ilerleyen bölümlerde detaylı bir şekilde anlatılacaktır.

1.3. Literatüre Katkıları

Geliştirilen model ile, günümüzde yaygın olarak hem küçük hem de büyük firmalar tarafından aktif bir şekilde yapılan pazarlama stratejileri ve bu stratejiler sonucu alınan aksiyonlara katkı sağlanarak geliştirilen strateji ve alınan aksiyonların gelişmeleri sağlanabilecektir. Firmalar, makine öğrenmesi algoritmalarının kullanılması ve kullanılan algoritmaların sonucunda ortaya çıkacak bilgilere dayanarak oluşturulacak modeller ile müşteri ve kullanıcı memnuniyetlerini

arttıracaklar, yeni müşteriler kazanabilecekler ve ürünlerini en doğru müşterilere sunarak başarı oranlarını arttıracaklardır. Bu sayede hem zamandan tasarruf edecekler hem de var olan gelirlerini koruyarak daha fazlasını elde etme imkânı bulacaklardır.

Katkı 1. Ele alınan çalışmada amaçlardan biri de maliyetlerin düşürülmesidir. Veri seti incelendiğinde yaklaşık yüzde 90 gibi bir başarısızlık durumu söz konusudur. Bu çalışmadan çıkarılacak sonuçlar ile müşteri aranmadan, nasıl bir davranış göstereceği tahmin edilebileceğinden, daha yüksek tutarlılıkla aramalar yapılabilecek ve bunun sonucunda maliyetlerde düşüş gerçekleşmesi amaçlanmıştır.

Katkı 2. Katkı 1 içerisinde anlatılan durum aynı zamanda zaman açısından da tasarruf sağlayacaktır. Buradaki doğru müşteri seçimi, seçenek sayısını düşürdüğünden daha az zamanda daha fazla başarıya ulaşmayı sağlayacaktır.

Katkı 3. Ele alınan çalışmada amaçlardan biri de farklı derin öğrenme ve makine öğrenmesi algoritmalarının aynı veri seti üzerinde nasıl sonuçlar verdiğinin gösterilmesidir.

1.4. Tezin Organizasyonu

Bu tez çalışması 5 bölümden oluşmaktadır.

- Birinci bölümde, problem tanımı yapılmıştır. Yapılan çalışma tanıtılmıştır, amacı ve önemi anlatılmıştır ve literatüre katkısından söz edilmiştir.
- İkinci bölümde, tez çalışmasının ana kaynağı olan makine öğrenmesi ve derin öğrenme algoritmalarından ve kullanılan veri setinden bahsedilmiştir.
- Üçüncü bölümde, çalışmada kullanılan yöntemler, çalışmada kullanılan teoriler, yaklaşımlardan ve bunların nasıl uygulandığından, amaçlarından bahsedilmiştir.
- Dördüncü bölümde, yapılan çalışma önerilen yöntemle ilgili detaylardan bahsedilmiştir.
- Beşinci yani son bölümde ise alınan sonuçlar ve sonuçlar üzerinde detaylı değerlendirmeler yer almaktadır.

2. ÖN BİLGİLER ve LİTERATÜR ARAŞTIRMASI

Bu bölümde araştırması yapılan çalışmanın temel bilgileri, tanımlamaları ve tez çalışması doğrultusunda bilgi edinmek için yapılan literatür araştırmalarından bahsedilmiştir. İlk olarak finansal anlamda pazarlamanın tanımının ne olduğu, hangi yollar ile yapılabileceği anlatılmıştır. Sonraki aşamada makine öğrenmesinin detay tanımlaması yapılmıştır ve tez çalışması esnasında kullanılan makine öğrenmesi algoritmaları incelenmiştir. En son aşamada ise konuya benzer ya da ilişkin olarak daha önce yapılmış olan akademik çalışmalar özetlenmiştir.

2.1. Bankacılıkta Pazarlama ve Pazarlama Türleri

2.1.1. Bankacılıkta Pazarlama

Bankacılıkta pazarlama aşamaları önemli bir yere sahiptir. Bankalar pazarlama yöntemleri sayesinde bilinirliklerini arttırabilir, yeni müşteriler elde edilebilir aynı zamanda mevcuttaki müşterilere yeni ürünler kullanılabılır. Bu bağlamda bankalar pazarlama işlemleri için farklı türlerde pazarlama yöntemleri kullanmaktadırlar.

2.1.2. Pazarlama Türleri

Bankacılıkta pazarlama türleri en geniş kapsamıyla ikiye ayrılmaktadır. Bu türler sonraki bölümde detaylandırılacaktır.

2.1.2.1. Dijital Pazarlama Türleri

Dijital pazarlama türleri kendi aralarında 5 gruba ayrılmaktadır. Bunlar; web sitesi aracılığı ile pazarlama, sosyal medya üzerinden pazarlama, dijital ortamdaki reklamlar ile pazarlama, mail bültenleri ile pazarlama ve SMS ile pazarlama olarak karşımıza çıkmaktadır.

Web Sitesi Aracılığı ile Pazarlama: Bankalar kendi web siteleri ile pazarlamalarını yapabilmektedirler. Bunu yaparken hem kendi müşterilerine hem de bankayı araştıran potansiyel müşterilere hitap etmektedirler. Kendi web siteleri aracılığı ile mevcutta ya da gelecekte yapmak istediği veya yaptığı kampanyaları veya ürünlerini kullanıcılara gösterebilmekte, aynı zamanda hazırladıkları kampanyaları kendi müşterilerine özelleştirerek sunabilmektedirler. Bu durumda ürün kullanımını ve kampanya katılım oranlarını pozitif etkilemektedir.

Sosyal Medya Aracılığı ile Pazarlama: Bankalar sosyal medya hesapları üzerinden mevcut müşterilerine veya potansiyeli olan müşterilerine ulaşabilmektedirler. Sosyal medya üzerinden bankalar kampanyalarını hitap ettiği kitlelere göre ayırarak farklı şekilde stratejiler oluşturabilmektedirler. Örneğin bir sosyal medyanın kullanımı orta yaşlılara hitap ederken bir başkası daha genç müşterilere hitap edebilmektedir.

Dijital Reklamlar Aracılığı ile Pazarlama: Bankalar kullanıcılarına internet ve mobil ortamlarda yayınladıkları reklamlar ile de ulaşabilmektedirler. Burada yine bankalar tarafından hedef kitlenin iyi analiz edilmekte ve bu doğrultuda farklı kampanya çeşitleri uygulanmaktadır. Burada her hedef kitleye en doğru reklamı göstermek ana hedef niteliği taşımaktadır. Bu ortamlarda verilen reklamlar genellikle kullanıcıları bankanın kendi web sitesine ya da mobil uygulamasına yönlendirip sonrasında ilgili kampanyaya katılmasını sağlamak ya da ürünü kullanmasını sağlamaya yöneliktir.

Mail Bültenleri Aracılığı ile Pazarlama: Bankalar e-posta üzerinden pazarlama yöntemini de aktif bir biçimde kullanmaktadırlar. E-posta aracılığı ile kredi ve kredi kartı gibi ürünlerin tanıtımı ve yönlendirmeleri yapılmaktadır. Yine E-posta aracılığı ile her bir müşteri kitlesine farklı ürün veya kampanya gönderimi sağlanabilmektedir. Bu durumda pozitif anlamda bir katkı sağlamaktadır.

SMS Aracılığı ile Pazarlama: SMS aracılığı ile pazarlama banka içerisinde yapılan en aktif pazarlama yöntemlerinden biridir. SMS pazarlaması ile kullanıcılara yeni ürünler tanıtılabilmekte ve bu ürünler için yönlendirmeler yapılabilmektedir. Genellikle SMS üzerinden kullanıcılar bankanın web sitesini ya da mobil uygulamasını kullanmaya teşvik edilmektedir.

2.1.2.2. Doğrudan Pazarlama Türleri

Doğrudan pazarlama türleri kendi aralarında 3 gruba ayrılmaktadır. Bunlar; şubeler aracılığı ile pazarlama, bankamatikler üzerinden pazarlama, telefon bankacılığı ile pazarlama olarak karşımıza çıkmaktadır.

Şubeler aracılığı ile: Bankalar kendi şubeleri üzerinden var olan müşterilerine ve şube üzerinden işlem yapmak isteyen herhangi bir potansiyel müşteriye ürün ve kampanya pazarlaması yapabilmektedir. Şubeler üzerinden yapılan ürün ve kampanya satışları müşteri doğrudan yetkili birisi ile konuşabildiğinden, müşteri için güven verici olup önemli bir pazarlama yöntemi olarak yer almaktadır.

Bankamatikler aracılığı ile: Bankalar kendi bankamatikleri üzerinden müşterilerine yada anlık olarak bankamatiği kullanan fakat müşterisi olmayan kişilere ürün ve kampanya tanıtımı yapabilmektedir. Tek tuşa basarak herhangi bir üründen ya da kampanyadan yararlanmak kullanıcıları cezbetmektedir.

Telefon Bankacılığı ile: Bankaların müşterilerine telefon aracılığı ile ulaşarak onlara kişiselleştirilmiş ürün ve kampanya teklifleri sunmaktadırlar. Burada sunulan teklifler kişiye özel olarak değişiklik gösterdiğinden ve müşterinin karşısında her türlü soruyu sorabileceği bir muhattap bulmasından dolayı etkili bir pazarlama yöntemidir. Aynı zamanda tez çalışması

içerisinde kullanılan veri seti telefon bankacılığı ile elde edilen sonuçlar üzerinden hazırlanmıştır dolayısı ile tez içerisinde yararlanılan pazarlama yöntemi telefon bankacılığı ile pazarlama olmuştur.

2.2. Makine Öğrenmesi

2.2.1. Makine Öğrenmesi Tanımı ve Amacı

Makine öğrenimi, genel anlamı ile önceki deneyimlerinden elde ettiği bilgileri kullanarak yeni gelen bilgileri bu doğrultuda inceleyip sürekli kendi performansını iyileştirmeye çalışan bir bilgisayar algoritmasıdır.[7]

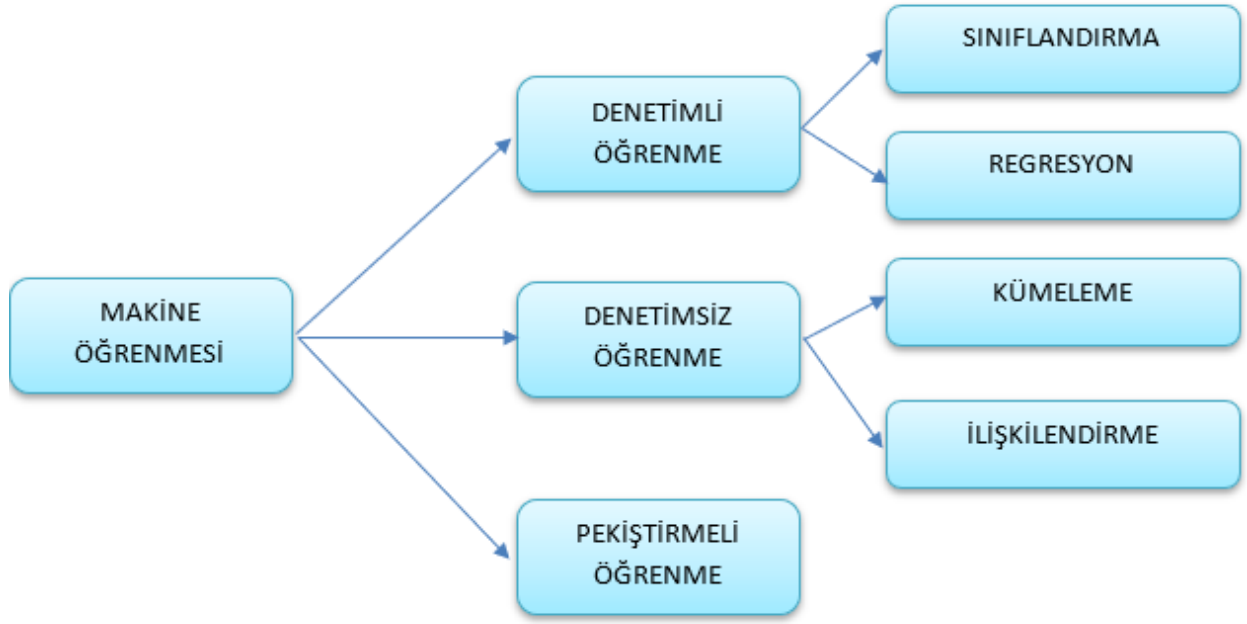
Makine öğrenimi, algoritmasını efektif bir şekilde tahmin yapmak ve kararlar almak için iki adet veri topluluğuna ihtiyaç duymaktadır. Bunlardan birincisi örnek veriler, ikincisi ise eğitilecek veriler olarak adlandırılmaktadır.[8]

Makine öğrenmesinin temel amacı, dışarıdan herhangi bir ek müdahale olmadan, hali hazırda var olan bilgileri kullanarak bu bilgilerden çıkarım yapması ve bu çıkarımların ileriki tahmin etme durumlarında veya yeri geldiği zaman kullanılmaya hazır edilmesidir. Makine öğrenmesinde öğrenim süreci sürekli devamlılık gösteren bir süreç olmakla birlikte, sürekli performans artışı sağlamaya yönelik bir eğilim gösteren bir süreç olarak karşımıza çıkmaktadır.

2.2.2. Makine Öğrenmesi Yöntemleri

Bu bölümde makine öğrenmesi algoritmalarının üst ve alt başlıkları anlatılmıştır. Şekil 2.1’de makine öğrenmeleri yöntemleri ve alt başlıkları kategorize edilerek gösterilmiştir.

- Denetimli Öğrenme
- Denetimsiz Öğrenme
- Yarı-Denetimli Öğrenme
- Pekiştirmeli Öğrenme

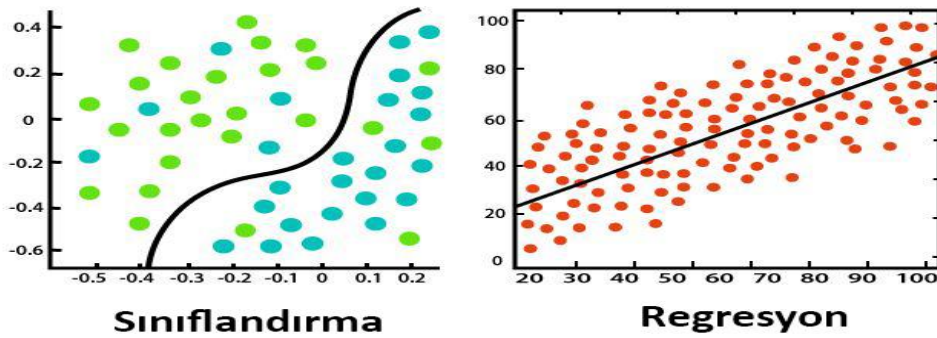


Şekil 2.1 - Makine Öğrenmesi Yöntemleri

Denetimli Öğrenme: Denetimli öğrenme, örnek olarak verilen etiketlenmiş girdi çıktı kombinasyonlarını inceleyerek, bu kombinasyonlara göre yeni girilen bir girdiye uygun olarak bir çıktı üreten makine öğrenimi yaklaşımıdır.[9] Bu yaklaşımın doğru sonuçlar üretebilmesi için öncelikle iyi tanımlanmış ve içerisinde çok sayıda veri bulunan bir veri setine sahip olmak gerekmektedir.[10] Denetimli öğrenme yaklaşımı kendi içerisinde iki ana gruba ayrılmaktadır. Bunlar;

- Sınıflandırma
- Regresyon

Şekil 2.2 üzerinden sınıflandırma ve regresyon yaklaşımlarının çalışma biçimi görülebilmektedir.



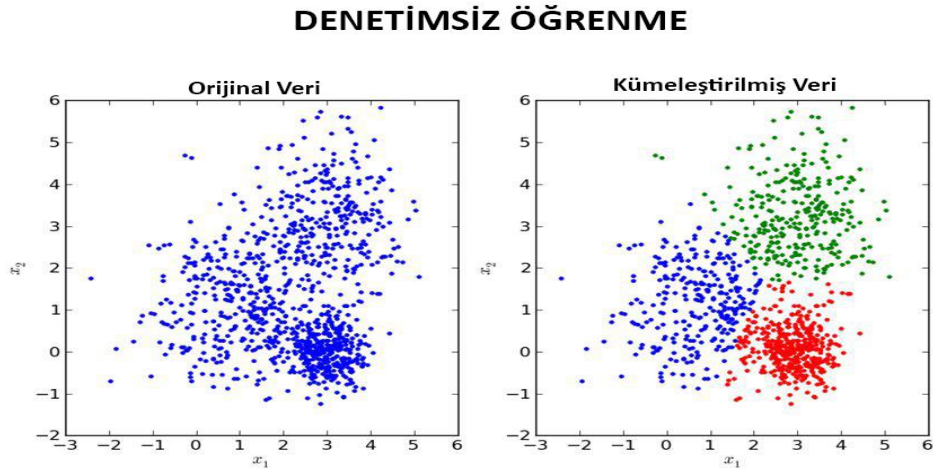
Şekil 2.2 - Sınıflandırma ve Regresyon

Günümüzde yaygın olarak problemlerin çözümünde kullanılan denetimli öğrenme algoritmaları Destek Vektör Makineleri (SVM), Karar ağaçları (DT), K-En Yakın Komşu Algoritması (KNN), Naif Bayes (NB) ve Regülasyon olarak karşımıza çıkmaktadır. Bu algoritmalarından tez çalışması dahilinde kullanılanlar Yararlanılan Makine Öğrenmesi Algoritmaları kısmında detaylandırılacaklardır.

Denetimsiz Öğrenme: Daha önceden etiketlenmiş verilerin bulunmadığı veri toplulukları içerisinde daha önce tespit edilmemiş kalıpları arayarak veriler arasındaki ilişkileri ortaya çıkarmaya çalışan makine öğrenmesi tekniğidir.[11] Denetimli öğrenmeye kıyasla daha fazla yanılma payı olan sonuçlar ortaya çıkarmaktadır çünkü girilen verilerin kullanılan algoritmalar sonucu verdiği çıktıyı doğrulayabilecek bir etiketli veri topluluğu bulunmamaktadır.

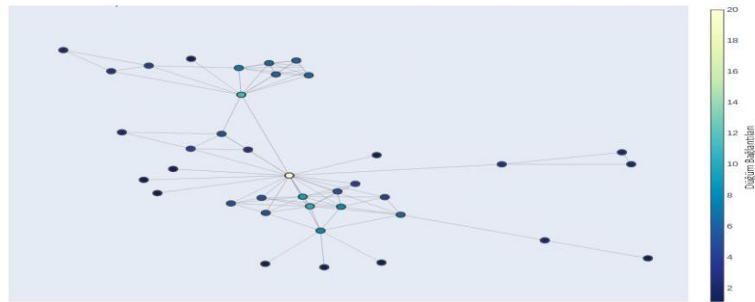
Denetimsiz öğrenme altında çalışmalar genellikle kümeleme, ilişkilendirme yöntemleri ile yapılmaktadır.

Şekil 2.3 üzerinden kümeleme yönteminin çalışma biçimi görülebilir.



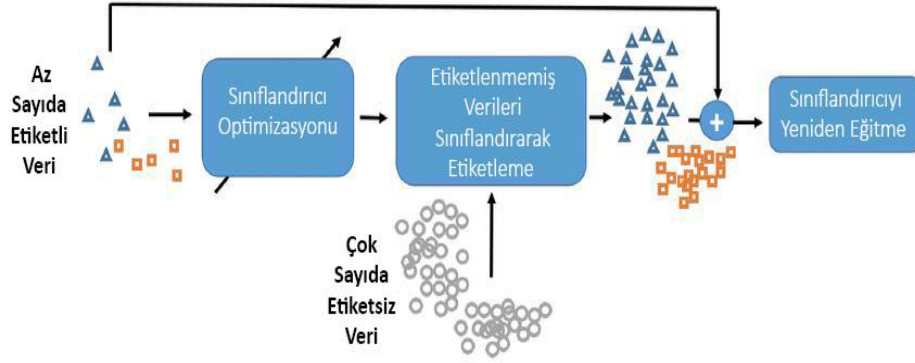
Şekil 2.3 - Kümeleme

Şekil 2.4 üzerinden ilişkilendirme yönteminin çalışma biçimi görülebilir.



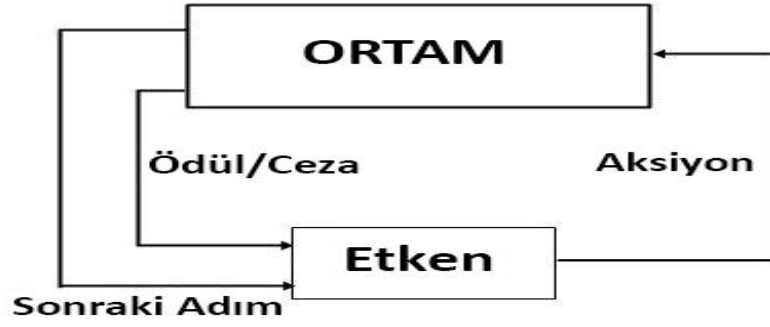
Şekil 2.4 - İlişkilendirme

Yarı Denetimli Öğrenme: İçerisinde hem etiketli hem etiketsiz verileri kullanarak, iki veri tipini de bir araya getirerek bu durumun öğrenme davranışlarında ne gibi etkiler oluşturduğunu gözlemleyen ve bu gözlemler doğrultusunda yeni algoritmalar tasarlayan makine öğrenmesi biçimidir.[12] Etiketlenmemiş verilerin az miktarda etiketlenmiş veriler ile birlikte kullanılması öğrenme doğruluğunda önemli gelişmeler sağlamaktadır.[13] Şekil 2.5 üzerinden yarı denetimli öğrenme yönteminin çalışma biçimi görülebilir.



Şekil 2.5 - Yarı Denetimli Öğrenme [24]

Pekiştirmeli Öğrenme: Bu öğrenme biçiminde herhangi bir ön bilgiye ihtiyaç duyulmadan, sadece durumların gözlemlenmesi ve bu gözlemler üzerinden en doğru seçimi yapma eğilimli bir öğrenme süreci mevcut. Bu öğrenme biçiminde, algoritmaların içerisinde kullanıldığı sistemler tek bir duruma odaklandırılıyor bu da en yüksek değerli ödül olarak adlandırılıyor. İçerisinde bulunduğu sistem eğer en yüksek değerden daha az bir değer ya da hiç ödül değeri almadıysa bu durum ceza alınmış gibi algılanıyor. Buradan da anlaşıldığı gibi algoritma her zaman en iyi değeri yakalamaya çalışmakta ve yakalayamadığı hiçbir durumu değerlendirmeye almamakta. Bu durum neticesinde buradaki öğrenme durumu sonsuza yakın olarak adlandırılmaktadır. Çünkü, her ortam değişikliğinde algoritmanın yapması gereken, çevre koşullarını tekrardan değerlendirerek en yüksek değerliği olan adımlar üzerinden ilerlemesi gerekmektedir. Şekil 2.6 pekiştirmeli öğrenmenin çalışmasını özetlemektedir.



Şekil 2.6 - Pekiştirmeli Öğrenme

2.2.3. Yararlanılan Makine Öğrenmesi Algoritmaları

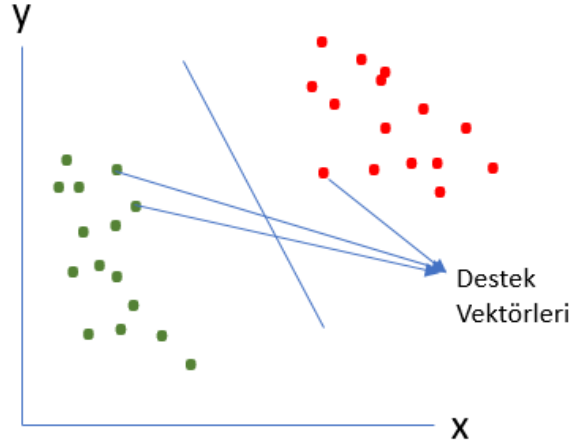
Bu makalede sadece denetimli öğrenmede sınıflandırma tekniğinden yararlandığından bu başlık altında sadece kullanılan denetimli öğrenme algoritmalarından bahsedilecektir.

2.2.3.1. Destek Vektör Makinesi Algoritması (SVM)

SVM, sınıflandırma ve regresyon problemlerini çözmek için kullanılan bir denetimli makine öğrenimi algoritmasıdır. İstatiksel öğrenme teorisine ve yapısal risk minimizasyonuna dayalı bir algoritmadır. Bu tez çalışması içerisinde sınıflandırma ihtiyacı olduğundan SVM algoritması sınıflandırma alanında kullanılacaktır. Sınıflandırma için bu algoritma, sınıflandırılacak iki grup arasında bir sınır çizerek bu grupları birbirinden ayırmaya odaklanmaktadır.[18] Bu sınır, iki sınıfta elemanlarına en uzak noktadan çizilmektedir. SVM algoritmasının kullanılmasında sağlanabilecek potansiyel avantajlar şunlardır;

- Yüksek doğruluk oranı
- Karmaşık karar sınırları modelleyebilme
- Çok sayıda bağımsız değişken ile çalışabilme
- Diğer birçok yöntemle göre aşırı oturma sorununun az olması.

Şekil 2.7 SVM algoritmasının sınıflandırma için çalışma şeklini göstermektedir.



Şekil 2.7 - Destek Vektör Makinesi Çalışma Biçimi

SVM algoritması veri seti içerisindeki gürültüden etkilenebilen bir algoritmadır. Eğer veri seti içerisindeki gürültü oranı yüksek ise, veriler birbirleri ile çakışacağından, algoritma gruplar arası sınırı çizmekte zorlanacağı için başarı oranında düşüş meydana gelecektir.

2.2.3.2. Karar Ağacı Algoritması (DT)

Karar ağaçları, verilerin sürekli olarak bir parametreye göre bölüldüğü bir denetimli makine öğrenimi algoritmasıdır. Karar ağacı içerisinde bulunan üç temel bileşen mevcuttur. Bu bileşenler yapraklar, temel düğüm ve karar düğümleri olarak adlandırılmaktadır. Yapraklar kararları temsil ederken, karar düğümleri verilerin bölüldüğü yerleri temsil etmekte, temel düğüm ise veri setini en çok açıklayacak veri üzerinden tüm algoritmayı başlatacak noktayı temsil etmektedir. Karar Ağaçları kendi başına bir yöntem olarak kullanılabilmekte, bunun yanında, Rastsal orman, Gradyen Güçlendirme gibi farklı algoritmalar içerisinde de kullanılmaktadır. Şekil 2.8 Karar ağacının yapısını göstermektedir.



Şekil 2.8 - Karar Ağacı Yapısı

2.2.3.3. Naif Bayes Algoritması (NB)

Naif Bayes, Bayes teoremine dayalı olarak geliştirilmiş bir makine öğrenimi algoritmasıdır.[20] Bayes teoremi bir değişken için geçerli olan olasılık dağılımları içinde koşullu olasılıklar ve marjinal olasılıklar arasındaki ilişkiyi göstermektedir. NB, veri kümesi içerisindeki bir eleman için, her durumun olasılığını hesaplamakta ve bu hesaplanan olasılıklar üzerinden değerlendirme yaparak en yüksek değerlikte olana göre sınıflandırma işlemini yapmaktadır. NB sınıflandırmasının formüle dökülmüş hali aşağıdaki gibidir.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

Yukarıdaki denkleme göre;

$P(A|B)$: B durumu gerçekleştiği zaman A durumunun gerçekleşme ihtimalini göstermektedir.

$P(B|A)$: A durumu gerçekleştiği zaman B durumunun gerçekleşme ihtimalini göstermektedir.

$P(A)$: A durumunun gerçekleşme ihtimalini göstermektedir.

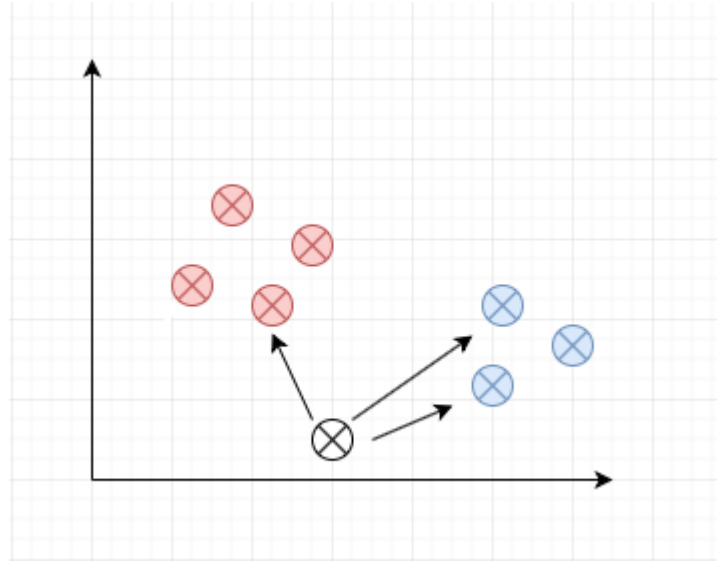
$P(B)$: B durumunun gerçekleşme ihtimalini göstermektedir.

2.2.3.4. K-En Yakın Komşu Algoritması (KNN)

KNN algoritması, benzer verilerin birbirine yakın olduğunu varsayarak bu varsayım üzerinden ilerleyen bir denetimli makine öğrenimi algoritmasıdır. Bu algoritma, veri setine eklenecek bir verinin hali hazırda veri setinde bulunan diğer verilere olan uzaklığını hesaplamakta ve bu hesaplama sonucunda veriyi sınıflandırmaktadır. Bu algoritmada izlenecek adımlar şu şekildedir;

- K parametresinin değeri belirlenir.
- Yeni verinin, mevcut verilere olan uzaklıkları tek tek hesaplanır,
- Hesaplanan uzaklıklar içerisinde en yakın k komşu ele alınır.
- Yeni veri bu komşulara göre etiketlenir

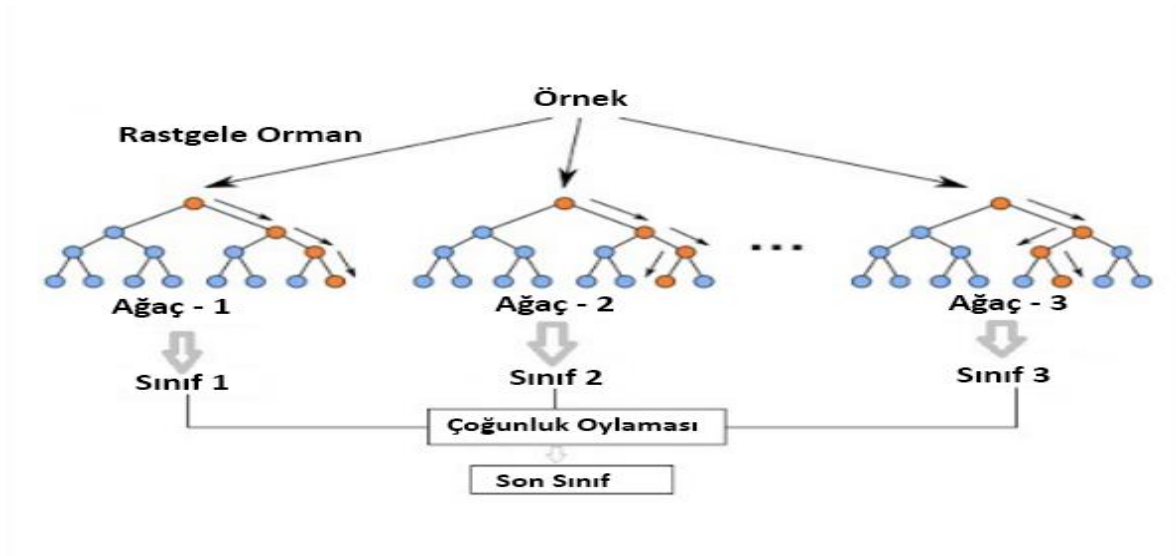
Yukarıda sıralanmış olan algoritma adımlarını Şekil 2.9 görsel olarak ifade etmiştir.



Şekil 2.9- KNN Algoritması

2.2.3.5. Rastsal Orman Algoritması (RF)

Rastsal Orman Algoritması, içerisinde birden fazla karar ağacını kullanarak sınıflandırma sonucunu iyileştirme hedefi güden bir algoritmadır. Algoritma içerisinde kullanılacak ağaç sayısı parametrikdir.[16] Bu parametre kapsamında oluşan karar ağaçları veri seti içerisinde Rastsal olarak seçilmiş veri kümeleridir. Rastsal olarak seçilen alt veri kümeleri üzerinde eğitim gerçekleşir ve her bir karar ağacı üzerinde tahmin etme yapılır. Yapılan bu tahminler sonucunda en yüksek başarı oranı olan karar ağacı ve tahmini sonuç olarak seçilmektedir. Şekil 2.10 üzerinde RF algoritmasının çalışma yöntemi gösterilmiştir



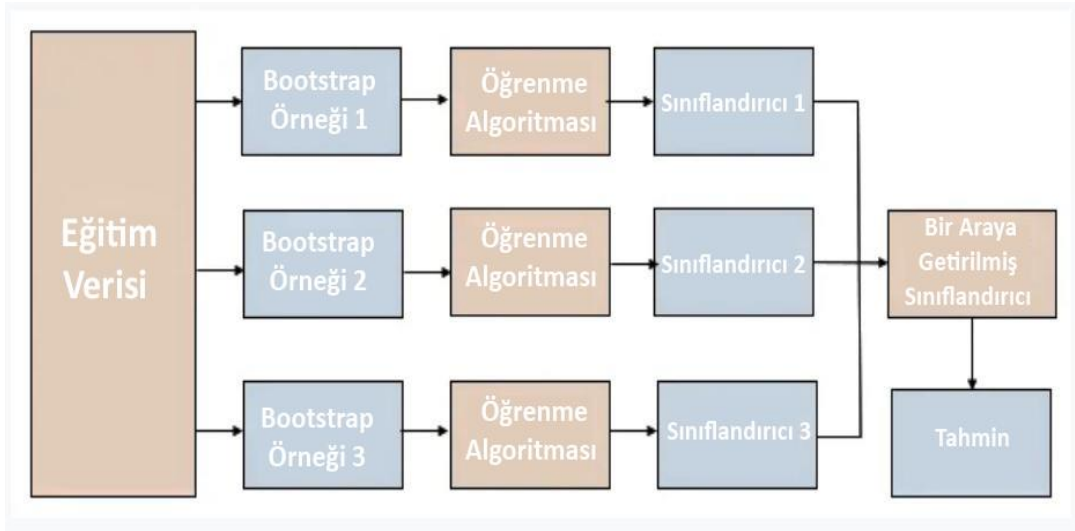
Şekil 2.10 - Rastsal Orman Çalışma Yöntemi

2.2.3.6. Ekstra Ağaç Algoritması (ET)

Ekstra Ağaç Algoritması, sınıflandırma sonucunu çıkarmak için birden fazla karar ağacından gelen tahminleri birleştirerek, bu birleşim üzerinden sürecine devam eden bir algoritmadır. Rastsal Orman algoritmasına benzerliği yüksektir. Buradaki fark, ET içerisindeki karar ağaçlarının belirlenmesi ve yaratılması RF'e göre farklılık göstermesidir. ET, en iyi eşik değerlerini bulmaya çalışmak yerine, her özellik için rastsal eşikler kullanarak ağaçları daha da rastsal hale getirmeye odaklanmaktadır.

2.2.3.7. Torbalama Algoritması (BA)

Torbalama algoritması, aşırı uyumu önlemek amacı ile 1994 yılında Leo Breiman tarafından önerilen bir algoritma olarak literatüre girmiştir. Torbalama algoritması, mevcuttaki bir veri seti içerisinde seçilmiş olan eğitim seti üzerinden, yeni eğitim setleri üretmek üzere öğreniciyi tekrardan eğitime eğiliminde bulunan bir makine öğrenmesi algoritmasıdır. Eğitim kümesi, n adet örnekten oluşmuş bir eğitim setinden yine n örnekle bir eğitim seti yerine konularak rastgele bir biçimde üretilmektedir. Seçilmiş olan her bir örnek, tekrar geri eğitim setine konulmaktadır. Şekil 2.11 torbalama algoritmasının çalışma mantığını göstermektedir. Burada da görülebileceği gibi Her bir seçilmiş örnek en son bir aşamada toplanır, sonrasında toplanmış olan örnekler üzerinden tahmin etme yapılmaktadır.



Şekil 2.11 - Torbalama Algoritması

2.2.3.8. Gradyan Artırma Algoritması (GBA)

GBA, sınıflandırma ve regresyon problemlerini çözmek için kullanılabilen bir makine öğrenmesi algoritmasıdır. GBA, zayıf tahmin modellerini bir araya getirip, karar ağaçlarının oluşturduğu bir model oluşturur ve oluşturulan bu model üzerinden sonuçlar alır.

2.2.3.9. Yapay Sinir Ağı (ANN)

Yapay sinir ağları, insanlarda bulunan biyolojik sinir ağlarından esinlenen modellerdir. Bu modellerde tıpkı biyolojik sinir ağlarında olduğu gibi birbirine bağlı ve iletişim halinde olan düğümler kullanılmaktadır. Bu düğümler birbirleri arasında sinyaller göndererek iletişim kurar. Yapay sinir ağlarında kurulan yapıya göre sinyaller giriş katmanından son çıkış katmanına taşınır. Kurulan yapay sinir ağlarında, katman sayısı ve bu katmanların özellikleri özelleştirilebilir. Ağırlıklar, katmanların kaç düğümden oluşacağı ve hangi fonksiyonlarla hesaplanacağı gibi özellikler belirlenebilir.

2.3. Derin Öğrenme

2.3.1. Derin Öğrenme Tanımı ve Amacı

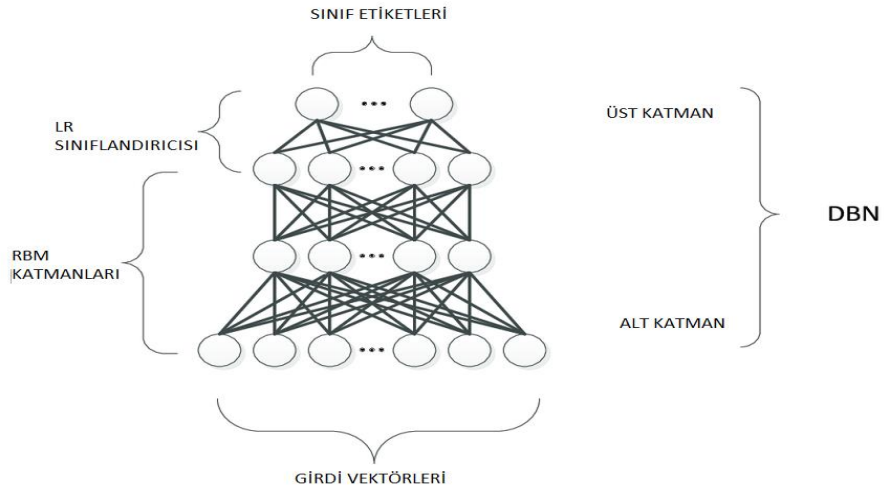
Derin öğrenme, makine öğrenmesinin bir alt sınıfı olarak karşımıza çıkmaktadır. Derin öğrenmeyi makine öğrenmesinden ayıran en temel fark, derin öğrenme yöntemlerinin büyük veriler için daha uygun olması ve büyük veriler üzerinde daha iyi sonuçlar vermesidir.[25]

2.3.2. Derin Öğrenme Mimarileri

Derin öğrenme yöntemleri, mimari açısından üçe ayrılmaktadır. Bu mimariler; Derin İnanç Ağları (DBN), Konvolüsyonel Sinir Ağları (CNN), Tekrarlayan Sinir Ağları (RNN) olarak karşımıza çıkmaktadır.

2.3.2.1. Derin İnanç Ağları (DBN)

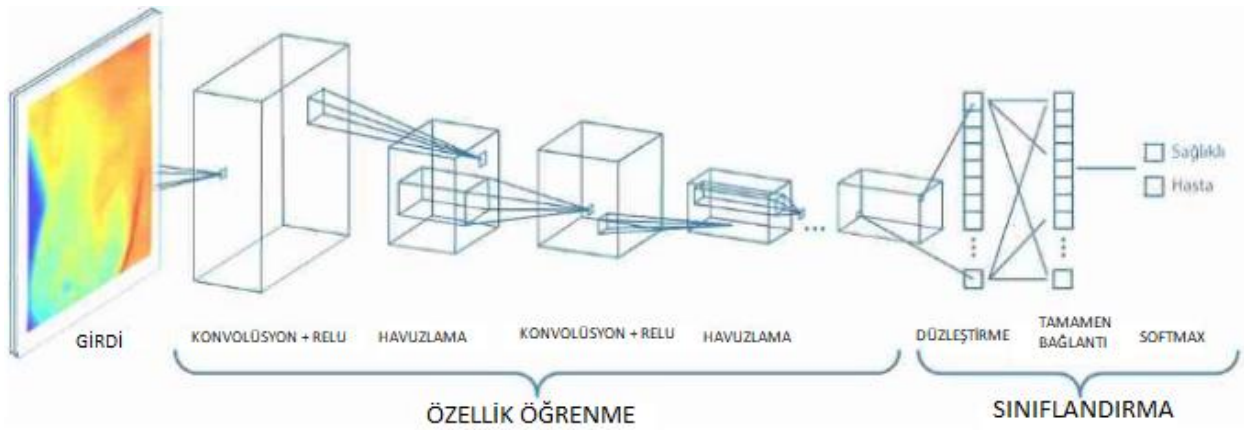
Kısıtlı Boltzmann Makinesi (RBM), sınıflandırma, özellik öğrenme, regresyon gibi işlemlerin yapılabilmesi için Hinton tarafından geliştirilen bir algoritmadır. Bu algoritma, derin inanç ağlarının yapı taşını oluşturmaktadır. RBM iki katmandan oluşmaktadır. Bunlardan ilki girdi katmanı, ikincisi ise gizli katmandır. Derin inanç ağları ise RBM yığımından oluşan bir tasarımdır. DBN, görüntü tanılama, video dizileri, hareket yakalama ve konuşma tanıma gibi farklı alanlarda kullanılabilir. Aşağıdaki Şekil 2.12 derin inanç ağlarının örnek bir mimarisini göstermektedir.



Şekil 2.12 - Derin İnanç Ağları Mimarisi

2.3.2.2. Konvolüsyonel Sinir Ağları (CNN)

Konvolüsyonel Sinir Ağları genellikle görüntü analizi için kullanılmaktadır. Görüntü içerisindeki nesnelerin ve objelerin ayrıştırılmasına yardımcı olmaktadır. Kenar bulma, piksel ekleme, kaydırma ve havuzlama gibi adımlardan oluşmaktadır. Konvolüsyonel sinir ağları daha çok yüz tanımlama, belge analizi, gri alanlar, reklamlar ve tarihi eserlerin ayrıştırılması gibi alanlarda kullanılmaktadır. Şekil 2.13 örnek bir konvolüsyonel sinir ağı mimarisini göstermektedir.

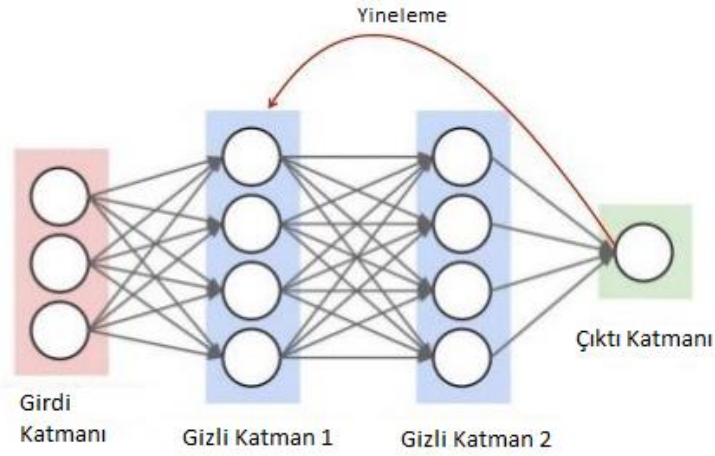


Şekil 2.13 - Konvolüsyonel Sinir Ağları Mimarisi [27]

2.3.2.3. Tekrarlayan Sinir Ağları (RNN)

Geleneksel tekrarlayan sinir ağları, önceki çıktıları tekrardan girdi olarak kullanan bir sinir ağları yapısıdır. Geçmiş bilgileri de dikkate alarak hesaplama yaptığından dolayı üretilen sonuçlar başarılıdır. Bu durum aynı zamanda işlem sayısını arttırdığından dolayı hesaplamaların yavaşlamasına neden olmaktadır. Genellikle konuşma tanıma, borsa tahmini, sınıflandırma gibi

alanlarda aktif olarak kullanılabilir. Şekil 2.14 üzerinde tekrarlayan sinir ağları mimarisi gösterilmiştir.



Şekil 2.14 - Tekrarlayan Sinir Ağları Mimarisi [33]

2.3.3. Yararlanılan Derin Öğrenme Yöntemleri

Bu tez çalışması kapsamında, derin öğrenme mimarilerinden biri olan RNN mimarisinin farklı yöntemleri ile çalışmalar yapılarak sonuçlar elde edilmiştir. RNN'ye ek olarak ANN yapısı kurularak bu yapı üzerinden de sonuçlar alınmıştır. Çalışma içerisinde kullanılan RNN yöntemleri şunlardır; Uzun Kısa-Süreli Bellek (LSTM), Geçitli Tekrarlayan Birim (GRU) ve Basit Tekrarlayan Sinir Ağı (SRNN).

2.3.3.1. Uzun Kısa-Süreli Bellek (LSTM)

LSTM algoritması, özel bir Tekrarlayan Sinir Ağı türüdür. Bu tür ilk olarak Hochreiter & Schmidhuber (1997) tarafından tanıtılmıştır. RNN yapıları kısa süreli hafızaya sahip oldukları için buradaki dezavantajı ortadan kaldırmak için ileri sürülen bir yapıdır. RNN ile karşılaştırıldığında, LSTM'ye giriş kapısı ve unutma kapısının eklendiği görülmektedir. [21] LSTM, daha önce var olan algoritmalara göre karmaşık problemleri çözme yeteneğine sahiptir. Aynı zamanda daha hızlı öğrenebilmekte ve daha başarılı sonuçlar vermektedir. [22] Bu çalışmada kullanılan LSTM yapısı 3 katmandan oluşmaktadır. Bunlardan ilki 64 boyutlu LSTM katmanıdır. İkincisi, 8 boyuttan oluşan yoğun bir katmandır. Son katman, çıktıları işlemek için yine sigmoid aktivasyon işlevini kullanan yoğun bir katmandır. Ağ kurulumu tamamlandıktan sonra, “adam” optimize edicisi ve “ikili çapraz entropi” kayıp fonksiyonu modele eklenip derleme işlemi yapılmıştır. Derleme adımı tamamlandıktan sonra, oluşturulan ağ, 100 period ve 32 yığın boyutu ile çalıştırılarak tahminler üretilmiştir.

2.3.3.2. Geçitli Tekrarlayan Birim (GRU)

GRU algoritması, standart Tekrarlayan Sinir Ağı modellerinin geliştirilmiş bir versiyonudur. Yapı olarak LSTM algoritmasına benzer. LSTM algoritmasının aksine, üç kapı kullanır. Bunlar, Güncelleme Geçidi, Sıfırlama Kapısı ve Mevcut Bellek Getirileridir. LSTM'ye göre daha az sayıda kapı kullanarak çalışması sayesinde LSTM'ye göre daha hızlı çalışır. [21] Bu çalışmada kullanılan GRU yapısı 3 katmandan oluşmaktadır. Bunlardan ilki 64 boyutlu GRU katmanıdır. Bu katman üzerinde aşırı oturmaya önlemek için 0,2 oranı ile seyreltme eklenmiştir. İkincisi, 8 boyuttan oluşan bir yoğun katmandır. Son katman, çıktıları eşlemek için yine sigmoid aktivasyon işlevini kullanmakta olan yoğun katmandır. Ağ kurulumu tamamlandıktan sonra, “adam” optimize edicisi ve “ikili çapraz entropi” kayıp fonksiyonu modele eklenip derleme işlemi yapılmıştır. Derleme adımı tamamlandıktan sonra, oluşturulan ağ, 100 period ve 32 yığın boyutu ile çalıştırılarak tahminler üretilmiştir.

2.3.3.3. Basit Tekrarlayan Sinir Ağı (SRNN)

Basit RNN algoritması, RNN modelinin standart bir algoritmasıdır. RNN modellerinde, çıkışlar tekrar giriş değerlerine geri beslenir. Basit RNN modelimiz bir tür çoktan bire RNN'dir. Birden fazla girdi alır ve tek çıktı üretir. Bu çalışmada kullanılan Basit RNN yapısı 3 katmandan oluşmaktadır. Bu katmanlardan ilki 64 boyutlu Basit RNN katmanıdır. Bu katmana aşırı oturmaya önlemek için 0,2 oranı ile seyreltme eklenmiştir. İkinci katman, 8 boyuttan oluşan yoğun bir katmandır. Son katman, çıktıları eşlemek için sigmoid aktivasyon işlevini kullanan yoğun bir katmandan oluşmaktadır. Basit RNN ağı tanımlandıktan sonra, “adam” optimize edicisi ve “ikili çapraz entropi” kayıp fonksiyonu modele eklenip derleme işlemi yapılmıştır. Derleme adımı tamamlandıktan sonra, oluşturulan ağ, 100 period ve 32 yığın boyutu ile çalıştırılarak tahminler üretilmiştir.

2.4. Literatür Araştırması

Bu bölümde tez çalışmasında yapılan literatür araştırmasına ve ilgili çalışmalara değinilmiştir.

Moro, Laureano ve Cortez pazarlama kampanyalarını yapmak için kendi iletişim merkezini kullanan bir Portekiz bankası üzerinden gerçek dünya verileri elde etmişlerdir. Burada 2008 ile 2010 yılları arasında toplamda 17 kampanya sonucunda elde edilen 79354 kişiye karşılık gelen bir veri seti elde edilmiştir ve bu veri seti içerisinde her bir kişi için 59 adet girdi mevcuttur. Elde edilen ilk veri seti içerisinde iki adet çıktı elde etmişlerdir, bunlar; iletişime geçilen kişinin kaydolup olmadığı ve ne kadar ücret yatırdığıdır. Fakat çalışmalarında kullanmak için ne kadar

ücret yatırdığı çıktısı dikkate almamışlar ve sadece vadeli mevduata kaydolup olmadıkları üzerinden ilerlenmiştir. Sonraki aşamalarında 79354 adet ulaşılan kişi bilgisi içerisinde kesinlik belirtmeyen veriler çıkartılmış ve bu sayı 55817 kişiye kadar düşürülmüştür. Bu veri seti üzerinde Naif Bayes ve Karar Ağacı algoritmalarını çalıştırmışlar ve bazı sonuçlara ulaşmışlardır. Fakat ulaşılan sonuçların daha da iyileştirilebileceğini düşünüp, girdi sayısını düşürmüşlerdir. Bu girdi sayılarını düşürürlerken, girdilerin sonuca olan etkilerine bakmışlardır. Bu çalışma sonrasında girdi sayısını yarı yarıya bir oranda düşürmüşlerdir. Sonrasında elde edilen yeni veri seti üzerinden, Naif Bayes, Karar Ağacı ve Destek Vektör Makinesi algoritmalarını çalıştırarak belirli sonuçlar elde etmişlerdir. Elde edilen sonuçların, veri setinin eski versiyonlarında elde edilen sonuçlar ile karşılaştırmasını yapıp, izledikleri modelin doğruluğunu ispatlamışlardır [6].

Moro, Cortez ve Rita, bu çalışmada Portekiz bankası üzerinden Mayıs 2008 ile Haziran 2013 yılları arasında yapılan toplamda 52944 adet telefon görüşmesi içeren bir veri seti hazırlamışlardır. Elde ettikleri veri setini, sosyal ve ekonomik etkenleri de göz önüne alarak örneğin, enflasyon oranı, işsizlik oranı gibi bilgilerle genişleterek 150 girdili bir veri seti haline getirmişlerdir. Daha sonra çalışmalarında algoritmalarını kullanmak için bu veri setini toplamda 22 özelliğe düşürmüşlerdir. Bu düşürme işlemi sırasında, öncelikle domain bilgisine dayalı olarak manuel bir eksiltme yoluna gitmişlerdir, bu işlemi 14 soru hazırlayarak bunların cevapları üzerinden sağlamışlardır. İkinci bir adım olarak, otomatik seçim yaklaşımını uygulamışlardır. Bu ikinci adım sonucunda ellerinde toplamda 22 adet özellik kalmıştır. Çalışmalarında toplamda 4 adet veri madenciliği yöntemi kullanmışlardır, bu yöntemler, Lojistik Regresyon, Karar Ağacı, Sinir Ağları ve Karar Destek Vektör algoritmalarıdır. Kullanılan bu algoritmaları AUC ve ALIFT sonuçları üzerinden birbirleri ile karşılaştırmışlar ve en başarılı yöntemin Sinir Ağları yöntemi olduğu sonucunu elde etmişlerdir [5].

Zeinulla, Bekbayeva ve Yazici, banka tele-pazarlama ile ilgili sınıflandırma modellerini karşılaştıran kapsamlı ve sınıflandırma yöntemlerinin birbirleri ile karşılaştırıldığı etkili bir çalışma sunmuşlardır. Bu çalışma içerisinde hem derin öğrenme algoritmaları hem de makine öğrenimi algoritmaları üzerinde çalışarak veri seti üzerinde her bir algoritma için başarı oranları elde etmişler ve bu başarı oranlarına göre kıyaslamalar yapmışlardır. Çalışma içerisinde kullandıkları algoritmalar şunlardır; Naif Bayes, Yapay Sinir Ağları, Rastsal Orman, Destek Vektör Makinesi, Lojistik Regresyon, K-En Yakın Komşu. Yaptıkları çalışma sonucunda, Rastsal Orman makine öğrenimi algoritmasının diğerlerine oranla daha iyi sonuç verdiği görülürken en düşük başarı oranını Lojistik Regresyon algoritması ile elde etmişlerdir [13].

Asare-Frempong ve Jayabalan, banka tele-pazarlama veri seti üzerinde çalışma yaparak farklı makine öğrenimi algoritmalarını bu veri seti üzerinde çalıştırmış ve başarı oranlarını gösteren bir

çalışma yapılmıştır. Bu çalışmada, Çok Katmanlı Algılayıcı Sinir Ağı, Karar Ağacı, Lojistik Regresyon ve Rastsal Orman makine öğrenimi algoritmaları üzerinde deneyler yapmışlar ve belirli başarı oranları elde edilmiştir. Bu doğrultuda bu makine öğrenimleri ile alınan sonuçların doğruluk değerlerine bakıldığında, en iyi sonucu Rastsal Orman makine öğrenimi algoritması verirken en düşük doğruluk payına ait yöntem Çok Katmanlı Algılayıcı Sinir Ağı algoritması olmuştur. Genel olarak doğruluk paylarına göre sıralandığında algoritmaların doğruluk oranları büyükten küçüğe sıralandığında sıralama şu şekilde olmaktadır. Rastsal Orman, Karar Ağacı, Lojistik Regresyon ve Çok Katmanlı Algılayıcı Sinir Ağı. Aynı zamanda bu çalışma içerisinde özellikler içerisindeki değerlerde incelenmiştir [15].

Cherif ve diğ [16], banka tele-pazarlama veri seti üzerinde çalışma yapan diğer makaleleri incelemişler ve diğer çalışmalara ek olarak yeni bir yaklaşım geliştirilmiştir. Bu yaklaşım doğrultusunda, ilk olarak veri seti içerisinde kullanılan özellikler, sayısal özellikler, kategorik özellikler ve ölçülebilir değerler olarak ayrılmıştır. Sayısal özelliklerin her biri için varyant, standart sapma ve ortalama gibi parametreleri hesaplanmış, kategorisel özellikleri kendi içerisinde üç gruba ayrılmıştır. Aynı zamanda ölçülebilir değerleri sıra numaralarına göre değiştirerek, sayısal özellikler de olduğu gibi hesaplama yapılmıştır. Son olarak bazı özellikler içerisinde bulunan “bilinmeyen” olarak tanımlanmış veriler, ortalama değerler ile değiştirilmiştir. Daha sonrasında elde edilen en son veriler üzerinden çeşitli makine öğrenmesi yöntemleri uygulanarak belirli sonuçlar elde edilmiştir.

Islam, Arifuzzaman ve Islam, çalışmalarında Moro ve diğerlerinin hazırladığı veri setini kullanmışlardır [5]. Bu veri setinin kendi içerisinde dengesiz bir veri seti olduğundan dolayı çalışmalarında SMOTE yöntemini bu veri seti üzerinde uygulamışlar ve veri setini dengeli bir hale getirmişlerdir. Aynı zamanda veri seti içerisinde var olan özelliklerin önem derecelerini analiz etmişlerdir. Daha sonra elde ettikleri sonuçları, farklı Naif Bayes algoritmaları ile çalıştırarak doğruluk, duyarlılık, kesinlik ve f1-ölçüm değerleri elde etmişlerdir. Elde edilen değerler sonucunda Gaussian Naif Bayes algoritmasının, Çok terimli Naif Bayes ve Bernoulli Naif Bayes algoritmalarına kıyasla daha yüksek doğruluk değeri verdiği görülebilmektedir [17].

Kim, Lee, Jo ve Cho çalışmalarında Moro ve diğerleri [5] tarafından hazırlanmış olan veri seti üzerindeki nitelikleri ve hiyerarşik özellikleri baz alarak bir Derin Evrişimli Sinir Ağı hazırlamışlardır. Hazırladıkları bu yapı ile aldıkları sonucu, karar ağacı, Destek Vektör Makinesi, Naif Bayes gibi çeşitli makine öğrenmesi yöntemleri üzerinden alınan sonuçlar ile karşılaştırmışlardır. Yaptıkları çalışma sonucunda hazırladıkları Derin Evrişimli Sinir Ağı yapısının diğer makine öğrenmesi yöntemlerine kıyasla daha iyi sonuçlar verdiğini görmüşler ve belgelemişlerdir. Hazırladıkları bu yeni modelin, finansal alandaki çalışmalarda diğer algoritmalar

ile birlikte kullanılabileceğini öne sürmüşlerdir [28].

Hassan, Rodan, ve Salem çalışmalarında Moro ve diğerleri [5] tarafından hazırlanan veri seti üzerinde veri madenciliği tekniklerini kullanarak kapsamlı bir çalışma yapmışlardır. Çalışmalarında Destek Vektör Makinesi, Naif Bayes, Karar Ağacı, Lojistik Regresyon, K en Yakın Komşu ve Sinir Ağı gibi farklı makine öğrenmesi yöntemlerini kullanmışlardır. Modellerini değerlendirmek için karmaşıklık matrisini kullanarak sonuçları not etmişlerdir. Elde edilen sonuçları karşılaştırdıklarında, Lojistik Regresyon algoritmasının en yüksek sonucu verdiğini, Naif Bayes ve K en Yakın Komşu algoritmalarının ise en düşük sonuçları verdikleri gözlemlenmiştir [14].

Puteni, Dewiani ve Tahir, çalışmalarında Moro ve diğerleri [5] tarafından hazırlanan veri setini baz alarak, Çok Katmanlı Algılayıcı Sinir Ağları (MLPNN) ve Radyal Temel Fonksiyon Sinir Ağı (RBFNN) yöntemlerini kullanarak belirli sonuçlara ulaşan bir çalışma hazırlamışlardır. Yöntemleri kullanmadan önce veri setini dengelemek amacı ile 41188 olan veri sayısını 15713'e düşürdükleri görülmektedir. Veri kümesi içerisindeki veri sayısı düşürüldükten sonra yine veri seti üzerinde özellik seçimi yöntemini uygulamışlar ve özellik sayısını 14'e düşürmüşlerdir. Bu işlemi uygularken en yüksek ağırlığı olan 14 özellik seçilmiştir. Ön işleme adımları bitirildikten sonra MLPNN ve RBFNN yöntemleri ile sonuçlar etmişlerdir. Çalışma sonucunda elde edilen sonuçlar, RBFNN yönteminin MLPNN yöntemine kıyasla daha iyi sonuçlar verdiğini göstermektedir [29].

Bu ve diğerleri, çalışmalarında tele-pazarlama kampanyaları kapsamında optimum tutma süresini incelemişlerdir. İncelemelerini yaparlarken altı ana durum üzerine yönelmişlerdir. Bu durumlardan bazıları şu şekildedir; müşterinin çalan telefonu hemen açması, hemen reddetmesi, uzun süre çaldıktan sonra açması, telefonu açıp hiç cevap vermeden operatörün kapatmasını beklemesi vs. Yaptıkları çalışma içerisinde operatörün toplam çalışma süresi, toplam yapılan arama sayısı, çağrı süresi, bekleme süresi, konuşma süresi gibi parametreleri göz önünde bulundurmışlardır. Çalışma sonucunda geliştirdikleri model ile optimal tutma süresini 20 ile 40 saniye arasında hesaplamışlardır ve tele-pazarlama ile uğraşan firmalara bu aralıkta tutma sürelerinin daha efektif olduğunu ve bu aralıkları kullanmaları yönünde tavsiyede bulunmuşlardır [30].

Cetiner ve Sahingoz, yazılım hata tahminleri üretmek için farklı makine öğrenmesi yöntemlerinin kullanıldığı ve başarı oranlarının birbirleri ile karşılaştırıldığı bir çalışma yapmışlardır. Bu çalışma kapsamında bir model geliştirmişlerdir. Bu model dahilinde, veri setleri üzerine PCA yöntemini kullanarak ve kullanmadan sonuçlar elde etmişler sonrasında elde edilen sonuçları birbirleri ile karşılaştırarak PCA'nın etkisini de çalışmalarında göstermişlerdir.

Çalışmalarında kullandıkları makine öğrenmesi yöntemleri şu şekildedir; Karar Ağacı, Destek Vektör Makinesi, Naif Bayes, K en Yakın Komşu, Rastsal Orman, Ekstra Ağaç, Adaboost, Torbalama ve Çok Katmanlı Algılayıcı. Çalışmalarında birden fazla veri seti kullanmışlar ve hepsi üzerinde ilgili makine öğrenmesi yöntemleri ile sonuçlar elde etmişlerdir. Çalışma sonuçlarında görülmüştür ki, en iyi doğruluk değerini Rastsal Orman algoritması göstermiştir. Aynı zamanda verisetleri üzerinde PCA yönteminin uygulanması da doğruluk değerlerini arttırmıştır [31].

2.5. Veri Seti

Veri, belirli ya da belirsiz konular üzerinde gözlem yolu ile elde edilen özellikler veya bilgilerdir. Örneğin, bir kişinin sosyal medya üzerinde yaptığı her hareket bir veridir. Veriler tek başlarına anlam ifade edebilir ya da diğer veriler ile birleştirilerek belirli anlamlar çıkartılmasında yardımcı olabilir. Veri seti ise bir konu üzerinde biriktirilen veri topluluğudur.

2.5.1. Yararlanılan Veri Seti

Veri seti, bir konu hakkında toplanan işlenmiş ya da işlenmemiş verilerin bir araya getirilmesiyle oluşan veriler bütünüdür. İnternet aracılığı ile her konu özelinde farklı farklı veri setlerine ulaşmak mümkündür. Sıradan bir kişi bir konu hakkında veri seti oluşturabileceği gibi bir kurum ya da firma kendi çalıştığı alanda veri setleri oluşturabilmektedir. Günümüzde veri madenciliğinin öneminin artması ile her kurum veya firma kendi veri setlerini hazırlayıp bu veri setlerini anlamlandırarak bunlar doğrultusunda yeni ürünler ya da kampanyalar üretip doğru kişilere pazarlamaya çalışmaktadırlar. Bu sayede hem zamandan tasarruf hem de iş gücünden tasarruf sağlamaktadırlar.

Bu tez çalışması içerisinde Kaliforniya Üniversitesi, Makine öğrenmesi havuzu içerisinde bulunan gerçek datalar kullanılmıştır. Bu datalar Portekiz Bankacılık Kurumu'nun direkt pazarlama yöntemleri ile 2008 ile 2013 yılları arasında aldığı sonuçları içerisinde barındırmaktadır. Bu veri seti kullanıcılar ile yapılan telefon görüşmelerine dayanmakta olup, veri setinin amacı kullanıcının bir vadeli mevduata üye olup olmayacağını sınıflandırmaktır [5].

Kullanılan veri setinin özellikleri Tablo 2-1'de gösterilmiştir.

Tablo 2-1-Veri Seti Bilgisi

ÖZELLİK SAYISI(GİRDİ)	ÖZELLİK SAYISI(ÇIKTI)	VERİ SAYISI	BAŞARILI VERİ SAYISI	BAŞARISIZ VERİ SAYISI	BAŞARI ORANI (%)
20	1	41118	4640	36548	11,28

Tablo 2.1 incelendiği zaman veri seti içerisinde toplamda yirmi bir özellik bulunduğunu ve bu yirmi bir özellikten yirmi tanesinin girdi kalan bir tanesinin ise sınıflandırma sonucunun

gösterildiği çıktı olduğu görülmektedir. Veri seti içerisinde toplamda 41118 veri bulunmakta olup bunlardan 4640 tanesi başarılı olarak sınıflandırılmış geri kalan 36548 tanesi başarısız olarak sınıflandırılmıştır. Genel tabloya bakıldığında başarı oranının yüzde 11,28 olduğu göze çarpmaktadır.

2.5.2. Yararlanılan Veri Setindeki Özellikler

Bir önceki bölümde genel bilgileri verilen veri seti içerisinde kullanılan özellikler ve özelliklerin açıklamaları Tablo 2-2 içerisinde gösterilmektedir.

Tablo 2-2 Veri Seti Özellik Bilgileri

ÖZELLİK	AÇIKLAMA
Age	Yaş Bilgisi.
Job	Aranan kişinin iş bilgisi
Marital	Aranan kişinin medeni durum bilgisi.
Education	Aranan kişinin eğitim bilgisi
Default	Aranan kişinin hali hazırda kredi kullanımı var mı?
Housing	Aranan kişinin konut kredisi kullanımı var mı?
Loan	Aranan kişinin bireysel kredi kullanımı var mı?
Contact	Kişi ile nasıl iletişim kurulduğu bilgisi
Month	Kişi ile en son hangi ay iletişim kurulduğu bilgisi
Day_of_week	Kişi ile haftanın hangi gününde iletişim kurulduğu bilgisi
Duration	Kişi ile kurulan iletişimin saniye cinsinden süresi
Campaign	İlgili kişi ile, kampanya bazında kaç kere iletişim kurulduğu bilgisi
Pdays	Kişi ile en son iletişim kurulduğundan beri geçen süre
Previous	Kişi ile daha önce en son kampanya harici kaç kere iletişim kurulduğunun bilgisi
Poutcome	Kişinin daha önceki kampanyalardan faydalanma durumu
Emp.var.rate	Üç aylık sürede istihdam değişim oranı
Cons.price.idx	Aylık tüketici fiyat endeksi
Cons.conf.idx	Aylık tüketici güven endeksi
Euribor3m	Günlük bankalar arası faiz oranı
Nr.employed	Üç aylık çalışan sayısı
y	Kişinin abone olup olmadığının bilgisi

Tablo 2 üzerinde veri seti içerisinde kullanılan özellikler ve bu özelliklerin kısa açıklamaları bulunmaktadır. Özelliklerin detaylı açıklamaları yine bu bölüm altında verilecektir.

- Age: Bu sayısal özellik, temas kurulan müşterinin yaş bilgilerini temsil eder.

- Job: Bu kategorik özellik, temas kurulan müşterinin "teknisyen", "hizmetler", "yönetici" gibi iş bilgilerini temsil eder.
- Marital: Bu kategorik özellik, temas kurulan müşterinin "evli", "boşanmış", "bekar", "bilinmiyor" gibi medeni durumunu temsil eder.
- Education: Bu kategorik özellik, temas kurulan müşterinin "temel 4 yıllık", "lise", "üniversite diploması" gibi eğitim durumunu temsil eder.
- Default: Bu özellik, iletişime geçilen müşterinin ödemekte başarısız olduğu herhangi bir krediye sahip olduğunu gösterir.
- Housing: Bu özellik, müşterinin herhangi bir aktif konut kredisine sahip olduğunu gösterir.
- Loan: Bu özellik, müşterinin herhangi bir aktif kişisel kredisi olduğunu gösterir.
- Contact: Bu özellik, müşteri ve çağrı merkezi temsilcisi arasındaki iletişim yolunu temsil eder.
- Month: Bu özellik, müşteriyle yılın son iletişim ayını temsil eder.
- Day_of_week: Bu özellik, haftanın son iletişim gününü temsil eder.
- Duration: Bu sayısal özellik, müşteri ile son temas süresini saniye cinsinden temsil eder.
- Campaign: Bu sayısal özellik, ilgili kampanya için müşteri ile kaç kez iletişime geçildiğini gösterir.
- Pdays: Bu sayısal özellik, mevcut kampanyadan önce müşteriyle en son iletişime geçildikten sonra kaç gün geçtiğini temsil eder. 999 değeri, müşteriyle daha önce hiçbir iletişim olmadığı anlamına gelir.
- Previous: Bu sayısal özellik, mevcut kampanyadan önce müşteri ile kaç kez iletişim kurulduğunu gösterir.
- Poutcome: Bu kategorik özellik, müşterinin önceki kampanyadan fayda sağladığını temsil ediyor.
- Emp.var.rate: Bu sayısal özellik, müşteri ile iletişime geçildiğinde yılın son çeyreğinde istihdam değişim oranındaki değişim miktarını temsil eder.
- Cons.price.idx: Bu sayısal özellik, müşteri ile iletişime geçildiğinde yılın son ayı için tüketici fiyat endeksini temsil eder. Bu özellik, müşteriler tarafından bir sepet mal için ödenen fiyatlardaki değişiklikleri gösterir.

3. YÖNTEM

Bu tez içerisinde yapılan çalışma sıralı şekilde gerçekleşen adımlardan meydana gelmektedir. İlk yöntem verilerin elde edilmesi olarak karşımıza çıkmaktadır. Bu çalışma doğrultusunda kullanılan verilerin elde edilmesi ve detayları bir önceki bölüm olan “Veri Seti” bölümünde anlatılmıştır. Burada veri seti kullanıma alınmadan önce veri setinin tutarlılığı, içerisindeki verilerin kalitesi ve tez çalışması ile örtüşmesi gibi konular göz önünde bulundurulmuştur.

Veri seti bulunduktan sonraki adım, veri setini makine öğrenmesi algoritmalarının çalışabileceği formata çevirmek olmuştur. Yine burada yapılan çalışmalar bir önceki bölüm olan “Veri Seti” bölümü içerisinde anlatılmış olup, yapılan çalışma Tablo 3 üzerinde görülebilmektedir.

Veri setinin hazırlanması aşaması bitirdikten sonra, veri seti bir “python” kütüphanesi olan “pandas” sayesinde okunarak sisteme aktarılır. Bu aktarım yapıldıktan sonra, veri seti içerisindeki özellikler girdi veya çıktı olarak sisteme tanıtılarak işleme devam edilir.

Sonraki adımda girdi ve çıktı olarak sisteme tanıtılan veriler normalizasyon işlemine alınmıştır. Burada normalizasyon yapılmasının amacı parametreler arasındaki genel bütünlüğün sağlanmasının istenmesidir.

Normalizasyon işlemi sonrasında, tez çalışması için kullanılacak olan makine öğrenimi algoritmalarının uygulamaları yapılmıştır.

Makine öğrenimi uygulamalarının başarılı bir şekilde bitmesi sonrasında, K-Katlamalı Çapraz Doğrulama yöntemi kullanılarak, tez içerisinde kullanılacak olan makine öğrenmesi algoritmaları ile birlikte eğitime alınmıştır. Burada kullanılan ‘K’ dinamik değeri 5 olarak belirlenmiştir.

Bu aşamadan sonra elde edilen sonuçlar Karışıklık Matrisi üzerinde gösterilmiş ve sonrasında doğruluk, kesinlik ve duyarlılık puanları not edilmiştir. Not edilen değerler En son aşamada birbirleri ile karşılaştırılarak başarı oranları gözlemlenmiştir.

3.1. Normalizasyon İşlemi

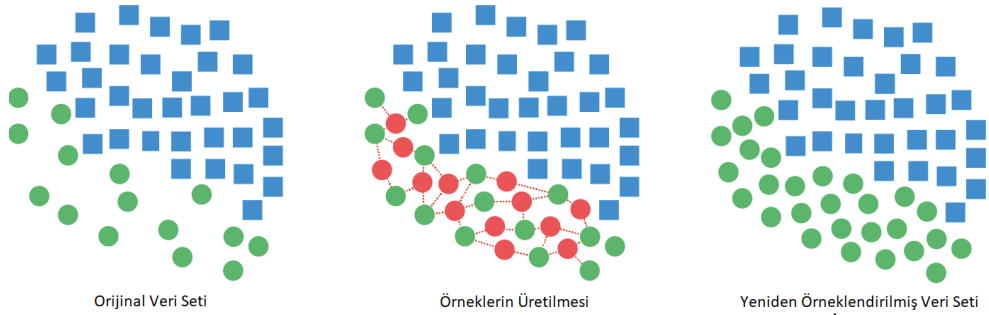
Normalizasyon işlemi, veriler arasındaki aykırı verilerin işlemler sırasında bozulmalara yol açmasını engellemek amacıyla yapılan bir işlemdir. Normalizasyon işleminin yapılabilmesi için çeşitli yollar bulunmaktadır. Bu çalışma içerisinde bu yollardan min-max normalizasyonu yöntemi kullanılmıştır. Min-max normalizasyonu yöntemi kullanılan en yaygın normalizasyon yöntemlerinden biridir. Bu normalizasyon yöntemi ile, en düşük değer 0 en yüksek değer 1 olacak şekilde bütün veriler bu aralığa yerleştirilmekte, sonrasındaki işlemler hesaplamalarına bu normalize edilmiş veriler üzerinden devam etmektedirler. Min-max normalizasyonunun denklemi aşağıdaki gibidir;

$$Xy = (Xo + Xmin)/(Xmax + Xmin) \quad (2)$$

Formül içerisinde gösterilen ‘Xo’ değeri veri seti içerisindeki işleme alınacak her bir özelliğin mevcuttaki değerini temsil etmektedir. ‘Xmin’ değeri, veri seti içerisinde ilgili özellik ile alakalı en küçük değerliğe sahip olan verinin değerini temsil ederken, ‘Xmax’ değeri, veri seti içerisinde ilgili özellik ile alakalı en büyük değere sahip olan veriyi temsil etmektedir. Formülde yapılan işlem, mevcuttaki özellik değerinden minimum özellik değeri çıkartılıp, maksimum değer ile minimum değer farkına bölünmesi gerektiğini göstermektedir. Bu işlem sonucunda çıkan değer, normalize edilmiş olan yeni özellik değeri olarak kaydedilmektedir.

3.2. SMOTE Uygulaması

SMOTE tekniği, azınlık sınıfına yeni örnekler ekleyerek azınlık sınıfını çoğunluk sınıfıyla eşitleme yaklaşımı sergileyen, küme tabanlı bir yüksek hızda örnekleme tekniğidir [19]. Bu yaklaşımın sonunda veri setinde azınlık olmadığı için veri seti dengelenmektedir. Bu teknik ilk olarak, Chawla ve diğerleri [34] tarafından tanıtılmıştır. Ayrıca bu teknik, dengesiz veri setleriyle başa çıkmak için en yaygın olarak kullanılan tekniktir [35]. Veri setinin dengeli olması, üzerinde çalıştırılacak algoritmalarından daha doğru sonuçlar alınmasını sağlamaktadır. Bu tez çalışması içerisinde SMOTE tekniği, veri seti eğitim ve test verilerine bölündükten sonra eğitim verilerine uygulanmıştır. Smote tekniğinin basit çalışma mantığı Şekil 3.1 üzerinde gösterilmektedir.



Şekil 3.1 - SMOTE Tekniği

3.3. K-Katlamalı Çapraz Doğrulama (CV)

Bu çalışma içerisinde veri setini eğitim ve test verilerine ayırmak için K-Katlamalı Çapraz doğrulama tekniği kullanılmıştır. Bunun sebebi başka bir alternatif olan ayırma yönteminin büyük sayıda veri içeren veri setleri için kullanışlı olması ve test ve eğitim oranlarının manuel olarak belirlenmesidir. Bu manuel belirlemelerde modelin eğitim ve testlerinde bazı sapmalar meydana gelebilmektedir. Bu çalışmadaki veri setinin büyüklüğü göz önüne alındığında K-Katlamalı Çapraz Doğrulama yöntemi en doğru yöntem olarak karşımıza çıkmaktadır. K-Katlamalı Çapraz doğrulama bir veri seti içerisindeki verilerin rastgele bir şekilde K adet eşit parçaya bölünerek,

Gerçek Pozitif (TP): Algoritma sonucunda alınan değer, orijinal değer ile aynıdır. Sınıflandırma işlemi doğru bir şekilde yapılmıştır.

Yanlış Negatif (FN): Algoritma sonucunda alınan değer, orijinal değerden farklıdır. Sınıflandırma işlemi yanlış bir şekilde yapılmıştır.

Yanlış Pozitif (FP): Algoritma sonucunda alınan pozitif değer, orijinal değerden farklıdır. Burada yanlış bir sınıflandırma işlemi yapılmıştır.

Gerçek Negatif (TN): Algoritma sonucunda elde edilen negatif değer, orijinal değer ile aynıdır. Sınıflandırma işlemi doğru bir şekilde yapılmıştır.

Karışıklık matrisi üzerinden elde edilen sonuçlar ile, kesinlik, duyarlılık, doğruluk ve f-puanı hesaplamaları yapılabilmektedir.

Doğruluk: Doğruluk değeri, toplamda doğru olarak tahmin edilen verilerin, veri seti üzerindeki toplam veri sayısına oranı ile bulunmaktadır.

Kesinlik: Kesinlik değeri, pozitif olarak tahmin edilen değerlerin gerçekten ne kadarının pozitif olduğunun oranını gösteren metriktir.

Duyarlılık: Duyarlılık değeri, gerçekte pozitif olan değerlerin tahminler sonucunda ne kadarının doğru bir şekilde pozitif olarak tahmin edildiğinin oranını gösteren metriktir.

F1-Ölçüm: F1-Ölçüm değeri, duyarlılık ve kesinlik metriklerinin harmonik ortalamasını gösteren metriktir. Bu metrik dengesiz veri setleri için kullanışlıdır.

Bu değerlerin nasıl hesaplandıkları, aşağıdaki formüller ile gösterilmiştir.

$$\text{Doğruluk} = (TP + TN)/(TP + FP + TN + FN) \quad (3)$$

$$\text{Kesinlik} = TP/(TP + FP) \quad (4)$$

$$\text{Duyarlılık} = TP/(TP + FN) \quad (5)$$

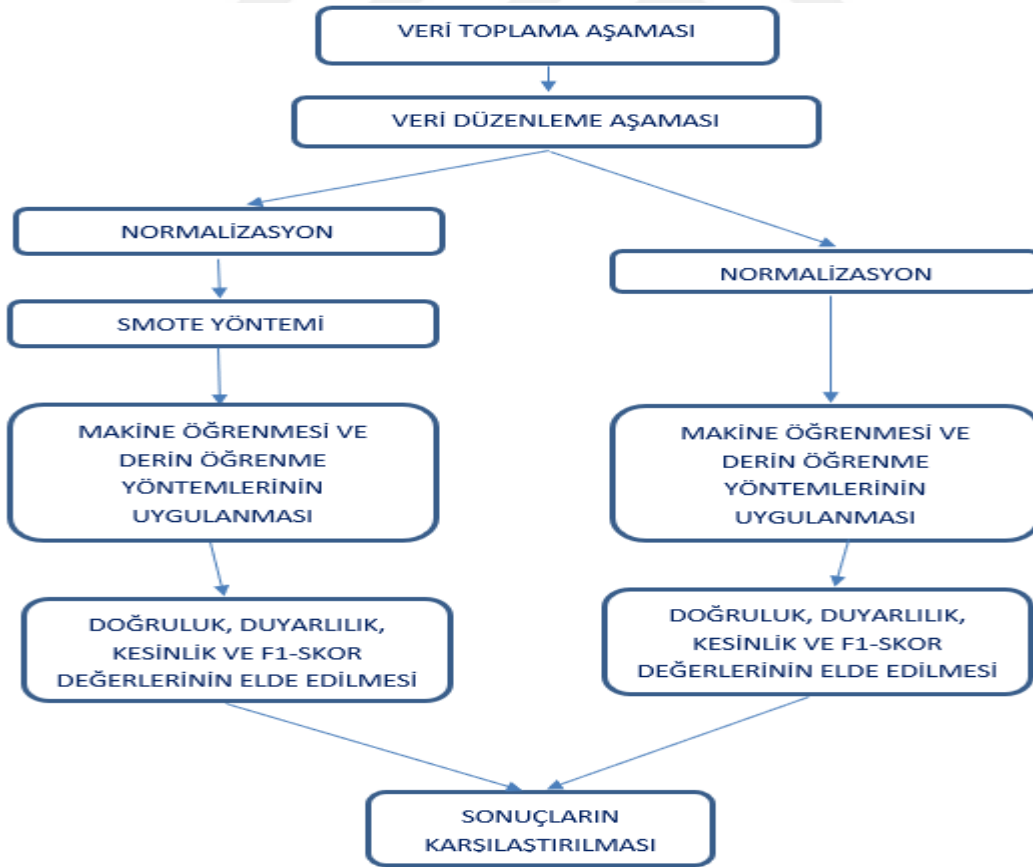
$$F1 - \text{Ölçüm} = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (6)$$

4. ÖNERİLEN MODEL

Yapılan çalışmada, önceki çalışmalar incelenmiş, diğer çalışmalarda kullanılan ve kullanılmayan çeşitli makine öğrenmesi ve derin öğrenme yöntemleri uygulanmış ve belirli sonuçlar alınmıştır. Aynı zamanda dengesiz olan veri seti, SMOTE tekniği uygulanarak dengeli bir hale getirilmiş ve dengeli hali üzerinden tekrar aynı yöntemler uygulanarak tekrardan sonuçlar alınmıştır. Burada amaçlanan, derin öğrenme ve makine öğrenmesi yöntemlerinin birbirleri ile karşılaştırılarak, bu tarz veri setleri üzerinde nasıl sonuçlar aldığını görmektir ve bu sonuçları birbirleri ile karşılaştırarak en başarılı yöntemi ve başarısız yöntemi bulmaktır. Bu sayede ileride yapılan çalışmalara yol göstererek, başarılı olan algoritmalar üzerinden alınan sonuçların nasıl daha iyi noktalara getirilebileceğine odaklanılabilir. Bu bölüm altında bu tez çalışması kapsamında yapılanlar ve genel işleyiş anlatılacaktır.

4.1. Modelin Akış Diyagramı

Yapılan çalışmalarda, çeşitli teknikler kullanılmış olup, bu tekniklerin uygulanma sırası Şekil 4.1 üzerinde gösterilmiştir.



Şekil 4.1 - Kullanılan Model

4.2. Modelde Kullanılan Teknoloji

Tez çalışması yapılırken, yazılım dili olarak makine öğrenmesi ve derin öğrenme yöntemleri ile ilgili çalışmalarda kütüphane desteklerinden dolayı yaygın olarak kullanılan “Python” dilinden ve çeşitli kütüphanelerinden yararlanılmıştır. Geliştirme ortamı olarak “Anaconda Navigator” aracılığı ile “Spyder” kullanılmıştır. Kullanılan kodlama ve mimari özellikleri tensorflow alt yapısına uygun bir şekilde kullanılmıştır. Çalışma içerisinde kullanılan kütüphaneler şu şekildedir;

- Makine öğrenmesi algoritmaları için “scikit-learn” kütüphanesi,
- Dizi işlemleri için “numpy” kütüphanesi,
- Veri görselleştirmeleri için “matplotlib” ve “seaborn” kütüphaneleri,
- Veri üzerinde işlem yapmak için “pandas” kütüphanesi,
- Derin öğrenme yöntemleri için “keras” kütüphanesi kullanılmıştır.



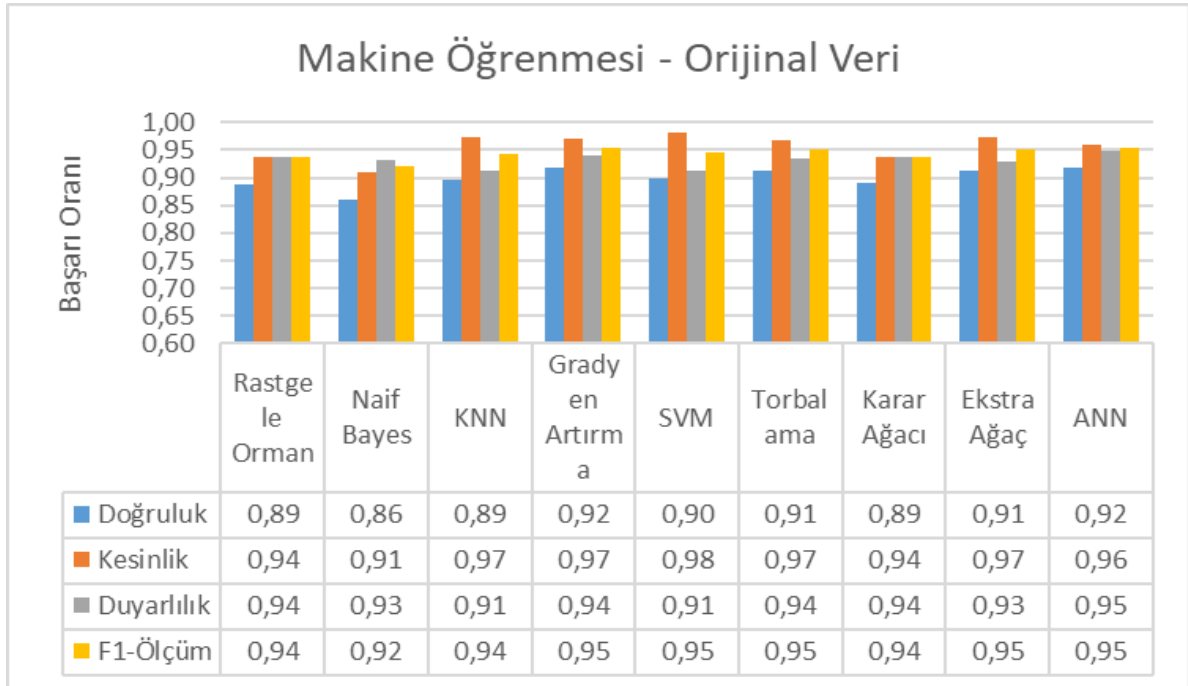
5. ALINAN SONUÇLAR VE DEĞERLENDİRMELERİ

Bu bölüm altında veri seti üzerinde öncelikle SMOTE tekniği uygulanarak ve uygulanmadan alınan sonuçlar ayrı ayrı ele alınarak kendi içlerinde karşılaştırılmıştır. Sonrasında genel olarak alınan sonuçlar üzerinde SMOTE tekniğinin etkisi yorumlanacaktır.

Sonuç – 1

SMOTE yöntemi uygulanmadan orijinal veri seti üzerine uygulanan çeşitli makine öğrenmesi yöntemleri ile alınan sonuçların grafiği Şekil 5.1 üzerinde gösterilmektedir. Oluşan grafik üzerinden sonuçlar şu şekilde yorumlanabilir;

- En yüksek doğruluğu (0,92) veren algoritmalar, Gradyen Artırma ve Yapay Sinir Ağları algoritmaları olmuştur.
- En yüksek kesinlik değeri (0,98) Destek Vektör Makinesi algoritması ile elde edilmiştir.
- En yüksek duyarlılık değeri (0,95) Yapay Sinir Ağları algoritması ile elde edilmiştir.
- En yüksek F1-Ölçüm değeri (0,95) Gradyen Artırma, Destek Vektör Makinesi, Ekstra Ağaç ve Yapay Sinir Ağları yöntemleri ile elde edilmiştir.
- En düşük doğruluk değeri (0,86) Naif Bayes algoritması ile elde edilmiştir.
- En düşük kesinlik değeri (0,91) Naif Bayes algoritması ile elde edilmiştir.
- En düşük duyarlılık değeri (0,91) K-En Yakın Komşu ve Destek Vektör Makinesi algoritmaları ile alınmıştır.
- En düşük F1-Ölçüm değeri (0,92) Naif Bayes algoritması ile elde edilmiştir.

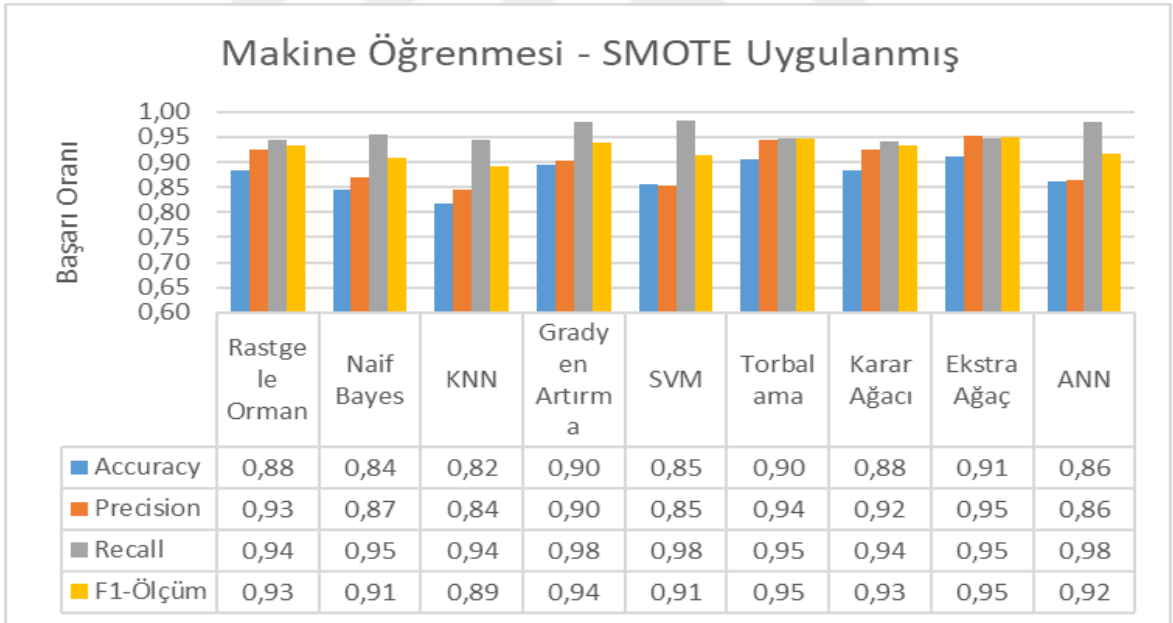


Şekil 5.1 - SMOTE Kullanılmadan Alınan Sonuçlar – Makine Öğrenmesi

Sonuç – 2

SMOTE yöntemi uygulandıktan sonra çeşitli makine öğrenmesi yöntemleri ile alınan sonuçların grafiği Şekil 5.2 üzerinde gösterilmektedir. Oluşan grafik üzerinden sonuçlar şu şekilde yorumlanabilir;

- En yüksek doğruluğu (0,91) veren algoritma, Ekstra Ağaç algoritmasıdır.
- En yüksek kesinlik değeri (0,95) Ekstra Ağaç algoritması ile elde edilmiştir.
- En yüksek duyarlılık değeri (0,98), Destek Vektör Makinesi algoritması uygulanarak elde edilmiştir.
- En yüksek F1-Skor değeri (0,95), Torbalama ve Ekstra Ağaç algoritmaları ile alınmıştır.
- En düşük doğruluk değeri (0,82) K-En Yakın Komşu algoritması ile elde edilmiştir.
- En düşük kesinlik değeri (0,84) K-En Yakın Komşu algoritması ile elde edilmiştir.
- En düşük duyarlılık değeri (0,94) Karar Ağacı algoritması ile elde edilmiştir
- En düşük F1-Skor değeri (0,89), K-En Yakın Komşu algoritması ile elde edilmiştir.



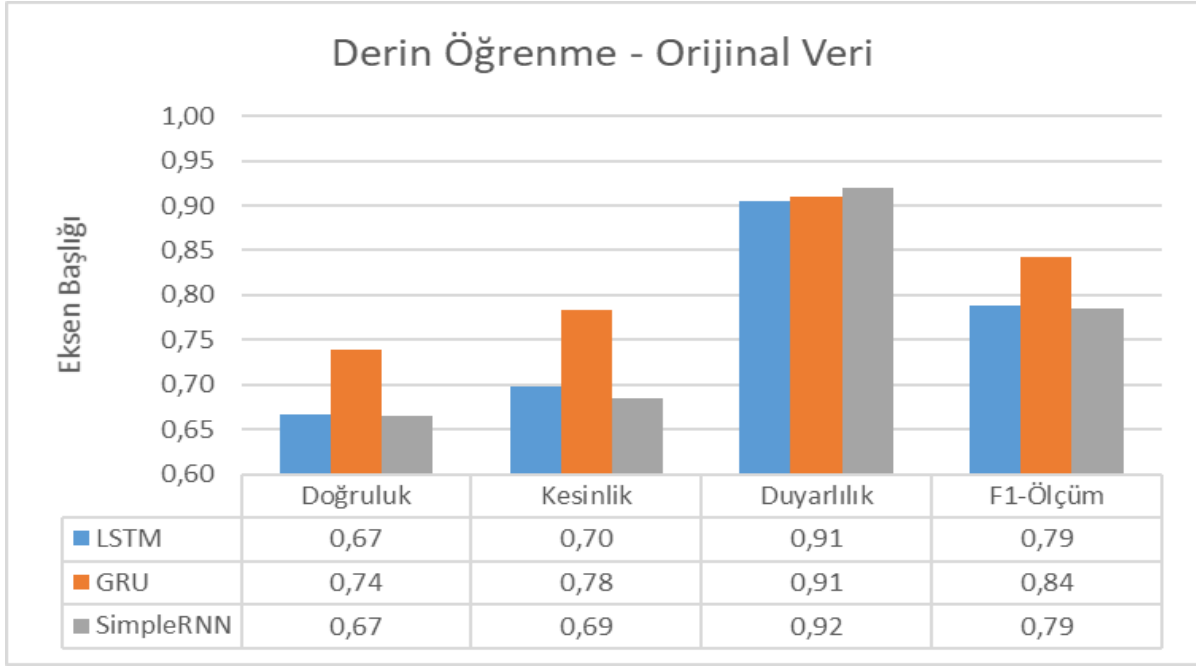
Şekil 5.2 – SMOTE Kullanılarak Alınan Sonuçlar – Makine Öğrenmesi

Sonuç – 3

SMOTE yöntemi uygulanmadan orijinal veri seti üzerine uygulanan çeşitli derin öğrenme yöntemleri ile alınan sonuçların grafiği Şekil 5.3 üzerinde gösterilmektedir. Oluşan grafik üzerinden sonuçlar şu şekilde yorumlanabilir;

- En yüksek doğruluğu (0,74) veren derin öğrenme yöntemi GRU olmuştur.
- En yüksek kesinlik değeri (0,78) veren derin öğrenme yöntemi GRU olmuştur.

- En yüksek duyarlılık değeri (0,92) veren derin öğrenme yöntemi Basit RNN olmuştur.
- En yüksek F1-Ölçüm değeri (0,84) GRU derin öğrenme yöntemi ile elde edilmiştir.
- En düşük doğruluk değeri (0,67) Basit RNN ve LSTM yöntemleri ile elde edilmiştir.
- En düşük kesinlik değeri (0,69) Basit RNN yöntemi ile elde edilmiştir.
- En düşük duyarlılık değeri (0,91) LSTM ve GRU yöntemleri ile elde edilmiştir.
- En düşük F1-Ölçüm değeri (0,79) LSTM ve Basit RNN yöntemleri ile elde edilmiştir.

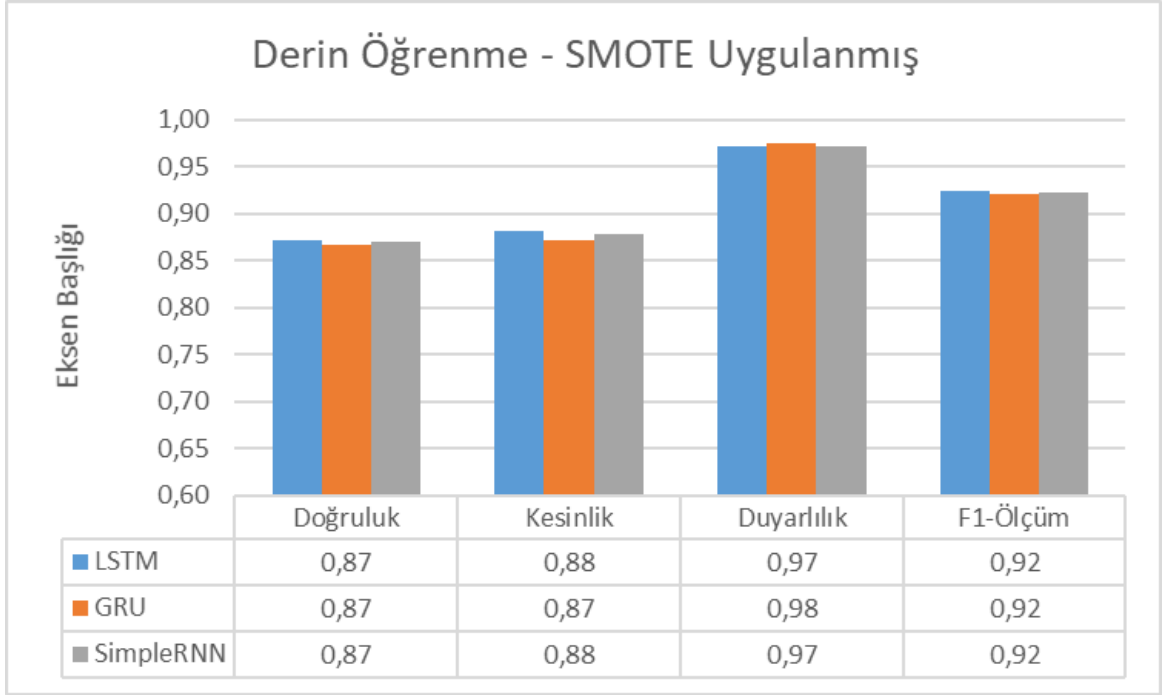


Şekil 5.3 - SMOTE Kullanılmadan Alınan Sonuçlar- Derin Öğrenme

Sonuç – 4

SMOTE yöntemi uygulandıktan sonra çeşitli derin öğrenme yöntemleri ile alınan sonuçların grafiği Şekil 5.4 üzerinde gösterilmektedir. Oluşan grafik üzerinden sonuçlar şu şekilde yorumlanabilir;

- Tüm derin öğrenme yöntemleri 0,87 ile aynı doğruluk değerini üretmiştir.
- En yüksek kesinlik değeri 0,88 ile LSTM ve Basit RNN yöntemleri ile alınmıştır.
- En yüksek duyarlılık değeri 0,98 ile GRU yöntemi ile alınmıştır.
- Tüm derin öğrenme yöntemleri 0,92 ile aynı f1-ölçüm değerini üretmiştir.
- En düşük kesinlik değeri 0,87 ile GRU yöntemi ile alınmıştır.
- En düşük duyarlılık değeri 0,97 ile LSTM ve Basit RNN yöntemleri ile alınmıştır.



Şekil 5.4 - SMOTE Kullanılarak Alınan Sonuçlar- Derin Öğrenme

Sonuç – 5

SMOTE yöntemi uygulandıktan sonra alınan sonuçlar literatürde aynı veri seti üzerinde yapılmış olan diğer çalışmaların sonuçları Tablo 5-1 üzerinde karşılaştırılmıştır. Ortaya çıkan tablo şu şekilde yorumlanabilir;

- Literatürdeki diğer çalışmalar üzerinde SMOTE tekniğini uygulayan tek çalışma Islam ve diğerleri [17] tarafından yapılan çalışmadır. Burada tek bir makine öğrenmesi yöntemi ile sonuç alınmıştır. Bu veri seti üzerinde SMOTE yöntemi uygulayan ikinci çalışma bu tez çalışması olmuştur. Bu çalışma içerisinde Islam ve diğerlerinin çalışmasına göre çok daha fazla makine öğrenmesi ve derin öğrenme yöntemleri ile sonuç alınmıştır.
- Yapılan tez çalışması LSTM, GRU ve Simple RNN yöntemleri ile sonuç alan ilk çalışmadır.
- Tez çalışması içerisinde alınan sonuçlar incelendiğinde makine öğrenmesi yöntemi ile alınan sonuçlarda diğer çalışmalara benzer sonuçlar aldığı görülmüştür. Derin öğrenme alanında yapılan çalışmalarda karşılaştırılacak, daha önce yapılmış olan bir çalışma olmadığından başarı oranlarında bir karşılaştırma yapılamamıştır.
- Tabloya eklenmeyen fakat bu tez çalışması kapsamında yapılan farklı makine öğrenmesi yöntemleri de mevcuttur. Bu yöntemler; Ekstra Ağaçlar, Torbalama ve Gradyen Artırma Algoritmasıdır.

Tablo 5-1 Literatürdeki Çalışmalar ile Alınan Sonuçların Karşılaştırılması

Bildiri	Yılı	KNN	RF	LR	ANN	NB	SVM	DT	CNN	RFNN	LSTM	GRU	SIMPLE RNN
Zeinulla ve diğerleri [13]	2018	86	91	86	90	87	89	-	-	-	-	-	-
Frempong ve Jayabalan [15]	2017	-	87	84	-	-	85	83			-	-	-
Islam ve diğerleri [17]	2019	-	-	-	-	89	-	-	-	-	-	-	-
Kim ve diğerleri [28]	2015	71		75		72	73	70	76	-	-	-	-
Hassan ve diğerleri [14]	2019	87	-	90	88	87	89	-	-	-	-	-	-
Puteri ve diğerleri [29]	2019	-	-	-	-	-	-	88	-	95	-	-	-
Tez Çalışması İçerisinde Alınan Sonuçlar	2021	82	88	-	86	84	85	88	-	-	87	87	87

6. SONUÇLAR

Pazarlama tahmini tüm sektördeki firmalar için önem arz etmektedir. Doğru pazarlama tahminleri şirketleri hem fazla zaman harcamaktan kurtarabilmekte hem de doğru müşteriye doğru ürün sunmalarını, bu sayede başarı oranlarında artış yakalamalarını sağlayabilmektedir. Özellikle bankalar için yeni müşteriler kazanmak ve var olan müşterilerine yeni ürünler pazarlamak önem arz etmektedir.

Bu tez çalışması içerisinde, banka tele-pazarlama verilerine uygulanan makine öğrenme yöntemleri ve derin öğrenme yöntemleri ile elde edilen sonuçlar karşılaştırılmıştır. Bu karşılaştırmalar yapılırken Karar Ağacı (DT), Destek Vektör Makinesi (SVM), Rastsal Orman (RF), Ekstra Ağaçlar (ET), Torbalama, Gradyen Artırma (GBA), K-En Yakın Komşu (KNN), Uzun-Kısa Vadeli Bellek (LSTM), Geçitli Tekrarlayan Birim (GRU), Yapay Sinir Ağı (ANN) ve Basit RNN yöntemleri kullanılarak sonuçlar elde edilmiştir. Sonuçlar incelendiğinde, derin öğrenme yöntemlerinin dengeli veri setlerinde dengesiz veri setlerine göre daha yüksek sonuçlar verdiği sonucuna varılmıştır. Aynı zamanda bütün derin öğrenme yöntemlerinin dengeli veri seti üzerinde benzer sonuçlar verdiği görülmüştür. Makine öğrenmesi yöntemleri incelendiğinde, çalışmada kullanılan veri seti üzerinde en iyi sonucu Gradyan Artırma Algoritması ile Yapay Sinir Ağları yapısının verdiği görülmüştür. Veri seti dengeli hale getirildiğinde ise en iyi doğruluk değeri Ekstra Ağaçlar algoritması ile alınmıştır.

İleride yapılacak çalışmalarda, veri setindeki özelliklerin sonuçlar üzerindeki etkisinin incelenmesi ve benzer karşılaştırmalara göre daha az etkili olan özelliklerin veri setinden tekrar edilmesi amaçlanmaktadır.

KAYNAKÇA

- [1] J. Kristin, P. Frank and C. Jennifer, *Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law*, New York, 2019.
- [2] M. Haenlein and A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present and Future of Artificial Intelligence," vol. 4, no. 61, pp. 5-14, 2019.
- [3] Oracle, [Online]. Available:
https://docs.oracle.com/cd/E18283_01/datamine.112/e16808/process.htm. [Accessed 2021].
- [4] F. Balsoz, *Bankacılıkta değişen pazarlama anlayışı*, Ankara: Ankara Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 2017.
- [5] S. Moro, P. Cortez and P. Rita, *A Data-Driven Approach to Predict the Success of Bank Telemarketing*, Elsevier, June 2014.
- [6] S. Moro, R. Laureano and P. Cortez, "Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodologht," in *Proceedings of the European Simulation and Modelling Conference-ESM'2011*, Guimaraes, Portugal, 2011.
- [7] T. Mitchell, *Machine Learning*, New York: McGraw Hill, 1997.
- [8] J. R. Koza, H. Bennett, D. Andre and M. A. Keane, *Automated Design of Both Topology and Sizing of Analog Electrical Circuits Using Generic Programming*, Artificial Intelligence in Design, 1996.
- [9] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, Third Edition*, Prentice Hall, 2010.
- [10] M. Mohri, A. Rostamizadeh and A. Talwalkar, *Foundations of Machine Learning*, The MIT Press, 2012.
- [11] G. Hinton and T. Sejnowski, *Unsupervised Learning: Foundations of Neural Computation*, MIT Press, 1999.
- [12] X. Xhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2009, p. 130.

- [13] E. Zeinulla, K. Bekbayeva and A. Yazici, "Comparative study of the classification models for prediction of bank telemarketing," in *IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, Almaty, Kazakhstan, 2018.
- [14] D. Hassan, A. Rodan, M. Salem and M. Mohammad, "Comparative Study of using Data Mining Techniques for Bank Telemarketing Data," in *2019 Sixth HCT Information Technology Trends (ITT)*, Ras Al Kaimah, United Arab Emirates, 2019.
- [15] J. Asare-Frempong and M. Jayabalan, "Predicting customer response to bank direct telemarketing campaign," in *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*, Kuala Lumpur, 2017.
- [16] L. Breiman, Random Forests. *Machine Learning*, 2001, pp. 5-32.
- [17] M. S. Islam, M. Arifuzzaman and M. S. Islam, "SMOTE Approach for Predicting the Success of Bank Telemarketing," in *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, Bangkok, Thailand, 2019.
- [18] L. Mohan, J. Pant, P. Suyal and A. Kumar, "Support Vector Machine Accuracy Improvement with Classification," in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Bhimtal, India, 2020.
- [19] T. E. Tallo and A. Musdholifah, "The Implementation of Genetic Algorithm in Smote(Synthetic Minority Oversampling Technique) for Handling Imbalanced Dataset Problem," in *2018 4th International Conference on Science and Technology (ICST)*, Yogyakarta, 2018.
- [20] A. M. Rahat, A. Kahir and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, Moradabad, India, 2019.
- [21] S. Yang, X. Yu and Y. Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," in *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, Shanghai, China, 2020.
- [22] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997.

- [23] M. Jeevan, "How is machine learning different from AI and Data Science," [Online]. Available: <https://laptrinhx.com/how-is-machine-learning-different-from-ai-and-data-science-4229307150/>. [Accessed January 2021].
- [24] F. Lotte, "Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain-Computer Interfaces," [Online]. Available: https://www.researchgate.net/figure/Principle-of-semi-supervised-learning-1-a-model-eg-CSP-LDA-classifier-is-first_fig4_277605013. [Accessed January 2021].
- [25] Y. N. Dong and G. S. Liang, "Research and Discussion on Image Recognition and Classification Algorithm Based on Deep Learning," in *2019 International Conference on Machine Learning, Big Data and Big Intelligence (MLBDBI)*, Taiyuan, China, 2019.
- [26] Li, Chenming&Wang, Tongchang&Zhang, Xiaoke&Gao, Hongmin&Yang, Yao&Wang and Jiawei, "Deep Belief Network for Spectral-Spatial Classification for Hyperspectral Remote Sensor Data," 2019.
- [27] P. Samudre, P. Shende and V. Jaiswal, "Optimizing Performance of Convolutional Neural Network Using Computing Technique," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India, 2019.
- [28] K. Kim, C. Lee, S. Jo and S. Cho, "Predicting the success of bank telemarketing using deep convolutional neural network," in *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, Fukuoka, 2015.
- [29] A. N. Puteri, Dewiani and Z. Tahir, "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods.," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2019.
- [30] D. Bu, Y. Liu, J. Guo, Q. Chen and T. Zheng, "Optimal Holding Time in Telemarketing," in *2010 International Conference on Management and Service Science*, Wuhan, 2010.
- [31] J. Kozak and P. Juszczuk, "The ACDF Algorithm in Stream Data Analysis for the Bank Telemarketing Campaign," in *2018 5th International Conference on Soft Computing & Machine Intelligence (ISCFMI)*, Nairobi, Kenya, 2018.

- [32] M. Cetiner and O. K. Sahingoz, "A Comparative Analysis for Machine Learning based Software Defect Prediction Systems," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020.
- [33] K. Spirina, "How Artificial Neural Networks Can Code Smarter Than GUI Programmer," [Online]. Available: <https://hackernoon.com/how-artificial-neural-networks-can-code-smarter-than-gui-programmer-1cdfaeb4851>. [Accessed 1 January 2021].
- [34] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, Journal Of Artificial Intelligence Research, 2002.
- [35] A. Patil, A. Framewala and F. Kazi, "Explainability of SMOTE Based Oversampling for Imbalanced Dataset Problems," in *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, San Jose, CA, USA, 2020.