

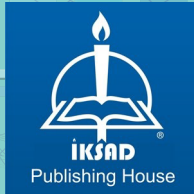
GAYRİMENKUL YATIRIM ORTAKLIKLARINDA KÂRLILIK BELİRLEYİCİLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNLEMESİ

Yazan:

Dr. Abdurrahman Coşkuner

Editör:

Doç. Dr. Ömer Faruk Rençber



**GAYRİMENKUL YATIRIM ORTAKLIKLARINDA
KÂRLILIK BELİRLEYİCİLERİNİN VERİ
MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNLEMESİ**

Yazan:

Dr. Abdurrahman Coşkuner

Editör:

Doç. Dr. Ömer Faruk Rençber

DOI: <https://dx.doi.org/10.5281/zenodo.11623921>



Copyright © 2024 by iksad publishing house

All rights reserved. No part of this publication may be reproduced, distributed or transmitted in any form or by any means, including photocopying, recording or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law. Institution of Economic Development and Social

Researches Publications®

(The Licence Number of Publicator: 2014/31220)

TÜRKİYE TR: +90 342 606 06 75

USA: +1 631 685 0 853

E mail: iksadyayinevi@gmail.com

www.iksadyayinevi.com

It is responsibility of the author to abide by the publishing ethics rules.

Iksad Publications – 2024©

ISBN: 978-625-367-723-7

Cover Design: İbrahim KAYA

June / 2024

Ankara / Türkiye

Size = 16x24 cm

ÖNSÖZ

Hayatımızın büyük bir kısmını geçirdiğimiz, güvenli ve huzurlu bir ortam sağlayan konutlar, bireylerin ve toplumların refahı için kritik bir rol oynamaktadır. . Gayrimenkul yatırım ortaklıkları, günümüzde yatırımcılar için önemli birer araç haline gelmiştir. Bu ortaklıklar, konut ve ticari gayrimenkul projelerine yatırım yaparak hem bireylerin hem de kurumsal yatırımcıların portföylerini çeşitlendirmelerine ve gayrimenkul sektöründeki fırsatları değerlendirmelerine olanak tanır. Bu kitap, gayrimenkul yatırım ortaklıklarının (GYO) önemini vurgulanması amacıyla hazırlanmıştır. GYO'ların karlılığını etkileyen faktörlerin detaylı bir şekilde incelenmesiyle, kitap özgün bir bakış açısı sunmaktadır. Kitap, gayrimenkul yatırım ortaklıklarının işleyişini ve karlılık faktörlerini anlamak isteyen herkes için bir kılavuz niteliği taşımaktadır. Hem konutun insan yaşamındaki rolünü anlamak hem de GYO'ların kârlılığını etkileyen faktörleri incelemek isteyen okuyuculara yol gösterici olmayı amaçlamaktadır.

Abdurrahman Coşkuner

İÇİNDEKİLER

ÖNSÖZ	i
TABLolar LİSTESİ.....	v
ŞEKİLLER LİSTESİ	vi
GİRİŞ	1
2. LİTERATÜR İNCELEMESİ.....	6
2.1. GYO Firmalarını Konu Alan Çalışmalar	6
2.2. Veri Madenciliği Yöntemlerini Konu Alan Çalışmalar	18
3. KAVRAMSAL ÇERÇEVE	22
3.1. Gayrimenkul Kavramı ve Önemi	22
3.2 Konut Finansmanı ve GYO.....	24
3.3. GYO'ların Tarihsel Gelişimi.....	26
3.4. Dünya'da GYO Sektörünün Mevcut Görünümü	27
3.4.1. ABD'de GYO Sektörünün Mevcut Görünümü	27
3.4.2. Avrupa'da GYO Sektörünün Mevcut Görünümü	29
3.4.3. Asya'da GYO Sektörünün Mevcut Görünümü	30
3.4.4. Türkiye'de GYO Sektörünün Mevcut Görünümü.....	30
3.5. GYO'ların Ekonomik Katkıları.....	31
3.6. GYO Kuruluşları	32
3.6.1.Ulusal Gayrimenkul Yatırım Ortaklıkları Birliği (NAREIT).....	32
3.6.2. Avrupa Kamu Gayrimenkul Birliği (EPRA)	33
3.6.4. Gayrimenkul Yatırımcıları Derneği (GYODER)	34
3.7. GYO Türleri	35
3.7.1. Konut GYO.....	37
3.8. GYO'lara Yönelik Yasal Düzenleme ve Kısıtlamalar	38
3.8.1. Avusturalya GYO Firmalarına Yönelik Yasal Düzenlemeler.....	38

3.8.2. Belçika GYO Firmalarına Yönelik Yasal Düzenlemeler	39
3.8.3. Kanada GYO Firmalarına Yönelik Yasal Düzenlemeler	39
3.8.4. Japonya GYO Firmalarına Yönelik Yasal Düzenlemeler ...	40
3.8.5. Türkiye GYO Firmalarına Yönelik Yasal Düzenlemeler....	41
3.8.6. Birleşik Krallık GYO Firmalarına Yönelik Yasal Düzenlemeler.....	42
3.8.7. ABD GYO Firmalarına Yönelik Yasal Düzenlemeler	43
3.8.8. Yasal Düzenleme ve Kısıtlamaların Karşılaştırılması.....	44
4.1. Veri Madenciliği Temel Bilgiler	46
4.1.1. Veri Madenciliği Uygulama Süreçleri.....	48
4.1.1.1. Veri tabanlarında Bilgi Keşfi Süreci (KDD)	48
4.1.1.2. Örneklem, Keşfetme, Değiştirme, Modelleme, Değerlendirme Süreci (SEMMA)	50
4.1.1.3. Veri Madenciliği için Sektörler Arası Standart Süreç (CRISP-DM).....	52
4.1.2. Veri Madenciliğinde Veri Seti Türleri.....	54
4.1.3. Veri Madenciliği Uygulamalarında Karşılaşılan Problemler	59
4.2. Çalışmada Uygulanan Yöntemler	63
4.2.1. Regresyon Analizi	66
4.2.1.1. Regresyon Analizi Performans Ölçütleri.....	67
4.2.2. Topluluk Öğrenme Yöntemleri	72
4.2.3. Rassal Orman Yöntemi.....	77
4.2.4. XGBoost Yöntemi	82
4.2.5. CatBoost Yöntemi	88
4.2.6. Sentetik Azınlık Aşırı Örneklem (SMOTE) Yöntemi	93
4.2.7. Değişken Önem Derecelerinin Tespitinde Kullanılan Yöntemler	98

5.1. Çalışmanın Amacı	100
5.2. Çalışmanın Kapsamı	100
5.3. Çalışmanın Özgünlüğü	102
5.4. Çalışmanın Sınırlılıkları	103
5.5. Çalışmada Kullanılan Değişkenler	103
5.6. Çalışmada Yer Alan Ülke Grupları	105
5.7. Veri Seti Hazırlık Uygulamaları ve Yöntemler	106
5.8. Rassal Orman Regresyon Uygulaması	108
5.8.1. Rassal Orman Regresyon Uygulamaları (1. Grup)	108
5.8.2. Rassal Orman Regresyon Uygulamaları (2. Grup)	110
5.9. CatBoost Regresyon Uygulaması	113
5.9.1. CatBoost Regresyon Uygulamaları (1. Grup)	113
5.9.2. CatBoost Regresyon Uygulamaları (2. Grup)	116
5.10. XGBoost Regresyon Uygulaması	118
5.10.2. XGBoost Regresyon Uygulamaları(2. Grup)	121
5.11. Tartışma	124
KAYNAKÇA	132

TABLolar LİSTESİ

Tablo 1: ABD’de piyasa değerleri büyüklüğüne göre GYO firmaları	28
Tablo 2: ABD GYO işçi sayıları	29
Tablo 3: Avrupa’da piyasa değerleri büyüklüğüne GYO firmaları	29
Tablo 4: Asya’da piyasa değerleri büyüklüğüne GYO firmaları	30
Tablo 5: Türkiye’de piyasa değerleri büyüklüğüne GYO firmaları	30
Tablo 6: Ülkelerde faaliyet gösteren konut GYO firma sayıları	37
Tablo 7: Asgari sermaye gerekliliğine göre ülkeler	44
Tablo 8: Temettü dağıtma zorunluluğuna göre ülkeler	45
Tablo 9: Borçlanma kısıtlamalarına göre ülkeler	45
Tablo 10: Sınıflandırma ve regresyon modelleri algoritmaları	65
Tablo 11: Rassal orman algoritması	81
Tablo 12: Rassal orman regresyon model parametreleri	82
Tablo 13: XGBoost model parametreleri	87
Tablo 14: CatBoost model parametreleri	92
Tablo 15: SMOTE algoritması	96
Tablo 16: Konut GYO firmaları	101
Tablo 17: Değişkenler	105
Tablo 18: Ülke grupları	106
Tablo 19: Yöntemler	107
Tablo 20: RF optimum model parametreleri (1. Grup)	109
Tablo 21: RF performans ölçütleri (1. Grup)	109
Tablo 22: RF optimum model parametreleri (2. Grup)	111
Tablo 23: RF performans ölçütleri (2. Grup)	111
Tablo 24: CB optimum model parametreleri (1. Grup)	114
Tablo 25: CB performans ölçütleri (1. Grup)	114
Tablo 26: CB optimum model parametreleri (2. Grup)	117
Tablo 27: CB performans ölçütleri (2. Grup)	117
Tablo 28: XGB optimum model parametreleri (1. Grup)	119
Tablo 29: XGB performans ölçütleri (1. Grup)	120
Tablo 30: XGB optimum model parametreleri (2. Grup)	122
Tablo 31: XGB performans ölçütleri (2. Grup)	122
Tablo 32: Yöntem performanslarının karşılaştırılması (1. Grup)	124
Tablo 33: Yöntem performanslarının karşılaştırılması (2. Grup)	124
Tablo 34: Değişken önem sıraları	125

ŞEKİLLER LİSTESİ

Şekil 1: GYO'ların ekonomik katkıları	32
Şekil 2: GYO türleri	36
Şekil 3: KDD süreci aşamaları	49
Şekil 4: SEMMA süreci aşamaları	51
Şekil 5: CRISP-DM	53
Şekil 6: Veri seti türleri	55
Şekil 7: Hold-Out doğrulama yöntemi.....	57
Şekil 8: Çapraz doğrulama (CV) yöntemi	58
Şekil 9: Optimum model	60
Şekil 10: Optimum model	62
Şekil 11: Veri madenciliği yöntemleri.....	64
Şekil 12: Denetimli öğrenme.....	65
Şekil 13: Topluluk öğrenme yöntemleri	74
Şekil 14: Karar ağaçları	75
Şekil 15: Rassal orman yöntemi	78
Şekil 16: XGBoost yöntemi akış şeması	87
Şekil 17: CatBoost yöntemi akış şeması	92
Şekil 18: Sınıf dengeleme yaklaşımları	94
Şekil 19: SMOTE yöntemine göre sentetik veri noktaları	95
Şekil 20: SMOTE yöntemi	96
Şekil 21: RF ağaç sayısı – MSE ilişkisi (1. Grup)	108
Şekil 22: RF model grafiği (1. Grup).....	109
Şekil 23: RF değişken önem dereceleri (1. Grup).....	110
Şekil 24: RF ağaç sayısı – MSE ilişkisi (Grup 2)	111
Şekil 25: RF model grafiği (2. Grup).....	112
Şekil 26: RF değişken önem dereceleri (2. Grup).....	112
Şekil 27: CB hiperparametre – MSE ilişkisi (1. Grup)	114
Şekil 28: CB model grafiği (1. Grup)	115
Şekil 29: CB değişken önem dereceleri (1. Grup)	115
Şekil 30: CB hiperparametre – MSE ilişkisi (2. Grup)	116
Şekil 31: CB model grafiği (2. Grup)	117
Şekil 32: CB değişken önem dereceleri (2. Grup)	118
Şekil 33: CB değişken önem dereceleri (2. Grup)	119
Şekil 34: XGB model grafiği (1. Grup)	120

Şekil 35: XGB deęişken önem dereceleri (1. Grup)	121
Şekil 36: XGB ağaç sayısı – MSE ilişkisi (2. Grup).....	121
Şekil 37: XGB model grafięi (2. Grup)	123
Şekil 38: XGB deęişken önem dereceleri (2. Grup)	123

BÖLÜM 1

GİRİŞ

Gayrimenkul, bir yerde sabit olarak duran, taşınamayan arsa, bina gibi mallar olarak tanımlanabilmektedir. Gayrimenkul, arsanın üzerinde yapı bulunup bulunmaması durumuna göre ikiye ayrılabilir. Arsanın üzerinde yapı bulunuyor ise geliştirilmiş gayrimenkul, yapı bulunmuyor ise geliştirilememiş gayrimenkul olarak nitelendirilmektedir (Seçkin, 1998). Gayrimenkul sektörü, gelişmekte olan ekonomiler için büyük bir öneme sahip ve aynı zamanda çimento, mobilya, dekorasyon malzemeleri, ev tekstili gibi çeşitli sektörlerin büyümesine de katkıda bulunmaktadır (Demirdöven, 2009). Konutlar, ticari binalar, sanayi tesisleri, tarım arazileri ve daha birçok gayrimenkul türü bulunmaktadır (Hepşen, 2009). Gayrimenkul türleri içerisinde “Konut” insanlar için büyük önem taşımaktadır. Konut, insanların en temel ihtiyaçlarından biri olan barınma ihtiyacını karşılayan önemli bir gayrimenkuldür. 1948 İnsan Hakları Beyannamesi'nde, konut bir insan hakkı olarak tanımlanmıştır (United Nations, 1948). Birleşmiş Milletler, BM-HABITAT programı aracılığıyla herkes için yeterli konut sağlanması hedefini benimsemektedir (United Nations, 1996).

Türkiye'nin nüfusu, 2025 yılında 85.569.125, 2030 yılında 88.427.604, ve 2035 yılında 90.680.302 kişiye ulaşması tahmin edilmektedir (TÜİK, 2022). Nüfus artışı, muhtemelen konut ihtiyacının da artacağı anlamına gelmektedir. Bununla birlikte, Türkiye'de hane halkı sayısında sürekli bir düşüş eğilimi gözlenmektedir. 2016 yılında ortalama hane halkı sayısı 3.48 iken, 2021 yılında bu rakam 3.23'e düşmüştür (TÜİK, 2022). Nüfus artışına rağmen hane halkı sayısındaki bu azalış, konut ihtiyacının önemli bir göstergesi olarak kabul edilebilmektedir. Ayrıca, doğal afetler de konut talebinin artmasına neden olabilen bir diğer faktör olarak gösterilmektedir. Bu tür olaylar, mevcut konut stokunun zarar görmesi veya yıkılmasıyla sonuçlanabilmekte, böylece konut talebini artırabilmektedir (Félix vd., 2013). Deprem gibi doğal afetlerin ardından evsiz kalan insanlar için acil konut ihtiyacı ortaya çıkmakta ve bu durumun genel konut talebini artırması beklenmektedir. Bu nedenle, doğal afetler ülkemizde konut ihtiyacını artıran önemli faktörler arasında yer almaktadır. Özellikle 6 Şubat depremleri gibi afetler, konut talebini önemli ölçüde artırabilmektedir. Türkiye Cumhuriyeti 11. Kalkınma Planı'nda, konut

ihtiyacıyla ilgili olarak Madde 685 ve 686 önemli hedefler yer almaktadır. Madde 685'te, dar gelirli başta olmak üzere herkesin, yeterli, yaşanabilir, dayanıklı, güvenli, kapsayıcı, ekonomik olarak karşılanabilir, sürdürülebilir, iklim değişikliğine dirençli ve temel altyapı hizmetlerine sahip konutlara erişiminin sağlanmasını temel amaç olarak belirlemektedir. Madde 686'da ise kentleşme, nüfus artışı, yenileme ve afetten kaynaklanan konut ihtiyacının arz-talep dengesi gözetilerek karşılanacağına vurgu yapmaktadır (T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı, 2019). Konut ihtiyacı, sadece barınma açısından değil, aynı zamanda sosyal, ekonomik ve çevresel faktörler açısından da büyük öneme sahiptir. Türkiye'nin nüfus artışı ve doğal afetlerle birlikte, konut politikalarının etkin bir şekilde yönetilmesi ve çözüm odaklı yaklaşımların benimsenmesi gerekmektedir. Bu politikaların, konut ihtiyacının karşılanmasında adil, sürdürülebilir ve toplumun ihtiyaçlarına uygun çözümler sunması büyük önem taşımaktadır.

Gayrimenkul sektöründeki, finansman zorlukları, değişen ekonomik denge ve kayıt dışı ekonomi gibi sorunlar, sektörün gelişimini olumsuz etkileyen unsurlar arasında yer almaktadır. Bu sorunların üstesinden gelmek ve sektöre yeni bir ivme kazandırmak amacıyla, gayrimenkul yatırım ortaklıkları (GYO) yeni bir model olarak ortaya çıkmıştır (Mert, 2012). GYO'lar, tüm süreçlerde resmî belgelendirme, tapu kaydı, vergi beyanları ve diğer yasal prosedürleri titizlikle takip etmektedir. Bu sayede, işlemler gerçek değerler üzerinden yürütülmekte ve kayıt dışı işlemlerin önüne geçilmektedir. Böylece, GYO'lar sektörde şeffaflığı ve güveni artırmaya yönelik bir rol oynamaktadır. GYO'lar ticari olarak farklı gayrimenkul türlerine yatırım yapmaktadırlar. Bu gayrimenkuller arasında ofisler, konutlar, depolar, perakende satış merkezleri, tıbbi tesisler, veri merkezleri, baz istasyonları, altyapı ve oteller gibi çeşitli türler bulunmaktadır (REİT, 2022).

Çalışmanın amacı, konut gayrimenkul yatırım ortaklıkları firmalarının kârlılığını etkileyen finansal göstergelerin önem sıralamalarını belirlemek ve bu amaç doğrultusunda kullanılan üç farklı veri madenciliği yöntemini (Rassal Orman, XGBoost, CatBoost) karşılaştırmaktır. Çalışmada konut GYO firmalarının örneklem olarak yer alma nedeni, aynı türdeki GYO firmalarının karşılaştırılması durumunda daha sağlıklı tespitler yapılabileceği ve insanların en temel ihtiyaçlarından olan barınma sorununa katkı sağlama potansiyeline

sahip olmasıdır. Dolayısıyla, elde edilecek sonuçların daha somut ve kapsayıcı olması beklenmektedir.

Ayrıca çalışmada, konut GYO firmalarının kârlılığını etkileyen finansal göstergelerin önem sırasını tespit etmek ve en uygun yöntemi belirlemek amacıyla çeşitli veri madenciliği yöntemlerinin kullanılması hedeflenmektedir. Çalışmanın istatistiki amacı, konut GYO firmalarının kârlılıklarını etkileyen finansal göstergelerin önem derecelerini en iyi tespit eden veri madenciliği yönteminin belirlenmesidir. Veri madenciliği, büyük miktardaki verilerden anlamlı bilgiler çıkarılmasını sağlayan bir dizi analiz ve modelleme tekniklerini içeren bir alan olarak tanımlanabilir (Hand, 2001). Bu çalışmada, Rassel Orman Regresyonu, XGBoost ve CatBoost gibi ağaç tabanlı topluluk öğrenme yöntemleri kullanılarak kârlılığı etkileyen finansal göstergelerin önem sıralarını tespit etmek için modeller oluşturulmuş ve karşılaştırılmıştır.

Karar ağaçları, veri madenciliği alanında oldukça etkili bir yöntem olarak kabul edilmektedir. Bu teknik, veri setlerini belirli özellikler ve karar kuralları temelinde analiz ederek, verileri hiyerarşik bir yapıya dönüştürme yeteneğine sahiptir (Rajan ve Krishnan, 2022). Tek bir karar ağacı, desenleri tespit etmek ve sınıflandırma veya tahmin yapmak için yetersiz kalabilmekte ve aşırı öğrenme eğilimi gösterebilmektedir (Chatpatanasiri, 2005). Rassel orman yöntemi ise birden fazla karar ağacını birleştirerek karar vermektedir (Buschjager ve Morik, 2018). XGBoost, yüksek performanslı karar ağacı modellerinin etkisini artırmak amacıyla ardışık olarak birden fazla karar ağacı oluşturmakta ve her bir karar ağacı, bir önceki ağacın hatalarını en aza indirmeye odaklanarak gelişmektedir (Kumar vd., 2022). Böylece, her bir ağaç hataları düzeltirken, genel performansı iyileştiren model oluşturulmaktadır. CatBoost, eğitim sürecinde, ardışık olarak bir dizi karar ağacı oluşturulmakta ve her bir sonraki ağaç, önceki ağaçtan elde edilen hataları azaltarak öğrenmektedir. Her bir karar ağacı, önceki ağacın öğrendiklerinden yararlanmakta ve bir sonraki ağacı etkileyerek modelin performansını artırmaktadır. Böylece güçlü bir model oluşturulmaktadır (Prokhorenkova et al., 2018).

Çalışmanın veri seti, InvestingPro web sayfasından elde edilmiş olup 2013.D1 - 2022.D1 dönemlerini kapsamaktadır. Çalışmada, sadece konut veya ağırlıklı olarak Konut sektöründe faaliyet gösteren, 32 konut GYO firmasına ait veriler kullanılmıştır. 32 firma 7 farklı ülkede faaliyet göstermektedir.

Gayrimenkul yatırım ortaklıkları kurumlar vergisinden muaftırlar. Ancak bu muafiyetin kazanılabilmesi için birtakım gereklilikler bulunmaktadır. Her yıl belirli oranda temettü dağıtım gerekliliği tüm ülkelerde öne plana çıkan önemli bir gereklilik olarak görülmektedir. Çoğu ülkelerde bu oran %90 seviyelerinde iken bazı ülkelerde söz konusu zorunluluk bulunamamakta bu oranın tespiti firma yöneticilerine bırakılmaktadır. Her yıl belirli bir oranda temettü dağıtma zorunluluğunun olmamasının, şirketlere iç finansman oluşturma konusunda önemli avantajlar sağlayacağı öngörülmektedir. İç fon yaratma avantajına sahip ve daha kısıtlı iç fon yaratma potansiyeline sahip firmaların kârlılıklarını etkileyen faktörlerin önem sıralarını karşılaştırmak amacı ile 32 firma, yıllık temettü dağıtma oranlarına göre iki gruba ayrılmıştır. Bağımlı değişken olarak, faiz amortisman vergi öncesi kâr (FAVÖK) oranı kullanılmıştır. Bağımsız değişkenler ise firma büyüklüğü (TA), net varlık değeri (NAV), kaldıraç oranı (LR), varlık devir hızı (AT) ve cari oran (CR)'dan oluşmaktadır.

Öncelikle, temettü dağıtma zorunluluğuna göre ülkeler gruplara ayrılmış ve veri setlerinin boyutları arasında dengesizlikler tespit edilmiştir. Bu dengesizliği gidermek amacıyla Sentetik Azınlık Aşırı Örneklemleme Tekniği (SMOTE) kullanılmıştır. Veri setindeki gruplar arası boyut dengesizlikleri giderildikten sonra, veri setinde bulunan veriler 0-1 aralığında normalleştirilmiştir.

Veri setinin hazırlanmasının ardından XGBoost, CatBoost ve Rassal Orman Regresyonu yöntemleri uygulanmış ve sonuçlar karşılaştırılmıştır. Elde edilen sonuçlar, konut GYO firmalarının kârlılığı üzerinde varlıkların önemli bir etkiye sahip olduğu tespit edilmiştir. İstatistiki açıdan, XGBoost ve CatBoost yöntemlerinin konut GYO firmalarının kârlılığını etkileyen faktörlerin önem derecelerini tespit etmede, Rassal Orman Regresyonu yöntemine göre daha başarılı sonuçlar elde ettiği tespit edilmiştir. Bu tespit veri madenciliği tekniklerinin kârlılık analizinde etkin bir şekilde kullanılabileceğini, XGBoost ve CatBoost yöntemlerinin konut GYO firmalarının kârlılıklarını tespit etmede başarılı yöntemler olduğunu göstermektedir. Elde edilen tespitlere göre yasa yapıcılara, firma yöneticilerine ve araştırmacılara çeşitli önerilerde bulunulmuştur.

Kitap dört bölümden oluşmaktadır. Birinci bölümde giriş, ikinci bölüm literatür incelemesinden oluşmaktadır. Üçüncü bölümde GYO sektörü hakkında kavramsal çerçeve sunulmuştur. Dördüncü bölümde veri madenciliği

ve yöntemlerine ilişkin bilgiler sunulmuştur. Beşinci bölümde üç yönteme göre uygulamalar gerçekleştirilmiştir ve uygulama sonuçları karşılaştırılmıştır.

BÖLÜM 2

LİTERATÜR İNCELEMESİ

2. LİTERATÜR İNCELEMESİ

Çalışmanın amacı, konut gayrimenkul yatırım ortaklıkları firmalarının kârlılığını etkileyen finansal göstergelerin önem sıralamalarını belirlemek ve bu amaç doğrultusunda kullanılan üç farklı veri madenciliği yöntemini karşılaştırmaktır. Bu çerçevede, literatür incelemesi bölümü, GYO Firmalarını Konu Alan Çalışmalar ve Veri Madenciliği Yöntemlerini Konu Alan Çalışmalar olmak üzere iki alt bölümden oluşmaktadır.

2.1. GYO Firmalarını Konu Alan Çalışmalar

Bu bölümde Gayrimenkul Yatırım Ortaklıkları (GYO) konu alan çalışmalar incelenmektedir. Bu çerçevede GYO'ları konu alan çalışmaların temel amaçları, çalışmalarda kullanılan yöntemler ve çalışmalardan elde edilen sonuçlar yer almaktadır.

Lu ve So (2001), GYO getirileri ile reel faaliyetler, para politikası ve enflasyon arasındaki ilişkiyi araştırmışlardır. Çalışmada, Ocak 1972 – Aralık 1995 dönemleri arası verileri ve Vektör Hata Düzeltme Modeli ile Granger Nedensellik testi yöntemlerini kullanmışlardır. Veriler, ABD'de faaliyet gösteren Ulusal Gayrimenkul Yatırım Ortaklıkları Derneği (NAREIT) REIT El Kitabı'ndan elde edilmiştir. Değişken olarak; GYO'ların toplam getiri endeksi, Federal Fon Oranı, Tüketici Fiyat Endeksi ve Sanayi Üretiminde değişim oranlarını kullanmışlardır. Çalışma sonucunda, GYO getirileri ile enflasyon arasında negatif yönlü bir ilişki olduğunu tespit etmişlerdir.

Kim vd. (2002), otel gayrimenkul GYO firmalarını genel GYO Sektörü ve altı GYO sektörü ile karşılaştırmış ve performanslarını incelemiştir. Çalışmanın örnekleme, New York Borsası, Amerikan Borsası ve Ulusal Menkul Kıymet Satıcıları

Birliği Otomatik Kotasyon sisteminde işlem gören 183 GYO'dan oluşmaktadır. Çalışmada, 1993 – 1999 dönemleri arası verileri ve tek yönlü varyans analizi ile ukey çoklu karşılaştırma yöntemini kullanmışlardır. Otel GYO'larının diğer GYO sektörlerine göre nispeten daha yüksek piyasa riskine sahip olduğunu tespit etmişlerdir.

Ambrose vd. (2005), New York Menkul Kıymetler Borsası, Nasdaq ve Amerikan Menkul Kıymetler Borsası'nda işlem gören gayrimenkul yatırım ortaklığı (GYO) şirketlerinin ölçek ekonomilerini araştırmışlardır. Çalışmada, Ocak 1990 - Aralık 2001 dönemleri arası verileri ve regresyon analizi yöntemini kullanmışlardır. Değişken olarak piyasa değeri, ödeme oranı, değişim oranında aralarında bulunduğu 27 değişkeni kullanmışlardır. Çalışma sonucunda, maliyetleri düşürmede büyük firmaların başarılı olduğunu tespit etmişlerdir.

Yetgin (2006), ABD, AB ve İslam ülkelerinde GYO sektörünün gelişimi hakkında bilgiler vermiştir. Çalışmada, GYO firmalarının ilk olarak, 1960 yılında ABD'de ortaya çıktığını, Türkiye'de 1995 yılında yapılan yasal düzenlemeler ile sermaye piyasası aracı olarak faaliyet gösterdiklerini belirtilmiştir. Ayrıca, gelecek yıllarda GYO modelinin AB ve İslam ülkelerinde yaygınlaşacağı belirtilmiştir.

Nishigaki (2007), ABD finans piyasasında gayrimenkul yatırım ortaklığı (GYO) getirileri ile konut fiyatı arasındaki uzun vadeli ilişkiyi araştırmıştır. Çalışmada, Ocak 1980 - Kasım 2006 dönemleri arası verileri ve Koentegrasyon Testi ile Vektör Hata Düzeltme Modelini kullanmıştır. Araştırmanın sonucunda, gayrimenkul yatırım ortaklığı (GYO) getirileri ile enflasyon arasında negatif yönlü bir ilişki, konut fiyatları ile ise pozitif yönlü bir ilişki tespit etmiştir.

Feng vd. (2007), GYO firmalarının, piyasa defter oranları ile kaldıraç oranları arasındaki ilişkiyi araştırmışlardır. Çalışmada, 1991 – 2003 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Çalışma sonucunda, oranlar arasında ilişki olduğunu, yüksek piyasa defter oranına sahip GYO'ların yüksek kaldıraç oranına sahip olma eğiliminde olduğunu tespit etmişlerdir.

Yat-Hung vd. (2008), ABD, Avustralya, Japonya ve Singapur GYO firmalarının getirileri ile diğer finansal ve reel menkul kıymetleştirilmemiş varlıkların uluslararası getirileri arasındaki dinamik ilişkileri araştırmışlardır. Çalışmada, 1995 – 2005 yılları arası verileri ve esnek en küçük kareler (FLS) yöntemini kullanmışlardır. Değişken olarak; üç aylık yerel büyük sermayeli hisse senedi endeksi getirileri, üç aylık yerel küçük sermayeli hisse senedi endeksi getirileri, üç aylık uzun vadeli tahvil fiyat endeksi getirileri ve menkul kıymetleştirilmemiş gayrimenkul yatırımının üç aylık getirileri kullanmışlardır. Araştırmanın sonucunda, gayrimenkul yatırım ortaklığı (GYO) getirileri ile

diğer finansal ve reel varlıkların getirileri arasındaki ilişkinin sadece ülkeler arasında değil aynı zamanda dönemler arasında da değiştiği belirlenmiştir.

Hayta (2009), GYO'ların performanslarını dört farklı regresyon modeli ile incelemiştir. Çalışmada, İstanbul Menkul Kıymetler Borsası'nda işlem gören GYO firmalarının 2002-2008 yılları arası verileri kullanmıştır. Çalışma sonucunda, dört farklı regresyon modeline göre GYO'ların başarılı performans gösterdiklerine ilişkin bulguya rastlamamıştır. Modellere ilişkin R2 değerlerinde, farklılık görülmemiştir. Ayrıca, büyük piyasa değerli firmalar ve küçük piyasa değerli firmalardan oluşan portföyler arasındaki getiri farklarına karşı GYO'ların duyarlı olduklarını tespit etmiştir.

Harrison vd. (2011), New York Menkul Kıymetler Borsası , Nasdaq ve Amerikan Menkul Kıymetler Borsası'nda faaliyet gösteren Gayrimenkul Yatırım Ortaklığı (GYO) firmalarının sermaye yapısı kararlarının belirleyicilerini araştırmışlardır. Çalışmada, 1990 – 2008 yılları arası verileri ve regresyon analizi yöntemini kullanmışlardır. Bağımlı değişken olarak kaldıraç oranını; bağımsız değişkenler olarak sabit varlıklar, kârlılık oranları ve piyasa defter değerlerini kullanmışlardır. Çalışma sonucunda, kaldıraç oranı ile sabit varlıklar arasında pozitif; kârlılık ve piyasa defter değerleri oranları ile negatif yönlü bir etki olduğunu tespit etmişlerdir.

Chang vd. (2011), para politikasının varlık piyasalarına aktarım mekanizmasını incelemek için Amerika Birleşik Devletleri'ndeki ana para politikası aracı olan Federal Fon Oranındaki değişikliklerin sırasıyla Hisse Senedi GYO'ları, konut ve hisse senedi getirileri üzerindeki etkisini araştırmışlardır. Çalışmada, 1975.Q2 – 2008Q1 dönemleri arası verileri ve VAR modelini kullanmışlardır. Çalışma sonucunda, Federal Fon Oranındaki değişikliğin GYO getirileri üzerinde etkili olduğunu tespit etmişlerdir.

Goebel vd. (2013), GYO getirileri ile firmaya özgü faktörler arasındaki ilişkiyi araştırmışlardır. Çalışmada, 1993 – 2009 yılları arası verileri ve çoklu doğrusal regresyon analizi yöntemini kullanmışlardır. Bağımlı değişken olarak, GYO getirileri; bağımsız değişkenler olarak piyasa değeri / defter değeri, öz sermaye, likidite azlığı, kurumsal mülkiyeti kullanmışlardır. Çalışma sonucunda, kurumsal sahiplik ve likidite azlığı ile GYO getirileri arasında güçlü bir ilişki olduğunu tespit etmişlerdir.

Jamal (2013), Malezya'da faaliyet gösteren GYO firmalarının performanslarını etkileyebilecek faktörleri ve faiz oranlarının GYO'ların

performanslarına etkisini araştırmıştır. Çalışmada, 2008- 2012 yılları arası verileri ve çoklu doğrusal regresyon analizi yöntemini kullanmıştır. Bağımlı değişken olarak Tasarruf hesabındaki mevduatları, bağımsız değişkenler olarak; İslami tasarruf hesabı için kâr oranları, Kuala Lumpur Bileşik Endeksi, Baz Borç Verme Oranı ve Tüketici Fiyat Endeksini kullanmıştır. Çalışma sonucunda, faiz oranı, temettü getirisi ve mülkiyet ortamı olan bağımsız değişkenin, bağımlı değişkenle iyi ve güçlü bir ilişkiye sahip olduğunu tespit etmiştir.

Deran vd., (2013), Türkiye’de faaliyet gösteren GYO firmaları ile menkul kıymet yatırım ortaklıkları firmalarının performanslarını karşılaştırmışlardır. Çalışmada, 2009-2011 yılları arası verileri kullanılmış ve oran analizi yapmışlardır. Çalışmada, aktif kârlılık, özsermaye kârlılık, hisse başına kâr ve piyasa değeri / Defter değeri oranını kullanmışlardır. Çalışma sonucunda, GYO’ların aktif kârlılık, özsermaye kârlılık ve hisse başına kâr oranlarının daha fazla olduğunu tespit etmişlerdir.

Fatnassi vd. (2014), para politikalarının GYO getirileri üzerindeki etkisini araştırmışlardır. Çalışmada, 1987.Q1 – 2011.Q3 dönemleri arası verileri ve Markov-anahtarlama modeli yöntemini kullanmışlardır. Bağımlı değişken olarak, GYO getirileri; bağımsız değişkenler olarak, enflasyon oranı, para arzı ve çıktı açığı kullanılmıştır. Çalışma sonucunda, para politikalarının GYO getirileri üzerinde doğrusal olmayan etkileri olduğunu tespit etmişlerdir.

Binti Mohamad ve Bin Zolkifli (2014), Asya GYO firmalarının performanslarını etkileyen faktörleri araştırmışlardır. Çalışmada, 2007-2011 yılları arası verileri ve çoklu doğrusal regresyon analizi yöntemini kullanmışlardır. Bağımlı değişken olarak, net varlık değeri ve getirileri; bağımsız değişken olarak risk, temettü geliri, net gelir ve firma büyüklüğünü kullanmışlardır. Çalışma sonucunda, getiri ile firma büyüklüğü; net varlık değeri ile firma büyüklüğü ve temettü getirisi arasında pozitif yönlü bir ilişki olduğunu tespit etmişlerdir.

Şahin (2014), GYO endeksi ile portföy büyüklüğü, fiyat / kazanç oranı ve piyasa değeri / defter değeri oranları arasındaki ilişkiyi tespit etmeyi amaçlamıştır. Çalışmada, BİST’de işlem gören firmaların Ocak 2002- Aralık 2011 dönemleri arası verilerini ve Çoklu regresyon yöntemini kullanmıştır. Çalışma sonucunda, portföy büyüklüğü ile GYO getirisi arasında pozitif, piyasa değeri / defter değeri oranı ile GYO getirisi arasında ilişki olmadığını tespit

etmiştir. Fiyat / kazanç değeri ile GYO getirisi arasında ise negatif yönlü bir ilişki tespit etmiştir.

Zügül ve Şahin (2015), GYO Firmalarının borsa getirileri ile enflasyon ve faiz oranları arasındaki ilişkiyi incelemişlerdir. Çalışmada, Ocak 2002 – Aralık 2012 dönemleri arası verileri ve en küçük kareler yöntemini kullanmışlardır. Bağımlı değişken olarak, GYO borsa getirilerini; bağımsız değişken olarak, enflasyon ve mevduat faiz oranlarını kullanmışlardır. Çalışma sonucunda, enflasyon oranı ile GYO getirileri arasında ilişki olmadığını tespit etmişlerdir. Faiz oranı ile GYO getirileri arasında ise negatif yönlü bir ilişki olduğunu tespit etmişlerdir.

Aytekin ve Kahraman (2015), BİST'te işlem gören GYO firmalarının performanslarını ölçmeyi amaçlamışlardır. Çalışmada, 2008 – 2012 yılları arası verileri ve Veri Zarflama analizi yöntemini kullanmışlardır. Girdi değişkeni olarak genel yönetim giderleri, finansman giderleri ve ödenmiş sermaye tutarlarını kullanmışlardır. Çıktı değişkeni olarak net aktif değeri, net dönem kârı ve piyasa değeri tutarlarını kullanmışlardır. Çalışma sonucunda, 2008 – 2012 yılları için etkin bulunan GYO firma sayılarını sırasıyla; 13, 13, 20, 22 ve 23 adet olarak tespit etmişlerdir.

Loo vd. (2016), Asya GYO piyasaları ile makroekonomik değişkenler arasındaki uzun ve kısa vadeli ilişkiyi araştırmışlardır. Çalışma, Japonya, Hong Kong, Singapur, Malezya, Tayland, Tayvan ve Güney Kore ülke firmalarına ait veriler ile yürütülmüştür.. Çalışmada, Johansen eş bütünleşme testi ve Granger nedensellik testini kullanmışlardır. Değişken olarak: GYO endeksi, uzun vadeli faiz oranı, gayri safi yurtiçi hasıla, kısa vadeli faiz oranı sanayi üretimi, enflasyon oranı, para politikası, maliye politikasını kullanmışlardır. Araştırmanın sonuçlarına göre, bazı gelişmekte olan gayrimenkul yatırım ortaklığı (GYO) piyasalarının uzun vadede makroekonomik değişkenlerle daha yüksek düzeyde entegrasyon sergilediği tespit edilmiştir. Bu durum, gelişmekte olan GYO piyasalarının, gelişmiş GYO piyasalarına kıyasla makroekonomik ortamdaki değişikliklere karşı daha hassas bir şekilde tepki verdiğini göstermektedir.

Fang vd. (2016) Japonya, Singapur ve Çin ülkelerindeki makroekonomik faktörlerin GYO endeksi üzerindeki etkisini araştırmışlardır. Çalışmada, Mart 2008 - Temmuz 2012 dönemleri arası verileri ve ARDL sınır testi yöntemini kullanmışlardır. Bağımlı değişken olarak, GYO endeksini; bağımsız değişken

olarak, hisse senedi endeksini, faiz ve enflasyon oranlarını kullanmışlardır. Çalışma sonucunda, GYO Endeksi ile hisse senedi endeksinin üç ülke için pozitif; enflasyon oranının ise olumsuz bir ilişkisi olduğunu ve tespit etmişlerdir. Ayrıca, faiz oranı ile Japonya ve Singapur GYO endeksi arasında negatif ilişki olduğunu tespit etmişlerdir.

Ma'In vd. (2016), Asya ve Orta Doğu ülkelerinde faaliyet gösteren İslami GYO firmalarının performanslarını etkileyen faktörleri tespit etmeyi amaçlamışlardır. Çalışmada, 2011 – 2015 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Bağımlı değişken olarak, Net Aktif Değerini kullanmışlardır. Bağımsız değişken olarak, faiz ve enflasyon oranları, firma büyüklüğü ve riskten oluşan makroekonomik faktörleri kullanmışlardır. Çalışma sonucunda, İslami GYO'ların performansı ile firma büyüklüğü arasında pozitif ilişkili olduğunu tespit etmişlerdir.

Glascok ve Lu-Andrews, (2016), makroekonomik değişkenlerin GYO hisse senetleri üzerindeki etkisini araştırmışlardır. Çalışmada, 1965.Q1 – 2011.Q1 dönemleri arası verileri ve panel regresyon analizi yöntemini kullanmışlardır. Bağımlı değişkenler olarak borç hizmet kapsamı oranı, kredi / değer oranı ve kredi sayısını kullanmışlardır. Bağımsız değişkenler olarak; Sanayi Üretimindeki büyüme oranı, yüzde olarak gerçekleşen enflasyon oranındaki değişim, işsizlik oranındaki değişim temerrüt primi, vade primi, inşaatta istihdamdaki değişim oranları kullanmışlardır. Çalışma sonucunda, makroekonomik değişkenlerin GYO hisse senetleri likiditesi üzerinde etkili olduğunu tespit etmişlerdir.

Nazlioglu vd. (2016), ABD'de halka açık olarak faaliyet gösteren; konut, otel, sağlık, perakende, ipotek ve depo/endüstriyel GYO'ları üzerinde petrol fiyatı şokları ve oynaklığının rolünü araştırmışlardır. Çalışmada, Ocak 2005–Aralık 2013 dönemleri arası verileri ve Fourier Toda-Yamamoto nedensellik testini kullanmışlardır. Çalışma sonucunda, petrol fiyatları ile sağlık, otel, depo/endüstriyel, konut ve perakende GYO'ları arasında önemli bir etkileşim olduğunu, ipotekli GYO'larda bir uygunsuzluk olduğunu tespit etmişlerdir.

Morri ve Parri (2017), ABD GYO firmalarının, sermaye yapısını ve ekonomik krizlerin firmalara etkisini analiz etmişlerdir. Çalışmada, 2005-2014 yılları arası verileri kullanmışlardır. Bağımlı değişken olarak finansal kaldıraç oranını; bağımsız değişkenler olarak kârlılık, varlıkların maddi olma durumu,

işletme risk, firma büyüklüğü, büyüme fırsatları ve krizleri kullanmışlardır. Çalışma sonucunda, finansal kaldıraç oranı ile varlıkların maddi olma durumu ile krizlerin pozitif; işletme riski ve büyüme fırsatlarının negatif yönde bir ilişki gösterdiğini tespit etmişlerdir.

Sha (2017), Endonezya Menkul Kıymetler Borsası'nda işlem gören GYO firmalarının fiyat kazanç oranı, hisse başına kazanç, defter/piyasa oranı ve gayri safi yurtiçi hasılanın hisse senedi fiyatı üzerindeki etkilerini araştırmıştır. Çalışmada, 2012 -2014 yılları arası verileri kullanmıştır. Hipotezlerin test edilmesinde t-testi ve F- testini kullanmıştır. Çalışma sonucunda, hisse senedi değeri ile hisse başına kazanç, defter/piyasa oranı ve gayri safi yurtiçi hasıla arasında anlamlı bir ilişki olduğunu tespit etmiştir.

Jakpar vd. (2018), Malezya'da faaliyet gösteren GYO firmalarının kârlılıklarını etkileyen faktörleri araştırmışlardır. Çalışmada, 2008 – 2015 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Bağımlı değişken olarak özkaynak kârlılık oranıdır; bağımsız değişkenler olarak faktörler temettü verimi, net varlık değeri ve fiyat kazanç oranını kullanılmışlardır. Çalışma sonucunda, sadece temettü verimi GYO kârlılığı arasında ilişki olduğunu tespit etmişlerdir.

Reddy ve Wong (2018), Avusturalya'da faaliyet gösteren GYO'ların faiz oranlarındaki değişimlere duyarlılıklarını araştırmışlardır. Çalışmada, 1995 – 2016 yılları arası verileri ve panel kantil regresyon yöntemini kullanmışlardır. Bağımlı değişkenler olarak kısa ve uzun vadeli faiz oranlarındaki değişimleri; bağımsız değişkenler olarak çeşitlendirilmiş, endüstriyel, perakende, ofis ve uzmanlaşmış fon getirilerini kullanmışlardır. Çalışma sonucunda, yükselen kısa vadeli faiz oranlarının pozitif getirilere, yükselen uzun vadeli faiz oranlarının daha düşük getirilere neden olduğunu tespit etmişlerdir.

Ma'ın vd. (2018), makroekonomik faktörlerin ve firma özelliklerinin Malezya'daki geleneksel GYO'ların performansı üzerindeki etkisini incelemişlerdir. Çalışmada, 2007 – 2014 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Bağımlı değişken olarak net varlık değeri; bağımsız değişken olarak enflasyon oranı, GSYİH ve piyasa değerini kullanmışlardır. Çalışma sonucunda, net varlık değeri ile GSYİH arasında pozitif; enflasyon ve piyasa değeri ile negatif yönlü bir ilişki olduğunu tespit etmişlerdir.

Çelik ve Manan (2018), Borsa İstanbul'da işlem gören GYO firmalarının risk ve performanslarını ölçmeyi amaçlamışlardır. Çalışmada, 2007-2016 yılları arası verileri ve yatay kesit çoklu regresyon modelini kullanmışlardır. Bağımlı değişken olarak, hisse getirisi ve piyasa değeri defter değerini; bağımsız değişken olarak, firma yaşı, toplam varlıklar, net kâr/aktif, cari oran, Toplam Borç / Toplam Varlıklar, Faiz ve vergi öncesi kâr/aktif oranlarını kullanmışlardır. Çalışma sonucunda, risk ve performans arasında ilişki tespit etmişlerdir.

Ünalı (2018), Borsa İstanbul'da işlem gören 10 GYO firmasının performansını incelemiştir. Çalışmada, 2011 – 2017 yılları arası verileri kullanmıştır. Bağımlı değişken olarak, Piyasa Değeri/Defter Değeri oranını; bağımsız değişken olarak, likidite, mali yapı, faaliyet, kârlılık oranlarını kullanmıştır. Çalışma sonucunda, Çalışma sonucunda, özsermaye çarpanı, özsermaye kârlılık oranı ve duran varlıklar/ özsermaye oranlarının Piyasa Değeri/Defter Değeri oranı üzerinde anlamlı bir etkiye sahip olduğunu tespit etmiştir.

Gümüş ve Can Öziç (2018), Borsa İstanbul'da işlem gören GYO firmalarının etkinliklerini ölçmeyi amaçlamışlardır. Çalışmada, 2013-2017 yılları arası veriler ve veri zarflama yöntemi kullanılmıştır. Girdi değişkenleri olarak cari oran, nakit oranı, alacak devir hızı, özsermaye devir hızı ve aktif devir hızı; çıktı değişkenleri olarak özsermaye kârlılık oranı, net kâr marjı, brüt kâr marjı ve borçlanma oranını kullanmışlardır. Çalışma sonucunda, 2013 – 2017 yılları için etkin bulunan GYO firma sayıları sırası ile; 13, 8, 9, 6 ve 17 adet olarak tespit etmişlerdir.

Gülyüz (2019), BİST'te işlem gören GYO firmalarının performanslarını etkileyen finansal oranları tespit etmeyi amaçlamıştır. Çalışmada, 2010.Q2 – 2019.Q3 dönemleri arası verileri ve panel regresyon analizi yöntemi kullanmıştır. Bağımlı değişken olarak, fiyat / kazanç ve hisse başına kâr oranlarını; bağımsız değişken olarak, Asit-test Oranı, Öz Sermaye Çarpanı, Finansal Kaldıraç Oranı, Uzun Vadeli Borçlar/Aktifler, Dönen Varlık Devir Hızı, Aktif Devir Hızı, Öz Sermaye Kârlılığı, Aktif Kârlılığı Oranı, FAVÖK Marjı ve PD/DD oranını kullanmışlardır. Çalışma sonucunda, firmaların performanslarına etki eden en önemli faktörlerin kârlılık oranları ve finansal yapı oranları olduğunu tespit etmiştir.

Sırma (2019), şirket varlıkları ve piyasa değişkeninin, GYO firmalarının pay senedi getirileri üzerindeki etkisini araştırmışlardır. Çalışmada, Borsa İstanbul'da işlem gören 30 firmanın 2008 – 2017 yılları arası verileri ve Panel regresyon analizi yöntemini kullanmışlardır. Çalışmanın bağımlı değişkeni, firmaların yıllık getirileridir. Bağımsız değişkenler, firmaların net aktifindeki ve uzun vadeli borçlarındaki değişimlerden oluşmaktadır. Çalışma sonucunda, firmaların varlıklarında meydana gelen artışların pay senedi getirileri üzerinde diğer faktörlerden daha az etkili olduğunu tespit etmişlerdir.

Günay ve Timur (2019), BİST'te işlem gören GYO firmalarının Ekonomik Katma Değer ölçütü ile performanslarını karşılaştırmışlardır. Çalışmada, 2008 – 2017 yılları arası verileri ve panel veri analizi yöntemi kullanılmıştır. Bağımlı değişken olarak, Standardize Edilmiş Ekonomik Katma Değer katsayısını kullanmışlardır. Bağımsız değişken olarak; Fiyat/Kazanç, Fiyat Nakit Akış, Hisse Başına kâr, Piyasa Değeri / Defter Değeri, Piyasa Değeri / Net Satışlar, Temettü Verimi, TOBIN Q ve Piyasa Değeri / Net Satış Büyüme oranlarını kullanmışlardır. Çalışma sonucunda, Fiyat/Kazanç Oranı, Fiyat Nakit Akış Oranı, Hisse Başına kâr ve Temettü Verimi ile bağımlı değişken arasında ilişki olduğunu tespit etmişlerdir.

Ambrose vd. (2019), Avrupa GYO firmalarının büyüklüğünün gelir, gider, kârlılık oranları ve sermaye maliyetleri üzerindeki etkilerini araştırmışlardır. Çalışmada, 2001 – 2015 yılları arası verileri ve panel veri regresyon analizi yöntemini kullanmışlardır. Çalışma sonucunda, Avrupa'daki daha büyük gayrimenkul şirketlerinin, şirket büyüklüğü başına daha yüksek gelir elde edebildiğini, daha düşük maliyetlere katlanabildiğini ve daha yüksek getiri elde edebildiğini tespit etmişlerdir.

Tarazi ve Hasan (2019), makroekonomik faktörlerin, Avustralya GYO firmaları üzerindeki etkisini araştırmışlardır. Çalışmada, 2006 – 2016 yılları arası verileri ve panel veri modeli ile vektör otoregresif (VAR) modelini kullanmışlardır. Değişken olarak; dört emlak faktörü, ortalama ev fiyatı, ortalama birim fiyat, ortalama kira getirisi, ortalama açık artırma temizleme oranı ve ortalama piyasa stokunu kullanmışlardır. Çalışma sonucunda, kira getirileri ile GYO firmalarının getirileri arasında olumlu ilişki olduğunu tespit etmişlerdir.

Dogan vd. (2019), Dünya çapında on iki ülkeden (Avustralya, Belçika, Kanada, Fransa, Hong Kong, Japonya, Hollanda, Singapur, Güney Afrika,

Türkiye, ABD, İngiltere) GYO firmalarının sermaye yapısının belirleyicilerini incelemişlerdir. Çalışmada, 2002- 2013 yılları arası verileri ve OLS regresyon modeli yöntemini kullanmışlardır. Bağımlı değişken olarak piyasa kaldıraç oranını; bağımsız değişkenler olarak varlıkların somutluğu, kârlılık, büyüme fırsatları, firma büyüklüğü ve faiz oranını kullanmışlardır. Çalışma sonucunda, GYO'ların borç finansmanını öz sermaye finansmanına tercih ettiğini ve ülkeye özgü değişkenlerin GYO kaldıraç üzerinde önemli bir etkisi olmadığını tespit etmişlerdir.

Hsieh vd. (2020), sermaye piyasası katılımcılarının yeşil bina algılarını ve öz sermaye maliyetini incelemişlerdir. Çalışmada, ABD'de faaliyet gösteren GYO firmalarının 2000 – 2016 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Değişken olarak; özsermaye maliyeti, yeşillik puanı, aktif kârlılık oranı, firma büyüklüğü, piyasa-defter oranı, kaldıraç oranını kullanmışlardır. Çalışma sonucunda, yeşil bina algısı ile özsermaye maliyeti arasında olumsuz ilişki olduğunu tespit etmişlerdir.

Özcan ve Gürol (2020), BİST'te işlem gören GYO firmalarının performanslarını ölçmeyi amaçlamışlardır. Çalışmada, firmaların 2013-2017 yılları arası verileri ve Çok Kriterli Karar Verme Tekniklerinden olan TOPSİS yöntemini kullanmışlardır. Çalışmada, çeşitli finansal oranları kullanmış ve yüksek performans için güçlü özkaynak yapısının gerekliliğini tespit etmişlerdir.

Aktaş ve Darwish (2020), Türkiye'de faaliyet gösteren GYO firmaları ile girişim sermayesi yatırım ortaklık firmalarının performanslarını karşılaştırmışlardır. Çalışmada, 2014-2019 yılları arası verileri kullanmış ve oran analizi yapmışlardır. Çalışmada yatırım, mali yapı, kârlılık ve piyasa performans oranlarını kullanmışlardır. Çalışma sonucunda, GYO firmalarının riskli yatırım, uzun vadeli borç, özsermaye kârlılık, aktif kârlılık ve hisse başına kârlarının daha fazla olduğunu tespit etmişlerdir.

Sarı ve Başakın (2020), BİST'te işlem gören GYO Endeksi getirilerini tahmin etmeyi amaçlamışlardır. Çalışmada, 2010 Ocak – 2019 Ocak dönemleri arası verileri kullanmışlardır. Çalışmanın yöntemleri; Rassal Orman Yöntemi, Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS) Yöntemi ve Sınıflandırma ve Regresyon Ağaçları (CART) Yöntemlerinden oluşmaktadır. Bağımlı değişken olarak, Gayrimenkul Yatırım Ortaklıkları Endeksi'ni; bağımsız değişken olarak, BİST Tüm ve Konut Fiyat Endekslerini

kullanmışlardır. Çalışma sonucunda yöntemlerin başarıları sıralarının; Rassal Orman Yöntemi, Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS) Yöntemi ve Sınıflandırma ve Regresyon Ağaçları (CART) Yöntemleri olduğunu tespit etmişlerdir.

Morri vd. (2021), Avrupa faaliyet gösteren 50 GYO firmasının yeşillik ve performansları arasındaki ilişkisini araştırmışlardır. Çalışmada, 2012-2016 yılları arası verileri ve en küçük kareler yöntemini kullanmışlardır. Değişken olarak, Yeşil Gayrimenkul Sürdürülebilirlik derecelendirmesi, Aktif kârlılık, özsermaye kârlılık oranlarını kullanmışlardır. Çalışma sonucunda, yeşillik göstergeleri ile işletme faaliyetleri arasında pozitif bir ilişki olduğunu tespit etmişlerdir.

Khairulanuwar ve Chuweni (2021), Malezya'da faaliyet gösteren GYO firmalarının performanslarını incelemişlerdir. Çalışmada, 2014 – 2018 yılları arası verileri kullanmışlardır. Değişken olarak, cari oran, kaldıraç oranı, aktif kârlılık oranı ve özsermaye kârlılık oranını kullanmışlardır. Çalışma sonucunda, cari orandaki düşüşler nedeni ile çeşitli GYO sektörlerinin likidite riski altında olabileceğini tespit etmişlerdir.

Karakuş ve Öksüz (2021), BİST GYO Endeksi ile faiz oranı, konut fiyat endeksi ve enflasyon oranı arasındaki ilişkiyi araştırmışlardır. Çalışmada, Ocak 2010 – Aralık 2020 dönemleri arası verileri ve ARDL Sınır Testi yöntemini kullanmışlardır. Çalışmanın bağımlı değişkeni, GYO Endeksi'dir. Bağımsız değişkenler, faiz oranı, konut fiyat endeksi ve enflasyon oranından oluşmaktadır. Çalışma sonucunda, konut fiyat endeksinde meydana gelen artışın uzun dönemde GYO endeksini arttırdığını, kira tüketici fiyat endeksi ve faiz oranlarındaki artışın uzun dönemde GYO endeksini düşürdüğünü tespit etmişlerdir.

Ching vd. (2021), Malezya ve Hong Kong GYO firmalarının kurumsal yönetim mekanizmaları ile firma performansı arasındaki ilişkiyi araştırmışlardır. Çalışmada, 2010 – 2019 yılları arası verileri ve panel veri analizi yöntemini kullanmışlardır. Bağımlı değişken olarak; aktif kârlılık oranı, özsermaye kârlılık oranı ve Tobin's Q katsayısını kullanmışlardır. Bağımsız değişkenler olarak; yönetim kurulu boyutu, bağımsız yönetici sayısı, kadın yönetici sayısı ve yönetim kurulu toplantı sayısını kullanmışlardır. Çalışma sonucunda, kurumsal yönetim mekanizmalarının, Malezya firmalarından ziyade Hong Kong firmaları ile anlamlı bir ilişkisi olduğunu tespit etmişlerdir.

Durer ve Berrak Köten (2021), Borsa İstanbul'da işlem gören GYO firmalarının performans ve kârlılıklarını etkileyen faktörleri araştırmıştır. Çalışmada, 2014.Q1 – 2021.Q1 dönemleri arası verileri ve panel regresyon analizi yöntemi kullanmıştır. Bağımlı değişken olarak Fiyat / Kazanç, Piyasa Değeri/Defter Değeri, Fiyat/Hisse Başına Kazanç, Fiyat/Hisse Başına Satış, Fiyat/Nakit Akışı oranları, İşlem hacimleri ve nakit döngüsü oranını kullanmıştır. Bağımsız değişken olarak finansal yapı oranları, kârlılık oranları ve makroekonomik değişkenleri (Kişi başına kullanılabilir gelir, enflasyon oranı, konut faiz oranı) kullanmıştır. Çalışma sonucunda, kişi başına kullanılabilir gelir, Maddi Duran Varlıklar (Net) / Aktif Toplamı Oranı ve “Finansman Giderleri ve Vergiden Önceki kâr / Kaynak (Pasif) Toplamı Oranı en fazla etkiye sahip oranlar olduğunu tespit etmiştir.

Njagi (2021), makroekonomik değişkenlerin Nairobi Menkul Kıymetler Borsası'nda işlem gören GYO'ların performansı üzerindeki etkisini araştırmıştır. Çalışmada, 2016 – 2020 yılları arası aylık verileri ve çoklu doğrusal regresyon analizi yöntemini kullanmıştır. Tek yönlü varyans analizi ile modelin anlamlılığını test etmiştir. Bağımlı değişken olarak firmaların aylık getirileri; bağımsız değişkenler olarak, döviz kuru, faiz oranı ve enflasyon oranını kullanmıştır. Çalışma sonucunda, GYO firmalarının performansları ile döviz kuru arasında zayıf bir ilişki, enflasyon oranı ile negatif yönde bir ilişki olduğunu tespit etmiştir.

Bosman vd. (2022), Covid-19 pandemisinin GYO firmaları üzerindeki etkisini araştırmışlardır. Çalışmada, Avustralya, Kanada, Çin, Fransa, Almanya, Hong Kong, Japonya, Hollanda, Yeni Zelanda, Singapur, Birleşik Krallık ve ABD, GYO firmalarının 02 Şubat 2020 - 24 Ocak 2022 dönemleri arası verileri kullanılmıştır. Nicelik üzerine nicel regresyon ve nicelik-içinde nedensellik yöntemleri kullanılmıştır. Değişken olarak; vaka sayısı, ölüm sayısı, korku endeksleri, medyada yer alma vb. değişkenleri kullanmışlardır. Çalışma sonucunda, COVID-19 vakalarının alt (üst) koşullu niceliklerde GYO getirileri üzerinde güçlü (zayıf) bir tahmin gücü olduğunu tespit etmişlerdir.

Öndeş ve Barakalı (2023), faiz oranları ile GYO firmalarının kârlılık oranları arasındaki ilişkiyi araştırmışlardır. Çalışmada, 2011.Q4 – 2021.Q3 dönemleri arası veriler ve panel veri analizi yöntemini kullanmışlardır. Çalışma sonucunda, firmaların aktif kârlılığı üzerinde konut ve ticari faiz oranlarının etkisi olduğunu tespit etmişlerdir.

Rahmawati vd. (2023), GYO'ların mülkiyet geliştirme projelerindeki finansman üzerindeki etkisini incelemişlerdir. Çalışmada, iki farklı araştırma yöntemi kullanılmışlardır. İlk yöntem olarak meta-analiz yapılarak daha GYO kullanımının sonuçları belirlenmiştir. İkinci yöntem olarak ise anket çalışması ile mülkiyet geliştirici meta-analiz sonuçları doğrulanmıştır. Veriler, çoklu doğrusal regresyon analizi kullanılarak analiz edilmiştir. Çalışma sonucunda, sermaye piyasasına giriş veya gayrimenkul edinme ihtiyacı ile GYO firmalarına yatırım yapıldığını tespit etmişlerdir.

2.2. Veri Madenciliği Yöntemlerini Konu Alan Çalışmalar

Bu bölümde çalışmanın uygulama bölümünde kullanılacak olan yöntemleri (Rassal Orman, XGBoost, CatBoost) konu alan çalışmalar incelenmektedir. Bu çerçevede, bu yöntemleri konu alan çalışmaların temel amaçları, çalışmalarda kullanılan yöntemler ve çalışmalardan elde edilen sonuçlar yer almaktadır.

Fang vd. (2016), Sigortacılık sektöründe yeni bir müşterinin kârlılığa olan etkisini belirlemek için doğrusal regresyon, karar ağaçları, rassal orman regresyonu, destek vektör makinaları ve güçlendirilmiş model gibi çeşitli yöntemler kullanılmıştır. Çalışma sonucunda, rassal orman regresyon yönteminin diğer yöntemlerden daha başarılı bir performans gösterdiğini tespit etmişlerdir.

Gumus ve Kiran (2017), ham petrol fiyatlarını etkileyen faktörleri XGBoost yöntemi ile tahmin etmeyi amaçlamışlardır. Çalışma sonucunda, XGBoost yönteminin başarılı sonuçlar elde ettiğini tespit etmişlerdir.

Wang vd. (2018), binaların saatlik enerji tüketimlerini tahmin etmeyi amaçlamışlardır. Çalışmada, rassal orman ve destek vektör makinası yöntemleri karşılaştırılmıştır. Çalışma sonucunda, rassal orman regresyon yönteminin daha başarılı olduğu tespit edilmiştir. Ayrıca, rassal orman regresyon yönteminin değişken sayısına çok duyarlı olmadığını belirtmişlerdir.

Jhaveri vd. (2019), Kickstarter kampanyalarının etkinliğini öngörmek ve kampanya sahiplerine geçmiş verilerden türetilen trendler ve analizler sunarak etkili bir kampanya oluşturmalarına yardımcı olmayı hedeflemişlerdir. Çalışmada, Rassal Orman, XGBoost, CatBoost ve AdaBoost yöntemlerini kullanmışlardır. Çalışma sonucunda, Rassal Orman ve AdaBoost yöntemlerinin daha başarılı sonuçlar elde ettiğini tespit etmişlerdir.

Tiwari vd. (2021), erken aşamalarda Parkinson hastalığını tespit etmek için Efee adlı bir ses veri kümesi kullanmışlardır. Çalışmada, rassal orman, CatBoost ve XGBoost gibi çeşitli yöntemleri karşılaştırmışlardır. Yapılan karşılaştırma sonuçlarına göre, bu modeller arasında CatBoost'un daha yüksek doğruluk oranına sahip olduğunu tespit etmişlerdir.

Bentéjac vd. (2021), XGBoost, LightGBM ve CatBoost yöntemlerinin performanslarını karşılaştırmışlardır. Çalışma sonucunda, CatBoost'un genel olarak genelleme doğruluğu açısından en iyi sonuçları elde ettiğini, LightGBM'in ise yöntemler arasında en hızlı olduğunu tespit etmişlerdir. XGBoost ise hem doğruluk hem de hız açısından ikinci sırada yer almıştır.

Luo vd. (2021), orman ekolojisi alanında yer üstü biyokütle tahmininde kullanılan bir tahmin aracının geliştirmeyi amaçlamışlardır. Çalışmada, CatBoost XGBoost, LightGBM ve Rassal Orman yöntemlerini kullanışlardır. Çalışma sonucunda, CatBoost yöntemi daha yüksek bir doğruluk oranı sağlamıştır.

Almaskati (2022), banka riski ve kârlılığını etkileyen faktörleri araştırmıştır. Çalışmada, Rassal Orman yöntemini kullanmış ve risk ve kârlılığı etkileyen faktörlerin önem derecesini tespit etmiştir. Çalışma sonucunda, banka kârlılığının büyük ölçüde bankaya özel faktörler tarafından belirlendiğini, banka riskinin daha çok ülke düzeyindeki faktörlerden etkilendiğini tespit etmiştir.

Demir ve Sahin (2022), konik penetrasyon testi verilerine dayanan zeminin sıvılaşma potansiyelini tahmin etmek için üç ağaç tabanlı yöntem kullanmışlardır. Bu yöntemler, Kanonik Korelasyon Ormanı, Rotasyon Ormanı ve Rassal Orman yöntemlerinden oluşmaktadır. Çalışma sonucunda, Kanonik Korelasyon Ormanı ve Rotasyon Ormanı yöntemlerinin Rassal Orman yöntemine göre daha başarılı sonuçlar elde ettiğini tespit etmişlerdir.

Rezapour ve Hansen, (2022), COVID-19 salgını sırasında sağlık çalışanlarının zihinsel sağlığını tahmin etmek için istatistiksel ve makine öğrenimi modellerini kullanmışlardır. Çalışmada, Karar Ağaçları, Lojistik Regresyon, Naive Bayes, k-En Yakın Komşu, Destek Vektör Makineleri, Yapay Sinir Ağları, Rassal Orman, XGBoost, CatBoost, LightGBM gibi çeşitli yöntemler kullanmışlardır. Çalışma sonucunda, XGBoost, CatBoost ve LightGBM yöntemlerinin daha iyi performans gösterdiğini tespit etmişlerdir.

Lee vd. (2022), yüksek performanslı beton dayanıklılığını tahmin etmek için veri madenciliği yöntemlerini kullanmışlardır. Bu yöntemler, AdaBoost, GBM, Rassel Orman, XGBoost, LightGBM ve CatBoost'dan oluşmaktadır. Çalışma sonucunda, tüm yöntemlerin iyi performans gösterdiğini tespit etmişlerdir.

Szczepanek (2022), dağlık havzalarda günlük akarsu akımını tahmin etmek için veri madenciliği yöntemlerini kullanmıştır. Bu yöntemler, XGBoost, LightGBM ve CatBoost'dan oluşmaktadır. Çalışma sonucunda, hiperparametre optimizasyonu sonrasında LightGBM yönteminin diğer yöntemlere göre daha başarılı sonuçlar elde ettiğini tespit etmiştir.

Jaiswal ve Gupta (2022), diyabet hastalığının tespiti ve öngörülmesi için veri madenciliği yöntemlerini kullanmışlardır. Çalışmada, CATBoost, LightGBM ve XGBoost yöntemlerini kullanmışlardır. Çalışma sonucunda, LightGBM yönteminin diğer yöntemlere göre daha iyi sonuçlar elde ettiğini tespit etmişlerdir.

Shobana Bai (2023), tek silindirli bir motordan kaynaklanan emisyonları azaltmak için çeşitli yöntem kombinasyonlarını araştırmıştır. Çalışmada, XGBoost, LightGBM, CatBoost ve Rassel Orman yöntemlerini kullanmıştır. Çalışma sonucunda, en iyi sonuçlar CatBoost yöntemi ile edilmiştir.

Sadaf, (2023), zararlı web sitelerini tespit etmeyi amaçlamıştır. Çalışmada, XGBoost ve CatBoost yöntemlerinin performansları karşılaştırılmıştır. Karşılaştırma sonuçlarına göre, XGBoost yönteminin CatBoost yöntemine göre biraz daha üstün performans sergilediği tespit edilmiştir.

Latha ve Bommi (2023), ağ trafiği üzerinden gerçekleşen sızma saldırılarının varlığını tespit etmek amacıyla CatBoost regresyon modelini kullanmışlardır. Çalışmanın sonuçlarına göre, önerilen sistem %92,5 doğruluk oranı elde etmiştir.

Chen (2023) zincir işletmelerin pazarlamasını ve müşteri tüketimini etkileyen faktörlerin analizine dayanarak, pazar ortamı, pazarlama faaliyetleri ve mali durumdan oluşan bir tahmin indeksi sistemi oluşturulmaktaydı amaçlamışlardır. Çalışmada, Destek vektör makineleri, lojistik regresyon, Rassel orman ve XGBoost yöntemlerini kullanmıştır. Çalışma sonucunda, XGBoost yönteminin diğer yöntemlere göre daha başarılı olduğunu tespit etmiştir.

Muhasshanah vd. (2023), veri madenciliği yöntemleri ile hamile kadınların doğum sürecini tahmin etmeyi amaçlamışlardır. Çalışmada, C4.5 ve rassal orman yöntemlerini kullanmışlardır. Çalışma sonucunda, her iki yönteminde yüksek başarı oranına sahip olduğunu tespit etmişlerdir.

Abdi vd. (2023), kaya elastik modülünün değerlendirilmesi için basit kaya indeks testi sonuçlarından yararlanarak yapay zekâ tabanlı öngörü tekniklerinin geliştirilmesini amaçlamışlardır. Bu amaç doğrultusunda Rassal Orman, AdaBoost, XGBoost ve CatBoost gibi ağaç temelli yöntemler kullanılmıştır. Yapılan çalışma sonucunda, Rassal Orman modeli, laboratuvar ortamında zayıf kaya örneklerinin ölçülen elastik modüllerine oldukça yakın tahmin sonuçları elde etmiştir.

Emami vd. (2023), çalışmalarında, COVID-19 vakalarının teşhisinde hastaların semptomlarını kullanarak yapay zekâ tabanlı bir yaklaşım geliştirmeyi amaçlamışlardır. Araştırmada, Destek Vektör Makinesi, Lojistik Regresyon, K En Yakın Komşu, Basit Bayes, Rassal Orman, LightGBM, XGBoost ve CatBoost gibi farklı yöntemler kullanılmıştır. Elde edilen bulgulara göre, CatBoost yönteminin en iyi sonuçları verdiği belirlenmiştir.

Kanaparthi (2023), finansal kurumlarda kredi riskini tahmin etmek için makine öğrenimi modellerini kullanarak yeni bir yaklaşım önermektedir. Çalışmada, XGBoost, ve CatBoost yöntemlerini kullanmıştır. Çalışma sonucunda, CatBoost'un XGBoost'a kıyasla biraz daha başarılı olduğunu tespit etmiştir.

Sari vd. (2023), akciğer kanserini erken tespit etmeyi amaçlamışlardır. Çalışmada, Naïve Bayes ve Rassal orman yöntemlerini kullanmışlardır. Çalışma sonucunda Rassal orman yönteminin daha başarılı sonuç elde ettiğini tespit etmişlerdir.

BÖLÜM 3

KAVRAMSAL ÇERÇEVE

3. KAVRAMSAL ÇERÇEVE

3.1. Gayrimenkul Kavramı ve Önemi

Gayrimenkul, bir arazi ve arazi üzerine inşa edilen her şey olarak tanımlanabilmektedir. Bu tanım dahilinde gayrimenkuller, oturmaya elverişli olan ve olmayan olmak üzere ikiye ayrılabilir (Seçkin, 1998). Oturmaya elverişli olan yapılara konut örnek olarak verilebilir. Oturmaya elverişli olmayan yapılara ise alışveriş merkezleri, ofisler örnek olarak verilebilir. Bunlar içerisinde en fazla önem verilmesi gereken gayrimenkul türü konut olarak görülmektedir. Konut, barınma ihtiyacını gidermekten yatırım aracı olmaya kadar geniş bir anlam ifade etmektedir (Heşşen, 2009). Ayrıca gayrimenkul sektörünü, çeşitli şekilde sınıflandırılabilir. Bu sınıflandırmalar, işlevine göre konut, ticari, endüstriyel, tarımsal; mekâna göre yerel, bölgesel, ulusal, uluslararası; mülkiyet hakkına göre kira, emanet; kamu müdahalesine göre kamu, özel şeklinde yapılabilir (Kazak vd., 2017).

Gayrimenkul yatırımı, gayrimenkul ticareti, geliştirme çalışmaları ve istikrarlı bir getiri sağlamak amacıyla uzun vadeli bir perspektifle gayrimenkule yatırım yapmayı içeren çeşitli faaliyetleri içermektedir. Gayrimenkul yatırımı, varlıkları doğrudan satın almak ve dolaylı olarak satın almak şeklinde iki farklı şekilde gerçekleştirmek mümkündür. Doğrudan satın alma, varlığı doğrudan satın almayı ifade ederken dolaylı olarak satın alma borsada işlem gören veya görmeyen bir kuruluş aracılığıyla satın almayı ifade etmektedir (Georgiev, 2002). Gayrimenkul, genellikle konut, ofis, perakende ve endüstriyel mülklerle ilişkilendirilse de konaklama, eğitim, sağlık ve eğlence altyapısı gibi çok daha geniş bir varlık sınıfını kapsayan bir kavramdır (Mert, 2012). Bu varlıkları konut ve ticari amaçlı gayrimenkul olmak üzere ikiye ayırabiliriz. Konut amaçlı gayrimenkul, konutlar için kullanılan mülkleri ifade ederken, ticari gayrimenkul ise ticari amaçlarla kullanılan mülkleri temsil etmektedir (PWC, 2022).

Gelişmekte olan ekonomiler için büyük önem arz eden gayrimenkul sektörü, çimento, mobilya, dekorasyon malzemeleri, ev tekstili gibi çeşitli

sektörlerin de gelişmesine katkı sağlamaktadır. Tüm bunlar istihdam ve milli gelir üzerinde önemli etkiler yaratmaktadır (Demirdöven, 2009).

Gayrimenkul sektörü ekonomik açıdan incelendiğinde, gayrimenkul geliştirme ve gayrimenkul yatırımı terimleri önem arz etmektedir. Gayrimenkul geliştirme yeni binalar geliştirme, mevcut binaları onarma, yenileme, geliştirmeyi ifade etmektedir. Gayrimenkul yatırımı ise bir firmanın gelir elde etmek amacı veya sermayesini artırmak amacı ile gayrimenkul edinmesini ifade etmektedir (PWC, 2022).

Gayrimenkule yatırım yapmanın bir dizi avantajı bulunmaktadır. Bunlar pasif gelir elde etme, istikrarlı nakit akışı sağlama, vergi avantajlarından yararlanma, portföyü çeşitlendirme imkânı sunma ve kaldıraç kullanma yeteneği gibi unsurlardır. Bu avantajlar aşağıdaki şekilde açıklanabilir (Palmer, 2022):

- Nakit akışı, gayrimenkul yatırımdan ipotek ödemeleri ve işletme giderleri düşüldükten sonra elde edilen net gelirdir. Gayrimenkul yatırımının en önemli avantajlarından biri nakit akışı yaratabilmesidir .

- Gayrimenkul yatırımcıları, çeşitli vergi muafiyetlerinden ve indirimlerden yararlanma fırsatına sahiptir. Genel olarak, bir mülke sahip olmanın, işletmenin ve yönetmenin maliyetlerini makul düzeyde azaltma potansiyeline sahiptirler.

- Gayrimenkul yatırımcıları, kira geliri, mülke bağlı ticari faaliyetlerden elde edilen kârlar ve değer artışı gibi kaynaklar aracılığıyla finansal getiriler elde etmektedir.

- Bir mülkün ipoteğini ödedikçe, net değer bir bileşeni olan öz sermaye oluşturulmaktadır. Özkaynak birikimi arttıkça, kaldıraç etkisinden faydalanarak daha fazla mülk edinme ve nakit akışını, serveti daha da artırma imkânı elde edilir.

- Gayrimenkul, diğer varlık sınıflarıyla genellikle düşük korelasyona sahiptir. Bu durum, çeşitlendirilmiş bir portföyün gayrimenkulü dahil etmesinin, portföyün oynaklığını azaltabileceği ve birim risk başına daha yüksek getiri potansiyeli sunabileceği anlamına gelmektedir.

- Kaldıraç, potansiyel getiriyi artırmak amacıyla finansal araçların veya alınan borç sermayesinin kullanılmasıdır. Gayrimenkul, somut bir varlık olması ve teminat olarak kullanılabilmesi nedeniyle finansman imkanlarını kolaylıkla

sağlayabilir. Bu sayede yatırımcılar, sınırlı öz sermayelerini kullanarak daha büyük bir varlık portföyü oluşturabilir ve potansiyel getiriye artırabilirler.

• Gayrimenkulün enflasyondan korunma yeteneği, gayrimenkule olan talebin GSYİH büyümesi ile pozitif bir ilişkisi olduğundan kaynaklanmaktadır. Ekonomik büyüme, gayrimenkul talebinin artmasına ve buna bağlı olarak kiraların yükselmesine neden olur. Bu durum, gayrimenkulün sermaye değerlerinin artmasına yol açar. Dolayısıyla, gayrimenkul enflasyonist etkileri, kiracılara kısmen aktararak ve sermaye değerlerinde artış şeklinde yansıtarak sermayenin satın alma gücünü koruma eğilimindedir. Bu özellik, gayrimenkulün yatırımcılara enflasyon riskine karşı bir koruma sağlama yeteneğini vurgulamaktadır.

3.2 Konut Finansmanı ve GYO

Yerleşim yeri, bölge veya ülkede belirli bir süre sonunda hane halkının gereksinim duyduğu konut sayısı, konut ihtiyacı olarak tanımlanabilmektedir. Türkiye nüfusunun 2025 yılında 85,5 milyon, 2030 yılında 88 milyon ve 2035 yılında 90 milyon düzeylerine ulaşacağı tahmin edilmektedir (TÜİK, 2022). Bu nüfus artışıyla birlikte, konut ihtiyacının da artması muhtemeldir. Ayrıca, hane halkı sayısındaki azalışlar da konut ihtiyacının bir göstergesi olarak kabul edilebilmektedir. Türkiye'de hane halkı sayısı yıllar itibariyle sürekli bir düşüş eğilimi göstermektedir. 2016 yılında Türkiye'nin hane halkı sayısı 3,48 iken, 2021 yılında bu rakam 3,23'e düşmüştür (TÜİK, 2022). Türkiye'de nüfus artışıyla birlikte hane halkı sayısında düşüşün olması, konut ihtiyacının önemli bir göstergesi olarak düşünülmektedir. Ayrıca, 6 Şubat depremi gibi doğal afetlerin konut ihtiyacını artırması da göz önünde bulundurulmalıdır. Depremler gibi olaylar, mevcut konut stoğunun zarar görmesine veya yıkılmasına neden olabilmektedir. Bu durum, konut talebinin artmasına yol açmaktadır. Depremlerin sonucunda evsiz kalan insanlar için acil konut ihtiyacı ortaya çıkmakta ve genel konut talebi artırmaktadır. Dolayısıyla, 6 Şubat depremi gibi afetler de konut ihtiyacını artıran bir diğer etken olarak gösterilebilmektedir.

Konut finansmanı, proje geliştiricilere yönelik sağlanan fonlar ve nihai kullanıcılara yönelik sağlanan fonlar olarak iki şekilde sınıflandırılabilir. Proje geliştirenlere sunulan fonlar içerisinde, Gayrimenkul Yatırım Ortaklıkları (GYO) ve yapımcılara yönelik krediler yer

almaktadır. Nihai kullanıcılara yönelik fonlar içerisinde ise doğrudan finansman yöntemi, sözleşme karşılığı birikim yöntemi, mevduat finansman yöntemi, ipotekli konut finansman yöntemi yer almaktadır (Bahadır ve Haznedaroğlu, 2011).

Gayrimenkul yatırımında finansal araç kullanmadaki temel amaç; ilgili gayrimenkul yatırımının sürekli olarak alınıp satılabilmesini sağlamaktır (Özyeşil, 2022). Gayrimenkul Yatırım Ortaklıkları, büyük ölçekli gayrimenkul, ipotek veya gayrimenkul ile alakalı menkul kıymetlere yatırım yapan şirket veya iş ortaklığı olarak tanımlanmaktadır (Han, 2004). GYO'lar ticari olarak ofis, konut, depo, perakende satış merkezi, tıbbi tesis, veri merkezleri, baz istasyonları, altyapı ve oteller dahil olmak üzere çok çeşitli gayrimenkullere yatırım yapmaktadırlar (REİT, 2022). Tebliğde de ifade edildiği üzere, GYO'lar, gayrimenkuller ve bunlara bağlı sermaye piyasası araçlarını portföyü işletmek mantığı ile kurulan sermaye piyasası kurumudur (Tebliğ, Seri III-48.1, Md. 4 (Değişik: RG-23/1/2014-28891)).

GYO'lar, gayrimenkul yatırımı için düşük maliyetli, etkili ve likit araçlardır. GYO yatırımlarının, hisse senetleri ve tahvillere göre belirli bir ekonomik döngüsünün olması, enflasyona karşı koruyucu olması (Simpson vd., 2007) ve güvenilir getirisinin olması gibi birtakım avantajları bulunmaktadır. Enflasyon karşısında korunmak isteyen yatırımcılar için GYO'lar, getiri bakımından iyi performans göstermektedir. GYO'lar, ABD'de yüksek ve orta enflasyon dönemlerinde, S&P 500'dan daha yüksek; düşük enflasyon dönemlerinde biraz daha düşük performans göstermişlerdir (John, 2022).

Yasal mevzuatlara göre inşaat işleriyle uğraşmayan GYO firmaları, inşaat firmalarına iş sağlayabilmektedirler (Hepsen vd., 2017). GYO firmalarının ülke ekonomisine en önemli katkıları; işsizliğin azalması ve üretimin artması olarak gösterilmektedir (Aktaş ve Darwish, 2020). Bunların yanı sıra bir ülke ekonomisine bir takım başka yararları da bulunmaktadır (Güven, 2006):

1. Gayrimenkul piyasalarının gelişmesine imkân sağlamaktadır.
2. Gayrimenkul, menkulleştirme ile likit bir varlığa dönüşmektedir. Gayrimenkullerin likit hale gelmesi ile sermaye piyasası ve ülke ekonomisine katkı sağlanmaktadır.
3. GYO'lar büyük projeler için finansal kaynak sağlamaktadırlar. Firmalar için temel problemlerden biri finansal kaynağa ulaşımıdır. GYO

firmaları aracılığı ile firmaların finansal kaynağa ulaşımı kolaylaşmakta ve bu durum projelerin hayata geçirilmesine imkân sağlamaktadır.

4. Yabancı piyasalardan ülkeye sermaye girişi sağlanır. Uzun vadeli yatırım yapmak isteyen yabancı yatırımcılar, GYO firmaları aracılığı ile yatırımlarını yapma imkanına sahip olurlar.

5. GYO firmalarının faaliyet göstermesi ile sektörün kayıt altına alınması imkânı oluşur. Firmaların yayınladıkları raporlarda, projenin maliyeti, kârlılığı gibi birçok konuda bilgi yer almaktadır. Bu durum, kayıt dışı ekonominin oluşmasına engel olmaktadır.

3.3. GYO'ların Tarihsel Gelişimi

Gayrimenkul Yatırım Ortaklıkları ilk olarak ABD'de ortaya çıkmıştır. ABD Gayrimenkul Yatırım Ortaklıkları diğer ülkeler için bir temel oluşturmaktadır. Diğer ülke ve bölgelerde kurulan GYO'lar genel olarak ABD GYO'ya benzer yapılarda kurulmuşlardır. ABD'de sektörün gelişimi açısından önemli gelişmeler meydana gelmiştir. Bu gelişmeler kronolojik sırasına göre aşağıda maddeler halinde özetlenmiştir (NAREİT, 2022):

- 1960 yılında Başkan Eisenhower tarafından ilk defa Gayrimenkul Yatırım Ortaklıkları yasası imzalanmıştır. Yasanın oluşturulma amacı, yatırımcılara gelir getiren gayrimenkullerden oluşan büyük ve çeşitlendirilmiş portföylere yatırım yapabilme imkânı sağlamaktır.

- Ulusal Gayrimenkul Yatırım Fonları Birliği, günümüzde Ulusal Gayrimenkul Yatırım Ortaklıkları Birliği (NAREIT) olarak bilinen kuruluş, 15 Eylül 1960 tarihinde kurulmuştur.

- Ocak 1972'de, gayrimenkul yatırım ortaklıklarının (GYO) fiyat ve toplam getiri yatırım performansını karşılaştırmak amacıyla ilk GYO endeksi piyasaya sunulmuştur.

- Kasım 1976 yılında Başkan Ford tarafından imzalanan Vergi Reformu'nun bir parçası olarak GYO'ların şirket olarak kurulmalarına ve faaliyet göstermelerine imkân tanınmıştır.

- Eylül 1991'de Kimco Realty Corporation, gayrimenkul yatırım ortaklığı (GYO) alanında ilk başarılı hisse senedi halka arzını gerçekleştirmiştir. Modern GYO döneminin başlangıcı olarak gösterilen bu olay, GYO sektöründe canlanmanın temelini oluşturmaktadır.

- Ekim 1991'de Faaliyetlerden Kaynaklanan Fon (FFO) kavramı, Ulusal Gayrimenkul Yatırım Ortaklıkları Birliği (NAREIT) tarafından duyurulmuştur.
- Aralık 1991 New Plan, 1 milyar dolarlık öz sermaye piyasası değerine ulaşan ilk halka açık GYO oldu. 2007 yılında New Plan, Avustralya merkezli Centro Properties Group tarafından satın alındı.
- 1993 yılında Başkan Clinton, Omnibus Bütçe Uzlaşma Yasası'nın bir parçası olarak emeklilik planlarının GYO'lara yatırım yapmasını kolaylaştıran "Beş veya Daha Az" kuralındaki bir değişikliği yasalaştırdı.
- Ekim 1999 Halka açık gayrimenkule artan ilginin bir işareti olarak Avrupa Halka Açık Gayrimenkul Birliği (EPRA) kuruldu. NAREIT ve EPRA, küresel çapta halka açık gayrimenkul şirketlerinin ve üyelerinin büyümesini teşvik etmek amacıyla sürekli bir işbirliği içerisinde bulunmaktadır.
- Ocak 2009'da, FTSE, NAREIT ve EPRA, gelişmekte olan piyasalardaki gayrimenkul yatırım ortaklıklarının ve borsaya kote edilmiş emlak şirketlerinin FTSE EPRA/NAREIT Küresel Emlak Endeksi'ne dahil edildiğini duyurmuştur. Bu gelişme, küresel emlak piyasasının genişlemesi ve bu piyasada faaliyet gösteren kuruluşların uluslararası ölçekte daha fazla tanınmasını sağlamaktadır.

3.4. Dünya'da GYO Sektörünün Mevcut Görünümü

Gayrimenkul Yatırım Ortaklığı (GYO) sektörü, dünya çapında hızla gelişen bir yapıya sahip olup, farklı coğrafyalarda önemli gelişmeler göstermektedir. Bu bölümde, Amerika Birleşik Devletleri, Avrupa, Asya ve Türkiye'deki GYO sektörüne ait güncel bilgiler alt başlıklar halinde yer almaktadır.

3.4.1. ABD'de GYO Sektörünün Mevcut Görünümü

Günümüzde ABD GYO firmaları yaklaşık 4.5 Trilyon Dolar değerinde brüt gayrimenkule sahiptirler. 2020 yılında GYO firmaları 88.8 milyar dolarlık temettü dağıtmışlardır ve 30 GYO firması, S&P500 üyesidir(NAREIT, 2022). ABD'de 183 GYO firması faaliyet göstermekte olup piyasa değerleri toplamı 1.293.610 Milyon Eurodur (EPRA, 2022). ABD'de 2021 yılı sonu itibariyle, Özsermaye GYO'ları, ABD'de halka açık tüm GYO brüt varlıklarının yaklaşık %82'sini oluşturmakta ve yaklaşık 535.000'den fazla mülke sahiptirler. Bu mülklerin dağılımı şu şekildedir (EY, 2021):

- 2.400'den fazla ofis binası
- 500'den fazla bölgesel alışveriş merkezi; yaklaşık 3.000 alışveriş merkezi; 1.700'den fazla restoran lokasyonu ve 26.000'den fazla diğer perakende mülk
- Yaklaşık 8.000 endüstriyel tesis
- Yaklaşık 2.800'den fazla çok aileli kiralık mülk
- Yaklaşık 1.800 otel
- Yaklaşık 8.500 tıbbi tesis
- Yaklaşık 7.200 kendi kendine depolama tesisi
- 300'den fazla veri merkezi
- Yaklaşık 100.000 telekomünikasyon kulesi
- Yaklaşık 157.000 müstakil ev kiralık mülk
- 212.000'den fazla reklam panosu ve açık hava reklamı.

ABD'de piyasa değerleri büyüklüğüne göre öne çıkan GYO firmaları Tablo 1'de gösterilmektedir (EPRA, 2022).

Tablo 1: ABD'de piyasa değerleri büyüklüğüne göre GYO firmaları

Firma Adı	Piyasa Değeri (Euro)
Prologis	83,306.15
Digital Realty Trust	57,203.26
Public Storage	52,496.66
Realty Income	39,279.83
Welltower Inc.	35,758.94

GYO'lar istihdama önemli katkılar sağlamaktadırlar. ABD Özsermaye GYO'ları 2021 yılında toplamda 278.700 kişi istihdam etmiştir. İpotek GYO firmalarında ise 27.900 kişi istihdam edilmiştir. Tablo 2'de ABD GYO firmalarının iş gücü istihdam rakamları ve oranları mülk türüne göre özetlenmektedir (EY, 2021).

Tablo 2: ABD GYO işçi sayıları

Mülk Türü	İşçi Sayısı	%
Veri Merkezi	63.100	21
Konut	49.400	16
Endüstriyel	33.200	11
Depolama	27.000	9
Alt Yapı ve Telekomünikasyon Kuleleri	22.300	9
Ormanlık Alan	17.500	6
Ofis	16.500	5
Perakende	16.400	5
Çeşitlendirilmiş	16.200	5
Uzmanlık	9.200	3
Sağlık Hizmetleri	4.600	2
Konaklama	3.300	1
ÖZSERMAYE GYO TOPLAM	278.700	91
İPOTEK GYO TOPLAM	27.900	9
GENEL TOPLAM	306.500	100

3.4.2. Avrupa'da GYO Sektörünün Mevcut Görünümü

Dünya genelinde olduğu gibi Avrupa'da da GYO'lar önem arz etmektedir. Küresel olarak, GYO'ların toplam değerinin 3 trilyon Euro'ya yakın olduğu tahmin edilmektedir. Avrupa genelinde istihdam yaratarak GSYİH ve toplumu olumlu yönde etkilemektedirler (PWC, 2022). Avrupa genelinde 199 GYO firması bulunmakta olup piyasa değerleri toplamı 238,80 milyar dolardır. 2021 yılı sonu itibariyle, 14 ülke gayrimenkul yatırımlarını teşvik etmek için kamu yararlarını tanıyan GYO mevzuatını yürürlüğe koymuştur. (PWC, 2022). Avrupa'da piyasa değerleri büyüklüğüne göre öne çıkan GYO firmaları Tablo 3'te gösterilmektedir (EPRA, 2022).

Tablo 3: Avrupa'da piyasa değerleri büyüklüğüne GYO firmaları

Ülke	Firma Adı	Piyasa Değeri (Euro)
B. Krallık	Segro	13,644.29
Hollanda	Unibail Rodamco Westfield	6,750.46
Fransa	Gecina	6,560.55
B. Krallık	Land Securities Group	5,719.40
Belçika	Warehouses De Pauw	5,577.94

3.4.3. Asya'da GYO Sektörünün Mevcut Görünümü

2021 yılı sonu itibari ile Asya'da toplam 198 GYO firması bulunmaktadır. Bu firmaların piyasa değeri 304,1 milyar ABD dolardır. Japonya, Singapur ve Hong Kong Özel İdari Bölgesi, Asya'nın en büyük üç GYO pazarını oluşturarak bölgenin gayrimenkul yatırım ortaklıkları sektörünün yaklaşık %90'lık bir kısmını temsil etmektedir. 198 gayrimenkul yatırım ortaklığının 89'u karma kullanımlıdır, yani birden fazla mülk türünü içermektedir. Bunlardan, 30'u ofis, 17'si otel sektörüne, 25'i endüstriyel/lojistik mülklere, 7'si sağlık sektörüne, 24'ü perakende sektörüne ve 4'ü çoklu aile konutlarına odaklanmıştır. İki GYO veri merkezi pazarına odaklanmıştır. En büyük Asya ve ikinci en büyük küresel GYO pazarı olan Japonya'da 2021 yılı sonunda toplam piyasa değeri yaklaşık 147,8 milyar ABD doları olan 61 GYO bulunmaktadır (G. Dong vd., 2022). Asya'da piyasa değerleri büyüklüğüne göre öne çıkan GYO firmaları Tablo 4'te gösterilmektedir (EPRA, 2022).

Tablo 4: Asya'da piyasa değerleri büyüklüğüne GYO firmaları

Ülke	Firma Adı	Piyasa Değeri (Euro)
Hong Kong	Link Real Estate Investment Trust	16,462.59
Singapur	CapitaLand Integrated Commercial Trust	9,881.32
Japonya	Nippon Building Fund Inc	8,095.99
Japonya	Nippon Prologis REIT	6,279.65

3.4.4. Türkiye'de GYO Sektörünün Mevcut Görünümü

Türkiye'de 2022'nin 3.çeyreğinde toplam GYO sayısı 38 ve halka açıklık oranı %45,84'tir. 2022'nin üçüncü çeyreğinde toplam piyasa değeri 135,9 milyar TL, dolar bazında 7,73 milyar dolardır (GYODER, 2022). Türkiye'de piyasa değerleri büyüklüğüne göre öne çıkan GYO firmaları Tablo 5'te gösterilmektedir (EPRA, 2022).

Tablo 5: Türkiye'de piyasa değerleri büyüklüğüne GYO firmaları

Firma Adı	Piyasa Değeri (Euro)
İş Gayrimenkul Yatırım Ortaklığı	781.59
Emlak Konut Gayrimenkul Yatırım Ortaklığı AŞ	553.75
Özak Gayrimenkul Yatırım Ortaklığı AŞ	229.11

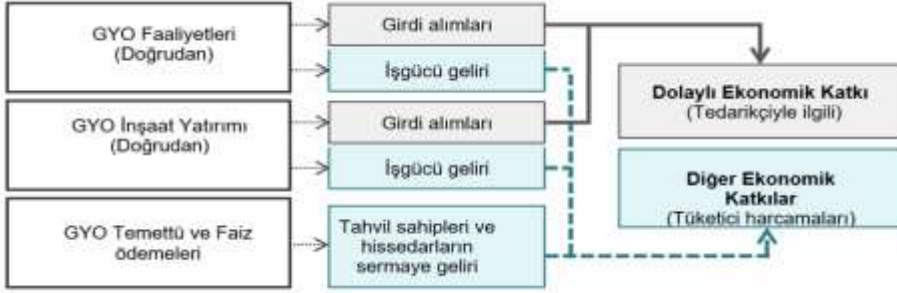
3.5. GYO'ların Ekonomik Katkıları

GYO'ların ekonomik etkileri, doğrudan, dolaylı ve diğer ekonomik etkiler olmak üzere üç kategoride sınıflandırılabilir. Ernst & Young'un (EY) 2021 yılında yayınladığı raporda, bu etkiler şu şekilde açıklanmaktadır:

Doğrudan Ekonomik Katkı: GYO firmalarının faaliyetleri ile birlikte oluşan istihdam ve GYO çalışanları tarafından kazanılan işgücü gelirlerinden oluşmaktadır. Doğrudan katkı ayrıca mevcut yapılar için düzenli sermaye harcamalarını ve yeni yapılara yapılan yatırımları ve ilgili geçici inşaat istihdamını da içermektedir.

Dolaylı Ekonomik Katkı: GYO'lar diğer işletmelerinden mal ve hizmet satın alarak bu tedarikçi işletmelerdeki istihdamın ve gelirin artmasına katkı sağlarlar. GYO tedarikçilerinin kendi tedarikçilerinden işletme girdileri satın almaları nedeniyle ek ekonomik döngü oluşmaktadır.

Diğer Ekonomik Katkılar: GYO hissedarları tarafından temettü yoluyla finanse edilen harcamalar ve GYO tahvil sahipleri ile diğer alacaklıların faiz geliri harcamaları gibi unsurlar, tüketici tarafından yeniden harcanan gelirin GYO'lar ve ilgili tedarikçi işletmeler tarafından üretilen ekonomik faaliyeti yansıtmaktadır. GYO çalışanları, tahvil sahipleri, hissedarları ve GYO'lar tarafından dolaylı olarak desteklenen işletmelerin çalışanları, işletmelerde (örneğin marketler, perakendeciler, sinema salonları gibi) kazançlarını harcadıklarında, ek ekonomik faaliyeti desteklemektedir. Bu durumda, harcamaların artmasıyla birlikte işletmeler arasındaki etkileşimler ve ekonomik döngü daha da güçlenmektedir. Örneğin, bir restoranda yapılan harcamalar, sadece restoranın kendisine değil, aynı zamanda çiftlikler, nakliye şirketleri ve restoranın tedarik zinciri içinde yer alan diğer işletmelere de katkı sağlamaktadır. Şekil 1'de GYO'ların dolaylı ve diğer ekonomik katkıları gösterilmektedir.



Şekil 1: GYO'ların ekonomik katkıları (EY, 2021)

Birçok ülkede, daha sonraki bölümlerde bahsedeceğimiz üzere GYO'ların vergi muafiyetinden yararlanabilmesi için vergilendirilebilir gelirlerinin en az %90'ını dağıtmaları şartı bulunmaktadır. GYO'ların temettü ve faiz ödemelerinin ekonomiye önemli katkılar sağladığı düşünülmektedir. Baker vd. (2006), temettülerin, hissedarlar için biriken kazançlardan elde edilen gelirlerin, sürekli olarak daha yüksek cari harcamaları teşvik ettiğini tespit etmişlerdir.

Bu bölümde, GYO firmaların ülke ekonomilerine katkıları hakkında bilgilere yer verilmiştir. Diğer bölümde GYO sektörü için önem arz eden GYO kuruluşları hakkında bilgiler yer almaktadır.

3.6. GYO Kuruluşları

Dünya çapında GYO faaliyetlerinin yürütülmesine öncülük eden birçok GYO dernek ve kuruluşları bulunmaktadır. Bu bölümde bu kuruluşlar hakkında bilgiler yer almaktadır.

3.6.1. Ulusal Gayrimenkul Yatırım Ortaklıkları Birliği (NAREIT)

NAREIT, Gayrimenkul Yatırım Ortaklıkları ve diğer gayrimenkul piyasasıyla ilgilenen şirketlerin çıkarlarını temsil eden bir birlik olarak 1960 yılında kurulmuştur. 200'den fazla üyeye sahip olan, ABD merkezli kuruluş GYO'lar için araştırmalar yapmakta, eğitim ve konferanslar düzenlemektedir. (Chen, 2022). NAREIT'in çeşitli faaliyetleri bulunmaktadır (NAREIT, 2022):

1. Savunuculuk: NAREIT, GYO'lara katkı sağlayacak mevzuat ve politikaların oluşturulması ve geliştirilmesine yardımcı olan bir kuruluştur. ABD başta olmak üzere diğer ülkelere bu konuda görüşler sunabilmektedir.

2. Üyelik: NAREIT gayrimenkul yatırımına ilgi duyanları bir araya getirme faaliyeti yürütmektedir.

Bu bölümde NAREIT kuruluşu hakkında bilgilere yer verilmiştir. Diğer bölümde Avrupa'da faaliyet gösteren Avrupa Kamu Gayrimenkul Birliği (EPRA) kuruluşu hakkında bilgiler yer almaktadır.

3.6.2. Avrupa Kamu Gayrimenkul Birliği (EPRA)

EPRA, 1999 yılında Belçika'da kurulmuş, Avrupa gayrimenkul sektörünün temsilcisi olan bir dernektir. 280'den fazla üyesi bulunan EPRA, 790 milyar Euro'nun üzerinde gayrimenkul varlığına sahip olup FTSE EPRA Nareit Europe Endeksi'nin piyasa değerinin %94'ünü temsil etmektedir (EPRA, 2023).

EPRA, Avrupa kamu gayrimenkul sektörünü teşvik etmek, geliştirilmek ve temsil edilmesini sağlamayı hedeflemektedir. Bu hedefe, yatırımcılara ve paydaşlara bilgi sağlayarak, kamusal ve siyasi tartışmalara katılım sağlayarak, en iyi uygulamaları teşvik ederek ve sektörün uyumunu ve güçlenmesini sağlayarak ulaşılmaktadır. EPRA, gayrimenkul sektöründeki profesyoneller arasında bilgi paylaşımını ve iş birliğini destekleyerek, sektörün sürdürülebilir büyümesine katkıda bulunmayı amaçlamaktadır (EPRA, 2023).

Bu bölümde EPRA kuruluşu hakkında bilgilere yer verilmiştir. Diğer bölümde Asya'da faaliyet gösteren Asya Pasifik Gayrimenkul Birliği (APREA) kuruluşu hakkında bilgiler yer almaktadır.

3.6.3. Asya Pasifik Gayrimenkul Birliği (APREA)

APREA, 2005 yılında Asya Pasifik bölgesinde etkili olan gayrimenkul trendlerini belirlemeyi amaçlayan bir grup sektör lideri tarafından kurulmuştur. APREA, menkul kıymetleştirilmiş gerçek varlıklar sektöründe öncülük etmektedir ve değişim ve müzakerelerin gerçekleştiği bir topluluk olma özelliği taşımaktadır. APREA, bölge genelinde üyelerin sürdürülebilir büyümesi için dört temel hizmet sunmaktadır (APREA, 2023):

1. Savunuculuk: APREA, Asya Pasifik'teki politika yapımcılarla etkileşimler sürdürülerek menkul kıymetleştirilmiş gerçek varlıklar sektörünün

uzun vadeli büyümesini desteklemektedir. Ayrıca, paydaşlar arasında istihdamın, ekonomik büyümenin ve sürdürülebilir çevrenin teşvik edilmesi ve sektör standartlarının özellikle şeffaflık ve sürdürülebilirlik raporlamasında uyumlaştırılmasını hedeflemektedir.

2. Eğitim ve Araştırma Faaliyetleri: APREA bünyesindeki veri ve istatistikler Üyeler ile paylaşılmakta ve üyelere özel öncü araştırmaların yapma imkânı sağlanmaktadır. Asya Pasifik genelindeki kural ve düzenlemelerin, sektör haberlerinin ve pozisyon belgelerinin güncellenmesi sağlanmaktadır. Ayrıca, sadece üyelerin erişebildiği APREA Endeksleri bulunmaktadır.

3. Mesleki Gelişim ve İlerleme: APREA Real Assets Academy tarafından Asya Pasifik'teki önde gelen üniversiteler ve sektör uzmanlarıyla iş birliği içinde kurslar düzenlemektedir.

4. Yatırımcı Desteği: Üyelerin ihtiyaçlarına yönelik özelleştirilmiş programlar oluşturulmakta ve üyelerin görünürlüğü hem dijital hem de fiziksel platformlarda artırılmaktadır. Ayrıca, anlaşma yapmayı teşvik eden bir ağ oluşturularak başarıya yönelik destek sağlanmaktadır.

Bu bölümde APREA kuruluşu hakkında bilgilere yer verilmiştir. Diğer bölümde Türkiye'de faaliyet gösteren Gayrimenkul Yatırımcıları Derneği (GYODER) kuruluşu hakkında bilgiler yer almaktadır.

3.6.4. Gayrimenkul Yatırımcıları Derneği (GYODER)

GYODER, Türkiye'deki gayrimenkul yatırım ortaklıklarının mevcut ve gelecekteki temsilcileri tarafından 1999 yılında kurulmuştur. "Türkiye'nin Gayrimenkul Platformu" olarak bilinen GYODER, sektörün farklı alanlarını tek bir çatı altında birleştirme amacıyla faaliyet göstermektedir. GYODER'in genişletilmiş bir vizyona sahip olması gerektiği ve günümüz çizgisine uyumlu, çağdaş bir yapıya bürünmesi gerektiği düşüncesiyle, GYODER'in markalaşma sürecini kolaylaştırmak ve güçlendirmek amacıyla isim değişikliğine gitmiş ve "Gayrimenkul Yatırımcıları Derneği" adını, 22 Haziran 2021 tarihinde yapılan 21. Olağan Genel Kurul Toplantısı'nda almıştır. Bu dönüşüm süreci, gayrimenkul sektörünün algısını güçlendirerek gelişimine destek olmak, sektöre kalite, kontrol ve eğitim standartları getirmek amacıyla güçlenerek devam etmektedir. Gayrimenkul Yatırımcıları Derneği hedefleri şu şekilde sıralanmaktadır (GYODER, 2023):

- GYODER'in sahip olduğu uzmanlık, tecrübe ve birikim ile sektörün geleceğe taşınmasını sağlamak.
- Sektörü teknoloji ve tasarımla birlikte büyütmek.
- Türkiye gayrimenkul sektörünün uluslararası platformlarda temsil edilmesi ve yurt dışına açılan bir pencere oluşturulmasını sağlamak.
- Sektörün gelişimi için çözümler üretilmesi ve ilgili mevzuat ve düzenlemelerin yapılmasında aktif rol alarak kamu otoritesini doğru yönlendirmek.
- Yerli ve yabancı kuruluşlarla iş birliği ortamları oluşturmak.
- Sektör paydaşları ve üyeler arasında iletişimi artırmak,
- Güncel ve sağlıklı bilgi sistemlerinin ve veri havuzunun oluşturulmasıyla sektör paydaşlarının kullanımına sunmak.
- Şehirleşme sürecinde teknoloji, tasarım ve doğa ile uyumlu çağdaş ve sürdürülebilir şehircilik unsurlarının gelişimine katkıda bulunmak.
- Gayrimenkul ve teknolojinin kesiştiği alanlarda farkındalık yaratma ve sektörde iş birliklerini ve yatırımları artırmak.

Bu bölümde GYO sektörü için önemli faaliyetler sürdüren GYO Kuruluşları hakkında bilgilere yer verilmiştir. Diğer bölümde GYO türleri hakkında bilgiler yer almaktadır.

3.7. GYO Türleri

GYO faaliyeti özellikle ticari gayrimenkuller olmak üzere gelir getiren gayrimenkullere yatırım yapmaya yöneliktir (Grupe ve DiRocco, 1999). GYO'lar gayrimenkul veya ipotek araçlarına sahip olma durumlarına göre özsermaye, ipotek ve hibrit olmak üzere üç türe ayrılmaktadır (Colestock, 2021). Bu bölümde, GYO türleri hakkında bilgiler yer almaktadır. Şekil 2'de GYO Türleri gösterilmektedir.



Şekil 2: GYO türleri

Özsermaye GYO: Özsermaye GYO firmaları , yatırım amaçlı gayrimenkul inşa etmek, geliştirmek, yönetmek, onarmak ve bazen satmaya odaklanmaktadır. Özsermaye GYO firmaları, yatırımcıların bütçelerini aşan, karşılayamayacakları mülk ve projelere yatırım yapmalarına olanak tanımaktadır. Özsermaye GYO firmalarına ait gayrimenkuller konut veya ticari nitelikte olabilmektedir. Ayrıca iki türün bir karışımı da olabilmektedir. Özsermaye GYO'ları konutlar, ofisler, oteller, alışveriş merkezleri, veri merkezleri, sağlık tesisleri, telekomünikasyon kuleleri, endüstriyel tesisler ve diğer mülkler dahil olmak üzere gelir getiren gayrimenkullere sahiptirler (EY, 2021).

İpotek GYO: Herhangi bir fiziksel varlığa sahip olmayan, ipotek veya ipoteye dayalı menkul kıymetler üretip veya satın alan GYO türüdür. Bankalardan ve diğer kaynak kuruluşlardan ipotek kredileri satın almaktadırlar. Daha sonra işletmelere ve ev alıcılarına kredi vermek için kullanılacak sermaye sağlamaktadırlar. Ticari ipotek GYO'ları ticari gayrimenkul finansmanını destekleyen kredilerin oluşturulmasına, servis edilmesine ve

menkul kıymetleştirilmesine de dahil olmak üzere bir dizi başka faaliyette de bulunmaktadır (EY, 2021).

Hibrit GYO: Özsermaye ve ipotek GYO türlerinin karışımından yer alan GYO türüdür (Glascock ve Lu-Andrews, 2015).

3.7.1. Konut GYO

Konut GYO'lar, çeşitli konut türlerini bünyesinde barındırarak bu mülklerdeki mekânları kiracılara tahsis etmektedirler. Konut GYO'ları; apartman binaları, öğrenci konaklama birimleri, prefabrik evler ve tek ailelik konutlara odaklanan GYO'ları içermektedir. Bu pazar segmentlerinde faaliyet gösteren bazı konut GYO'ları ise belirli coğrafi pazarlara veya mülk sınıflarına özel bir odaklanma stratejisi benimseyebilmektedirler (NAREIT, 2023). Ev sahipliğindeki düşüş ve bireylerin kararsızlık yaşadığı belirsiz konut piyasası, konut GYO'ları için olumlu bir bakış açısı oluşturabilir (Case, 2008). Yatırımcılar, konut GYO'lara daha fazla ilgi göstermektedir (Schwartz-Driver, 2008)

Özsermaye GYO'ları içerisinde konut türü önemli bir yere sahiptir. Tablo 6'da ülkelerde faaliyet gösteren konut GYO firma sayıları yer almaktadır.

Tablo 6: Ülkelerde faaliyet gösteren konut GYO firma sayıları

Ülke	Konut GYO Firma Sayısı
Avusturalya	1
Belçika	1
Kanada	2
Japonya	5
Türkiye	2
Birleşik Krallık	2
ABD	19
Toplam	32

Bu bölümde GYO türleri ve konut GYO hakkında bilgilere yer verilmiştir. Diğer bölümde GYO firmalarına yönelik ülkeler bazında yasal düzenleme ve kısıtlamalar hakkında ayrıntılı bilgiler almaktadır.

3.8. GYO'lara Yönelik Yasal Düzenleme ve Kısıtlamalar

Bu bölümde, GYO'ların işleyişini etkileyen yasal düzenleme ve bu düzenlemelerin getirdiği kısıtlamalar yer almaktadır. GYO'ların faaliyetlerine yön veren yasal çerçevenin anlaşılması, bu sektörün işleyişinin ve potansiyelinin daha iyi kavranmasını sağlayacaktır. Bu çerçevede, konut GYO firmalarının faaliyet gösterdiği ülkelerde yer alan yasal düzenleme ve kısıtlamalar ayrı başlıklar halinde bu bölümde yer almaktadır.

3.8.1. Avustralya GYO Firmalarına Yönelik Yasal Düzenlemeler

Avustralya GYO firmalarına yönelik öne çıkan yasal düzenleme ve kısıtlamalar aşağıda maddeler halinde yer almaktadır (Farrell, 2023):

1. Sermaye Gereksinimi: Avustralya'da, GYO firmaları için özel bir asgari sermaye gereksinimi bulunmamaktadır. GYO'ların sermaye yapısı ve sermaye miktarı, her bir GYO'nun iş stratejilerine bağlı olarak belirlenmektedir.

2. Yatırımcı Kısıtlamaları: Yabancı bireylerin bir GYO firmasının %10 veya daha fazla paya sahip olamamalarına ilişkin bir kısıtlama bulunmaktadır. Bu kısıtlama, Avustralyalı yatırımcıların yabancı bireyler tarafından gerçekleştirilen hisse edinme işlemlerinin potansiyel etkilerinden korunmasını amaçlamaktadır.

3. Borçlanma: Avustralya'da GYO firmaları brüt varlıklarının %60'ına kadar borçlanabilmektedirler.

4. Borsada İşlem Görme: Avustralya GYO firmaları için listelenme zorunluluğu bulunmamaktadır.

5. Temettü Dağıtımı: Avustralya'da GYO sektöründe, dağıtılmamış gelirin vergilendirilmesi önemli bir konudur. Dağıtılmamış gelir, GYO'nun elde ettiği gelirin temettü olarak yatırımcılara dağıtılmadan kalan kısmıdır. Dağıtılmamış gelir, en yüksek vergi oranı olan %49 üzerinden vergilendirilmektedir. Bu durumda, GYO'lar elde ettikleri gelirin tamamını (%100) yatırımcılara dağıtmayı tercih etmektedirler.

Bu bölümde, Avustralya GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde Belçika GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.2. Belçika GYO Firmalarına Yönelik Yasal Düzenlemeler

Belçika'da GYO firmalarının faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemlerini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir (Jurion, 2021):

1. Kuruluş Yeri ve Kanun: Belçika GYO firmaları, Belçika'da kurulmalı ve Mayıs 2014 tarihli yasalar da dahil olmak üzere geçerli olan Belçika yasalarına uygun olmalıdır.

2. Borçlanma: Belçika GYO firmalarının toplam kredileri, varlıklarının %65'ini aşmamalı ve yıllık faiz maliyetleri, toplam operasyonel ve finansal gelirlerinin %80'ini geçmemelidir.

3. Temettü Dağıtımı: Belçika GYO firmaları her yıl kârının %80'ini ve borç azalışının net miktarını dağıtmak zorundadır. Ancak, borç oranının %65'i aşması durumunda veya dağıtım sonucunda bu sınırı aşması durumunda dağıtım yapılamamaktadır.

4. Sermaye Gereksinimi: Belçika GYO firmalarının, ödenmiş sermayesi en az 1.2 Milyon Euro olmalıdır.

5. Yabancı Varlıklar Üzerindeki Kısıtlamalar: Belçika GYO firmalarına yönelik, başka ülkelerdeki varlıklara yatırım yapma konusunda herhangi bir kısıtlama bulunmamaktadır.

Bu bölümde, Belçika GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde Kanada GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.3. Kanada GYO Firmalarına Yönelik Yasal Düzenlemeler

Kanada'da GYO firmalarının faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemlerini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir (Vangou vd., 2021):

1. Yasal Düzenlemeler: Kanada, 2007 yılından bu yana gelir vergisi mevzuatında özel bir düzenlemeye sahiptir. Bu düzenleme, halka açık olarak işlem gören GYO'lar için belirli bir kural setini içermektedir ve bu kurallara "GYO Kuralları" adı verilmektedir.

2. Borsada İşlem Görme: Kanada GYO firmaları için kapalı uçlu fonların borsaya kote olma zorunluluğu bulunmaktadır. Ancak açık uçlu fonlar için böyle bir zorunluluk bulunmamaktadır.

3. Temettü Dağıtımı: Kanada GYO firmaları için asgari bir temettü dağıtım zorunluluğu bulunmamaktadır. Ancak vergi yükümlülüklerinden kaçınmak amacıyla vergilendirilebilir gelirin tamamı, vergilendirilebilir sermaye kazançları da dahil olmak üzere her yıl hisse sahiplerine ödenmeli veya ödenebilir hale getirilmelidir.

4. Gelir Kaynakları: Kanada GYO firmalarının, gelirlerinin çeşitli kaynaklardan elde edilmesini sağlamak amacıyla belirli bir kurala tabidir. Buna göre, GYO gelirlerinin en az %90'ı, gayrimenkullerden elde edilen kira gelirleri, sermaye mülklerinin elden çıkarılmasından kaynaklanan gelirler, faiz gelirleri, temettü gelirleri, telif hakları ve uygun yeniden satış mülklerinin elden çıkarılmasından kaynaklanan gelirlerin herhangi bir kombinasyonundan oluşmalıdır.

5. Mülk Dağılımı: Kanada GYO firmalarının, mülklerinin en az %80'i Kanada'da bulunan gayrimenkuller, nakit, hisse senetleri, pazarlanabilir menkul kıymetler, tahviller, borç senetleri ve diğer bazı varlıklardan oluşmalıdır.

Bu bölümde, Kanada GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde Japonya GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.4. Japonya GYO Firmalarına Yönelik Yasal Düzenlemeler

Japonya'da GYO firmalarının faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemlerini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir (Takagi ve Handler, 2021):

1. Yasal Düzenleme ve Kanun: Japon GYO firmaları, "Yatırım Ortaklıkları ve Yatırım Şirketleri Kanunu" kapsamında faaliyet göstermekte ve belirli varlıklara yatırım yaparak bu varlıkları yönetmeyi amaçlayan kuruluşlardır.

2. Asgari Sermaye: Japon GYO firmaları, Yatırım Ortaklıkları ve Yatırım Şirketleri Kanunu gereğince 100 milyon JPY (Japon Yeni) sermaye tutarıyla kurulmalıdır.

3. Borsada İşlem Görme: Japon GYO firmalarının bir borsaya kote olma zorunlu bulunmamaktadır.

4. Varlıklar ile İlgili Düzenlemeler: Japon GYO firmaları için varlıklar ile ilgili belirli kriterler mevcuttur. Bu kriterler arasında, yönetilen varlıkların en az %70'inin gayrimenkul olması, gayrimenkul, gayrimenkulle ilgili varlıklar ve likit varlıkların toplamının yönetilen toplam varlıklara oranının en az %95 olması, net varlıkların en az 1 Milyar JPY olması ve toplam varlıkların en az 5 Milyar JPY olması bulunmaktadır.

5. Yabancı Varlıklar: Japon GYO firmalarının yabancı varlıklara yatırım yapma konusundaki kısıtlamalar 12 Mayıs 2008 tarihli bir düzenleme ile kaldırılmıştır. Bu düzenleme ile Japon GYO firmalarının yabancı piyasalardaki fırsatları değerlendirmeleri ve portföylerini çeşitlendirmelerine imkân sağlanmaktadır.

6. Temettü Dağıtımı: Bir Japon GYO firmasının temettü dağıtım indiriminden yararlanabilmesi için, dağıtılabilir kârının en az %90'ını aşacak oranda temettü dağıtması gerekmektedir.

Bu bölümde, Japonya GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde Türkiye GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.5. Türkiye GYO Firmalarına Yönelik Yasal Düzenlemeler

Türkiye'deki GYO firmaların faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemesini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir (Bayraktaroğlu, 2023):

1. SPK Düzenlemeleri: GYO'ların faaliyetlerini ve işleyişini düzenlemek amacıyla Sermaye Piyasası Kurulu (SPK) tarafından çıkarılan düzenlemeler, sektörün etkin ve şeffaf bir şekilde işlemesini temin etmektedir. Bu düzenlemeler, SPK tarafından yayımlanan ve "Gayrimenkul Yatırım Ortaklıklarına İlişkin Esaslar Tebliği" adını taşıyan belge aracılığıyla GYO'ların faaliyetlerine ilişkin kuralları içermektedir.

2. Yatırım Amaçları : Türkiye GYO firmaları, gayrimenkul sektöründe çeşitli faaliyetlerde bulunarak, gayrimenkul projelerine yatırım yapma, gayrimenkul edinme, kiralama, geliştirme ve yönetme gibi işlemleri gerçekleştirirler. Bu işlemler, GYO'ların gayrimenkul portföyünü çeşitlendirme, değer artırma ve gelir elde etmeleri amacıyla yapılmaktadır.

3. Finansal Raporlama ve Şeffaflık: GYO'ların, SPK tarafından belirlenen finansal raporlama standartlarına uyma ve düzenli olarak faaliyet raporları sunma zorunluluğu bulunmaktadır.

4. Temettü Dağıtım: GYO'lar, kendi kâr dağıtım politikalarını belirleme yetkisine sahiptirler. Bu karar, GYO yönetim kurulu tarafından belirlenmektedir.

5. Borçlanma: GYO'lar, kısa vadeli fon taleplerini veya portföye bağlı maliyetleri karşılamak amacıyla, sahip olduğu özkaynakların beş katı oranında kredi kullanabilmektedirler.

Bu bölümde, Türkiye GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde Birleşik Krallık GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.6. Birleşik Krallık GYO Firmalarına Yönelik Yasal Düzenlemeler

Birleşik Krallıktaki GYO firmalarının faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemlerini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir (Clements vd., 2021):

1. Asgari Sermaye: Birleşik Krallık GYO firmalarının sermaye gereksinimlerine ilişkin herhangi bir zorunluluk bulunmamaktadır.

2. Borsada İşlem Görme: Birleşik Krallık GYO firmaları, tanınmış bir borsada işlem görmeye kabul edilmeli ve Londra Menkul Kıymetler Borsası veya uluslararası düzeyde eşdeğer bir borsada listelenmelidir.

3. Faaliyetler: Birleşik Krallık GYO firmalarının faaliyetleri, mülk kiralama işine ve diğer faaliyetlere sahip olabilme özelliği taşımaktadır. Bu çerçevede, GYO'nun elde ettiği kârın en az %75'i ve toplam varlıkların %75'i, mülk kiralama işiyle doğrudan ilişkili olmalıdır.

4. Yabancı Varlıklar: Yabancı varlıklara ilişkin ek bir kısıtlama bulunmamaktadır. Bu durum, Birleşik Krallık GYO firmalarının uluslararası gayrimenkul yatırımlarına esneklik sağlamaktadır.

5. Temettü Dağıtımı: Birleşik Krallık GYO firmaları, muafılığı olan kira kârının (finansman maliyetleri, genel giderler ve vergi amortismanı düşüldükten sonra kalan kira geliri) en az %90'ını, Gayrimenkul Gelir Temettüsü olarak dağıtması gerekmektedir.

Bu bölümde, Birleşik Krallık GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde ABD GYO firmalarına yönelik yasal düzenlemeler hakkında bilgiler yer almaktadır.

3.8.7. ABD GYO Firmalarına Yönelik Yasal Düzenlemeler

ABD GYO firmalarının faaliyetlerini denetlemek, yatırımcıların korunmasını sağlamak ve sektörün düzenli bir şekilde işlemlerini sağlamak amacıyla bir takım yasal düzenleme ve kısıtlamalar bulunmaktadır. Ön plana çıkan yasal düzenleme ve kısıtlamalar aşağıda özetlenmektedir

1. Gelir Kaynakları: ABD GYO firmalarının, yıllık olarak elde ettikleri brüt vergilendirilebilir gelirlerinin en az %75'i, gayrimenkul kiralari, gayrimenkul ipoteklerinden kaynaklanan faiz gelirleri, gayrimenkul ve ipotek kredilerinin satışından elde edilen kazançlar, diğer ABD GYO'larından gelen temettüler ve kazançlar gibi gayrimenkulle ilgili gelirlerden oluşmalıdır.

2. Yabancı Varlıklar: Yabancı varlıklara ilişkin ek bir kısıtlama bulunmamaktadır. Bu durum, ABD GYO firmalarının uluslararası gayrimenkul yatırımlarına esneklik sağlamaktadır.

3. Temettü Dağıtımı: ABD GYO firmaları, yıllık olarak elde ettikleri olağan vergilendirilebilir gelirlerinin en az %90'ını dağıtmakla yükümlüdürler.

4. Borsada İşlem Görme: ABD GYO firmaları, bir borsada işlem görme zorunluluğu bulunmamaktadır.

5. Yatırımcı Sayısı: ABD GYO firmalarının en az 100 hisse sahibine sahip olması gerekmektedir. Ancak her bir hissedar için belirlenmiş bir asgari değer bulunmamaktadır.

Bu bölümde, konut GYO firmalarının faaliyet gösterdikleri ülkelerde uygulanan yasal düzenlemeler hakkında bilgiler verilmiştir. Diğer bölümde bu

yasal düzenleme ve kısıtlamaların ülkeler bazında karşılaştırılması yapılmaktadır.

3.8.8. Yasal Düzenleme ve Kısıtlamaların Karşılaştırılması

Önceki bölümlerde belirtildiği üzere, GYO şirketlerine ilişkin yasal düzenlemeler ve kısıtlamalar ülkeler arasında farklılık göstermektedir. Bu bölümde, çalışma kapsamında bulunan 7 ülkede geçerli olan yasal düzenlemeler ve kısıtlamalar karşılaştırılmaktadır.

Yasal düzenleme ve kısıtlamalar arasında ön plana çıkanlar; kuruluş aşamasındaki asgari sermaye gerekliliği, belirli bir oranda temettü dağıtma zorunluluğu ve dış finansmana yönelik kısıtlamalardır. Tablo 7’de, asgari sermaye gerekliliği düzenlemesine göre ülkeler hakkında bilgiler yer almaktadır.

Tablo 7: Asgari sermaye gerekliliğine göre ülkeler

Ülke	Asgari Sermaye
Avusturalya	-
Belçika	1.2 M Euro
Kanada	-
Japonya	100 M JPY
Türkiye	64.5 M TL
Birleşik Krallık	-
ABD	-

M: Milyon, **JPY:** Japon Yeni, **TL:** Türk Lirası

Tablo 7’de konut GYO firmalara sahip ülkelerde asgari sermaye gerekliliği yasal düzenlemesine göre bilgiler yer almaktadır. Buna göre, Avusturalya, Kanada, Birleşik Krallık ve ABD’de asgari bir sermaye gerekliliği bulunmamaktadır. Tablo 8’de, belirli bir oranda temettü dağıtma zorunluluğu düzenlemesine göre ülkeler hakkında bilgiler yer almaktadır.

Tablo 8: Temettü dağıtma zorunluluğuna göre ülkeler

Ülke	Temettü Dağıtım Oranı
Avustralya	% 100
Belçika	% 80
Kanada	-
Japan	% 90
Türkiye	-
Birleşik Krallık	% 90
ABD	% 90

Tablo 8’de konut GYO firmalara sahip ülkelerde belirli bir oranda temettü dağıtma zorunluluğu yasal düzenlemesine göre bilgiler yer almaktadır. Buna göre, Türkiye ve Kanada’da belirli bir oranda temettü dağıtma zorunluluğu bulunmamaktadır. Tablo 9’da, GYO firmalarının borçlanma ile ilgili kısıtlamalarına göre ülkeler hakkında bilgiler yer almaktadır.

Tablo 9: Borçlanma kısıtlamalarına göre ülkeler

Ülke	Borçlanma Kısıtlamaları
Avustralya	Toplam varlıklarının %60’ını aşamaz.
Belçika	Kaldıraç Oranı: %65’i aşamaz
Kanada	-
Japonya	-
Türkiye	Kısa vadeli krediler özkaynakların beş katı ile sınırlıdır
Birleşik Krallık	Kâr / finansman oranı en az %25 olmalı
ABD	-

Tablo 9’da konut GYO firmalara sahip ülkelerde borçlanma ile ilgili kısıtlamalarına göre bilgiler yer almaktadır. Buna göre, Kanada, Japonya ve ABD’de bir kısıtlama bulunmamaktadır.

BÖLÜM 4

YÖNTEM

4. YÖNTEM

Bu bölümde, veri madenciliği hakkında kapsamlı bilgiler yer almaktadır. İlk alt başlıkta, veri madenciliğine yönelik temel bilgiler yer almaktadır. Bu çerçevede, veri madenciliği uygulama süreçleri, veri madenciliğinde veri seti türleri ve veri madenciliği uygulamalarında karşılaşılan problemler hakkında detaylı bilgiler yer almaktadır. Daha sonraki alt başlıkta ise çalışmada uygulanan yöntemler (Rassal orman, XGBoost ve CatBoost) hakkında kapsamlı bilgiler yer almaktadır. Bu çerçevede, her bir yöntemin çalışma şekli, avantaj ve dezavantajları ayrıntılı olarak yer almaktadır.

4.1. Veri Madenciliği Temel Bilgiler

Günümüzdeki hızlı teknolojik gelişmeler ve veriye kolay erişim imkanları, veri madenciliğinin önemini daha da artırmaktadır. Veri madenciliği, büyük hacimli veri kümelerinin analiz edilerek, keşfedilerek ve değerli bilgilerin elde edilmesi amacıyla kullanılan bir süreç olarak tanımlanabilir (Hand, 2001). Veri madenciliği, meteoroloji (Bartok vd., 2010), genomik (Lee vd., 2008), karmaşık fizik simülasyonları (Kamath, 2001), biyolojik ve çevresel araştırmalar (Sowmya ve Suneetha, 2017), finans ve iş dünyası (Bose ve Mahapatra, 2001) gibi çeşitli sektörlere kadar uzanan geniş bir uygulama alanına sahip bir alandır İşletmeler, pazarlama stratejilerinin geliştirilmesi (Aydoğan vd., 2008), müşteri davranışlarının anlaşılması (Mach-Król ve Hadasik, 2021), verimliliklerin artırılması ve rekabet avantajı elde edilmesi gibi hedeflere ulaşmak (Mak vd., 2011) amacıyla; finansal kurumlar ise risk analizi (Jin vd., 2018) ve dolandırıcılık tespiti (Debreceny ve Gray, 2010) gibi konularda veri madenciliği tekniklerinden yararlanabilmektedirler. Sağlık sektöründe, hastalık teşhisi ve tedavi planlaması (Koçak ve Ergün, 2023) gibi kritik kararların desteklenmesi amacıyla veri madenciliği yöntemleri kullanılabilir (Gheorghe ve Petre, 2014). Veri madenciliği, büyük veri çağında rekabet edebilme ve etkili kararlar alma açısından önemli bir araç haline gelmiştir ve büyük verilere erişim imkanının kolaylaşmasıyla birlikte belirli bir sektöre odaklanarak bir araştırma gerçekleştirme fırsatı oluşturmuştur.

Veri madenciliği, istatistik, makine öğrenimi ve veri ambarları veya veri tabanlarını kullanarak verilerden anlamlı bilgiler çıkarma süreci olarak tanımlanabilmektedir (Ming-Syan Chen vd., 1996). Bu çerçevede, veri madenciliğini; istatistik, büyük veri ve makine öğrenmesi kavramları ile açıklamak mümkündür. İstatistik, verilerin analiz edilmesi ve sonuçların çıkarılması için temel yöntemleri sağlamaktadır. Klasik istatistiğin çıkarımsal ve tanımlayıcı olmayan kısımlarının kullanılmasıyla ilgilidir. Bu çerçevede; p değerleri, karelerin toplamı, ki-kare değerleri, serbestlik derecesi ve F-oranları gibi göstergeler kullanılmaktadır (Ratner, 2017). Büyük Veri; hacim, çeşitlilik ve hız gibi özellikleriyle geleneksel veri işleme yöntemlerinin sınırlarını zorlayan büyük ve karmaşık veri setlerini ifade etmektedir (Pyne et al., 2016). Büyük verinin kullanım amacı, büyük hacimli ve/veya yüksek boyutlu veri kümelerini kullanarak tahminlerde bulunmak için uygun yöntemler geliştirmek, değişkenler arasındaki ilişkileri anlamak, gizli yapıları araştırmak ve önemli ortak özelliklere sahip alt-popülasyonları ortaya çıkarmaktır. Günümüzün dijital ortamı nedeniyle büyük veriye özel bir önem verilmektedir (Badaoui vd., 2017). Makine Öğrenmesi, yapay zekâ, bilgisayar görüşü ve istatistiksel öğrenme gibi alanların birleştiği bir disiplin olarak tanımlanmaktadır ve günümüzde birçok akıllı hizmet, uygulamanın temelini oluşturmaktadır (Hastie vd., 2001). Makine öğrenmesi, verilerin örüntülerini keşfetmek, tahminlerde bulunmak ve karar verme süreçlerini otomatikleştirmek amacı ile algoritmaları ve modelleri kullanmaktadır (Kohavi, 1998). Veri madenciliği, bu üç kavramın bir araya gelmesiyle, veri setlerindeki gizli bilgileri ortaya çıkartarak bilgiye dönüştürmeyi amaçlamaktadır. Veri madenciliği, işletmelerde, bilimsel araştırmalarda, pazarlama analizinde ve daha birçok alanda bilgiye dayalı karar verme süreçlerine yardımcı olmak için önemli bir araç olarak görülmektedir (Shaw vd., 2001).

Bu bölümde, veri madenciliği hakkında temel bilgiler yer almaktadır. Bu çerçevede, alt başlıklarda veri madenciliği süreçleri, veri madenciliğinde veri seti türleri ve veri madenciliği uygulamalarında karşılaşılan problemler yer almaktadır. Veri madenciliği uygulama süreçleri başlığı altında, Veri tabanlarında Bilgi Keşfi (KDD), Örnekleme, Keşfetme, Değiştirme, Modelleme ve Değerlendirme Süreçleri (SEMMA) ve Veri Madenciliği için Sektörler Arası Standart Süreç (CRISP-DM) gibi süreçlerin açıklamaları yer almaktadır. Bir diğer başlık altında veri seti türleri olan eğitim, test ve

doğrulama veri setleri kavramları ele alınmıştır. Veri madenciliği uygulamalarında karşılaşılan problemler başlığı altında ise aşırı öğrenme ve eksik öğrenme kavramları ele alınmıştır.

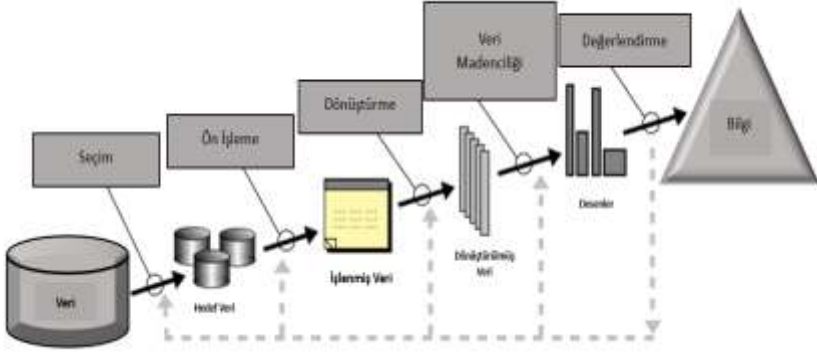
4.1.1. Veri Madenciliği Uygulama Süreçleri

Veri madenciliği süreçleri, disiplinli bir yaklaşımı temsil eden ve veri analizi alanında önemli bir çerçeve sunan araçlardır (Schulz vd., 2020). Bu süreçlerin kullanımının, doğru kararlar almak ve değerli bilgiler elde etmek için veri madenciliği projelerinin başarısını artırmada önemli bir rol oynaması beklenmektedir. Bu süreçler arasında, “Veri Tabanlarında Bilgi Keşfi (KDD)”, “Örnekleme, Keşfetme, Değiştirme, Modelleme ve Değerlendirme Süreçleri (SEMMA)” ve “Veri Madenciliği için Sektörler Arası Standart Süreç (CRISP-DM)” yer almaktadır (Shafique ve Qaiser, 2014). Bu süreçler, veri madenciliği projelerinde yol gösterici bir role sahiptirler ve başarılı sonuçlar elde etmek için bir planlama ve uygulama çerçevesi sağlamaktadırlar. KDD, genel bir bakış açısıyla veri madenciliği sürecine yaklaşırken, SEMMA ve CRISP-DM daha uygulama odaklı bir yaklaşım sunarak adım adım yönergeler ve aşamalar içermektedir.

Bu bölümde, veri madenciliği süreçleri (KDD, SEMMA ve CRISP-DM) hakkında bilgiler yer almaktadır. Bu süreçler, veri madenciliği projelerinin yönetilmesini sağlamakta ve veri setlerindeki bilgilerin değerli bilgilere dönüşmesine yardımcı olmaktadır.

4.1.1.1. Veri tabanlarında Bilgi Keşfi Süreci (KDD)

İşletmeler, diğer kuruluşlar ve araştırmacıların, verilerden anlamlı bilgiler elde etmeleri için verileri etkin bir şekilde kullanmaları beklenmektedir. Bu süreçte, KDD (Veri tabanlarında Bilgi Keşfi) veri toplama, veri hazırlama, veri analizi, model oluşturma ve sonuç değerlendirme gibi adımları içeren sistemli bir süreç sunarak, verilerden anlamlı ve kullanışlı bilgilerin çıkarılmasını desteklemektedir (Fayyad vd., 1996). Bu sayede, bilinçli kararlar almak ve stratejik adımlar atmak için verilerin derinlemesine analizi ve anlaşılması sağlanmaktadır (Maimon ve Rokach, 2005). KDD süreci 5 aşamadan (seçim, ön işleme, dönüştürme, veri madenciliği, yorumlama) oluşmaktadır. Şekil 3’te, bu aşamalar gösterilmektedir.



Şekil 3: KDD süreci aşamaları (Fayyad vd., 1996)

Şekil 3'te KDD sürecinin aşamaları gösterilmektedir. Bu aşamalar aşağıda maddeler halinde özetlenmektedir:

1. Seçim: İlk aşamada, KDD süreci için gerekli olan verilerin hedefleri belirlenmektedir. Daha sonra, keşif yapılacak veri kümesi oluşturulmaktadır. Bu veri kümesi farklı veri kaynaklarından elde edilebilmektedir. Bu nedenle entegrasyonu gerekebilmektedir. Verilerin entegre edilmesi, sonraki aşamalarda etkili bir keşif işlemi için önemli bir faktördür (Mavrogiorgou vd., 2021).

2. Ön işleme: Bu aşama, gürültülü veya aykırı değerleri çıkarmayı içermektedir. Bu sebeple bu aşamada, gürültüyü modellemek veya hesaba katmak için gerekli bilgilerin toplanması, eksik veri alanlarıyla başa çıkmak için stratejilerin belirlenmesi, zaman dizisi bilgisi ve bilinen değişikliklerin etkin bir şekilde ele alınması amaçlanmaktadır. Ayrıca, veri tabanı yönetim sistemiyle ilgili konulara (veri tipleri, şema düzeni ve eksik veya bilinmeyen değerlerin eşlenmesi gibi) kararlar alınması da bu aşamada gerçekleştirilmektedir (Mariscal vd., 2010).

3. Dönüşüm: Bu aşama, boyut azaltma veya dönüşüm yöntemleriyle verilerin dönüştürülmesini içermektedir. Veri örnekleme içinde yüzlerce değişken (nitelik) bulunabilmektedir. Ancak, KDD süreciyle ilgili sadece birkaç değişken önemli olabilmektedir. Veri dönüşüm aşamasında, önceden işlenmiş veriler, bilgi keşfi sürecini etkilemeyecek şekilde gereksiz

niteliklerden arındırılmakta ve dönüştürülmektedir (Zelevnikow ve Stranieri, 1997).

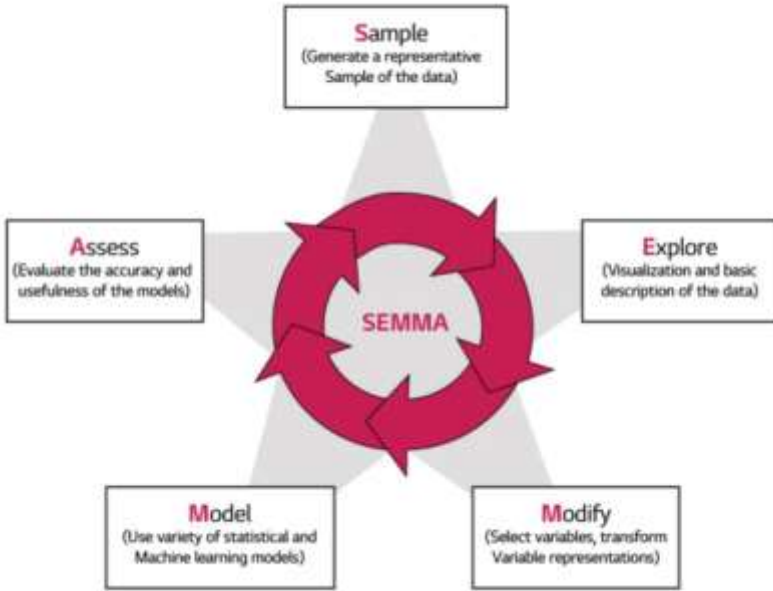
4. Veri Madenciliği: Veriler tamamen dönüştürüldüğünde, keşfedici analiz, model ve hipotezler seçilmektedir. Veri madenciliği sürecinde, altta yatan veri desenlerini çıkarmak için uygun veri madenciliği yöntemleri (kümeleme, ilişki kuralları, özetleme, sınıflandırma, regresyon gibi) ve algoritmalar (örneğin, K-MEANS, DBSCSAN, KNN gibi) seçilmekte ve uygulanmaktadır. Bu aşama, çıkarma işlemi için uygun modelleri ve parametreleri belirlemeyi içermektedir (Mavrogiorgou vd., 2021).

5. Yorumlama/Değerlendirme: Bu aşama, veri madenciliği sonuçlarının yorumlanmasını ve değerlendirilmesini içermektedir (Zelevnikow ve Stranieri, 1997).

KDD, etkileşimli ve yinelemeli bir süreç olup, verilerden örüntüleri keşfetmeyi, çıkarmayı ve yorumlamayı amaçlamaktadır (Elragal ve Klischewski, 2017). KDD, veri analizine dayanan bir dizi adımın sürekli olarak gerçekleştirilmesi olarak tanımlanabilmektedir. Bu adımlar, veri ön işleme, veri madenciliği algoritmalarının uygulanması, keşfedilen kalıpların yorumlanması ve sonuçların değerlendirilmesi gibi aşamaları içermektedir (Fayyad vd., 1996). Bu süreç, bilgiyi etkili bir şekilde ortaya çıkarmak ve anlamak için sistemli bir yaklaşım sağlamaktadır (Rossetti, 2022).

4.1.1.2. Örneklemeye, Keşfetme, Değiştirme, Modelleme, Değerlendirme Süreci (SEMMA)

Bir diğer süreç ise SEMMA (Örneklemeye, Keşfetme, Değiştirme, Modelleme, Değerlendirme) sürecidir. SEMMA süreci de KDD sürecinde olduğu gibi işletmeler, diğer kuruluşlar ve araştırmacılara verilerin etkin kullanımı için yardımcı olmaktadır. Süreç; Örneklemeye, keşfetme, değiştirme, modelleme ve değerlendirme aşamalarını içermekte ve veri madenciliği projelerinin sistematik bir şekilde ilerlemesini yardımcı olmaktadır (Olson ve Delen, 2008). SEMMA süreci, SAS Enstitü tarafından oluşturulmuş olup beş aşamadan oluşmaktadır. Bu aşamalar Şekil 4'te görsel olarak yer gösterilmektedir.



Şekil 4: SEMMA süreci aşamaları (Zhu, 2017)

Şekil 4’te SEMMA süreci aşamaları gösterilmektedir. Bu aşamalar aşağıda özetlenmiştir (Santos ve Azevedo, 2005):

1. Örneklem: Büyük veri kümesinin bir kısmının, önemli bilgiler içerecek büyüklükte örneklenmesi sürecidir. Seçilen örneklem ile eğitim, doğrulama ve test olmak üzere üç veri seti oluşturulabilmektedir. Eğitim seti, modelin uygunluğunu belirlemek, Doğrulama seti, değerlendirme ve aşırı uyumu önlemek ve Test seti, modelin genellemesini değerlendirmek amacıyla kullanılmaktadır (Obenshain, 2004).

2. Keşfetme: Bu aşama, verileri doğal eğilimleri veya gruplamaları açısından görsel veya sayısal olarak keşfetmeyi içermektedir. Keşif süreci, anlayış kazanmak ve fikirler elde etmek için kullanılmaktadır. Eğer görsel keşif net bir eğilim göstermezse, veriyi faktör analizi, uyum analizi ve kümeleme gibi istatistiksel tekniklerle inceleyerek keşif işlemine devam edilebilmektedir. Bu keşif süreci, tüm veri kümesi bir arada işlenirse tespit edilemeyebilecek daha zengin desenleri ortaya çıkarabilmektedir (Alawi vd., 2022).

3. Değiştirme: SEMMA sürecinin üçüncü aşamasıdır. Bu aşamada, model seçim sürecine odaklanmak için değişkenler oluşturulmakta, seçilmekte

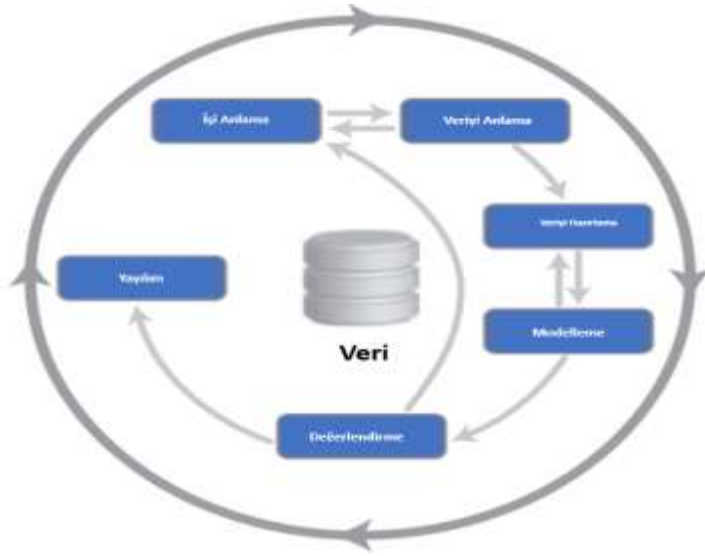
ve dönüştürülmektedir. Ayrıca, bu aşamada aykırı değerleri belirlemek ve değişken sayısını azaltmak amaçlanabilmektedir (Shafique ve Qaiser, 2014).

4. Modelleme: Bu aşama, veri modellemesiyle ilgilenmektedir. Yazılım, aranan istenen sonuçları güvenilir bir şekilde tahmin edebilecek veri kombinasyonlarını otomatik olarak aramaktadır. Her bir model türü, özel durum ve problemlere bağlı olarak kendi işlevselliğine ve kullanım alanına sahiptir (Shah vd., 2017).

5. Değerlendirme: SEMMA yönteminin son aşaması olan değerlendirme aşamasında, modelin değerlendirilmesi yapılmaktadır. Model değerlendirmesi, modelin performansının tahmin edilmesi için kullanıcı tarafından gerçekleştirilmektedir. Bir modelin değerlendirilmesinde yaygın bir strateji, verileri eğitim ve test verileri olarak bölmektir; eğitim verileri modelin oluşturulmasında kullanılırken test verileri ile modelin performansı değerlendirilmektedir. Değerlendirme aşamasının amacı, modelin doğruluğunu ve eğitim veri setinde elde edilen sonuçlarla tutarlılığını değerlendirmektir (Alawi vd., 2022).

4.1.1.3. Veri Madenciliği için Sektörler Arası Standart Süreç (CRISP-DM)

Üçüncü süreç ise CRISP-DM sürecidir (Veri Madenciliği için Sektörler Arası Standart Süreç). CRISP-DM, veri madenciliği projelerinde kullanılan süreç modelidir. Bu süreç, sektörler arasında ortak bir dil ve yaklaşım sağlamaktadır (Wirth ve Hipp, 2000). Bu süreç, veri madenciliği projelerinin planlanması, yönetilmesi ve uygulanması için sistematik bir çerçeve sunmaktadır (Chapman vd., 2000). Projelerin daha yapılandırılmış ve kontrol edilebilir bir şekilde ilerlemesini sağlaması bakımından CRISP-DM'nin oldukça önemli olduğu düşünülmektedir. Ayrıca, projelerin tekrarlanabilirlik ve ölçeklenebilirlik açısından avantaj kazanmasına ve veri madenciliği ekipleri arasında bilgi ve deneyim paylaşımını kolaylaştırmasına katkıda bulunabilmektedir. Bu sayede, veri madenciliği projelerinin başarı şansının artması ve bilgi keşfi sürecinde daha etkili sonuçların elde edilmesi beklenmektedir. Şekil 5'te CRISP-DM süreci gösterilmektedir.



Şekil 5: CRISP-DM (Hui, 2018)

Şekil 5'te gösterilen CRISP-DM süreci, veri madenciliği projeleri için standart bir süreç oluşturmaktadır. CRISP-DM süreci altı bölümden oluşmaktadır (Kelleher vd., 2015):

1. İş Anlayışı: Veri analistinin başlıca hedefi, iş açısından müşterinin gerçekten neyi başarmak istediğini tam olarak anlamaktır. Müşteri genellikle çeşitli hedeflere ve kısıtlamalara sahip olabilmektedir. Bu durum, dikkatli bir dengeleme yaklaşımını gerektirmektedir. Analist, projenin sonucunu etkileyebilecek önemli faktörleri başlangıçta ortaya çıkarmayı amaçlamaktadır (Chapman vd., 2000).

2. Veri Anlama: Veri analistinin, kuruluş içerisinde yer alan farklı veri kaynaklarını ve bu kaynaklardaki çeşitli veri türlerini anlamak için çaba sarf ettiği aşamadır. Bu aşamada, veri madenciliği proje hedefine ilişkin gizli bilgiler için hipotezler, deneyim ve nitelikli varsayımlara temel alınarak oluşturulmaktadır (Huber vd., 2019).

3. Hazırlama: Tahmine dayalı modellerin oluşturulması için belirli bir yapıda düzenlenen verilere ihtiyaç duyulmaktadır. Bu aşamada, farklı veri kaynaklarını dönüştürmek için gerekli tüm faaliyetler yer almaktadır (Kelleher vd., 2015).

4. Modelleme: Modelleme sürecindeki ilk olarak kullanılacak gerçek modelleme tekniğinin seçimi yapılmaktadır. İş anlayışında muhtemelen zaten bir araç seçilmiş olsa da bu görev özel modelleme tekniğinin belirlenmesine yöneliktir. Örneğin, C4.5 ile karar ağacı oluşturma veya geri yayılım ile sinir ağı gibi tekniklerin belirlenmesi. Birden fazla teknik uygulanacaksa, her bir tekniğe yönelik olarak bu adım ayrı ayrı gerçekleştirilmektedir (Chapman vd., 2000).

5. Değerlendirme: Modellerin kullanılmadan önce değerlendirilmesi ve amaca uygun olduklarının kanıtlanması önemlidir. Bu aşama, tahmin modelinin doğru tahminlerde bulunabileceğine yönelik tüm değerlendirme görevlerini kapsamaktadır. Genellikle, çeşitli modeller birbiriyle rekabet etmektedir. Veri bilimci, model sonuçlarını alan bilgisine, başarı kriterlerine ve test tasarımına dayanarak yorumlamak zorundadır (Saltz, 2021).

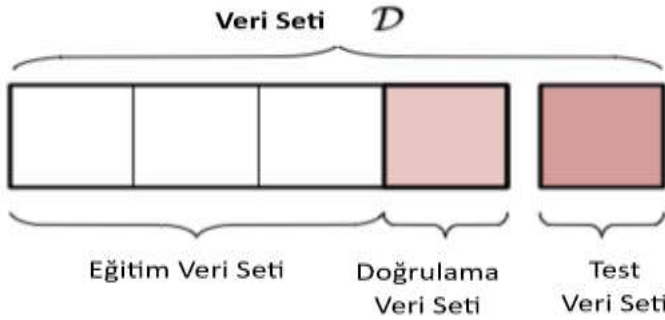
6. Dağıtım: CRISP-DM sürecinin altıncı ve son aşamasıdır. Bu aşamada, elde edilen bilgi ve sonuçların kullanımına odaklanılmaktadır. Aynı zamanda bu aşama, elde edilen bilginin organize edilmesine, raporlanmasına ve ihtiyaç duyulduğunda sunumuna yönelik çalışmaları içermektedir (Shafique ve Qaiser, 2014).

CRISP-DM yöntemi, farklı veri bilimi projelerinin farklı gereksinimlere sahip olabileceği için adım adım takip edilmesi zorunlu bir metodoloji değildir. Bunun yerine, CRISP-DM metodolojisini, projenize özgü farklı yönleri dikkate aldığınızdan emin olmak için bir şablon olarak kullanılabilir. Bu yöntem, projenin analiz aşamalarını kapsayan bir çerçeve sunmakta ve projenin başarılı bir şekilde yürütülmesine rehberlik etmektedir. Ancak, her projenin benzersiz gereksinimleri olduğu için, CRISP-DM yöntemini esnek bir şekilde uygulamak ve projenin özel ihtiyaçlarını karşılamak önemlidir (Luna, 2021). Bir sonraki bölümde veri madenciliğinde veri seti türleri hakkında bilgiler yer almaktadır.

4.1.2. Veri Madenciliğinde Veri Seti Türleri

Bu bölümde veri seti türleri hakkında bilgiler yer almaktadır. Bu aşama, veri setlerinin doğru ve dikkatli bir şekilde oluşturulması, analiz edilecek veri örneklerinin temsil gücünü artırmak ve modele genellemeyi sağlamak açısından kritik bir adımdır.

Veri madenciliğinde, büyük veri setinden eğitim ve test ve doğrulama örnekleri seçilerek veri seti parçalara ayrılmaktadır (George vd., 2016). Böyle bir parçalama işlemi yapılmaksızın tüm örnekler eğitimde kullanılması durumunda, aşırı öğrenme ve test seti için büyük tahmin hataları oluşabilmektedir (Montesinos López vd., 2022). Bu sebeple, veri madenciliğinde veri setleri üç gruba ayrılmaktadır. Bunlar eğitim, test ve doğrulama setleridir. Veri setinin %70 ile %80'lik kısmı ile eğitim veri seti oluşturulmaktadır. Eğitim veri setinin amacı, algoritmayı girdilerden çıktıları tahmin etmesi konusunda eğitmesidir. Veri setinin %20 ile %30'luk kısmı ile test veri seti oluşturulmaktadır (Neelamegam ve Ramaraj, 2013). Şekil 6'da veri setleri türleri görsel olarak gösterilmektedir.



Şekil 6: Veri seti türleri (Ziganto, 2018)

Şekil 6'da görüldüğü üzere veri setinin büyük bir kısmı eğitimi için kullanılmaktadır. Kalan kısımlar ise doğrulama ve test veri seti olarak kullanılmaktadır. Bu bölümde, eğitim, test ve doğrulama veri setleri hakkında bilgiler ayrı başlıklar altında yer almaktadır. Eğitim veri seti, bir modelin öğrenme sürecinde kullanılan verileri içermektedir. Model, bu veri seti üzerinde eğitilerek örüntüleri öğrenmektedir. Test veri seti, eğitilen modelin performansını değerlendirmek için kullanılmaktadır. Bu veri seti, modelin genelleme yeteneğini ölçmek için kullanılmaktadır. Doğrulama veri seti ise modelin genel performansını doğrulamak için kullanılmaktadır. Bu veri setleri, projelerde önemli bir rol oynamaktadırlar. Doğru bir şekilde kullanıldıklarında modelin güvenilirliğini ve etkinliğini artırmaktadırlar.

Eğitim veri seti, veri madenciliği sürecindeki en önemli unsurlardan biridir. Bu veri seti, modelin öğrenmesi için kullanılan örnek verileri

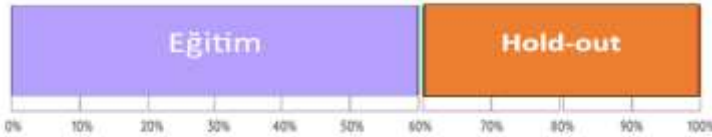
içerilmektedir. Modelin ilişkileri anlamasını, desenleri öğrenmesini ve gelecekteki verilere genelleme yapabilmesini sağlamaktadır (Amatriain vd., 2011). Eğitim veri setinin kalitesi, modelin doğruluğunu, performansını ve genel başarısını etkilemektedir. Bu nedenle, uygun büyüklükte, temsilci ve çeşitlilik gösteren bir eğitim veri setinin kullanılması, veri madenciliği modelinin başarılı bir şekilde oluşturulmasında kritik bir rol oynamaktadır (Kavzoglu, 2009). Eğitim veri setinin amacı, algoritmayı girdilerden çıktılar tahmin etmesi konusunda eğitmesidir (Rençber, 2021). Genellikle, bir veri setinin %70'i eğitim seti olarak ayrılmaktadır. Ancak, verinin %70'inin eğitim için ayrılması ve geri kalanının test için kullanılması zorunlu bir kural değildir. Bu, kullanılan veri setinin özelliklerine ve gerçekleştirilmesi gereken göreve bağlı olarak değişmektedir (Lhessani, 2019).

Test veri seti, veri madenciliği modellerinin performansını değerlendirmek ve genelleme yeteneklerini test etmek için kullanılan önemli bir unsurdur. Test veri seti, modelin görmediği ve daha önce tahmin etmediği yeni verilerden oluşmaktadır. Test veri setinin amacı, eğitim veri seti ile geliştirilen modelin değerlendirmesi için veri sağlamaktır (Amatriain vd., 2011). Veri setinin %20 ile %30'luk kısmı ile test veri seti oluşturulmaktadır. Başarılı sonuca ulaşamadığı takdirde algoritma yeniden eğitilmesi sağlanabilir (Rençber, 2021). Ayrıca, test veri seti modelin aşırı öğrenme (overfitting) durumlarını tespit etmek için de kullanılmaktadır. Test veri setinin, eğitim ve doğrulama veri setlerinden bağımsız olarak seçilmesi gerekmektedir. Bu, gerçek hata oranının güvenilir bir tahminini elde etmek için önem arz etmektedir. Test setinin bağımsızlığı, elde ettiğimiz sonuçların genellemesi ve güvenilirliği açısından büyük önem taşımaktadır (Han vd., 2011).

Doğrulama veri seti, genellikle veri setinin %10'luk kısmını oluşturmaktadır. Doğrulama veri seti, test veri setinden önce kullanılmaktadır. Böylece eğitimin daha başarılı yapılabilmesi amaçlanmaktadır (Rençber, 2021). Verilerin küçük bir kısmını doğrulama seti olarak ayrılmasının nedeni, yanlılığı azaltmak ve modelin genelleme yeteneğini artırmaktır. Doğrulama seti, modelin eğitim ve test verilerine göre tarafsız bir şekilde değerlendirilmesini sağlamaktadır. Yalnızca eğitim ve test verilerine sahip olduğunda model eğitim ve test verilerine aşırı uyum sağlamış olabilir. Doğrulama seti, modelin genel performansını daha objektif bir şekilde değerlendirmemize olanak tanımaktadır (Pickell, 2021).

Bir modelin bir veri kümesiyle eğitilmesi sırasında, optimal bir yöntem belirlemek önemlidir. Eğitim süreci, modelin yeterli örnekler üzerinde eğitilmesini sağlarken aynı zamanda aşırı öğrenme problemini de engellemelidir. Bununla birlikte, yeterli sayıda örnek bulunmadığında modelin yeterince eğitilemeyeceği ve test aşamasında düşük performans sergileyeceği göz önünde bulundurulmalıdır (Yadav ve Shukla, 2016). Model değerlendirme sürecinde çeşitli yöntemler kullanılmaktadır. Bu yöntemler arasında yaygın olarak kullanılanlar, Hold-out Yöntemi ve Çapraz Doğrulama (Cross-Validation) yöntemidir.

Hold-out doğrulaması, aşırı öğrenmeyi önlemek amacıyla kullanılmaktadır (Rice vd., 2020). Bu yöntemde, veri seti, eğitim ve test olarak iki ayrı parçaya bölünmekte ve Test amaçlı ayrılan veriye "hold-out" ismi verilmektedir (Way vd., 2010). Geri kalan kısım, modelin öğrenilmesi için kullanılmaktadır. Bu yöntem, verinin bir kısmını test etmek ve geri kalanını eğitim sürecinde kullanmak amacıyla tercih edilmektedir. Farklı yüzdelikler ile hold-out yöntemi kullanılabilir. Ancak, veri setinin dağılımı ve farklılık düzeyi dikkate alınmalıdır (Nurhayati vd., 2014). Hold-out doğrulama yöntemi, bir diğer model değerlendirme yöntemi olan çapraz doğrulamaya kıyasla daha kısa bir sürede modelin öğrenilmesine olanak sağlamaktadır (Yadav ve Shukla, 2016). Şekil 7’de hold-out yöntemi görsel olarak gösterilmektedir.



Şekil 7: Hold-Out doğrulama yöntemi

Holdout yöntemi, basit bir çapraz doğrulama yöntemidir ve hesaplamak için daha az zaman almaktadır. Bununla birlikte, değerlendirme sürecinde yüksek bir varyansa sahip olabilmektedir. Değerlendirme sonuçları, veri noktalarının eğitim setinde veya test setinde nasıl bölümlendirildiğine bağlı olarak önemli ölçüde farklılık gösterebilmektedir. Bu durum, değerlendirme sonuçlarının belirsizliklerle karşılaşabileceği anlamına gelmekte ve modelin performansının bu bölümlendirmeye bağlı olarak değişebileceğini göstermektedir (Schneider, 1997).

Model değerlendirme yöntemlerinden bir diğeri, Çapraz Doğrulama (CV) Yöntemidir. Çapraz doğrulama veya "k-kat çapraz doğrulama", veri setinin "k" adet gruba rastgele olarak bölüdüğü bir yöntemdir (Marcot ve Hanea, 2021). Gruplardan biri test geri kalanlar ise eğitim veri seti olarak kullanılmaktadır. Model, eğitim veri seti ile eğitilmektedir. Daha sonra, test veri seti ile değerlendirilmektedir. Bu süreç, her bir benzersiz grup test seti olarak kullanılana kadar tekrarlanmaktadır (Allibhai, 2018).

Çapraz doğrulama yöntemi olan k-kat çapraz doğrulama algoritması şu adımlardan oluşur (Yadav ve Shukla, 2016):

1. İlk olarak verinin 1/k'sini tutulan bir örnek olarak ayrılmaktadır.
2. Daha sonra, model kalan veri üzerinde eğitilmektedir.
3. Daha sonra, model 1/k tutulan örnek üzerine uygulanmakta ve gerekli model değerlendirme metriklerini kaydedilmektedir.
4. Diğer aşamada, ilk 1/k veriyi geri almakta ve bir sonraki 1/k'ı ayrılmaktadır (ilk seferde seçilen kayıtları hariç tutulmalıdır).
5. Daha sonra, 2. ve 3. Adımlar tekrarlanmaktadır.
6. Her kaydın tutulan kısmında kullanılana kadar tekrarlanmalıdır.
7. Son olarak, model değerlendirme metriklerinin ortalaması alınır veya başka bir şekilde birleştirilir. Şekil 8'de yöntem görsel olarak gösterilmektedir.



Şekil 8: Çapraz doğrulama (CV) yöntemi

Çapraz doğrulama, birden fazla eğitim-test ayrımı üzerinde eğitim yapma olanağı sağlamaktadır (Liu vd., 2018). Yaygın olarak kullanılan bir yöntemdir (Anguita vd., 2012). Bu, modelin gerçek veriler üzerindeki performansının daha iyi tahmin edilmesine yardımcı olmaktadır. Bekletme yöntemi ise yalnızca tek bir eğitim-test ayrımına dayanmakta ve sonuçları veri

bölünmesine bağlı hale getirmektedir. Büyük veri setleri, zaman kısıtları veya yeni bir veri bilimi projesi başlangıcında hold-out yöntemini tercih etmek faydalı olabilmektedir. Ancak çapraz doğrulama, daha fazla hesaplama gücü ve zaman gerektiren birden fazla eğitim-test ayrımını içerdiğinden daha kapsamlı bir yöntemdir (Allibhai, 2018).

Çapraz doğrulama, modelleri karşılaştırmak için basit bir yöntem olması ile birlikte, dikkat edilmesi gereken bazı hususlar bulunmaktadır (Hertzmann vd., 2015):

- Yöntem, çok sayıda eğitim çalışması gerektirdiği için zaman alıcı olabilmektedir. Birkaç parametreden oluşan modeller için çapraz doğrulama, yetersiz verimlilik nedeniyle kullanışsız olabilmektedir.

- Eğitim için azaltılmış bir veri seti kullanıldığından hem eğitim verilerinde hem de test verilerinde problemle ilgili tüm olguların bulunduğundan emin olacak kadar yeterli eğitim verisi içermelidir.

- Korelasyonların olasılığını önlemek için rastgele bir bölüm kullanmak önemlidir. Örneğin, veriler bir hafta boyunca toplandıysa, haftanın başındaki verilerin haftanın ilerleyen zamanlarındaki verilerden farklı bir yapıya sahip olma olasılığı bulunmaktadır.

- Çapraz doğrulama sonucunda, aşırı öğrenme (overfitting) ve eksik öğrenme (underfitting) hala oluşabilmektedir. Ancak olasılığı çok daha düşüktür. Örneğin, test seti çok küçükse, test verisindeki rasgele deseni uyan bir model tercih edebilir.

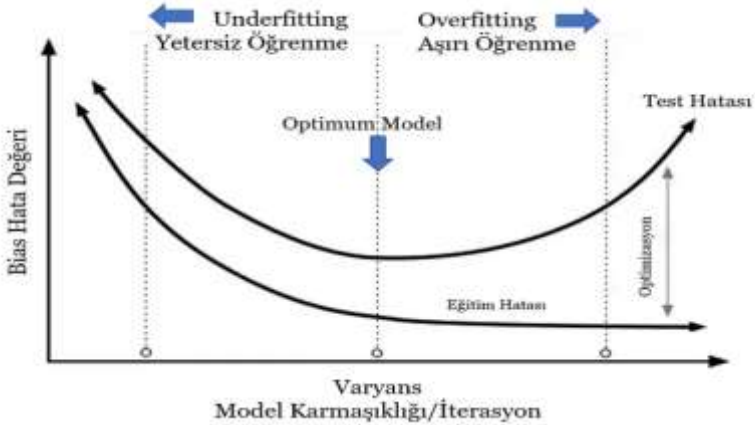
Bu bölümde, veri madenciliği yöntemleri kullanılarak oluşturulan modellerin değerli ve güvenilir olmasında önemli bir etkiye sahip olan veri seti türleri hakkında bilgilere yer verilmiştir. Özellikle eğitim veri setinin, tüm veri setinin temsil gücünü sağlaması başarılı modellerin oluşmasında büyük bir etken olarak görülmektedir. Çalışma, veri madenciliğinde karşılaşılan problemler hakkında bilgiler içeren bölüm ile devam etmektedir.

4.1.3. Veri Madenciliği Uygulamalarında Karşılaşılan Problemler

Model oluşturma, veri madenciliğinin temel adımlarından biri olarak görülmekte ve bir problem veya görev için en iyi tahmin veya sınıflandırma modelini oluşturmayı amaçlamaktadır. İyi bir model oluşturma süreci hem yanlılığı hem de varyansı dengede tutmaya çalışmaktadır. Optimal bir modelin,

hem veriye uygun bir şekilde yanlılık göstermesi hem de farklı veri kümeleri üzerinde tutarlı sonuçlar üretecek şekilde varyansa sahip olması beklenmektedir. Bu dengeyi sağlamak için, uygun model seçimi, veri ön işleme, hiperparametre ayarlama ve model değerlendirme yöntemleri kullanılmaktadır. Bu bölümde yanlılık ve varyans kavramları üzerinden optimal model kavramı açıklanmaktadır.

Veri madenciliğinde yanlılık (bias) ve varyans (variance) kavramları, modelin performansını ve genelleme yeteneğini anlamak için önemli iki kavram olarak görülmektedir. İkisi arasındaki denge, modelin doğru sonuçları tahmin edebilme yeteneği açısından kritik öneme sahiptir. Şekil 9'da yanlılık ve varyans denge noktası gösterilmektedir.



Şekil 9: Optimum model (Yiğit, 2020)

Yanlılık, modelin yanlış varsayımlar nedeniyle gizli kalıpları ortaya çıkaramayacağı hatalar olarak ifade edilmektedir. Yanlılık, modelin temel yapısının yetersizliğiyle ilişkilidir ve genellikle daha karmaşık modeller kullanılarak azaltılabilmektedir (Geurts, 2009). Varyans ise, eğitim verilerine aşırı duyarlılık nedeniyle ortaya çıkan hatalar olarak ifade edilmektedir. Yanlılık, modelin sürekli olarak aynı yanlış şeyi öğrenme eğiliminde olduğunu gösterirken, varyans rastgele şeyleri öğrenme eğiliminde olduğunu göstermektedir. Yanlılık ve varyans, modelin doğru tahminler yapabilmesi için gereken karmaşıklık düzeyi hakkında bilgi sağlamaktadırlar (Rocks ve Mehta, 2022). Daha karmaşık modeller, eğitim verilerine daha iyi uyum sağlayabilir

ancak gürültülü verileri aşırı öğrenme riski de bulunmaktadır. Karmaşık olmayan, basit modellerde ise tahmin yapmak için gerekli esneklik eksik olabilmektedir. Optimal bir model için yanlılık ve varyans arasında bir denge kurulması gerekmektedir (Brighton ve Gigerenzer, 2012). Doğrusal modeller genellikle yüksek yanlılığa sahiptir çünkü doğrusal bir düzlemlerle iki sınıfa ayırmak mümkün olmayabilmektedir. Karar ağaçları ise yüksek varyans sorunuyla karşılaşabilmektedir. Yüksek yanlılığa sahip modeller genellikle modele uymazken, düşük varyansa sahip olabilmektedirler (Domingos, 2012).

Topluluk Öğrenme Yöntemleri, yanlılık ve varyans arasındaki optimal dengeyi sağlamada önemli bir yere sahiptir (Harine Rajashree ve Hariharan, 2021). Topluluk yöntemleri, birden çok sınıflandırıcının çıktısını birleştirerek daha iyi bir genel performans elde etmek için yanlılığı ve varyansı azaltmaya yardımcı olmaktadır. Topluluk yöntemleri, farklı veri alt kümeleri ile eğitilen temel sınıflandırıcıları kullanılmaktadır. Eğitim setlerinin seçimi ve temel sınıflandırıcıların kararlılığı, topluluk sınıflandırıcılarının yanlılığı ve varyansı azaltmasında etkili olabilmektedir (Zaki ve Meira, 2014).

Varyans ve yanlılık kavramları, aşırı ve eksik öğrenme sorunlarıyla ilişkilidir. Varyans, modelin farklı eğitim verileri üzerinde değişen sonuçlar üretme eğiliminde olduğu durumlarda ortaya çıkmaktadır ve bu aşırı öğrenme ile ilişkilendirilmektedir. Yanlılık ise modelin veriye yetersiz uyum sağladığı durumları ifade etmektedir ve eksik öğrenme sorunlarına yol açmaktadır. Veri madenciliği uygulamalarında karşılaşılan en önemli problemler olarak aşırı öğrenme ve eksik öğrenme kavramları gösterilebilmektedir. Bu kavramlar, doğru model seçimi ve eğitim sürecinin başarısı açısından dikkate alınması gereken kritik faktör olarak gösterilebilmektedir. Aşırı öğrenme durumunda, model eğitim verilerine fazlasıyla uyum sağlamak ve bu durum genelleme yeteneğini zayıflatmaktadır. Eksik öğrenme durumunda ise model, verileri yeterince öğrenememekte ve düşük performans göstermektedir. Bu nedenle, aşırı öğrenme ve eksik öğrenme arasında bir denge kurulması gerekmektedir. Bu dengeyi sağlayarak, modelin eğitim verilerine uyum sağlama düzeyi optimize edilebilmekte ve aynı zamanda yeni verilere genelleme yapabilme yeteneği de geliştirilebilmektedir. Bu da doğru model seçimi ve eğitim sürecinin etkin bir şekilde yönetilmesini gerektirmektedir. Bu bölümde aşırı ve eksik öğrenme kavramlarının önemini içeren bilgiler yer almaktadır.

Bir modelin, düşük varyans yüksek yanlılığa sahip olması eksik öğrenme (underfitting) olarak tanımlanmaktadır. Bir modelin yüksek varyans, düşük önyargıya sahip olması ise aşırı öğrenme (overfitting) olarak tanımlanmaktadır. Her ikisi de regresyon ve sınıflandırma problemlerinde yaygın bir şekilde karşılaşılan sorunlardır (Lever vd., 2016). Şekil 10'da aşırı öğrenme ve eksik öğrenme kavramları grafik ile gösterilmektedir.



Şekil 10: Optimum model (Yiğit, 2020)

Aşırı öğrenme, bir modelin eğitim verileri üzerinde yüksek performans sergilemesine rağmen, test verileri veya yeni gözlemlerle karşılaşıldığında kötü sonuçlar verme durumunu ifade etmektedir (Bilbao ve Bilbao, 2017). Başka bir tanımda ise bir modelin eğitim verilerine aşırı derecede uyması sonucunda ortaya çıkan bir durum olarak tanımlanmaktadır. Bu durumda, model eğitim verileri üzerinde mükemmel bir performans sergilerken, gerçek dünya girdileri için düşük bir performans göstermektedir (Kim, 2017). Aşırı öğrenme, genelleme performansını olumsuz etkileyen temel faktörlerden biri olarak görülmektedir. Topluluk öğrenme yöntemleri, aşırı öğrenmeyi önlemek için tercih edilebilir bir yöntemdir. Topluluk öğrenme yöntemleri, farklı sınıflandırıcıları bir araya getirerek daha dengeli ve genelleyici bir performans elde etmeyi hedeflemektedirler (Sagi ve Rokach, 2018).

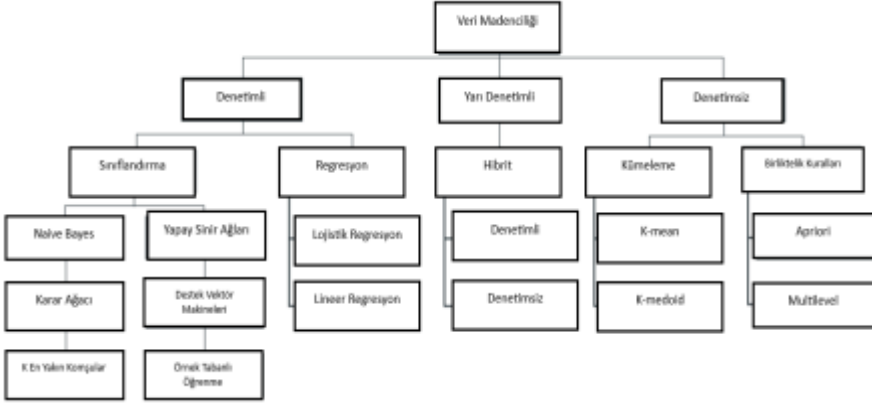
Eksik öğrenme, aşırı öğrenmenin tam tersidir ve modelin görevini yerine getirememesi durumunu ifade etmektedir (Jabbar ve Khan, 2015). Bu durumun; yanlış model seçimi, yanlış özelliklerin kullanılması veya yanlış dönüşümlerin uygulanması gibi birçok nedeni olabilmektedir. Eksik öğrenmenin net bir çözümü yoktur, ancak farklı modeller denemek, veri sayısını artırmak, yeni dönüşümler kullanmak gibi yöntemlerle eksik öğrenme düzeltilebilmektedir. Yüksek kaliteli bir modelin, aşırı öğrenme ve eksik öğrenme arasında bir denge

kurulması beklenmektedir. Modelin eğitim verilerine yeterli bir şekilde uyum sağlaması ve genelleyebilme yeteneği önem arz etmektedir. Bu dengeyi sağlamak için modelin karmaşıklığı, veri sayısı, özellik seçimi ve dönüşümler gibi faktörlerin dikkate alınması gerekmektedir (Maimon ve Rokach, 2009).

Bu bölümde, veri madenciliği uygulamalarında karşılaşılan problemler ele alınmıştır. Bu sorunların varlığı istenmeyen sonuçlara yol açabilmekte ve modelin gerçek dünya verileri üzerinde başarısını olumsuz yönde etkileyebilmektedir. Bu sebeple, oluşturulacak modellerde bu sorunlar engellenmeye çalışılmamıştır. Çalışmanın devam eden bölümünde, uygulama bölümünde kullanılan yöntemler hakkında detaylı bilgiler yer almaktadır.

4.2. Çalışmada Uygulanan Yöntemler

İstatistik, bilim ve endüstri tarafından sunulan sorunlarla sürekli olarak karşı karşıya kalmaktadır. Bilgisayarların ortaya çıkmasıyla istatistiksel sorunlar boyut ve karmaşıklık açısından artış göstermiştir. Bu durum, veri depolama, düzenleme ve arama gibi alanlarda yaşanan zorluklar sonucunda "veri madenciliği" adı verilen yeni bir disiplinin doğmasına yol açmıştır. Günümüzde birçok alanda büyük miktarda veri üretilmektedir. Bu verilerden anlamlı bir şekilde yararlanmak, önemli desenleri ve eğilimleri çıkarmak ve verinin içerdiği bilgileri anlamak istatistikçinin görevidir. Buna veriden öğrenme denilmektedir. Ele aldığımız öğrenme problemleri genellikle denetimli veya denetimsiz olarak kategorize edilmektedir (Hastie vd., 2001). Denetimli yöntemler, kesin bir hedef olduğunda kullanılmakta iken denetimsiz yöntemler, sonuç için bir belirsizlik söz konusu olduğunda kullanılmaktadır (Albayrak ve Akbulut, 2008). Şekil 11'de öğrenme yöntemleri ayrımı gösterilmektedir.

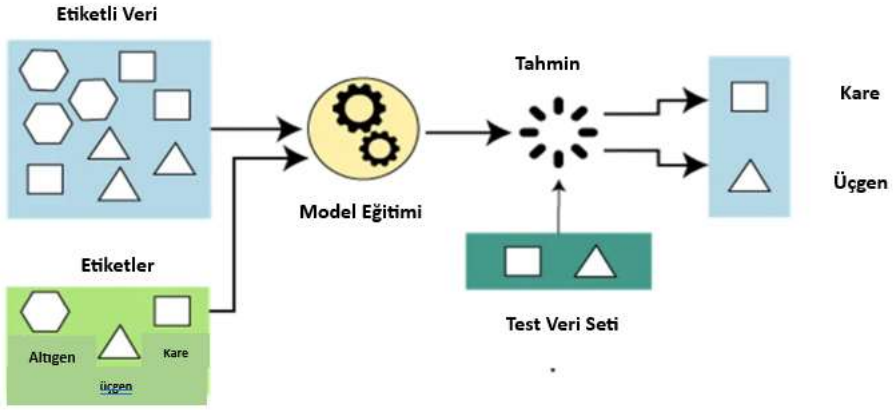


Şekil 11: Veri madenciliği yöntemleri (Ahmad Tarnizi vd., 2019)

Denetimli öğrenme, mevcut verilerdeki hedef veya etiketlenmiş durumlarla algoritmanın bilgiyi genelleme yeteneğini yansıtarak, algoritmanın yeni (etiketlenmemiş) durumları tahmin etmesi için kullanılmaktadır (Berry vd., 2020). Denetimli öğrenme yöntemleri, geçmiş örnekler veya veriler aracılığıyla belirli bir hedef özelliği ile ilişkili tanımlayıcı özellikler arasındaki ilişkiyi öğrenmek için kullanılmaktadır. Bu yöntemler, bir model oluşturarak ve modeli eğiterek ilişkiyi analiz etmekte ve gelecekteki veriler üzerinde tahminler yapmak için bu modeli kullanmaktadır (Kelleher vd., 2015). Denetimli öğrenme, bireylerin yeni bilgileri öğrenme süreciyle benzerlikler taşımaktadır. Örneğin, insanlar deneme problemlerini çözerken, bu süreçte yeni bilgiler kazanmaktadırlar. Bu süreci şu adımlarla açıklayabiliriz (Kim, 2017):

1. Bir deneme problemi seçilir ve mevcut bilgi bu problemi çözmek için uygulanır. Ardından, çözümle karşılaştırma yapılır.
2. Eğer elde edilen cevap yanlış ise, mevcut bilgi düzeltilir ve yeniden deneme yapılır.
3. Bu adımlar, tüm egzersiz problemleri için tekrarlanır.

Bu örnekte, deneme problemleri ve çözümleri eğitim verilerine, bilgi ise modelimize karşılık gelmektedir. Bu çerçevede, çözümlere ihtiyaç duymamız denetimli öğrenmenin hayati bir yönünü oluşturmaktadır. Şekil 12’de denetimli öğrenme süreci gösterilmektedir.



Şekil 12: Denetimli öğrenme (www. tutorialforbeginner.com)

Denetimli öğrenme, sınıflandırma ve regresyon olmak üzere iki alt başlığa ayrılmaktadır. Sınıflandırmada çıktılar ayrık etiketler şeklinde ifade edilmektedir. Regresyonda ise çıktılar gerçek değerlerden oluşmaktadır (Hertzmann vd., 2015). Tablo 10’da sınıflandırma ve regresyon yöntemlerinde kullanılan başlıca algoritmalar gösterilmektedir (Argüden ve Erşahin, 2008).

Tablo 10: Sınıflandırma ve regresyon modelleri algoritmaları

Algoritma	Model
Yapay Sinir Ağları	Sınıflandırma ve Regresyon
Bayes Sınıflandırması	Sınıflandırma
En Yakın Komşu	Sınıflandırma
Karar Destek Makineleri	Sınıflandırma ve Regresyon
Zaman Serisi Analizi	Sınıflandırma
Karar Ağaçları	Sınıflandırma ve Regresyon
Lojistik Regresyon	Sınıflandırma
Lineer Regresyon	Regresyon

Tablo 10’da sınıflandırma ve regresyon yöntemlerinde kullanılan başlıca algoritmalar gösterilmektedir. Bazı modeller sadece sınıflandırma veya regresyon problemlerinde kullanılabilirken, bazı modeller hem sınıflandırma hem de regresyon problemlerinde kullanılabilir.

Sonraki bölümlerde, çalışmada uygulanan yöntemler hakkında bilgilere yer almaktadır. Bu çerçevede, ilk olarak regresyon analizi hakkında detaylı bilgiler verildikten sonra çalışmada uygulanan regresyon yöntemleri hakkında detaylı bilgilere yer almaktadır.

4.2.1. Regresyon Analizi

Regresyon, deęer tahmininde bulunmak için kullanılan bir yöntemdir ve eğitim verilerinde doğru çıktı deęerlerini kullanmaktadır. Bu süreçte, regresyon doğru çıktı deęerlerine dayanarak tahminler gerçekleştirmektedir (Kim, 2017). Örneęin, pazarlama müdürü, belirli bir müşterinin AllElectronics'teki bir satış sırasında ne kadar harcayacağını tahmin etmek istedięinde regresyon yöntemlerini kullanması gerekmektedir (Argüden ve Erşahin, 2008). Bu bölümde regresyon analizi ve regresyon analizinde kullanılan performans ölçütleri hakkında bilgiler yer almaktadır.

Veri madencilięinde regresyon algoritmaları, baęımlı ve baęımsız deęişkenler arasındaki ilişkiyi tahmin etmek için kullanılan yöntemlerdir. Bu algoritmalar, veri seti üzerinde istatistiksel analizler yaparak ve matematiksel modeller oluşturarak tahminler yaparlar. Yaygın olarak kullanılan regresyon algoritmalarına; doğrusal, lasso, elastic net ve polinom regresyon örnek olarak gösterilebilir (Pandey, 2022).

Doğrusal regresyon yöntemi, en temel regresyon yöntemi olarak gösterilmektedir. Baęımlı deęişken ile baęımsız deęişkenler arasındaki doğrusal ilişkiyi modellemeye çalışmaktadır (Sharma vd., 2017). Temel olarak, doğrusal regresyon modelleri, iki deęişken veya faktör arasındaki ilişkiyi göstermek veya tahmin etmek için kullanılmaktadır. Çözümlemesi yapılan faktör, tahmin edilen faktör olarak kabul edilmektedir; bu faktör aynı zamanda baęımlı deęişken olarak isimlendirilmektedir. Baęımlı deęişkenin deęerini öngörmek için kullanılan faktörler ise baęımsız deęişkenler olarak isimlendirilmektedir (Gupta, 2015). Regresyon analizi, sadece bir tahmin edici deęişkenin bulunduğu bir modeli ifade ettięinde "basit regresyon" olarak adlandırılmaktadır. Regresyon modelinde en az iki yordayıcı deęişkenin bulunduğu durumlar ise "çoklu regresyon" olarak adlandırılmaktadır. Birden fazla yordayıcı deęişkenin kullanıldığı regresyon analizi, daha karmaşık ilişkileri incelemek ve tahminler yapmak için kullanılmaktadır (Gana ve Broc, 2019). Basit regresyon modeli aşağıda yer alan formül ile gösterilebilmekte ve modelde yer alan deęişkenler şu şekilde açıklanabilmektedir (Smith, 2015):

$$Y = \alpha + \beta X + e \quad (1)$$

Y: Baęımlı deęişkeni temsil etmektedir. Bu, tahmin etmeye çalıştığımız veya açıklamaya çalıştığımız deęişkendir.

α : Kesme terimidir veya doğrusal regresyon modelinde y eksenini kestiği noktadaki değeri temsil eden terimdir.

β : Eğim katsayısıdır ve X değişkeninin Y üzerindeki etkisini göstermektedir. Bu, X değişkeni bir birim arttığında Y'nin beklenen artış veya azalışını temsil etmektedir.

X: Bağımsız değişkeni temsil etmektedir. Bu, Y'nin tahmin edilmesinde kullanılan veya Y'nin değişimini açıklamaya çalıştığımız değişkendir.

e: Hata terimidir veya tahmin edilen değer ile gerçek değer arasındaki farkı temsil etmektedir. Bu, modelimizin tüm değişkenleri açıklamadığı veya diğer faktörlerin etkisini göz ardı ettiği durumları yansıtabilir.

Standartlaştırılmış bir β katsayısı, tahmin edicilerin kontrol edildiği durumda bağımlı değişken Y'deki standart sapma birimi cinsinden beklenen artışı ifade etmektedir (Nieminen, 2022). Öte yandan, standartlaştırılmamış regresyon katsayısı (B) değeri sınırsız olabilmekte ve dolayısıyla + sonsuzdan, - sonsuza kadar uzanabilmektedir. Bu durumda, yüksek bir B değeri, X'in Y'nin güçlü bir tahmincisi olduğunu doğrudan göstermemektedir. Bununla birlikte, standartlaştırılmış regresyon katsayısı (β) değerleri genellikle -1.00 ile +1.00 aralığında değişmekte ve bazen bu sınırları biraz aşabilmektedir. Standartlaştırılmış değişkenlerin standart sapması 1'e eşit olduğundan, ilgili β katsayılarını karşılaştırmak mümkündür (Gana ve Broc, 2019).

Bu bölümde, regresyon analizi hakkında genel bilgilere yer almaktadır. Alt başlıkta regresyon analizinde kullanılan performans ölçütleri açıklanmaktadır.

4.2.1.1. Regresyon Analizi Performans Ölçütleri

Regresyon yöntemleri ve algoritmaları hakkındaki bilgiler, oluşturulan modellerin en önemli beklentisinin gerçek dünyadaki verileri anlama, ilişkileri keşfetme ve gelecekteki değerleri tahmin etme yeteneği olduğunu göstermektedir. Önceki bölümlerde bahsedilen yöntemler modelde yer alan hatayı en aza indirerek doğru tahminler yapmayı hedeflemektedir. Regresyon yöntemlerine göre oluşturulan modelin başarısı, bir problemin çözümü için doğru ve güvenilir tahminler yapabilme yeteneğiyle ilişkilidir. Modelin başarısını değerlendirmek için çeşitli performans ölçütleri kullanılmaktadır. Regresyon yöntemleri performans ölçütleri, bir regresyon modelinin ne kadar iyi çalıştığını ve tahminlerinin ne kadar doğru olduğunu değerlendirmek için

önemlidir. Bu ölçütler, modelin gerçek değeri ne kadar yakın tahmin ettiğini ve hatalarının ne kadar düşük olduğunu göstermektedir. Dolayısıyla, model hatalarını değerlendirmede en uygun ölçütü seçmek, model performansının doğru bir şekilde anlaşılması için önemli bir karar olarak kabul edilmektedir. İyi bir performans ölçütü, modelin güvenilirliğini ve etkinliğini belirlerken, farklı modellerin veya değişken seçim yöntemlerinin karşılaştırılmasına da yardımcı olmaktadır. Böylece, regresyon analizinde performans ölçütlerini değerlendirmek, doğru ve güvenilir tahminler elde etmek ve en iyi modeli seçmek için önemlidir. Bu bölümde regresyon analizi yöntemlerinin başarılarının değerlendirilmesinde kullanılan ölçütler yer almaktadır. Determinasyon katsayısı, ortalama kare hatası, ortalama mutlak hata veri madenciliği regresyon uygulamalarında en fazla tercih edilen performans ölçütleri olarak gösterilmektedir.

Determinasyon katsayısı (R^2), bir regresyon modelinin performansını değerlendiren önemli bir ölçüttür. Diğer ölçütlerle birlikte kullanıldığında, modelin doğruluğu ve açıklama gücü hakkında daha kapsamlı bir anlayış sağlamaktadır. Bu bölüm, determinasyon katsayısı hakkında temel bilgileri içermekte ve ayrıca bu katsayının nasıl hesaplandığını açıklayan bir formül sunmaktadır.

Çoklu korelasyon katsayısı, regresyon modellerinin değerlendirilmesinde yaygın olarak kullanılan bir araçtır. Bu katsayı, genellikle "R" veya basit durumlarda "r" olarak gösterilmekte ve istatistiksel anlamlılık, R'nin değerlendirilmesinde temel bir ölçüt olarak kabul edilmektedir (Crocker, 1972). Determinasyon Katsayısı (R^2) ise bağımlı değişkenin varyansının açıklanabilir kısmını temsil etmektedir (Nagelkerke, 1991). Bu değerler, modelin uyumunu ve açıklama kapasitesini göstermektedir. R^2 , genellikle yüzde olarak ifade edilmekte ve 0 ile 1 arasında değer almaktadır. R^2 değerinin yüksek olması, tahmin modelinin doğruluğunun da yüksek olduğunu göstermektedir (Miles, 2014). R^2 değeri, regresyon analizinde model uyumunu ve açıklama gücünü değerlendirmek için önemli istatistiksel ölçüt olarak kabul edilmektedir (Gana ve Broc, 2019). R^2 ölçütüne ait formül aşağıda yer almaktadır (Chicco vd., 2021):

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}$$

Cohen (1988), R2 için geleneksel tanımlamaları sağlamıştır. Buna göre, küçük bir etki R2'nin 0.02'ye eşit olması, orta düzeyde bir etki R2'nin 0.13 olması ve büyük bir etki ise R2'nin 0.26 olması olarak tanımlanmaktadır (Miles, 2014).

Ortalama Kare Hatasıdır (MSE): MSE, tahmin modellerinin performansını değerlendirmek, hata miktarını yönetmek ve model karşılaştırmalarında kullanılan önemli bir ölçüttür. Bu bölüm, Ortalama Kare Hatası hakkında temel bilgileri içermektedir ve ayrıca bu katsayının nasıl hesaplandığını açıklayan bir formül sunmaktadır.

MSE, gerçek ve tahmin edilen değerler arasındaki farkın karesinin ortalamasını gösteren ölçüt olarak tanımlanmaktadır (Baasith, 2021). Başka bir ifade ile MSE, karesel hataların ortalamasını hesaplamaktadır (Khare vd., 2021). MSE ölçütüne ait formül aşağıda yer almaktadır (Chicco vd., 2021):

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)$$

Formülde yer alan, m tahmin sayısını, Xi gerçek değerleri ve Yi tahmin edilen değerleri ifade etmektedir (Khare vd., 2021).

Kök Ortalama Kare Hatasıdır (RMSE): RMSE, tahmin modellerinin performansını değerlendirmek, hata miktarını yönetmek ve model karşılaştırmalarında kullanılan önemli bir ölçüttür. Daha düşük RMSE değerleri, daha iyi performans ve daha doğru tahminler anlamına gelmektedir. Bu bölüm, Kök Ortalama Kare Hatası hakkında temel bilgileri içermektedir ve ayrıca bu katsayının nasıl hesaplandığını açıklayan bir formül sunmaktadır.

Kök Ortalama Kare Hatası (RMSE), yaygın bir şekilde kullanılan değerlendirme ölçütüdür. Bu ölçüt, Ortalama Kare Hatasının (MSE) karekökünü alarak hesaplanmaktadır (Khare et al., 2021). RMSE, gerçek ve tahmin edilen değerler arasındaki farkın karelerinin ortalamasının karekökünü gösteren ölçütüdür (Acharya, 2021). RMSE ölçütüne ait formül aşağıda yer almaktadır (Chicco et al., 2021):

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}$$

RMSE, gerçek değerler ile tahmin değerler arasındaki farkın hem yönünü hem de büyüklüğünü dikkate almaktadır. Aykırı değerlere karşı MAE'ye göre daha duyarlıdır (Hodson, 2022). RMSE'nin aykırı değerlere olan hassasiyeti, bu ölçütün kullanımıyla ilgili en sık karşılaşılan endişelerden biridir. Bu nedenle, RMSE hesaplarken aykırı değerlerin etkisi dikkate

alınmalıdır. Aykırı değerler, hata ölçüsünü büyük ölçüde etkileyebilmekte ve tahmin performansını yanıltıcı bir şekilde yüksek veya düşük gösterebilmektedir (Chai ve Draxler, 2014). Aykırı değerlere karşı duyarlı olması ve hesaplama zorluğu ölçütün dezavantajları arasında gösterilmektedir (Willmott ve Matsuura, 2005).

Ortalama Mutlak Hatadır (MAE): Ortalama Mutlak Hata (MAE), tahmin modellerinin performansını değerlendirmek, hata miktarını yönetmek ve model karşılaştırmalarında kullanılan önemli bir ölçüttür. Daha düşük MAE değerleri, daha iyi performans ve daha doğru tahminler anlamına gelmektedir. Bu bölüm, Ortalama Mutlak Hata hakkında temel bilgileri içermektedir ve ayrıca bu katsayının nasıl hesaplandığını açıklayan bir formül sunmaktadır.

Ortalama mutlak hata (MAE), model değerlendirmelerinde yaygın olarak kullanılan başka bir faydalı ölçüdür. MAE, gerçek ve tahmin edilen değerler arasındaki mutlak farkların ortalamasını göstermektedir (Chugh, 2020). Bu nedenle aykırı değerlerin etkisini daha dengeli bir şekilde ele almaktadır (Chai ve Draxler, 2014). RMSE'den farkı gerçek değer ile tahmin değerler arasındaki farkın yönlerini dikkate almamasıdır (Hodson, 2022). MAE ölçütüne ait formül aşağıda yer almaktadır (Chicco et al., 2021):

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

Kök Ortalama Kare Hatası (RMSE) ve Ortalama Mutlak Hata (MAE), farklı disiplinlerde yaygın olarak kullanılan iki mutlak hata ölçüsüdür. Bu ölçümler özellikle makine öğrenimi ve veri madenciliği gibi alanlarda sıklıkla kullanılmaktadır (Tang vd., 2017). Bu ölçümler, temel olarak model uydurma sürecinde (belirli bir modelin optimum parametre seçimi), model doğrulama işleminde, model seçiminde, model karşılaştırmalarında ve tahmin değerlendirmelerinde kullanılmaktadır. RMSE ve MAE, bu süreçlerde model performansının değerlendirilmesi ve karşılaştırılması için önemli araçlar olarak kullanılmaktadır. Bu ölçümler, modelin tahmin yeteneğini, hata miktarını ve doğruluğunu anlamak ve model seçiminde objektif bir karar vermek için kullanılmaktadır (Karunasingha, 2022).

Bu bölümde regresyon analizi hakkında bilgiler yer verilmiştir. Diğer bölümde topluluk öğrenme yöntemleri hakkında bilgiler yer almaktadır.

3.2.2. Topluluk Öğrenme Yöntemleri

Marquis de Condorcet'in, Jüri Teorimine göre oylamaya daha fazla katılımın çoğunluğun doğru olma olasılığını arttırmaktadır. Condorcet'in bu teoremi topluluk öğrenme modelleri için de geçerlidir (Ganaie vd., 2021). Topluluk öğrenme yöntemleri, birden çok yöntemin bir araya getirilerek ortak bir hedefi başarmak için iş birliği yaptığı bir öğrenme yaklaşımını ifade etmektedir (Dietterich, 2002). Topluluk öğrenme yöntemleri, aynı sorunu çözmek için tasarlanmış bir dizi öğrenme modelini bir araya getirerek daha etkili bir model oluşturmayı amaçlamaktadır (Ksieniewicz vd., 2018). Bu bölümde, çeşitli topluluk öğrenme yöntemleri ve kullanım alanları hakkında daha detaylı bilgiler yer almaktadır.

Topluluk öğrenimi, regresyon veya sınıflandırma problemlerinde tahmin yapmak için birleştirilen çeşitli modeller üreten yöntemleri ifade etmektedir (Xie ve Peng, 2019). Topluluk öğrenme süreci üç aşamaya ayrılabilir (Mendes-Moreira vd., 2012). İlk aşamada, bir dizi modeller ile topluluklar oluşturulmaktadır. İkinci aşama, budama aşamasıdır. Bu aşamada, bazı modeller topluluktan çıkartılarak budama işlemi yapılmaktadır. Son aşama ise, topluluk entegrasyonudur. Temel modelleri birleştirmek için bir strateji tanımlanmaktadır. Bu strateji daha sonra, temel modellerin tahminlerine dayanarak, yeni durumlar için topluluğun tahminini elde etmek için kullanılmaktadır.

Tahmin performansının iyileştirilmesinde topluluk öğrenme yöntemlerinin kullanılmasının birkaç nedeni bulunmaktadır (Polikar, 2006):

1. İstatistiksel Nedenler: Birden fazla sınıflandırıcının çıktılarını ortalama alarak birleştirmek, düşük performans gösteren bir sınıflandırıcının yanlışlıkla seçilme riskini azaltabilmektedir. Ortalama ile yanlış bir seçim yapma riski azalma eğilimindedir.

2. Büyük Hacimli Veriler: Analiz edilmesi gereken veri, tek bir sınıflandırıcı tarafından işlenemeyecek kadar büyük olabilmektedir. Bu durumlarda, topluluk öğrenmesi yöntemlerini kullanmak faydalı olabilmektedir.

3. Çok Az Veri: Topluluk sistemleri, yeterli eğitim verilerinin mevcut olmadığı durumlarda, yeniden örnekleme teknikleri kullanılarak her biri farklı bir sınıflandırıcıyı eğitmek ve topluluğu oluşturmak amacıyla kullanılabilirler.

4. Veri Birleştirme: Farklı kaynaklardan gelen verilerin birleştirildiği uygulamalara veri birleştirme uygulamaları denilmektedir. Farklı doğaya sahip özniteliklerin bulunduğu ve çeşitli kaynaklardan elde edilen birden fazla veri setinin olduğu durumlarda, tüm verilerin içerdiği bilgiyi öğrenmek için tek bir sınıflandırıcı yetersiz kalabilmektedir. Örneğin, nörolojik bir bozukluğu teşhis ederken, nörolog, MRI taraması, EEG kaydı, kan testleri vb. gibi birkaç test isteyebilir. Her test, toplu olarak eğitim için kullanılamayacak farklı sayıda ve özellikte veri üretmektedir. Bu tür durumlarda, her bir test yönteminden elde edilen veriler, çıktılarını birleştirebilmek için daha sonra eğitilebilen farklı bir sınıflandırıcı kullanılmak üzere kullanılabilir.

Bagging ve boosting en çok kullanılan topluluk öğrenme yöntemleri olarak gösterilmektedir (Malek vd., 2022). En çok kullanılma nedenleri, sınıflandırma doğruluk oranlarını iyileştirmede iyi bir performans göstermeleridir. Bu teknikler, eğitim veri kümesini manipüle ederek çeşitli sınıflandırıcı toplulukları oluşturmakta ve ardından yeni verileri bir oylama mekanizması yardımı ile sınıflandırmaktadırlar (Tuysuzoglu ve Birant, 2020). Şekil 13'te bu yöntemler gösterilmektedir.

4.2.2. Topluluk Öğrenme Yöntemleri

Marquis de Condorcet'in, Jüri Teorimine göre oylamaya daha fazla katılımın çoğunluğun doğru olma olasılığını arttırmaktadır. Condorcet'in bu teoremi topluluk öğrenme modelleri için de geçerlidir (Ganaie vd., 2021). Topluluk öğrenme yöntemleri, birden çok yöntemin bir araya getirilerek ortak bir hedefi başarmak için iş birliği yaptığı bir öğrenme yaklaşımını ifade etmektedir (Dietterich, 2002). Topluluk öğrenme yöntemleri, aynı sorunu çözmek için tasarlanmış bir dizi öğrenme modelini bir araya getirerek daha etkili bir model oluşturmayı amaçlamaktadır (Ksieniewicz vd., 2018). Bu bölümde, çeşitli topluluk öğrenme yöntemleri ve kullanım alanları hakkında daha detaylı bilgiler yer almaktadır.

Topluluk öğrenimi, regresyon veya sınıflandırma problemlerinde tahmin yapmak için birleştirilen çeşitli modeller üreten yöntemleri ifade etmektedir (Xie ve Peng, 2019). Topluluk öğrenme süreci üç aşamaya ayrılabilir (Mendes-Moreira vd., 2012). İlk aşamada, bir dizi modeller ile topluluklar oluşturulmaktadır. İkinci aşama, budama aşamasıdır. Bu aşamada, bazı modeller topluluktan çıkartılarak budama işlemi yapılmaktadır. Son aşama ise,

topluluk entegrasyonudur. Temel modelleri birleştirmek için bir strateji tanımlanmaktadır. Bu strateji daha sonra, temel modellerin tahminlerine dayanarak, yeni durumlar için topluluğun tahminini elde etmek için kullanılmaktadır.

Tahmin performansının iyileştirilmesinde topluluk öğrenme yöntemlerinin kullanılmasının birkaç nedeni bulunmaktadır (Polikar, 2006):

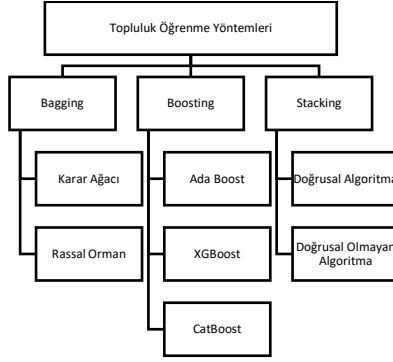
1.İstatistiksel Nedenler: Birden fazla sınıflandırıcının çıktılarını ortalama olarak birleştirmek, düşük performans gösteren bir sınıflandırıcının yanlışlıkla seçilme riskini azaltabilmektedir. Ortalama ile yanlış bir seçim yapma riski azalma eğilimindedir.

2.Büyük Hacimli Veriler: Analiz edilmesi gereken veri, tek bir sınıflandırıcı tarafından işlenemeyecek kadar büyük olabilmektedir. Bu durumlarda, topluluk öğrenmesi yöntemlerini kullanmak faydalı olabilmektedir.

3.Çok Az Veri: Topluluk sistemleri, yeterli eğitim verilerinin mevcut olmadığı durumlarda, yeniden örnekleme teknikleri kullanılarak her biri farklı bir sınıflandırıcıyı eğitmek ve topluluğu oluşturmak amacıyla kullanılabilirler.

4.Veriler Birleştirme: Farklı kaynaklardan gelen verilerin birleştirildiği uygulamalara veri birleştirme uygulamaları denilmektedir. Farklı doğaya sahip özniteliklerin bulunduğu ve çeşitli kaynaklardan elde edilen birden fazla veri setinin olduğu durumlarda, tüm verilerin içerdiği bilgiyi öğrenmek için tek bir sınıflandırıcı yetersiz kalabilmektedir. Örneğin, nörolojik bir bozukluğu teşhis ederken, nörolog, MRI taraması, EEG kaydı, kan testleri vb. gibi birkaç test isteyebilir. Her test, toplu olarak eğitim için kullanılamayacak farklı sayıda ve özellikte veri üretmektedir. Bu tür durumlarda, her bir test yönteminden elde edilen veriler, çıktılarını birleştirebilmek için daha sonra eğitilebilen farklı bir sınıflandırıcı kullanılmak üzere kullanılabilir.

Bagging ve boosting en çok kullanılan topluluk öğrenme yöntemleri olarak gösterilmektedir (Malek vd., 2022). En çok kullanılan nedenleri, sınıflandırma doğruluk oranlarını iyileştirmede iyi bir performans göstermeleridir. Bu teknikler, eğitim veri kümesini manipüle ederek çeşitli sınıflandırıcı toplulukları oluşturmakta ve ardından yeni verileri bir oylama mekanizması yardımı ile sınıflandırmaktadırlar (Tuysuzoglu ve Birant, 2020). Şekil 13'te bu yöntemler gösterilmektedir.



Şekil 13: Topluluk öğrenme yöntemleri

Şekil 13'te görüldüğü üzere topluluk öğrenme yöntemleri; bagging, boosting ve stacking olmak üzere üç gruba ayrılmaktadır. Topluluk öğrenme yöntemleri, veri madenciliği alanında önemli bir rol oynamakta ve çeşitli uygulama alanlarında başarılı sonuçlar elde etmektedirler. Bagging yöntemi içerisinde Rassal Orman Yöntemi, boosting yöntemleri içerisinde ise AdaBoost önemli bir yere sahiptir (Ngo vd., 2022). Diğer boosting yöntemleri ise XGBoost ve CatBoost yöntemleridir. Çalışmanın devam eden bölümlerinde, uygulama aşamasında kullanılacak olan bagging ve boosting yöntemleri hakkında ayrıntılı bilgilere yer almaktadır.

Bagging yöntemi, makine öğrenimi ve veri madenciliği alanında kullanılan bir tekniktir. Bagging, "bootstrap aggregating" kelimesinin kısaltmasından oluşmaktadır (Breiman, 1996). Bu yöntem, eğitim veri setinden rastgele örnekleme yaparak birden çok örnek alt kümesi oluşturmaktadır. Her bir örnek alt kümesi, temel modellerin eğitimi için kullanılmaktadır (Dong vd., 2020). Bootstrapping, istatistiksel bir tekniktir. Bu yöntemde örnekleme, gözlemlenen değerlerin bir kümesinden tekrar tekrar rastgele seçimlerle yapılmaktadır. Bu yöntem, örneklem tahminlerinin doğruluğunu ve güvenilirliğini değerlendirmek için kullanılmaktadır. Özellikle küçük örneklemeler veya dağılımı bilinmeyen veya normal dağılım olmayan örneklemeler için önerilmektedir (Egbert ve Plonsky, 2020). Bagging, bootstrapping yaklaşımını kullanarak eğitim verilerinin rastgele alt kümelerini oluşturmaktadır. Yöntem, temel bir öğrenme algoritmasını birden çok rastgele veri alt kümesi üzerinde eğiterek çalışır. Ardından, bu alt modellerin

tahminlerini birleştirir, genellikle bu tahminlerin ortalaması ile bir nihai tahmin oluşturur (Scikit-learn, 2023).

Karar ağaçları, veri madenciliğinde kullanılan güçlü bir yöntemdir. Karar ağaçları, verileri belirli özellikler ve karar kuralları temelinde sınıflandırarak veya tahminler yaparak veri setini hiyerarşik bir yapıya dönüştürmektedir (Rajan ve Krishnan, 2022). Bu hiyerarşik yapı, veri setinin anlaşılması, keşfedilmesi ve yorumlanması için değerli bir çerçeve sağlamaktadır. Bunun yanı sıra, karar ağaçlarının basit ve açık bir yapıya sahip olması, sonuçları kolayca yorumlanabilir ve anlaşılabilir kılmaktadır. Bu özellikleri sayesinde, veri madenciliği süreçlerinde bilgi keşfi, sınıflandırma, regresyon ve öznitelik seçimi gibi birçok uygulama alanında yaygın şekilde kullanılmaktadır.

Karar ağaçları, girdi verilerin sınıflandırmaya veya girdilerin çıktı değerlerini tahmin etmeye yarayan yapılardır. Daha çok sınıflandırma problemlerinde kullanılmaktadır (Enayati vd., 2022). Akış şemasına benzer yapılarının olması, görselleştirme ve yorumlanmalarını kolaylaştırmaktadır (Francois ve JJ, 2017). Gerçek ağaçlarda olduğu gibi karar ağaçları kök, dal, yapraklar, kök düğüm, ara düğüm ve yaprak düğümlerden oluşmaktadır (Kavuncuoğlu vd., 2022). Bir ağaç yapısında, yaprak düğümler dışındaki her düğüm (kök ve iç düğümler), belirli bir açıklayıcı özelliğe yönelik bir testi temsil etmektedir. Bir tanımlayıcı özelliğin alabileceği seviyelerin sayısı, yaprak düğümlerden aşağı doğru dalların sayısını belirlemektedir. Her yaprak düğümü, hedef özelliğin tahmini bir seviyesini ifade etmektedir (Kelleher et al., 2015). Şekil 14'te karar ağaçları yapısı gösterilmektedir.



Şekil 14: Karar ağaçları (medium.com)

Karar ağaçlarının iki amacı bulunmaktadır. Birincisi veri gruplarının olabildiği kadar homojen sınıflandırılmasını sağlamaktır. Bu durumda oluşan ağaçlar, Sınıflandırma Ağaçları olarak isimlendirilmektedir. İkinci amacı ise tahmin modellerinin kurulmasını sağlamaktır. Bu durumda oluşan ağaçlar ise Regresyon Ağaçları olarak isimlendirilmektedir. Karar ağaçları algoritmaları arasında farklılık gösteren çeşitli özellikler bulunmaktadır (Akpınar, 2014):

- Kullanılan ölçü skalası
- Her düğümden ortaya çıkan yeni düğüm sayıları
- Ağacın büyümesini durdurma kriteri
- En iyi bölen özelliğin seçilmesi
- Budama süreci

Boosting, Schapire (1990) tarafından "Strength of Weak Learning" adlı makalede tanıtılmıştır. Boosting yönteminin hedefi, rastgele tahminlerden daha iyi sonuçlar elde edebilen zayıf sınıflandırıcıların bir topluluğundan güçlü bir sınıflandırıcı oluşturmaktır (Zheng vd., 2018). Bu bölümde yöntemin avantajları ve dezavantajları, çalışma şekli ve önemli parametreleri ayrı başlıklar altında yer almaktadır. Ayrıca, çalışmanın yöntemlerinden olan XGBoost ve CatBoost yöntemleri hakkında detaylı bilgiler bu bölümde yer almaktadır.

Boosting, bagging gibi genel bir yaklaşımdır ve regresyon veya sınıflandırma gibi çeşitli problemlerde kullanılabilir (Dev ve Eden, 2019). Bu yöntemde, her bir zayıf sınıflandırıcı, eğitim verilerinin üzerinde çalışarak bir tahmin yapmakta ve ardından bu tahminin hatalı olan örneklerine daha fazla ağırlık verilerek tekrar eğitilmektedir (Zhang ve Zhang, 2008). Bu şekilde, zayıf sınıflandırıcılar, önceki hatalarına odaklanarak daha iyi sonuçlar üretmeyi öğrenmektedir. Boosting yöntemleri, bu aşamaları tekrarlayarak, düşük eğitim hatası elde etmek için zayıf sınıflandırıcıların bir araya getirilmesini sağlamaktadır (Skurichina ve Duin, 2002). Böylece, başlangıçta zayıf olan sınıflandırıcılar güçlendirilmekte ve daha sonra birleştirildiklerinde daha yüksek bir sınıflandırma performansı sergileyen bir güçlü sınıflandırıcı ortaya çıkmaktadır.

Boosting, zayıf sınıflandırıcılardan oluşan bir topluluğu birleştirirken, bagging yönteminden önemli ölçüde farklılık göstermektedir. Bagging yönteminde, eğitim için seçilen örnekler, eğitim verilerinin önyüklemeli kopyalarından oluşmaktadır. Bu durumda her bir örneğin her eğitim veri

kümesinde eşit bir şansa sahip olduğu kabul edilmektedir. Boosting yöntemi, zayıf sınıflandırıcıların hatalarını düzelterek ve basit çoğunluk oylamasını kullanarak güçlü bir sınıflandırıcı oluşturmaktadır (Polikar, 2012). Bu yaklaşım, bagging yönteminden ayrılarak, eğitim veri kümesindeki yanlış sınıflandırılan örneklerin üzerinde yoğunlaşmakta ve bu şekilde daha iyi bir sınıflandırma performansı elde edilmesini sağlamaktadır. Dağılım seçimi, zayıf öğrencilerin yanlış sınıflandırdığı örneklerin daha fazla ağırlığa sahip olmasını sağlayarak bu hatalı sınıflandırmaların düzeltilmesine odaklanmaktadır. Böylelikle, boosting yöntemi, öğrenciyi daha zor örnekler üzerinde daha fazla çalışmaya yönlendirmekte ve daha keskin bir karar sınırlayıcısı oluşturmaktadır (Schapire, 2003).

AdaBoost ve Gradient Boosting Machines (GBMs) en çok tercih edilen boosting yöntemleridir. AdaBoost ve GBM'ler (Gradient Boosting Machines), küçük veri kümeleri üzerinde etkili sonuçlar elde etmede başarılı olabilirler. Ancak, daha büyük veri kümeleriyle çalışırken, ölçeklenebilir algoritmalara ihtiyaç duyulmaktadır. Bu tür veri kümelerinde performans ve hesaplama süreleri önemli hale gelmektedir. Bu bağlamda, XGBoost, LightGBM ve CatBoost gibi algoritmalar, büyük ölçekli veri kümeleri için tasarlanmıştır ve ölçeklenebilirlik gereksinimini karşılamak amacıyla geliştirilmiştir (Dev ve Eden, 2019). Çalışmanın uygulama bölümünde, XGBoost ve CatBoost gibi algoritmalarından yararlanılacaktır. Bu iki yöntem hakkındaki bilgiler, çalışmanın devam eden bölümlerinde detaylı bir şekilde yer almaktadır.

Bu bölümde veri madenciliği için önemli bir yere sahip olan topluluk öğrenme yöntemleri hakkında bilgilere yer verilmiştir. Devam eden bölümlerde, uygulama bölümünde kullanılan Rassal Orman, XGBoost ve CatBoost yöntemleri hakkında ayrıntılı bilgiler yer almaktadır.

4.2.3. Rassal Orman Yöntemi

Bu bölümde Rassal orman yöntemi hakkında detaylı bilgiler yer almaktadır. Bu çerçevede yöntemin çalışma şekli, avantaj ve dezavantajları hakkında ayrıntılı bilgiler yer almaktadır.

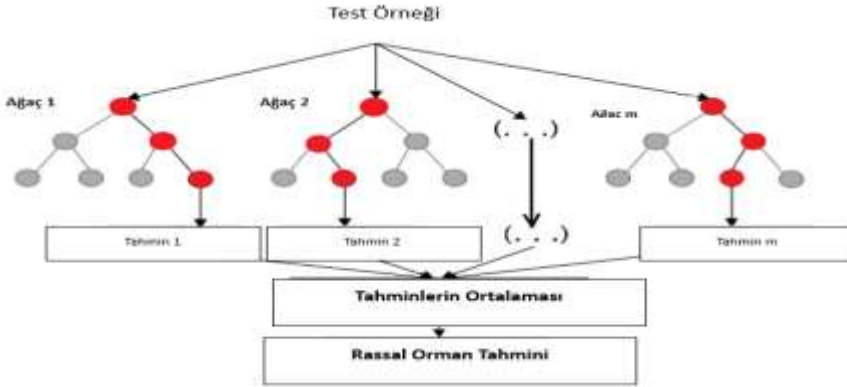
Tek bir karar ağacı, desenleri tespit etmek ve sınıflandırma veya tahmin yapmak için yetersiz kalabilmekte ve aşırı uyuma eğilimi gösterebilmektedir (Chatpatanasiri, 2005). Rassal orman yöntemi ise birden fazla karar ağacını birleştirmektedir (Buschjager ve Morik, 2018). Rassal ormanlar, her bir ağacın

rastgele bir vektörün değerlerine dayandığı ve tüm ağaçların aynı dağılıma sahip olduğu varsayımıyla oluşturulan bir yöntem olarak tanımlanmaktadır (Breiman, 2001). Yöntem, rastgele alt kümelerde eğitilen ağaçların, tahminlerini birleştirilerek daha güçlü ve istikrarlı sonuçlar elde etmeyi amaçlanmaktadır (Bernard vd., 2009). Ayrıca, Rassel Orman yöntemi, aşırı öğrenme (overfitting) eğilimine sahip olan karar ağaçlarında bu sorunun giderilmesi için etkili bir yöntem olarak görülmektedir.

Bu yöntemde, birden çok karar ağacı modeli, örnek ve özellik boyutlarından oluşan iki perspektiften eğitilmektedir (Dong et al., 2020). Rassel orman algoritmasının literatürde önem görmesini başlıca nedenleri olarak şunlar sıralanabilir (Zhang ve Ma, 2012):

- Hem sınıflandırma hem de regresyon modellerinde kullanılabilir olması,
- Hızlı eğitim ve tahmin gücüne sahip olması,
- Sadece bir iki ayar parametresine bağlı olması,
- Yerleşik bir genelleme hatası tahminine sahip olması,
- Yüksek boyutlu problemlere uygulanabilir olması.

Ayrıca, değişkenleri önem derecelerine göre sıralaması, eksik değerler yerine değer ataması ve görselleştirme kolaylığı tercih edilmesindeki bir diğer nedenlerdir. Şekil 15'te yöntemin çalışma şekli gösterilmektedir.



Şekil 15: Rassel orman yöntemi

Şekil 15'e göre Rassel orman yöntemi, çalışma adımları şu şekildedir (Liaw ve Wiener, 2002):

1. Orijinalden ntree önyükleme örnekleri çizilir.

2. Her önyükleme örneği için, kesilmemiş bir sınıflandırma veya regresyon ağacı oluşturulur. Her bir düğümde, tüm tahmin ediciler arasında en üstün bölümü seçmek yerine, tahmin edicilerin rastgele örnekleri alınır ve bu örneklerle arasından en iyi bölüm seçilir.

3. Ağaçların tahminlerinin ortalamasını (regresyon için) alarak yeni verileri tahmin edilir.

Bu bölümde, Rassal Orman yöntemi hakkında genel bilgilere yer verilmiştir. Yöntemin avantajları, dezavantajları, çalışma şekli ve önemli parametreleri alt başlıklar halinde devam eden bölümlerde yer almaktadır.

Rassal Orman yöntemi, karar ağaçlarının bir araya getirilmesi ile oluşturulan bir topluluk öğrenme yöntemidir. Yöntem, çeşitli uygulama alanlarında başarılı sonuçlar elde etmek için sıklıkla tercih edilmekte ve birçok avantajı bulunmaktadır (Vens ve Costa, 2011):

1.Hedef bilgisini dikkate alarak göreve özgü bir benzerlik kavramı oluşturmaktadır. Bu sayede daha spesifik ve ihtiyaçlara yönelik sonuçlar elde edilebilmektedir.

2. Rassal ormanlar, önyüklemeli bir topluluk yöntemi olduğu için verideki küçük değişikliklere karşı dirençlidir. Bu, veri setindeki gürültü vs. problemlere karşı güvenilir sonuçlar elde etmeyi sağlamaktadır.

3.Rassal ormanlar, özellik seçimi mekanizması ile verinin temsil gücünü artırmaktadır. Bu mekanizma sayesinde bilgisiz veya gürültülü özelliklerin etkisi azalmakta ve daha bilgilendirici özelliklere odaklanılmaktadır.

4.Rassal ormanlar, eksik değerlerle başa çıkmak için doğal bir yöntem sunmaktadır. Eksik ayırım değeri olan örnekler tüm dallara dağıtılmakta veya rastgele bir dal takip edilmektedir. Bu şekilde eksik veri durumlarına uygun bir işleme yapılmakta ve modelin performansı etkilenmemektedir.

Rassal Orman yöntemi, birçok avantajının yanı sıra bazı dezavantajlar da içermektedir. Rassal Orman yönteminin başlıca dezavantajları şunlardır (Prajwala, 2015):

1. Rassal Orman yöntemi, ağaçların her bir düğümünde, veri setinin rastgele bir örneği üzerinde özellik seçmektedir. Bu seçim sırasında, kategorik değişkenlerin her bir seviyesi ayrı ayrı ele alınmaktadır. Eğer bir kategorik değişken birden fazla seviyeye sahipse, daha fazla seviyeye sahip olanlar, modelin tahminlerinde daha fazla ağırlığa sahip olabilmektedir.

2. Rassal Orman yöntemi, özellik seçimi sırasında her bir düğümde rastgele bir özellik alt kümesi seçmektedir. Eğer veri setinizde birbirine yakın ilişkili özellik grupları varsa, yöntem daha küçük grupları daha büyük gruplara tercih edebilmektedir. Bu durumda özellikler arasındaki ilişkiler göz ardı edebilmektedir. Sonuç olarak, model özellikler arasındaki gerçek ilişkileri tam olarak yakalamakta zorlanabilmektedir.

Rassal Orman regresyon algoritması süreçleri şu şekildedir (Gao ve Liu, 2021):

"Orjinal veri kümesi içerisinde bulunan n vakasından, Bootstrap yöntemi kullanılarak b adet eğitim örnek kümesi tekrar tekrar çıkarılır ve her biri için b regresyon ağacı oluşturulur. Her bir eğitim örneği kümesi alındığında, bu örnekle seçilmeyen vakalardan oluşan b örnekle, "out-of-bag" (OOB) verisi olarak adlandırılır ve bu veri kümesi daha sonra test örneği kümesi olarak kullanılır.

Regresyon ağacı oluşturulurken, her bir ağacın alt düğümünde, k bağımsız değişken arasından rastgele m_{try} ($m_{try} < k$) sayısında bağımsız değişken seçilir. Bu seçilen değişkenler, alt düğümlerde dallanma değişkenleri olarak aday olarak kullanılır. Daha sonra, dallanma kalitesini değerlendiren bir kriter kullanılarak en iyi dallanma seçilir.

Her regresyon ağacı, baştan başlayarak sürekli olarak dallanır ve büyüme sürecini sürdürür. Bu büyüme sürecini sonlandırmak için ağaç sayısı, yani n_{tree} değeri, bir sonlandırma koşulu olarak belirlenir.

Oluşturulan b regresyon ağacı, bir rastgele orman regresyon modelini oluşturur ve modelin tahmin yeteneği, "out-of-bag" veri (OOB) tahminlerinin doğruluğu ile değerlendirilir; yani test kümesinin ortalama karesel hatası (MSE) ile değerlendirilir.

Rassal orman regresyon yönteminin özelliği şu şekildedir (Gao ve Liu, 2021):

Regresyon ağaçlarının sayısı sonsuza yaklaştığında,

$$E_{X,Y}(Y - av_j h(X, \theta_j))^2 \rightarrow E_{X,Y}(Y - E_\theta(X, \theta))^2$$

av ortalamayı ifade etmektedir. Bu nedenle, j yeterince büyük olduğunda $Y - [av]_j h(X, [\theta]_j)$ formülü, rastgele orman regresyon işlevi $E_\theta(X, \theta)$ 'a yaklaştırmak için kullanılmaktadır. Bu sırada, PE^* , rastgele orman modelinin genelleme hatasını temsil etmek için kullanılır ve bir regresyon ağacının ortalama genelleme hatası şöyledir (Gao ve Liu, 2021):

$$PE^*(tree) = E_{\theta} E_{X,Y} (Y - h(X - \theta))^2$$

Eğer tüm θ değerleri için $E(Y) = E[Xh(X, \theta)]$ geçerliyse, o zaman şu durum mevcuttur (Gao ve Liu, 2021):

$$\rho \leq \rho^*(tree)$$

Formülde, ρ , $Y - h(X, \theta)$ ve $Y - h(X, \theta)$ arasındaki korelasyon katsayısıdır ve θ ile θ birbirinden bağımsızdır. Bu durum, rastgele orman modelinin genelleme hatasının regresyon ağacının genelleme hatasından daha düşük olduğunu göstermektedir.

Algoritma, belirli bir görevi gerçekleştirmek için adım adım talimatlar içeren, açık ve düzenli bir yöntem olarak tanımlanabilmektedir. Tablo 11’de Rassal Orman Algoritması yer almaktadır (Vens ve Costa, 2011).

Tablo 11: Rassal orman algoritması

Rassal Orman Algoritması	
1:	for $k \leftarrow 1$ to M do
2:	$D_k \leftarrow \text{Bootstrap}(D)$
3:	$h_k \leftarrow \text{RandomTree}(D_k, \sqrt{f})$
4:	$\text{Forest} \leftarrow \cup h_k$
5:	for each $h_k \in \text{Forest}$ do
6:	for each $x_i \in D$ do
7:	$P_{hk}(x_i) \leftarrow \text{NodesOnPath}(x_i, h_k)$
8:	for each $n \in P_{hk}(x_i)$ do
9:	$\text{node_id} \leftarrow \text{GetNodeID}(n)$
10:	$\text{feature_id} \leftarrow \text{Hash}(\text{node_id}, F)$
11:	$x'_i.\text{feature_id} \leftarrow x'_i.\text{feature_id} + 1$
12:	return $D' \leftarrow \cup x'_i$

Rassal Orman algoritmaları genellikle ilk çalıştırıldıklarında yüksek doğruluk oranı ile çalışmaktadır. Ancak bazen bu doğruluk oranı yeterli olmamaktadır. Makul bir doğruluk oranı sağlamak için bazı ayarlama (tuning) işlemleri yapılmaktadır. Rassal Ormanların, daha iyi doğrulukta çalışabilmesi için şu üç parametrenin iyi ayarlanması gerekmektedir (Cutler vd., 2012):

1. m : her bir düğümde seçilen rastgele seçilmiş tahmin değişkenlerinin sayısıdır. Regresyon için varsayılan m değeri örneklem büyüklüğünün 3’te 1’i kadardır ($N/3$).

2. J: ormanda bulunan ağaç sayısını temsil etmektedir. Genel olarak, ağaç sayısı arttıkça, Rassal Ormanlar için genelleme hatası azalmaktadır.

3. Ağaç boyutu: Bölme için en küçük düğüm boyutu veya maksimum uç düğüm sayısı ile ölçülmektedir. Breiman (2001) çalışmasında çok büyük ağaçların yetiştirilmesini önermektedir.

Bu parametrelerin iyi ayarlanması, Rassal Ormanların daha iyi doğrulukta çalışmasını sağlayabilir ve istenmeyen sorunları önleyebilir. Bu nedenle, farklı parametre değerlerini deneyerek modelin performansını değerlendirmek önemlidir. Tablo 12’de Rassal Orman Regresyon yöntemine ait önemli parametreler gösterilmektedir (Sklearn, 2023).

Tablo 12: Rassal orman regresyon model parametreleri

Parametre	Tanım
<i>n_estimators</i>	Modelde yer alacak ağaç sayısını ifade etmektedir.
<i>max_depth</i>	Ağacın maksimum derinliğini ifade etmektedir.
<i>min_samples_split</i>	Bir düğümün bölünmesi için gereken minimum örnek sayısını ifade etmektedir.
<i>min_samples_leaf</i>	Bir yaprak düğümünde olması gereken minimum örnek sayısını ifade etmektedir.
<i>max_features</i>	Her ağaçta bölünme için kullanılacak olan maksimum özellik sayısını ifade etmektedir.
<i>random_state</i>	Modelin her çalıştırılmada aynı şekilde davranmasını sağlayan parametredir.

Rassal Orman yönteminin önemli parametrelerinin ayarlanması, modelin performansını, hızını, genelleme yeteneklerini etkilemektedir. Bu nedenle, bu önemli parametrelerin uygun bir şekilde ayarlanması, daha iyi sonuçlar elde etmeyi sağlamaktadır.

Bu bölümde, Rassal Orman yöntemi ayrıntılı bir şekilde ele alınmıştır. Yöntem hakkında temel bilgiler, yöntemin avantajları ve dezavantajlar, yöntemin algoritma ve parametreleri detaylı bir şekilde açıklanmıştır. Sonraki bölümde, XGBoost yöntemi hakkında bilgiler yer almaktadır.

4.2.4. XGBoost Yöntemi

Bu bölümde, boosting yöntemlerinden olan XGBoost yöntemi hakkında detaylı bilgiler yer almaktadır. Bu çerçevede, yöntem hakkında genel bilgiler, avantaj ve dezavantajları, çalışma şekli ve önemli parametreleri ayrı başlıklar altında yer almaktadır.

XGBoost, "extreme gradient boosting" (aşırı gradyan artırma) teriminin kısaltması olup, gradyan artırma karar ağacı algoritmasını uygulamaktadır (Li ve Zhang, 2020). Bu isim, her yeni modelin eklenmesinden sonra kaybı minimize etmek için gradyan iniş algoritmasının kullanılmasından gelmektedir (Sarkar ve Natarajan, 2019). XGBoost, yüksek performanslı karar ağacı modellerinin etkisini artırmak amacıyla ardışık olarak birden fazla karar ağacı oluşturmakta ve her bir karar ağacı, bir önceki ağacın hatalarını en aza indirmeye odaklanarak gelişmektedir. Bu amaçla, gradyan iniş optimizasyon algoritması kullanılmaktadır (G. Kumar vd., 2022). Gradyan iniş algoritması, her aşamada modelin hatasını azaltmak için gradyanları takip ederek optimal yöne ilerlemektedir (Elavarasan ve Vincent, 2020). Böylece, her bir ağaç hataları düzeltirken, genel performansı iyileştiren bir topluluk modeli oluşturulmaktadır. XGBoost'un bu yaklaşımı, karar ağaçlarından elde edilen tahminleri birleştirerek daha güçlü ve genelleştirilebilir bir model elde etmeyi amaçlamaktadır.

XGBoost, karar ağacı temelli bir topluluk öğrenme algoritması olarak, optimize edilmiş öğrenme stratejileri ve gradyan boosting yöntemiyle karmaşık veri analiz sorunlarını çözmek için etkili bir araç sağlamaktadır. XGBoost, kayıp fonksiyonunun ikinci derece gradyanlarını hesaplayarak kaybı minimize etmek ve gelişmiş düzenleme teknikleri (L1 ve L2) uygulamak suretiyle çalışmaktadır. Bu yaklaşım, aşırı öğrenmeyi azaltmakta, modelin genelleme yeteneğini artırmakta ve performansını iyileştirmektedir (Li vd., 2022). XGBoost'ta aşırı öğrenmeyi önlemek için kayıp fonksiyonu yer almaktadır (Li ve Chen, 2020b):

$$\mathcal{L}_K(F(x_i)) = \sum_{i=1}^n \Psi(y_i, F_K(x_i)) + \sum_{k=1}^K \Omega(f_k),,$$

$F(x_i)$, K'nci artırma adımında i'nci örneğin tahminini temsil etmektedir. Ψ , gerçek tahmin ile gerçek etiket arasındaki farkları ölçen kayıp fonksiyonudur. $\Omega(f_k)$, düzenleme terimidir ve aşağıdaki gibi ifade edilebilir (Kim vd., 2022):

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

Düzenleme teriminde, γ karmaşıklık parametresini temsil eder ve yaprakların karmaşıklığını göstermektedir. T yaprak sayısını belirtirken, λ ceza parametresini göstermektedir. $\|\omega\|^2$ ise her yaprak düğümünün çıktısını ifade

etmektedir. XGBoost ikinci dereceden Taylor serisini amaç fonksiyonu olarak benimsemektedir (Li ve Chen, 2020b):

$$\mathcal{L}_K \cong \sum_{i=1}^n [g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)] + \Omega(f_k)$$

Burada g_i ve h_i sırasıyla kayıp fonksiyonu üzerindeki birinci ve ikinci dereceden gradyan istatistiklerini temsil etmektedir. Ayrıca, XGBoost hızlı bir şekilde yorumlanabilmekte ve büyük boyutlu veri kümeleriyle etkili bir şekilde başa çıkabilmektedir. Bu özellikleri sayesinde XGBoost, karmaşık veri analiz problemlerini çözmek için yaygın olarak kullanılan bir yöntem olarak literatürde yer almaktadır (Sarker, 2021). XGBoost'un başarısının önemli faktörleri şu şekilde gösterilmektedir (Chen ve Guestrin, 2016):

1. Ölçeklenebilirlik: XGBoost, her senaryoda ölçeklenebilirlik sağlamaktadır. Tek bir makinede mevcut popüler çözümlere göre on kat daha hızlı çalışmakta ve dağıtılmış veya bellek sınırlı ortamlarda milyarlarca örneğe ölçeklenebilmektedir.

2. Sistem ve Algoritmik Optimizasyonlar: XGBoost'un ölçeklenebilirliği, birkaç önemli sistem ve algoritmik optimizasyona dayanmaktadır. Bu optimizasyonlar, seyrek verileri işlemek için yeni bir ağaç öğrenme algoritması ve yaklaşık ağaç öğrenmesinde örnek ağırlıklarını işlemek için teorik olarak geçerli bir ağırlıklı kantil özeti prosedürü gibi yenilikleri içermektedir.

3. Paralel ve Dağıtılmış Hesaplama: Paralel ve dağıtılmış hesaplama, öğrenmeyi hızlandırmakta ve daha hızlı model keşfi sağlamaktadır. XGBoost, bu teknikleri kullanarak out-of-core hesaplama yapmakta ve veri bilimcilerin bir masaüstünde yüz milyonlarca örneği işlemesine olanak tanımaktadır.

4. En Az Miktarda Küme Kaynağıyla Daha Büyük Veriye Ölçeklenebilme: XGBoost'un bu teknikleri birleştirerek daha büyük verilere en az miktarda küme kaynağıyla ölçeklenebilme yeteneği bulunmaktadır.

Özetle, XGBoost'un önemi, ölçeklenebilirlik sağlama, sistem ve algoritmik optimizasyonlar, paralel ve dağıtılmış hesaplama kullanma ve daha büyük verilere en az kaynakla ölçeklenebilme yeteneği gibi faktörlerden kaynaklanmaktadır. XGBoost yöntemi hız, kolay kullanım, özelleştirilebilirlik ve regresyon ile sınıflandırma tahminleme problemlerini destekleyen yetenekleri, araştırma ve endüstride geniş kabul görmesine sebep olan faktörler olarak gösterilebilmektedir. Bu nedenle, XGBoost, veri analizinde önemli bir

araç olarak kabul edilmekte ve yaygın olarak tercih edilen bir yöntemdir (Terko vd., 2019).

Her yöntemin kendine özgü avantajları ve dezavantajları bulunmaktadır. Yöntemin avantajlarından tam olarak yararlanmak için parametrelerin doğru şekilde yapılandırılması ve veri setinin özelliklerine uygun bir şekilde kullanılması gerekmektedir. XGBoost, genellikle yüksek performans, doğruluk ve esneklik gibi özellikleriyle dikkat çeken bir algoritmadır. Yüksek performansı sayesinde geniş veri kümelerinde etkili sonuçlar elde edilebilmektedir. Doğruluk açısından da XGBoost, birçok probleme karşı başarılı sonuçlar sunmaktadır. XGBoost yönteminin avantajları şu şekilde özetlenebilir (Kumar, 2019):

1. Düzenleştirme (Regularization): XGBoost algoritması, modelin aşırı uyuma eğilimini azaltmak için dahili olarak L1 (Lasso Regresyonu) ve L2 (Ridge Regresyonu) düzenliliği kullanmaktadır. Bu özellik sayesinde XGBoost, GBM'ye göre daha düzenli bir yapıya sahiptir.

2. Paralel İşleme: XGBoost algoritması, paralel işleme yeteneklerini kullanarak GBM'ye göre önemli ölçüde daha hızlı çalışmaktadır. XGBoost, modelin yürütülmesi için birden fazla CPU çekirdeğini etkin bir şekilde kullanmaktadır. Paralel işleme yaklaşımı, hesaplama yükünü çekirdekler arasında dengeli bir şekilde dağıtarak verimliliği artırmakta ve modelin daha hızlı sonuçlar üretmesini sağlamaktadır. Bu sayede büyük veri kümeleri üzerindeki işlemler daha verimli bir şekilde gerçekleştirilebilmektedir.

3. Eksik Değerleri İşleme: XGBoost, eksik verilerin işlenmesi için yerleşik bir yeteneğe sahiptir. Bir düğümde eksik bir değerle karşılaşıldığında, XGBoost hem sol hem de sağ alt düğüm bölme seçeneklerini değerlendirmekte ve her bir düğüm için daha yüksek bir kayba yol açan ayrımı öğrenmektedir. Bu yaklaşım, eksik değerlerin doğru şekilde işlenmesini sağlamak ve modelin veri setindeki eksik bilgilerle etkili bir şekilde başa çıkabilmesini sağlamaktadır. Aynı yöntem, test verileri üzerinde çalışırken de uygulanarak tutarlı sonuçlar üretilmektedir. Bu özellik, gerçek dünya veri setlerinde eksik verilerin yaygın olduğu durumlarda XGBoost algoritmasının güçlü bir yanını oluşturmaktadır.

4. Çapraz Doğrulama: XGBoost algoritması, boosting sürecinin her bir iterasyonunda çapraz doğrulama çalıştırılmasına olanak sağlamak ve bu sayede tek bir çalışma içerisinde tam optimum boosting iterasyon sayısını elde

etmek kolaylaşmaktadır. Bu özellik, GBM'den farklı olarak ızgara arama yapma zorunluluğunu ortadan kaldırmakta ve sınırlı değerlerin test edilebildiği kısıtlamalara tabi olmayı engellemektedir. XGBoost, kullanıcının modelin performansını en üst düzeye çıkarmak için en uygun iterasyon sayısını belirlemesine yardımcı olan esnek bir araç sağlamaktadır. Bu yöntem, modelin güvenilirliğini artırırken aynı zamanda hesaplama kaynaklarının daha verimli bir şekilde kullanılmasına olanak tanımaktadır.

5. Etkili Ağaç Budama: GBM algoritması, bir bölünme aşamasında negatif bir kayıpla karşılaştığında bir düğümü bölmeyi durdurma eğilimindedir. XGBoost algoritması ise belirtilen maksimum derinliğe kadar bölmeler yapar ve ardından ağacı geriye doğru budamaya başlar, pozitif bir kazanç sağlanmayan bölmeleri kaldırır. Bu yöntem, ağacın gereksiz dallarının ortadan kaldırılmasını sağlayarak, modelin daha sade ve genelleyici bir yapıya sahip olmasını sağlamaktadır. Bu sayede aşırı öğrenme riski azaltılırken, modelin daha iyi bir performans sergilemesi hedeflenmektedir. XGBoost, ağaç yapısının etkili bir şekilde budanmasını sağlayarak, veri setlerindeki gürültüyü en aza indirmekte ve daha güvenilir sonuçlar üretmektedir.

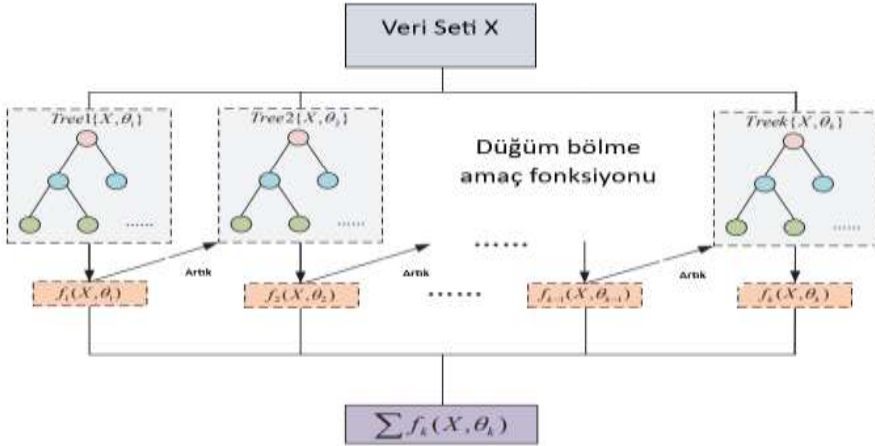
XGBoost yönteminin bu avantajlarına karşılık dezavantajları da bulunmaktadır (GITHUB, 2020):

1. Aşırı Öğrenme (Overfitting): XGBoost algoritması, parametreleri uygun şekilde ayarlanmazsa aşırı öğrenme riski taşımaktadır. Bu durumda model, eğitim verilerine aşırı derecede uyarlanabilmekte ve genelleme performansı düşebilmektedir. Bu nedenle, XGBoost algoritmasını kullanırken parametrelerin doğru bir şekilde ayarlanması önemli bir faktördür.

2. Eğitim Süresi: XGBoost, büyük veri kümeleri için eğitim süresi oldukça yüksek olabilmektedir. Bu, özellikle çok büyük ve karmaşık veri setleri üzerinde çalışırken dikkate alınması gereken bir dezavantaj olarak görülmektedir.

XGBoost, eğitim süreci boyunca sürekli olarak modelin düğüm kaybını hesaplamakta ve en büyük kazanç kaybına sahip yaprak düğümünü seçmektedir. Daha sonra, özellikleri sürekli olarak bölerek yeni ağaçlar eklemektedir. Her bir ağaç eklendiğinde aslında bir önceki tahminin artıklarını uyum sağlamak için yeni bir fonksiyon $f_k(X, \theta_k)$ öğrenilmektedir. Eğitim sonrasında K ağaç elde edildiğinde, tahmin örneklerinin özellikleri her bir ağaçta karşılık gelen bir yaprak düğümüne sahip olmakta ve her bir yaprak

düğümü bir puanla eşleşmektedir. Son olarak, her bir ağacın karşılık gelen puanları toplanmaktadır. Şekil 16’da bu süreç gösterilmektedir.



Şekil 16: XGBoost yöntemi akış şeması (R. Guo vd., 2020)

XGBoost yöntemi parametreleri, modelin performansını ve davranışını kontrol etmek için kullanılmaktadır. XGBoost yönteminin birçok parametresi bulunmaktadır. Tablo 13’te XGBoost yöntemine ait önemli parametreler gösterilmektedir (Ashwini vd., 2022).

Tablo 13: XGBoost model parametreleri

Parametre	Tanım
<i>objective</i>	Hedef fonksiyonunu belirleyen parametredir.
<i>alpha</i>	Ağırlıklar üzerinde L1 düzenleme terimidir.
<i>learning_rate</i>	Öğrenme hızını kontrol etmektedir.
<i>n_estimators</i>	Modelde yer alacak ağaç sayısını ifade etmektedir.
<i>colsample_bytree</i>	Her bir ağacın inşası sırasında sütunların alt örnekleme oranıdır.
<i>max_depth</i>	Ağacın maksimum derinliğini ifade etmektedir.

XGBoost’un önemli parametrelerinin ayarlanması, modelin performansını, hızını, genelleme yeteneklerini etkilemektedir. Bu nedenle, bu önemli parametrelerin uygun bir şekilde ayarlanması, daha iyi sonuçlar elde etmeyi sağlamaktadır.

Bu bölümde, XGBoost yöntemi ayrıntılı bir şekilde ele alınmıştır. Yöntem hakkında temel bilgiler, yöntemin avantajları ve dezavantajlar, yöntemin algoritma ve parametreleri detaylı bir şekilde açıklanmıştır. Sonraki bölümde, Catboost yöntemi hakkında bilgiler yer almaktadır.

4.2.5. CatBoost Yöntemi

Bu bölümde, boosting yöntemlerinden olan CatBoost yöntemi hakkında detaylı bilgiler yer almaktadır. Bu çerçevede, yöntem hakkında genel bilgiler, avantaj ve dezavantajları, çalışma şekli ve önemli parametreleri ayrı başlıklar altında yer almaktadır.

CatBoost yöntemi, Prokhorenkova vd. (2018) tarafından tanıtılmıştır. CatBoost, karar ağacı gradyanını güçlendirmek amacıyla geliştirilmiş yüksek performanslı bir açık kaynak yazılımdır. Yandex araştırmacıları ve mühendisleri tarafından oluşturulan CatBoost, çeşitli özelliklere sahiptir. Temel özelliği, kategorik verileri ön işleme ihtiyacı duymadan kullanılabilme yeteneğine sahip olmasıdır (Hancock ve Khoshgofaar, 2020). Buna ek olarak, bu tür verileri tamsayılara dönüştürmek için kodlamaya gerek duymamaktadır (Wei vd., 2023). Bu özellikleri sayesinde CatBoost, kategorik verilerin işlenmesi ve kullanılması sürecinde kullanıcıların zaman ve çaba harcamalarını azaltmayı hedeflemektedir (Najm vd., 2023). CatBoost, her iterasyonda zayıf bir öğrenici üretmekte ve her bir öğrenici bir önceki öğrenicinin gradyanına göre eğitilmektedir. Ardından, tüm öğrenicilerin sınıflandırma sonuçları birleştirilerek nihai çıktı elde edilmektedir. Bu yöntem, gradyanı kullanarak hataları azaltmayı ve modelin performansını artırmayı hedeflemektedir. CatBoost, bu gradyan tabanlı eğitim sürecini benimseyerek etkili sonuçlar elde etme kapasitesine sahiptir (Zhang vd., 2020).

$$h^t = \arg_{h \in H} E\{L(y, F^{t-1}(x) + h(x))\}$$

Burada y çıktıyı, h ise H 'den seçilen bir gradyan adım fonksiyonunu temsil etmektedir. Adım fonksiyonu aşağıdaki gibi hesaplanabilmektedir (Chehreh Chelgani vd., 2023):

$$h(x) = \sum_{j=1}^J b_j \mathbb{1}_{\{x \in R_j\}}$$

R_j , ağacın yapraklarına karşılık gelen ayrık bölgeleri ifade etmektedir. b_j , bölgenin tahmin değerini temsil etmek ve l bir gösterge işlevidir.

CatBoost'un belirgin özelliklerinden biri, diğer makine öğrenme modellerine göre geniş eğitim verisi gerekliliklerinden bağımsız olarak etkili sonuçlar elde edebilme kabiliyetidir. Ayrıca, CatBoost çeşitli veri formatlarında çalışabilme özelliğine sahip olup, bu esnada yüksek bir istikrar sunmaktadır. Bununla birlikte, CatBoost toplu değişkenleri otomatik olarak işleyebilmekte ve tür dönüşüm hatalarından kaçınmaktadır. Bu özellik, modelin

iyileştirilmesine odaklanmayı ve çeşitli hataların düzeltilmesiyle uğraşmak yerine performansı artırmayı kolaylaştırmaktadır (Dasi vd., 2023). CatBoost'un diğer önemli bir özelliği, kategorik özelliklerle nasıl başa çıktığıdır. CatBoost'ta, her kategori için beklenen hedef değerini ölçen sayısal bir özellik, kategorik özellikleri yerine koymak için kullanılır. Eğitim verilerine aşırı uyum sağlama riskini azaltmak amacıyla, bu sayısal özelliğin ideal olarak farklı bir veri seti kullanılarak hesaplanması gerekmektedir. Ancak, bu ideal durum genellikle mümkün olmamaktadır. CatBoost, bu yeni sayısal özelliğin hesaplanması için model oluşturma sürecinde benzer bir yöntem önermektedir. Bu yaklaşım, kategorik özellikleri doğru bir şekilde işlemek ve modelin aşırı uyum sorunlarıyla karşılaşmadan verimli sonuçlar üretmesine yardımcı olmaktadır (Bentéjac vd., 2021).

Yöntemin avantajları, dezavantajları, çalışma şekli ve önemli parametreleri bu bölümde yer almaktadır.

CatBoost veri madenciliğinde önemli bir yöntem olarak kabul edilir. Kategorik değişkenlerin etkili işlenmesi, yüksek sınıflandırma doğruluğu, kolay hiperparametre ayarlaması ve aşırı öğrenmeyi azaltma yetenekleri, CatBoost'un birçok gerçek dünya uygulamasında başarılı sonuçlar elde etmesini sağlayan faktörler olarak gösterilmektedir. Yöntemin sunduğu avantajlarla birlikte, bazı dezavantajları da dikkate alınmalıdır. Bu bölümde bu avantaj ve dezavantajlar yer almaktadır.

Catboost yönteminin avantajlarını maddeler halinde şu şekilde sıralayabiliriz (Ke vd., 2021):

1. Otomatik Kategorik Değişken İşleme: CatBoost, kategorik değişkenleri otomatik olarak işleyebilme yeteneğine sahiptir. Bu, kategorik değişkenlerin one-hot encoding yöntemine göre daha etkili bir şekilde işlenebilmesini sağlamaktadır. Geleneksel one-hot encoding yöntemi ağaç derinliği sorununa yol açabilirken, CatBoost istatistiksel yöntemler kullanarak kategorik değişkenleri işlemektedir. Bu, modelin daha iyi sınıflandırma doğruluğu elde etmesini sağlamaktadır.

2. Hiperparametre Ayarlamasında Kolaylık: CatBoost, hiperparametre ayarlama ihtiyacını azaltmak için bazı otomatik hiperparametre ayarlama stratejileri kullanmaktadır. Bu stratejiler, modelin hiperparametrelerini otomatik olarak optimize etmekte ve kullanıcının manuel olarak ayar yapma

gereksinimini azaltmaktadır. Bu, kullanıcının zaman ve çaba harcamasını azaltırken, modelin performansını iyileştirmektedir.

3. Aşırı Öğrenmeyi Azaltma: CatBoost, aşırı öğrenmeyi etkili bir şekilde azaltmak ve modellerin sağlığını artırmak için rastgele çoklu permütasyonlar üretmektedir. Bu yöntem, yeterli sayıda permütasyon elde etmek amacıyla örneklemeler yapmaktadır. Bu permütasyonlar, her bir özelliğin değerlerini rastgele karıştırarak oluşturulmaktadır. Ardından, bu permütasyonlar kullanılarak her bir özelliğin önemi değerlendirilmekte ve aşırı öğrenmeyi azaltmak için modelin gelişimi optimize edilmektedir. Bu yaklaşım, CatBoost'un aşırı uyumu etkili bir şekilde kontrol etme ve modelin sağlığını artırma yeteneğini sağlamaktadır (Prokhorenkova vd., 2018).

Bu avantajları Catboost'un popüler bir yöntem olması sağlamaktadır. Yöntemin bu avantajlarının yanı sıra birtakım dezavantajları da bulunmaktadır. Yöntemi kullananlar bu dezavantajları da dikkate almalıdır. Bu dezavantajlar maddeler halinde şu şekilde sıralanabilir (Ke et al., 2021):

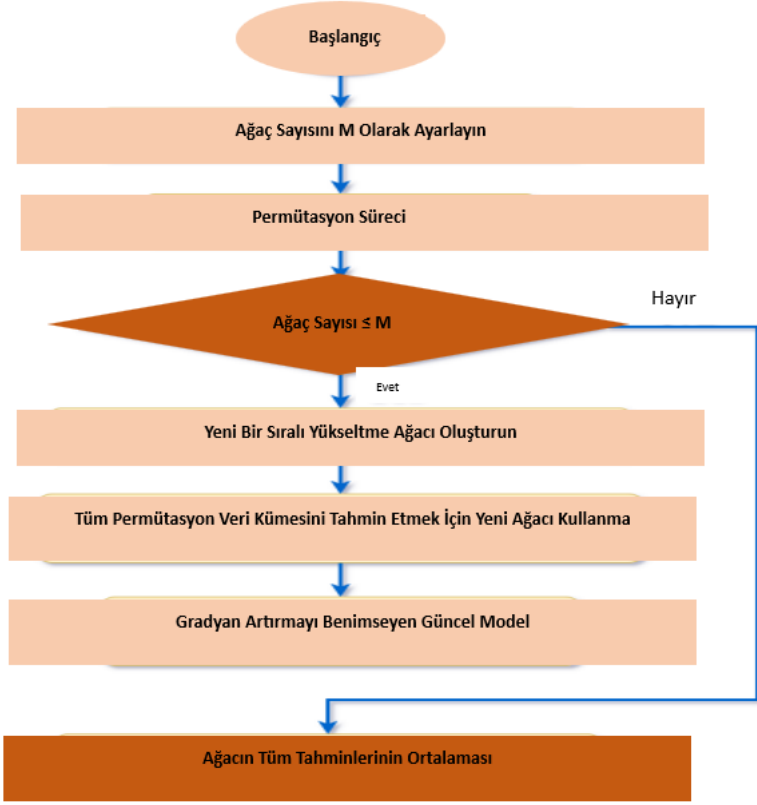
1. Bellek ve Zaman Gereksinimi: CatBoost'un bellek ve zaman gereksinimi, özellikle büyük veri kümeleriyle çalışıldığında kategorik özelliklerin işlenmesi sırasında artabilmektedir. Kategorik özelliklerin doğru bir şekilde işlenmesi, daha fazla bellek kullanımı ve daha uzun işlem süreleri gerektirebilmektedir.

2. Rastgele Sayıların Ayarının Etkisi: CatBoost, rastgele sayıların ayarlanmasının model tahmin sonuçları üzerinde belirli bir etkiye sahip olabileceğini belirtmektedir. Bu, modelin farklı çalıştırılmasında farklı sonuçlar üretebileceği anlamına gelmektedir.

CatBoost'un bu avantajları ve dezavantajları, modelin genel performansını etkileyebilecek önemli faktörlerdir. Ancak, bu faktörlerin etkisi uygulanacak senaryo ve veri setine bağlıdır. Dolayısıyla, CatBoost'u kullanırken bu avantajlar ve dezavantajlar göz önünde bulundurulmalıdır.

CatBoost, Gradyan Güçlendirilmiş Karar Ağacı (GBDT) algoritmalarının geliştirilmiş bir versiyonudur ve gradyan inişi çerçevesine dayanan karmaşık bir topluluk öğrenme tekniği kullanmaktadır. Model eğitimi sürecinde, ardışık olarak bir dizi karar ağacı oluşturulmakta ve her bir sonraki ağaç, önceki ağaçtan elde edilen hataları azaltarak öğrenmektedir. Yani, her bir karar ağacı, önceki ağacın öğrendiklerinden yararlanmakta ve bir sonraki ağacı etkileyerek modelin performansını artırmaktadır. Böylece güçlü bir öğrenici

oluşturulmaktadır (Prokhorenkova et al., 2018). CatBoost yönteminde, kategorik özellikler genellikle one-hot encoding yöntemini kullanılarak anlamlı bilgilere dönüştürülmektedir (Micci-Barreca, 2001). Bu dönüşüm, kategorik özelliklerin daha iyi anlaşılmasını ve CatBoost'un model eğitimi sürecinde daha etkili bir şekilde kullanılmasını sağlamaktadır. CatBoost algoritması, sıralı artırma mekanizmasıyla öne çıkan önemli bir özelliğe sahiptir. Geleneksel GBT algoritmalarında, birkaç artırma adımından sonra tüm eğitim örnekleri kullanılarak tahmin modeli oluşturulmaktadır. Ancak bu yaklaşım, modelde bir tahmin kayması (prediction shift) sorunu ortaya çıkartabilmekte ve özel bir hedef sızıntısı problemine yol açabilmektedir (Hancock ve Khoshgoftaar, 2020). CatBoost algoritması ise sıralı artırma mekanizmasını kullanarak bu sorunu etkin bir şekilde önlemektedir. Ayrıca, diğer geleneksel öğrenme sınıflayıcılarından farklı olarak, eğitim veri kümesinin çeşitli permütasyonlarını kullanarak aşırı uyum sorununu ele almaktadır. Bu nedenle, CatBoost algoritması, tahmin modellerinin güvenilirliğini ve performansını artırmak için tercih edilen bir seçenek olarak görülmektedir (Hussain vd., 2021). Şekil 17'de CatBoost yönteminin akış şeması gösterilmektedir.



Şekil 17: CatBoost yöntemi akış şeması (Yalçın vd., 2023)

CatBoost yöntemi parametreleri, modelin performansını ve davranışını kontrol etmek için kullanılmaktadır. CatBoost yönteminin birçok parametresi bulunmaktadır Tablo 14’te CatBoost yöntemine ait parametreler yer almaktadır (CATBOOST, 2023).

Tablo 14: CatBoost model parametreleri

Parametre	Tanım
<i>iterations</i>	Oluşturulacak ağaç sayısı ifade etmektedir.
<i>learning_rate</i>	Öğrenme hızını ifade etmektedir.
<i>depth</i>	Ağacın derinliğini ifade etmektedir.
<i>l2_leaf_reg</i>	L2 düzenleme parametresidir.
<i>random_strength</i>	Rastgelelik gücü parametresidir.
<i>border_count</i>	Sayısal özelliklerin bölme noktalarının sayısını ifade etmektedir.
<i>bagging_temperature</i>	Bayesyen bootstrap ayarlarını tanımlar.
<i>early_stopping_rounds</i>	Modelin doğrulama setindeki performansında bir iyileşme olmadığı durumda eğitimi otomatik olarak durduran parametredir.

CatBoost'un önemli parametrelerinin ayarlanması, modelin performansını, hızını, genelleme yeteneklerini etkilemektedir. Bu nedenle, bu önemli parametrelerin uygun bir şekilde ayarlanması, daha iyi sonuçlar elde etmeyi sağlamaktadır.

Bu bölümde, CatBoost yöntemi ayrıntılı bir şekilde ele alınmıştır. Yöntem hakkında temel bilgiler, yöntemin avantajları ve dezavantajlar, yöntemin algoritma ve parametreleri detaylı bir şekilde açıklanmıştır. Sonraki bölümde, sınıf dengesizliğini gidermek için kullanılan Sentetik Azınlık Aşırı Örnekleme (SMOTE) yöntemi hakkında bilgiler yer almaktadır.

4.2.6. Sentetik Azınlık Aşırı Örnekleme (SMOTE) Yöntemi

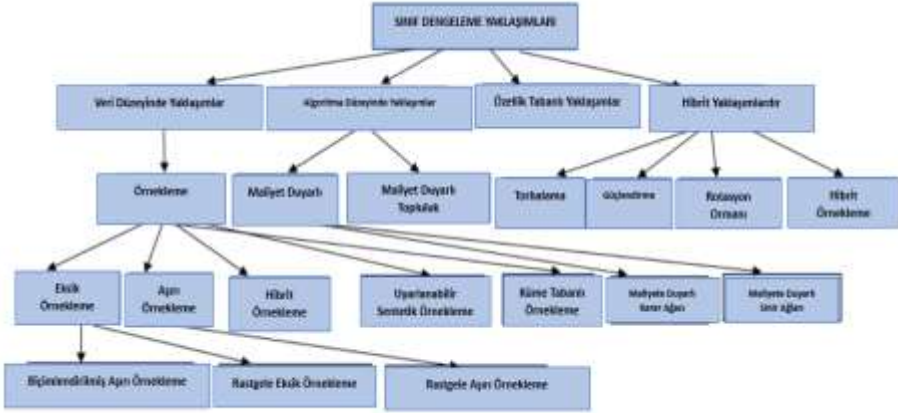
Sınıf dengesizliğiyle başa çıkmak için çeşitli yaklaşımlar geliştirilmiştir. Bu bölümde bu yaklaşımlar hakkında genel bilgiler yer alırken, uygulama kısmında kullanılan SMOTE yöntemi ile ilgili ayrıntılı bilgiler yer almaktadır.

Sınıf dengesizliği, sınıflandırma problemlerinde yaygın olarak karşılaşılan bir durumdur. Bir sınıfın diğerine kıyasla çok daha az temsil edildiği durumlarda ortaya çıkmakta ve algoritmaların doğruluğunu olumsuz etkileyebilmektedir (Sun vd., 2009). Çalışmada, SMOTE yöntemi veri hazırlık aşamasında kullanılmıştır. SMOTE yöntemi ile gruplar arasındaki veri boyutu farkı giderilmiş daha sonra yöntemler uygulanmıştır.

Dengeli veri seti, sınıflandırma modelinin daha doğru ve tarafsız sonuçlar üretmesine yardımcı olmaktadır. Dengesiz veri setlerinde ise küçük sınıfların temsil edilme oranı düşük olabilmektedir (Yavaş vd., 2020). Bu durum, modelin çoğunluk sınıfına odaklanmasına ve azınlık sınıfları doğru bir şekilde sınıflandırmakta zorluk yaşamasına neden olabilmektedir. Bu tür dengesizlikler, modelin yanlış veya yanıltıcı sonuçlar üretmesine yol açabilmektedir (Chawla, 2009). Bu nedenle, dengeli bir veri setinin, her sınıfın eşit derecede temsil edildiği ve modelin tüm sınıfları eşit derecede öğrenebildiği bir ortam sağlamadığı ifade edilebilir. Bunun sonucu olarak modelin doğru sınıflandırmalar yapma yeteneğinin artması ve genel performansını iyileşmesi beklenmektedir.

Sınıf dengesizliğini gidermek amacıyla birçok yöntem kullanılmaktadır. Bu yöntemler, Veri Düzeyinde Yaklaşımlar, Algoritma Düzeyinde Yaklaşımlar, Özellik Tabanlı Yaklaşımlar ve Hibrit Yaklaşımlar olmak üzere 4

ana başlık altında sınıflandırılabilir. Şekil 18’de bu yöntemler ve alt başlıkları yer almaktadır.



Şekil 18: Sınıf dengeleme yaklaşımları (Yadav ve Bhole, 2020)

Şekil 18’de görüldüğü üzere sınıflandırma problemlerinde sınıf dengesizliğini gidermek amacıyla birçok yaklaşım bulunmaktadır. Bu bölümde, çalışmada kullanılacak olan veri düzeyinde yaklaşımlara ayrıntılı olarak yer almakta, diğer yaklaşımlar ise genel bir bakış açısıyla yer almaktadır.

Dengesiz veri sorununu gidermek için kullanılan yaklaşımlardan ilki veri düzeyinde yaklaşımlardır. Eğitim örneklerinin, sınıflandırıcıların daha dengeli bir sınıf dağılımıyla performans göstermesine olanak tanıyan şekilde değiştirildiği yöntemlerdir. Bu yöntemler, veri kümesindeki sınıf dengesizliğini ele alarak, az sayıda örneğe sahip sınıflarla çok sayıda örneğe sahip sınıflar arasındaki dengesizliği azaltmayı hedeflemektedirler (Fernández vd., 2018).

Sınıfları dengelemek için basit bir veri seviyesi yöntemi, orijinal veri setini yeniden örnekleyerek azınlık sınıfını aşırı örnekleme veya çoğunluk sınıfını alt örnekleme yaparak sınıfları yaklaşık olarak eşit şekilde temsil edene kadar düzenlemeyi içermektedir. Bu bağlamda veri düzeyinde yaklaşımlar; aşırı örnekleme, eksik örnekleme ve hibrit örnekleme olmak üzere 3’e ayrılabilir (Yadav ve Bhole, 2020).

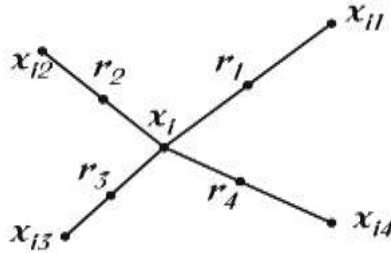
Son yıllarda, dengesiz veri kümeleriyle gerçekleştirilen veri madenciliği çalışmaları hem teorik hem de pratik açıdan giderek daha fazla ilgi çekmektedir. Dengesiz veri sorunu, bu çalışmalarda önemli bir zorluk

görülmektedir (Gu vd., 2022). Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) bu sorunu çözmek için kullanılan popüler bir yöntemdir (Czarnowski, 2022).

Chawla vd. (2002), "yeniden örnekleme ile aşırı örnekleme" yöntemine karşılık olarak, azınlık sınıfını sentetik örnekler oluşturarak aşırı örnekleme yapan SMOTE yöntemini önermişlerdir. Sentetik örnekler, veri uzayında yapılan işlemlerle üretilmektedir. Aşırı örnekleme işlemi, azınlık sınıfındaki her örneğin en yakın komşuları arasında yer alan çizgi parçaları boyunca sentetik örneklerin oluşturulmasıyla gerçekleştirilmektedir. Sentetik örnekler, özellik vektörleri ile en yakın komşu arasındaki farkın, rasgele bir katsayıyla çarpılıp bu dönüşüm sonucu elde edilen değerlerin özellik vektörüne eklenmesi yöntemiyle üretilmektedir. SMOTE, azınlık sınıfındaki örneklerin (x ve x_R) benzerlik ölçütü kullanılarak seçilmesini ve daha sonra bu örneklerin lineer bir kombinasyonu ile sentetik olarak oluşturulan örnekleri ifade etmektedir. SMOTE yöntemi şu şekilde formüle edilmektedir (Blagus ve Lusa, 2013):

$$s = x + u \cdot (x^R - x)$$

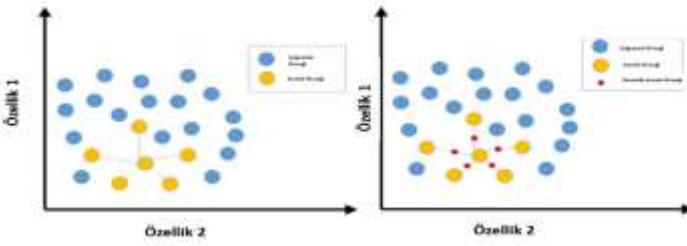
u , $0 \leq u \leq 1$ arasında rastgele bir sayıdır. x_R , x 'in azınlık sınıfına ait en yakın komşusundan rastgele seçilmektedir. Şekil 19'da, SMOTE yöntemine göre sentetik veri noktalarının nasıl oluşturulduğu gösterilmektedir.



Şekil 19: SMOTE yöntemine göre sentetik veri noktaları

Şekil 19'a göre, azınlık sınıfından bir x_i pozitif örneği seçilerek, yeni sentetik veri noktaları oluşturulmaktadır. Bu noktaları oluşturmak için, bir mesafe metriği kullanılmakta ve aynı sınıfa ait birkaç en yakın komşu (x_{i1} 'den x_{i4} 'e kadar olan noktalar) eğitim setinden seçilmektedir. Son olarak, rastgele bir ekleme yapılmakta ve yeni örnekler r_1 , r_2 , r_3 ve r_4 elde edilmektedir (Fernández vd., 2018). Bu yöntem, azınlık sınıfının karar bölgelerini

genişletirken aşırı öğrenme sorununa yol açmadan daha büyük ve daha genel bölgeler öğrenmektedir. Bu da öğrenilen sınıflandırıcının daha iyi genelleme yapabilmesine katkı sağlamaktadır (Maciejewski ve Stefanowski, 2011). Şekil 20’de SMOTE yöntemi gösterilmektedir.



Şekil 20: SMOTE yöntemi (Vijayvargiya vd., 2021)

Tablo 15’te SMOTE algoritmasının çalışma algoritması yer almaktadır. Algoritma 12 adım içeren bir sıra ile çalışmaktadır.

Tablo 15: SMOTE algoritması

SMOTE Algoritması	
1:	function SMOTE(T,N,k) Input: T ; N; k #minority class examples, Amount of oversampling, #NNs Output: (N/100) * T synthetic minority class samples Variables: Sample[][]: array for original minority class samples; newindex: keeps a count of number of synthetic samples generated, initialized to 0; Synthetic ic[][]: array for synthetic samples
2:	if N < 100 then
3:	Randomize the T minority class samples
4:	T = (N/100)*T
5:	N = 100
6:	end if
7:	N = (int)N/100 The amount of SMOTE is assumed to be in integral multiples of 100.
8:	for i = 1 to T do
9:	Compute KNN for i, and save the indices in the n
10:	POPULATE(N , i, narray)
11:	end for
12:	end function

SMOTE algoritmasının orijinal versiyonundan (Chawla vd., 2002) sonra, literatürde çeşitli değişiklikler önerilmiştir. SMOTE yaklaşımı, tüm nominal özelliklere sahip veri kümeleriyle başa çıkamamaktadır. Ancak, sürekli ve nominal özelliklerin karışık olduğu veri kümeleri için geliştirilmiştir. SMOTE-NC (Nominal Sürekli) ve SMOTE-N (Nominal) yöntemlerini önerilmiştir. Bu şekilde SMOTE, nominal özelliklerin de dahil olduğu durumlarda kullanılabilir hale gelmiştir (Ganganwar, 2012). SMOTE yönteminin bazı dezavantajları bulunmaktadır (Maciejewski ve Stefanowski, 2011):

1. Azınlık sınıfına ait tüm örnekler aşırı örnekleme için kullanılmaktadır. Ancak, bu örnekler sınıflandırıcılar için eşit derecede öneme sahip olmayabilmektedir. Karar sınırına yakın ve yakınında bulunan örnekler, hatalı sınıflandırmalara daha yatkın olabilirken, sınıf bölgesi içinde bulunan örnekler daha kolay öğrenilebilmektedir. Bu durum, modelin sınıflandırma performansını olumsuz etkileyebilmektedir.

2. Veri uzayında yapay örnekler oluşturularak aşırı örnekleme yapılmaktadır. Ancak, bu yapay örnekler gerçek verilerden türetilmiş olduğu için bazı durumlarda gerçekçiliklerini kaybedebilmektedir. Bu da modelin genelleme yeteneğini olumsuz etkileyebilmektedir.

3. Aşırı örnekleme yaparken sınıf etiketlerini değiştirmemektedir. Bu, aşırı örnekleme sonucunda sınıf dağılımında herhangi bir değişiklik olmadığı anlamına gelmektedir. Bazı durumlarda, sınıf etiketlerinin dengelenmesi de önemli olabilmekte ve sadece örnekleme yapmak yeterli olmayabilmektedir.

Bu bölümde SMOTE yöntemi hakkında bilgiler verilmiştir. Sentetik örnekler oluşturularak azınlık sınıfını dengelemek amacıyla kullanılan yöntem, özellik uzayında işlemler yaparak yeni örnekler yaratmaktadır (Pradipta vd., 2021). Bunun sonucunda, aşırı uydurma sorunu oluşturmadan daha iyi genelleme yapabilen bir sınıflandırıcı elde edilmektedir (Jadwal vd., 2022). Bununla birlikte, SMOTE'nin bazı dezavantajları da vardır. Örneğin, azınlık örneklerinin seçimi ve üretilen sentetik örneklerin dağılımı gibi konular, yöntemin etkinliğini etkileyebilmektedir. Ayrıca, yüksek boyutlu veri kümelerinde ek gürültü oluşturabilme potansiyeli bulunmaktadır (Radivojac vd., 2004). SMOTE yöntemi kullanılırken bu dezavantajlar göz önünde bulundurulmalı dikkate alınmalıdır. Çalışmada SMOTE yöntemi Regresyon yöntemi öncesinde veri hazırlık aşamasında kullanılacaktır. Bu yönüyle

literatüre farklı bir bakış açısı sunması beklenmektedir. Sonraki bölümde, değişken önem derecelerinin tespitinde kullanılan yöntemler hakkında bilgiler yer almaktadır.

4.2.7. Değişken Önem Derecelerinin Tespitinde Kullanılan Yöntemler

Veri madenciliği, verilerden anlamlı bilgiler çıkartma süreci olarak tanımlanabilmektedir. Bu süreçte veri setinde yer alan çeşitli değişkenler kullanılarak anlamlı bilgiler elde edilmeye çalışılmaktadır. Bu bilgilerin tespitinde, hangi değişkenin model üzerinde etkisinin nasıl olduğu önem arz etmektedir. Değişken önem derecelerinin tespitinde çeşitli yöntemler bulunmaktadır. Bunlar arasında, Ağaç tabanlı modellerde değişken önem derecesinin tespiti genellikle modelin doğruluğunda meydana gelen azalma veya Gini safsızlık oranındaki ortalama azalmaya göre yapılmaktadır (L.-P. Chen, 2021). Aşağıda bu iki yöntem ve diğer yöntemler açıklanmaktadır.

1. Gini Kriteri: Gini safsızlık oranındaki azalmaya göre, safsızlığı büyük ölçüde azaltan bölme önemli olarak kabul edilmektedir. Her bölünmede, bölünmeyi oluşturmak için kullanılan X_i değişkeni için Gini düşümü safsızlığındaki azalma kaydedilir. X_i 'nin bölünmeyi oluşturduğu ormandaki Gini safsızlığındaki tüm düşüşlerin ortalaması, Gini değişken önem ölçüsünü verir (Nembrini vd., 2018).

2. Modelin Doğruluğunda Meydana Gelen Azalma: Bir değişkenin önemi, değişkenin değerleri rastgele bir şekilde değiştirildiğinde tahmin doğruluğundaki değişim ile orijinal gözlemler karşılaştırarak hesaplanır (Guo vd., 2016).

3. CAR Scores: Açıklayıcı değişkenler arasındaki korelasyona göre düzeltilmiş marjinal korelasyonlar olarak tanımlanan CAR puanları, doğal bir değişken önem kriteridir. CAR puanları sadeliklerine rağmen hem küçük hem de büyük örneklem durumlarında etkili model seçimi yapabilmektedirler (Zuber ve Strimmer, 2011).

4. Shapley Değeri: Lipovetsky ve Conklin (2001) ve Conklin vd. (2004), oyun teorisinde Shapley değerini kullanarak LMG ölçüsünü yeniden düzenlemişlerdir. Shapley, "N kişilik oyunlar için bir değer" olarak adlandırılan bir kavramı geliştirmiştir ve bu kavram çok oyunculu bir kooperatif oyunda oyuncuların değerlerinin sıralamasını değerlendirmek için kullanılmaktadır.

Shapley değeri, oyun teorisi içindeki görelî önem için teorik bir temel sağlamak amacıyla sıralamaların ortalamasını alarak hesaplanmaktadır (BI, 2012).

5. Akaike Ağırlıkları: Tüm değişkenler için Akaike ağırlıkları hesaplandıktan sonra, gözlenen değerlerin sıralanmasıyla görelî önem değerlendirilir; bir değişkenin daha büyük toplamı, diğer değişkenlere göre nispeten daha önemlidir (Burnham ve Anderson, 2004).

BÖLÜM 5

UYGULAMA

Bu bölüm, çalışmanın uygulama aşamalarını ayrıntılı bir şekilde sunmayı amaçlamaktadır. Bu amaç doğrultusunda, öncelikli olarak çalışmanın hedefleri, kapsamı, özgünlüğü ve sınırlılıkları yer almaktadır. Ardından, uygulamalarda kullanılan değişkenler ve çalışmada yer alan ülke grupları, veri seti hazırlık uygulamaları. Daha sonra, Rassal Orman Regresyon, CatBoost Regresyon ve XGBoost Regresyon uygulamaları ve elde edilen sonuçların karşılaştırılmaktadır. Bu bölüm, tezin pratiğe yönelik yönünü ortaya koyarak, elde edilen bulguların değerlendirilmesini ve sonuçlara yönelik anlamlı çıkarımların yapılmasını sağlamaktadır.

5.1. Çalışmanın Amacı

Çalışmanın amacı, konut Gayrimenkul Yatırım Ortaklıkları (GYO) firmalarının kârlılığını etkileyen finansal göstergelerin önem sıralamalarını belirlemek ve bu amaç doğrultusunda kullanılan üç farklı veri madenciliği yöntemini karşılaştırmaktır. Farklı veri madenciliği yöntemlerinin uygulanmasıyla konut GYO firmalarının kârlılığını etkileyen finansal göstergeleri en iyi şekilde tahmin edebilen yöntemin tespit edilebilmesi amaçlanmaktadır. Bu amaç doğrultusunda, konut GYO firmalarına ait 2013-2022 yılları arasında çeyrek dönemlik veriler kullanılmıştır.

5.2. Çalışmanın Kapsamı

Çalışmanın ilk bölümü, GYO firmalarıyla ilgili yapılan araştırmaları ve kullanılan yöntemleri kapsamlı bir şekilde inceleyen bir literatür taramasını içermektedir. Bu bölüm, mevcut bilimsel çalışmaların derinlemesine analizini sunarak araştırma alanına sağlam bir temel oluşturmaktadır. İlgili literatürdeki makaleler, raporlar ve diğer kaynaklar taranarak, GYO firmalarının özellikleri, performans ölçütleri, finansal analiz yöntemleri ve veri madenciliği teknikleri gibi konular incelenmiştir. İkinci bölümde ise, gayrimenkul kavramından başlayarak GYO firmaları hakkında detaylı bilgiler sunulmaktadır. Üçüncü bölümde, veri madenciliği ve veri madenciliği yöntemleri hakkında detaylı bilgiler yer almaktadır. Bu bölümde ise, Rassal Orman, XGBoost ve CatBoost gibi çeşitli veri madenciliği yöntemleri kullanılarak pratik uygulamalar

gerçekleştirilecektir. Bu yöntemler, GYO firmalarıyla ilgili verilerin analiz edilmesi ve önemli faktörlerin keşfedilmesi amacıyla kullanılacaktır. Veri madenciliği yöntemlerinin seçimi, öncelikle literatürdeki başarıları ve uygulanabilirlikleri göz önünde bulundurularak yapılmıştır. Bu uygulamalar, GYO firmalarının kârlılığını etkileyen faktörleri daha iyi anlamak ve tahmin etmek için gelişmiş modelleme ve öngörü yeteneklerini kullanmayı hedeflemektedir. Bu şekilde, çalışma, akademik bir bakış açısıyla veri madenciliği yöntemlerinin GYO sektörü üzerindeki etkisini ve analitik potansiyelini araştırmaktadır.

Çalışmada, farklı ülkelerde faaliyet gösteren konut GYO firmalarının verileri kullanılmıştır. Konut GYO firmalarının tespiti için finance.yahoo web sayfasından yararlanılmıştır. Söz konusu web sayfasında konut GYO firmalarını içerecek şekilde filtreleme işlemi gerçekleştirildiğinde 157 firma listelenmektedir. Ancak, birden fazla borsada işlem gören firmaların her bir borsa için ayrı ayrı gösterildiği göz önüne alınarak, öncelikli olarak bu firmaların tespiti yapılmıştır. Ayrıca, verileri eksik olan firmalar da dikkate alınmamıştır. Sonuç olarak, 7 farklı ülkede faaliyet gösteren 32 konut GYO firması, çalışmanın kapsamında incelenmek üzere seçilmiştir. Tablo 16'da bu firmaların isimleri yer almaktadır.

Tablo 16: Konut GYO firmaları

Firma İsmi	Ülke
Ingenia Communities Group	Avustralya
Home Invest Belgium S.A.	Belçika
Canadian Apartment Properties Real Estate Investment Trust	Kanada
Killam Apartment REIT	Kanada
Advance Residence Investment Corporation	Japonya
Comforia Residential REIT, Inc	Japonya
Daiwa Securities Living Investment Corporation	Japonya
Kenedix Residential Next Investment Corporation	Japonya
Starts Proceed Investment Corporation	Japonya
Emlak Konut Gayrimenkul Yatırım Ortaklığı A.S.	Türkiye
Sinpas Gayrimenkul Yatırım Ortaklığı A.S.	Türkiye
Empiric Student Property plc	Birleşik Krallık
KCR Residential REIT plc	Birleşik Krallık

American Homes 4 Rent	ABD
Apartment Income REIT Corp.	ABD
AvalonBay Communities, Inc.	ABD
Boardwalk Real Estate Investment Trust	ABD
BRT Apartments Corp.	ABD
BSR Real Estate Investment Trust	ABD
Camden Property Trust	ABD
Centerspace	ABD
Clipper Realty Inc.	ABD
Equity LifeStyle Properties, Inc.	ABD
Equity Residential	ABD
Essex Property Trust, Inc.	ABD
Firm Capital Apartment Real Estate Investment Trust	ABD
Independence Realty Trust, Inc.	ABD
Invitation Homes Inc.	ABD
Mid-America Apartment Communities, Inc.	ABD
NexPoint Residential Trust, Inc.	ABD
Veris Residential, Inc.	ABD
InterRent Real Estate Investment Trust	ABD

Firmaların belirlenmesinin ardından, veriler InvestingPro web sayfası aracılığıyla 2013.D1 – 2022.D1 dönemlerini kapsayacak şekilde indirilmiş ve analizler Python programlama dili kullanılarak gerçekleştirilmiştir.

Elde edilen bulgular ile firma yöneticilerine, yasa yapıcılara ve araştırmacılara ayrı ayrı öneriler sunulmuştur. Bu öneriler, çalışmanın sonuçlarına dayanarak geliştirilmiş ve mevcut literatür ile uyumlu bir şekilde değerlendirilmiştir.

5.3. Çalışmanın Özgünlüğü

Çalışmanın üç bakış açısıyla özgünlüğe sahip olduğu söylenebilir. Bunlardan ilki, farklı ülkelerde faaliyet gösteren 32 GYO firmasının 2013.D1 - 2022.D1 dönemlerini kapsayan 37 dönemlik veri setinin kullanılmasıdır. Bu veri setinin kullanımı, sektörün kapsamlı bir şekilde incelenmesini sağlaması ve veri madenciliği yöntemlerinde karşılaşılan aşırı öğrenme sorununun da etkili bir şekilde engellenmesi amaçlanmaktadır.

Çalışmanın bir diğer özgün yanı, GYO firmalarının kârlılıklarını etkileyen faktörlerin tespitinde veri madenciliği yöntemlerinin kullanılmasıdır. Veri madenciliği teknikleri, veri setinin derinlemesine analiz edilmesi ve kârlılığı etkileyen faktörlerin tespiti için güçlü bir araç sağlamaktadır.

Çalışmanın üçüncü bir özgün yanı ise sınıflandırma problemlerinde kullanılan SMOTE (Sentetik Azınlık Örnekleme) yönteminin bir regresyon probleminin, veri hazırlık aşamasında uygulanmasıdır. Bu yöntem ile gruplar arasında dengeli bir veri setleri elde etmek amaçlanmaktadır. Bu yaklaşımın, literatüre farklı bir bakış açısı sağlayacağı düşünülmektedir.

5.4. Çalışmanın Sınırlılıkları

Çalışmada, konut GYO firmalarına ait son 10 yıla çeyreklik veriler ile analizler yapılmıştır. Bu durum, çalışma verilerinin elde edildiği InvestingPro web sayfasında en fazla son on yıla ait verilere ulaşılabilmesinden kaynaklanmaktadır. Çalışmanın bir diğer sınırlılığı, tek bir GYO türü verileri ile analizlerin yapılmasıdır. Her türün kendi içerisinde farklı faktörlere sahip olması tek bir GYO türü verileri ile çalışmanın yapılmasını gerektirmiştir. Böylelikle bulguların sağlıklı bir şekilde yorumlanması hedeflenmektedir.

5.5. Çalışmada Kullanılan Değişkenler

Regresyon uygulamalarında, en az bir bağımlı değişken ve bağımsız değişkenlere ihtiyaç vardır. GYO'ların varlık yapılarında duran varlıklar önemli bir yere sahiptir. Bu nedenle, amortismanın GYO firmalarının gerçek kârlılıklarını yansıtmasına engel olacağı muhtemeldir (Graham ve Knight, 2020). Bundan dolayı Ulusal Gayrimenkul Yatırım Ortaklıkları Birliği (NAREIT, 1991) tarafından net gelire alternatif olarak faaliyetlerden elde edilen fonlar (FFO) oranı tanıtılmıştır. FFO, amortisman, itfa payı ve tekrarlanmayan nakitsiz gider ve gelirlerin çoğunu düşmeyerek nakit akışlarının daha iyi tahmin edebilmesine imkân sağlamaktadır (Ben-Shahar vd., 2020). Ancak, FFO Genel Kabul Görmüş Muhasebe İlkeleri (GAAP) içerisinde yer almadığı için GYO yöneticileri belirli giderleri net gelirden seçici olarak hariç tutarak manipüle edebilirler. Manipülasyona açık olması ve uluslararası muhasebe standartları çerçevesinde bir geçerliliğinin olmaması nedeni ile FFO çalışmada kullanılmamıştır. Bunun yerine, uluslararası muhasebe standartları çerçevesinde geçerliliği olan faiz amortisman ve vergi

öncesi kâr (FAVÖK) oranının kârlılık göstergesi olarak kullanılmıştır. Bağımsız değişkenler ise firma büyüklüğü (TA), net varlık değeri (NAV), kaldıraç oranı (LR), varlık devir hızı (AT) ve cari oran (CR)'dan oluşmaktadır. Bu değişkenlere ait bilgiler ve seçilme nedenleri aşağıda yer almaktadır.

1. Firma Büyüklüğü: Firmaların sahip oldukları varlıklar firma büyüklüğünün bir göstergesi olarak kabul edilebilmektedir. B. Ambrose ve Linneman (2001), GYO firmalarının büyüklüğü ile kârlılıkları arasında ilişki olduğunu, daha büyük firmaların yüksek kâr marjlarına sahip olduklarını tespit etmişlerdir. Belirli bir oranda temettü dağıtma zorunluluğu olmayan firmaların bu fonları yeni varlıklar ediniminde kullanacağı düşünülmektedir.

2. Kaldıraç Oranı: Kaldıraç oranı firmaların borçlarının özkaynaklarına olan oranını göstermektedir. Firmaların yabancı kaynak kullanımı hakkında bilgiler içermektedir. Fama ve French (2002) ve Barclay vd. (2006), kaldıraç oranının kârlılık üzerinde azaltıcı bir etkisi olduğunu tespit etmişlerdir. Baker vd. (2006) ise borçların kârlılık üzerinde önemli bir etkiye sahip olduğunu belirtmektedirler. Yüksek oranda temettü dağıtma zorunluluğunun olması GYO firmalarının dış finansmana yönelmeleri üzerinde önemli bir etkiye sahip olacağı düşünülmektedir.

3. Net Varlık Değeri: Net varlık değeri bir firmanın finansal yapısını göstermesi bakımından önemlidir. Jakpar vd. (2018b) GYO firmalarının kârlılıkları üzerinde net varlık değerinin etkisi olduğunu tespit etmiştir.

4. Duran Varlık Devir Hızı: Temettü dağıtım zorunluluğu olmaması firmalara varlık finansmanı için iç fon yaratmaktadır. Bu iç fonlar ile edinilen varlıkların etkili kullanımının kârlılık üzerindeki etkisini tespit etmek amacıyla çalışmada bu oranın bağımsız değişkenlerden biri olarak tercih edilmiştir.

5. Cari Oran: Nakit akış sıkıntısı yaşayan firmalar kısa vadeli yabancı kaynaklara yönelmeyi tercih etmektedirler. Temettü dağıtma zorunluluğu bulunmayan GYO'ların ise bu akışı iç kaynaklardan karşılayabileceği düşünülmektedir. Bu sebeple bu oranın bağımsız değişkenlerden biri olarak kullanılması planlanmaktadır. Çelik ve Arslanlı (2020) bu oranın GYO firmaların kârlılıkları üzerinde etkisi olduğu tespit etmişlerdir.

Tablo 17'de bağımlı ve bağımsız değişkenler yer almaktadır. Değişkenlere ilişkin kısaltmalar İngilizce karşılıklarına göre yapılmıştır.

Tablo 17: Değişkenler

Değişken Türü	Değişken Adı	Kısaltma	Formül
Bağımlı Değişken	Faiz, Amortisman Vergi Öncesi Kâr	EBITDA	Esas Faaliyet Karı + Amortisman Giderleri
Bağımsız Değişkenler	Firma Büyüklüğü	TA	Toplam Varlıklar
	Net Varlık Değeri	NAV	Toplam Varlıklar- Toplam Yükümlülükler
	Kaldıraç Oranı	LR	Toplam Borçlar / Toplam Varlıklar
	Duran Varlık Devir Hızı	AT	Net Satışlar / Ortalama Duran Varlıklar
	Cari Oran	CR	Dönen Varlıklar / Kısa Vadeli Yabancı Kaynaklar

5.6. Çalışmada Yer Alan Ülke Grupları

GYO firmaları için temettü dağıtımına ilişkin düzenlemeler önem arz etmektedir. Temettü dağıtım oranları, firmaların vergi muafiyetinden yararlanabilmeleri için dağıtımaları gereken temettü miktarının asgari seviyesini belirtmektedir. Bu oran çoğu ülkede genellikle %90 civarında belirlenmiştir. Ancak Türkiye ve Kanada’da böyle bir sınırlama bulunmamaktadır. Bu bağlamda, çalışmada yer alacak ülkeler, temettü dağıtım zorunluluğuna göre iki guruba ayrılmıştır. Buna göre 1. Grupta, vergi muafiyetinden yararlanmak amacı ile her yıl belirli bir asgari oranda temettü dağıtım zorunluluğu bulunan ülkeler yer almaktadır. 2. Grupta ise her yıl belirli bir asgari oranda temettü dağıtım zorunluluğu bulunmayan ülkeler yer almaktadır. Tablo 18’de bu gruplar ve veri boyutuna ilişkin bilgiler yer almaktadır.

Tablo 18: Ülke grupları

Temettü Dağıtma Zorunluluğu Olan Ülkeler (1.GRUP)	Temettü Dağıtma Zorunluluğu Olmayan Ülkeler (2. GRUP)
Avusturya Belçika Japonya Birleşik Krallık ABD	Türkiye Kanada
Toplam Firma Sayısı: 28	Toplam Firma Sayısı: 4
Toplam Veri Satırı Sayısı: 1074	Toplam Veri Satırı Sayısı: 164

5.7. Veri Seti Hazırlık Uygulamaları ve Yöntemler

Firmalara ait finansal veriler, InvestingPro web sitesinden temin edilmiş ve uygulamalar, Python programlama dili kullanılarak gerçekleştirilmiştir.

Ham veriler çeşitli tutarsızlıklar, aralık dışı değerler, eksik değerler veya aşırılıklar gibi kusurlar içerebilmektedir. Bu gibi durumlar ortadan kaldırılmadığında yöntemlerin performansında düşüş yaşanması muhtemeldir (Çetin ve Yıldız, 2022). Veriler içerisinde yer alan tutarsız ve hatalı verilere gürültü denilmektedir. Veri seti içerisinde, eksik değer içeren verilerin çıkartılması, eksik değerler yerine sabit bir değer atanması, değişkenin ortalaması alınarak eksik değerlerin yerine bu ortalamanın yazılması, verilere uygun tahmin yapılarak (regresyon) veri setindeki gürültü temizlenebilir (Coşlu, 2013). Bu çalışmada firmalara ait eksik değerler her firma özelinde tespit edilmiştir. Eksik değerler her firmaya ait ortalamalar ayrı ayrı dikkate alınarak ortalama yöntemi ile giderilmiştir.

Tablo 18’de temettü dağıtma zorunluluğuna göre ülkeler gruplara ayrıldığında veri seti boyutları arasında ciddi bir fark olduğu, temettü dağıtma zorunluluğu olmayan firmalara ait verilerin azınlıkta kaldığı görülmektedir. Temettü dağıtma zorunluluğu bulunan firmalara ait toplam veri satır sayısı 1074, temettü dağıtma zorunluluğu bulunmayan bu firmalara ait toplam veri satır sayısı 164’tür. Veri setleri arasındaki bu dengesizliği gidermek amacı ile Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) kullanılmıştır. SMOTE

yöntemi uygulaması ile temettü dağıtma zorunluluğu bulunan ve bulunmaya gruplara ait veri boyutu eşitlenmiş ve satır sayısı 1074 olmuştur. Toplam veri satırı sayısı 2148 olmuştur. Bu sayede, kârlılığın gruplar arasındaki belirleyicilerini tespit etme imkânı oluşmuştur.

Daha sonra veri setinde yer alan değişkenlerin normalleştirilmesi işlemi yapılmıştır. Veri setinde yer alan değerler sürekli ve oran değişkenler içerdiği için böyle bir normalleştirmeye ihtiyaç duyulmuştur. Normalleştirme işlemi en az – en çok (min-max) normalleştirme yöntemi ile yapılmıştır. En az – en çok normalleştirme, veri değerlerini 0 ile 1 arasında ölçeklendiren bir yöntem olarak da bilinmektedir. Bu yöntemde, her bir veri noktasından en düşük değer çıkartılmakta ve ardından en düşük ve en yüksek değerler arasındaki farka bölünerek normalleştirme işlemi gerçekleştirilmektedir. Veri setinin hazırlanmasının ardından Rassal Orman Regresyon, CatBoost Regresyon ve XGBoost Regresyon yöntemlerine göre uygulamalar yapılmıştır. Daha sonra uygulama sonuçları karşılaştırılmıştır. Tablo 19’da çalışmada kullanılan yöntemler yer almaktadır.

Tablo 19: Yöntemler

Yöntem	Kısaltma
Rassal Orman Regresyon	RF
CatBoost Regresyon	CB
XGBoost Regresyon	XGB

Rassal Orman Regresyon, CatBoost Regresyon ve XGBoost Regresyon yöntemleri, benzer özelliklere sahip olmaları nedeniyle karşılaştırılabilirlik açısından tercih edilmektedir. Bu yöntemlerin ortak özelliği, ağaç tabanlı yaklaşımları benimsemeleridir. Her bir yöntem, veri setini ağaç yapısı şeklinde modeller ve her ağaç, veri örneklerini sınıflandırma veya tahmin etme işlemlerinde karar kurallarını kullanır. Bunun yanı sıra, bu yöntemler topluluk öğrenmesi prensiplerine dayanmaktadır. Her ağacın bağımsız olarak öğrenmesi ve sonuçlarının birleştirilmesiyle, daha güçlü ve kararlı tahminler elde etmek mümkün olur. Karşılaştırılabilirlik açısından, bu üç yöntem arasında benzer performans özellikleri bulunmaktadır. Her bir yöntem, doğrusal olmayan ilişkileri yakalama yeteneklerine sahip olup, karmaşık veri setlerinde iyi performans sergileyebilirler. Ayrıca, gürültülü veriye dayanıklıdır ve veri setlerindeki tutarsızlıkları ve eksik değerleri ele alma yeteneklerine sahiptirler. Ancak, her bir yöntemin kendine özgü özellikleri ve hiperparametre ayarları

bulunmaktadır. Bu nedenle, çalışmanın kapsamında bu yöntemlerin performansı ve uygulanabilirlikleri karşılaştırılacak ve en uygun yöntemin seçimi yapılacaktır.

5.8. Rassal Orman Regresyon Uygulaması

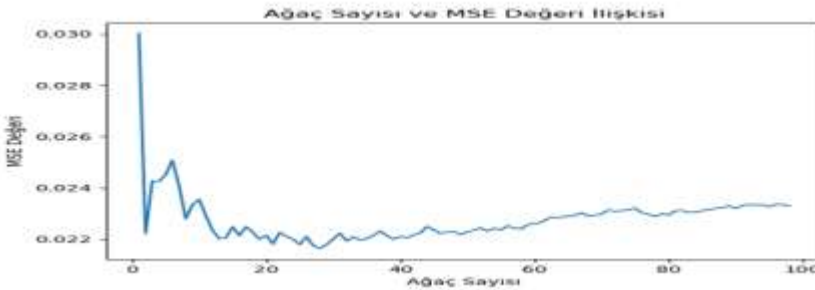
Rassal orman regresyon (RF) yöntemi ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna (1. Grup) daha sonra temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. Uygulama adımları her grup için şu şekildedir:

1. İlk olarak her bir yöntem için hiperparametre ayarlaması yapılması.
2. Tespit edilen optimum hiper parametreler ile modelin oluşturulması.
3. Modelin performans ölçütlerinin tespit edilmesi.
4. Modelde yer alan değişkenlerin önem derecelerinin tespit edilmesi.

Rassal Orman Regresyon yöntemi uygulaması, bu adımları kapsayacak şekilde her iki grup için gerçekleştirilmiştir. Her iki gruba ait uygulama sonuçları ayrı başlıklar altında yer bu bölümde yer almaktadır.

5.8.1. Rassal Orman Regresyon Uygulamaları (1. Grup)

Rassal orman regresyon yöntemi ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna (1. Grup) yönelik uygulanmıştır. İlk olarak parametre ayarlaması yapılmaksızın model oluşturulmuştur. Şekil 21’de bu modele ilişkin ağaç sayısına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 21: RF ağaç sayısı – MSE ilişkisi (1. Grup)

Şekil 21’de görüldüğü üzere ağaç sayısında meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak

amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 20’de en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 20: RF optimum model parametreleri (1. Grup)

Parametre	Optimum Değer
<i>max_depth</i>	26
<i>max_features</i>	4
<i>n_estimators</i>	25

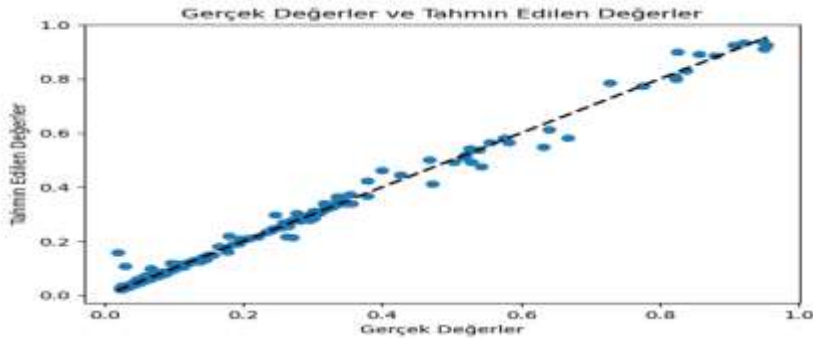
Tablo 20’de rassal orman regresyon modeli için optimum parametreler gösterilmektedir. Buna göre maksimum derinliği 26, maksimum özelliği 4 ve 25 adet karar ağacı ile optimum bir model elde edilebilmektedir. Bu değerlere göre model oluşturulmuş ve optimum model elde edilmiştir.

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile Rassal Orman Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 21’de modele ilişkin performans ölçütleri yer almaktadır.

Tablo 21: RF performans ölçütleri (1. Grup)

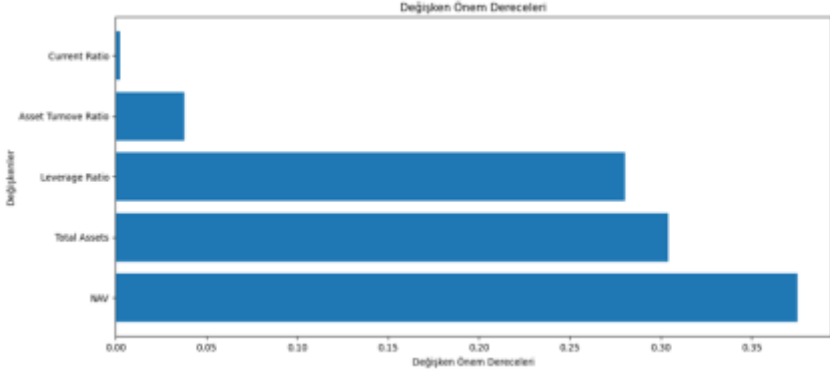
Ölçüt	Değer
R^2	0.8877
<i>MSE</i>	0.0215
<i>MAE</i>	0.0122

Tablo 21’e göre, R^2 değeri %89 olarak, hata değerleri ise 0’a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluştuğunu göstermektedir. Şekil 22’de modele ait grafik gösterilmektedir.



Şekil 22: RF model grafiği (1. Grup)

Rassal orman regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, Rassal orman regresyon yöntemine göre çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 23'te modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.

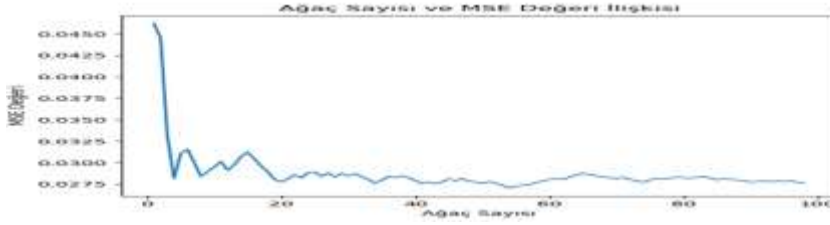


Şekil 23: RF değişken önem dereceleri (1. Grup)

Şekil 23'e göre temettü dağıtma zorunluluğu bulunan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem derecelerine göre gösterilmektedir. Buna göre kârlılık üzerinde en önemli etkiye sahip değişken net varlık değeridir (NAV). Firma büyüklüğü ve kaldıraç oranı da kârlılık üzerinde önemli etkiye sahip diğer değişkenlerdir. Duran varlık devir hızı ve cari oran değişkenlerinin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır. Diğer bölümde, Rassal Orman Regresyon uygulaması 2. Grup verileri ile gerçekleştirilmiştir.

5.8.2. Rassal Orman Regresyon Uygulamaları (2. Grup)

Bu bölümde, Rassal orman regresyon yöntemi belirli bir oranda temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. İlk olarak parametre ayarlaması yapılmaksızın model oluşturulmuştur. Şekil 24'te ağaç sayısına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 24: RF ağaç sayısı – MSE ilişkisi (Grup 2)

Şekil 24’te görüldüğü üzere ağaç sayısında meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 22’de en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 22: RF optimum model parametreleri (2. Grup)

Parametre	Optimum Değer
<i>max_depth</i>	28
<i>max_features</i>	5
<i>n_estimators</i>	28

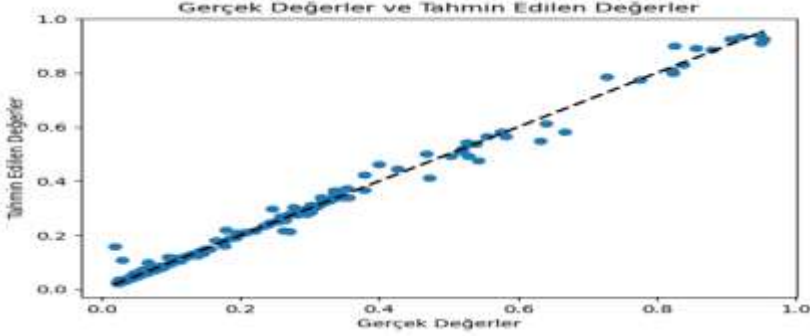
Tablo 22’de rassal orman regresyon modeli için optimum parametreler gösterilmektedir. Buna göre maksimum derinliği 28, maksimum özelliği 5 ve 28 adet karar ağacı ile optimum bir model elde edilebilmektedir. Bu değerlere göre model oluşturulmuş ve optimum model elde edilmiştir.

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile Rassal Orman Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 23’te modele ilişkin performans ölçütleri yer almaktadır.

Tablo 23: RF performans ölçütleri (2. Grup)

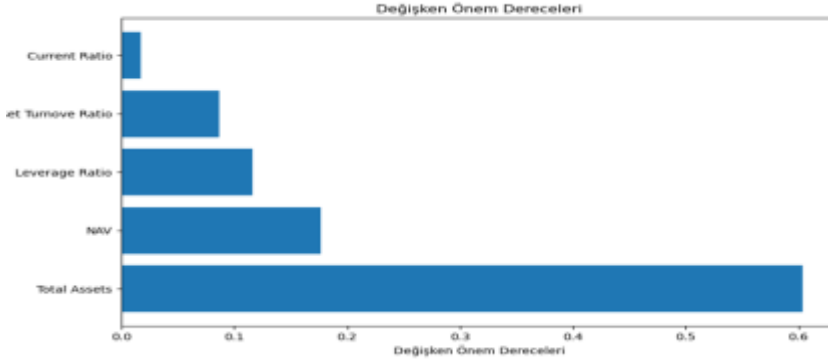
Ölçüt	Değer
R^2	0.98
<i>MSE</i>	0.0178
<i>MAE</i>	0.0091

Tablo 23’e göre, R^2 değeri %98 olarak, hata değerleri ise 0’a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluşturduğunu göstermektedir. Şekil 25’te modele ait grafik gösterilmektedir.



Şekil 25: RF model grafiği (2. Grup)

Rassal orman regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, Rassal orman regresyon yöntemine göre çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 26'da, modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.



Şekil 26: RF değişken önem dereceleri (2. Grup)

Şekil 26'ya göre temettü dağıtma zorunluluğu bulunmayan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem derecelerine göre gösterilmektedir. Buna göre kârlılık üzerinde en önemli etkiye sahip değişken firma büyüklüğüdür (TA). Cari oran değişkeninin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır. 1. ve 2. Grup sonuçları karşılaştırıldığında, firma kârlılıkları üzerinde varlıkların önemli bir etkiye sahip olduğu görülmektedir.

Bu bölümde, her iki gruba ait veriler ile Rassal Orman Regresyon uygulaması gerçekleştirilmiştir. Yöntem ile her iki grup için başarılı sonuçlar elde edilmiştir. Firma büyüklüğü (TA) her iki grupta da kârlılığın önemli bir belirleyicisi olarak öne çıkmaktadır. Diğer bölümde her iki grup için Catboost Regresyon uygulamaları yer almaktadır.

5.9. CatBoost Regresyon Uygulaması

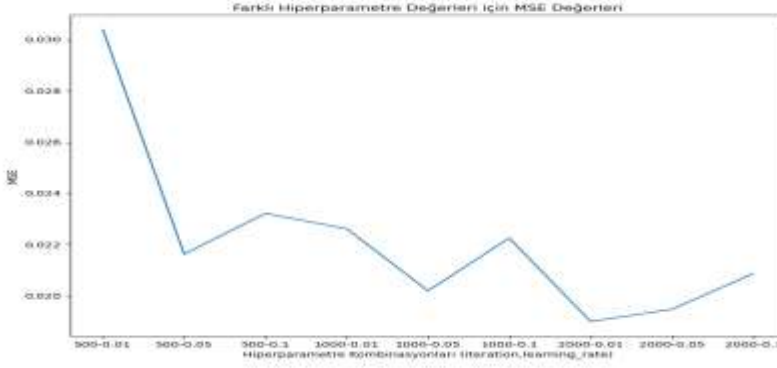
CatBoost regresyon yöntemi (CB) ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna (1. Grup) daha sonra temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. Uygulama adımları her grup için şu şekildedir:

1. İlk olarak her bir yöntem için hiperparametre ayarlaması yapılması.
2. Tespit edilen optimum hiperparametreler ile modelin oluşturulması.
3. Modelin performans ölçütlerinin tespit edilmesi.
4. Modelde yer alan değişkenlerin önem derecelerinin tespit edilmesi.

CatBoost yöntemi uygulaması, bu adımları kapsayacak şekilde her iki grup için gerçekleştirilmiştir. Her iki gruba ait uygulama sonuçları ayrı başlıklar altında yer bu bölümde yer almaktadır.

5.9.1. CatBoost Regresyon Uygulamaları (1. Grup)

CatBoost Regresyon yöntemi ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna (1. Grup) yönelik uygulanmıştır. İlk olarak parametre ayarlaması yapılmaksızın model oluşturulmuştur. Şekil 27’de farklı hiperparametre kombinasyonlarına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 27: CB hiperparametre – MSE ilişkisi (1. Grup)

Şekil 27’de görüldüğü üzere hiperparametrelerde meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 24’te en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 24: CB optimum model parametreleri (1. Grup)

Parametre	Optimum Değer
<i>depth</i>	4
<i>iterations</i>	2000
<i>learning_rate</i>	0.1
<i>loss_function</i>	MAE

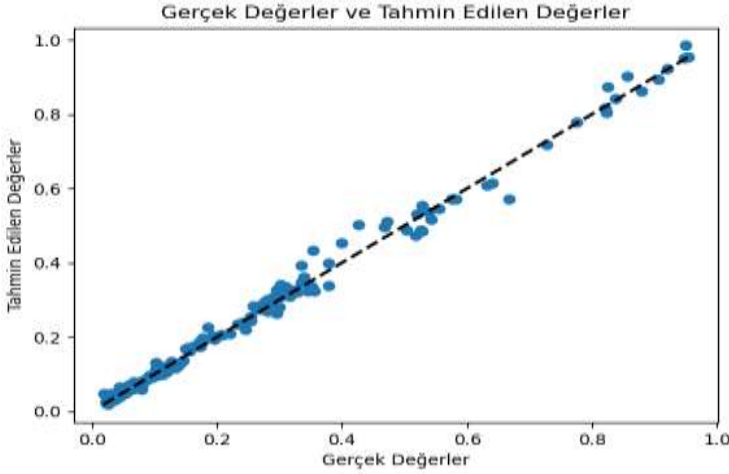
Tablo 24’e CatBoost Regresyon modeli için optimum parametreler gösterilmektedir. Bu değerlere göre model oluşturulmuş ve optimum model elde edilmiştir.

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile CatBoost Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 25’de modele ilişkin performans ölçütleri yer almaktadır.

Tablo 25: CB performans ölçütleri (1. Grup)

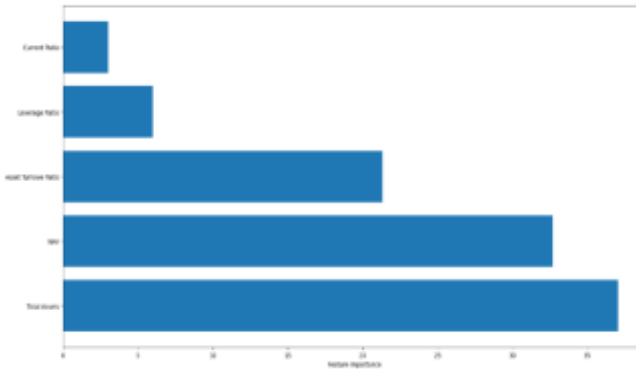
Ölçüt	Değer
R^2	0.9936
<i>MSE</i>	0.0179
<i>MAE</i>	0.0104

Tablo 25'e göre, R2 değeri %99 olarak, hata değerleri ise 0'a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluştuğunu göstermektedir. Şekil 28'da modele ait grafik gösterilmektedir.



Şekil 28: CB model grafiği (1. Grup)

CatBoost orman regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, CatBoost regresyon yöntemine göre çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 29'da modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.

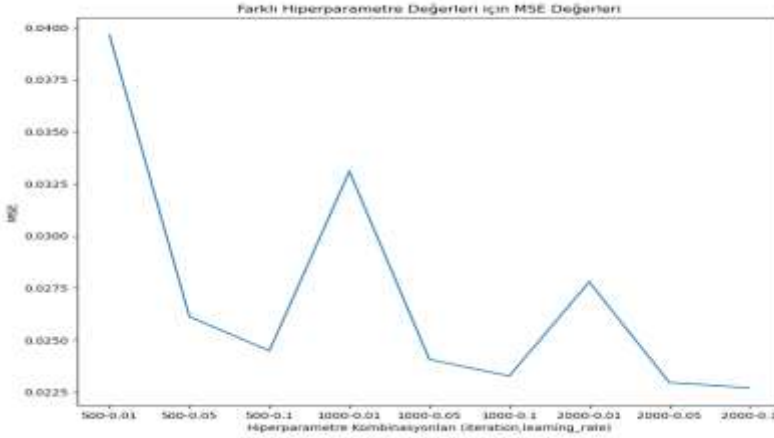


Şekil 29: CB değişken önem dereceleri (1. Grup)

Şekil 29'a göre temettü dağıtma zorunluluğu bulunan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem derecelerine göre gösterilmektedir. Buna göre kârlılık üzerinde en önemli etkiye sahip değişken firma büyüklüğüdür (TA). Net Varlık Değeri ve Duran varlık Devir Hızı değişkenleri de kârlılık üzerinde önemli etkiye sahip diğer değişkenlerdir. Kaldıraç oranı ve cari oran değişkenlerinin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır. Diğer bölümde, CatBoost Regresyon uygulaması 2. Grup verileri ile gerçekleştirilmiştir.

5.9.2. CatBoost Regresyon Uygulamaları (2. Grup)

Bu bölümde, CatBoost regresyon yöntemi belirli bir oranda temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. İlk olarak parametre ayarlaması yapılmaksızın model oluşturulmuştur. Şekil 30'da farklı hiperparametre kombinasyonlarına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 30: CB hiperparametre – MSE ilişkisi (2. Grup)

Şekil 30'da görüldüğü üzere hiperparametrelerde meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 26'da en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 26: CB optimum model parametreleri (2. Grup)

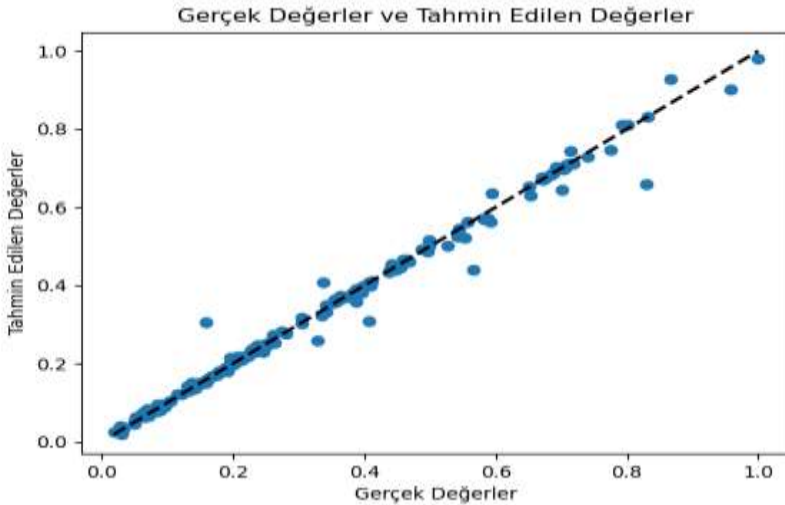
Parametre	Optimum Değer
<i>depth</i>	6
<i>iterations</i>	2000
<i>learning_rate</i>	0.1
<i>loss_function</i>	RMSE

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile CatBoost Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 27’de modele ilişkin performans ölçütleri yer almaktadır.

Tablo 27: CB performans ölçütleri (2. Grup)

Ölçüt	Değer
R^2	0.9896
<i>MSE</i>	0.0226
<i>MAE</i>	0.0090

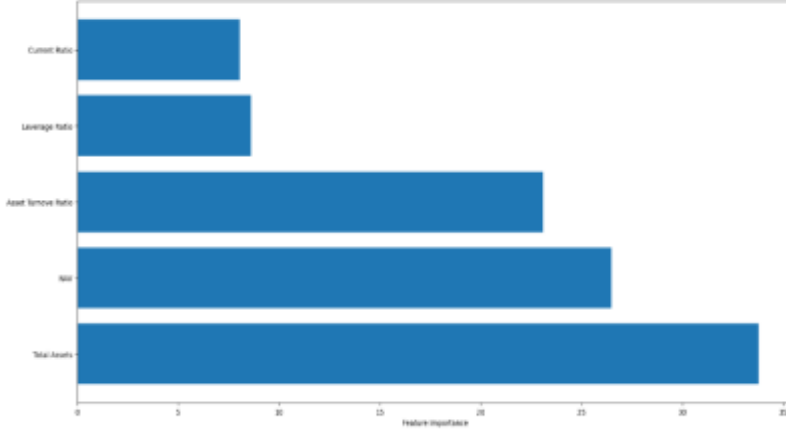
Tablo 27’ye göre, R^2 değeri %99 olarak, hata değerleri ise 0’a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluştuğunu göstermektedir. Şekil 31’de modele ait grafik gösterilmektedir.



Şekil 31: CB model grafiği (2. Grup)

CatBoost regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, CatBoost regresyon yöntemine göre

çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 32’de, modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.



Şekil 32: CB değişken önem dereceleri (2. Grup)

Şekil 32’ye göre temettü dağıtma zorunluluğu bulunmayan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem derecelerine göre gösterilmektedir. Buna göre kârlılık üzerinde en önemli etkiye sahip değişken firma büyüklüğüdür (TA). Net Varlık Değeri ve Duran varlık Devir Hızı değişkenleri de kârlılık üzerinde önemli etkiye sahip diğer değişkenlerdir. Kaldıraç oranı ve cari oran değişkenlerinin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır.

Bu bölümde, her iki gruba ait veriler ile CatBoost Regresyon uygulaması gerçekleştirilmiştir. Yöntem ile her iki grup için başarılı sonuçlar elde edilmiştir. Firma büyüklüğü (TA) ve Net Varlık Değeri (NAV) her iki grupta da kârlılığın önemli bir belirleyicileri olarak öne çıkmaktadır. Diğer bölümde her iki grup için XGBoost Regresyon uygulamaları yer almaktadır.

5.10. XGBoost Regresyon Uygulaması

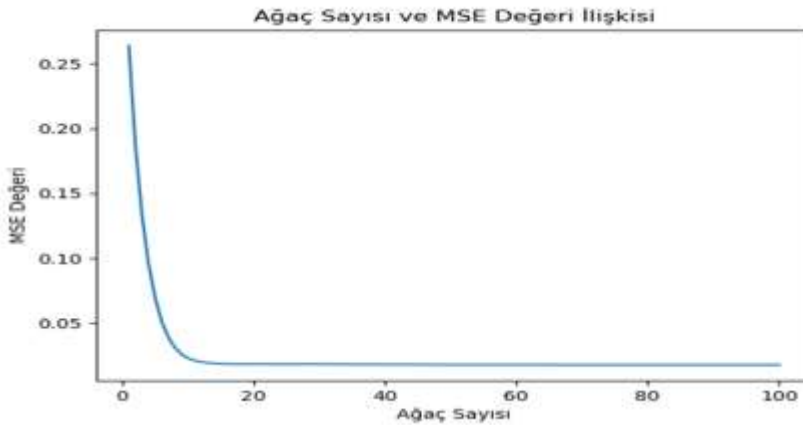
XGBoost regresyon yöntemi (XGB) ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna (1. Grup) daha sonra temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. Uygulama adımları her grup için şu şekildedir:

1. İlk olarak her bir yöntem için hiperparametre ayarlaması yapılması.
2. Tespit edilen optimum hiperparametreler ile modelin oluşturulması.
3. Modelin performans ölçütlerinin tespit edilmesi.
4. Modelde yer alan değişkenlerin önem derecelerinin tespit edilmesi.

XGBoost Regresyon yöntemi uygulaması, bu adımları kapsayacak şekilde her iki grup için gerçekleştirilmiştir. Her iki gruba ait uygulama sonuçları ayrı başlıklar altında yer bu bölümde yer almaktadır.

5.10.1. XGBoost Regresyon Uygulamaları(1. Grup)

XGBoost Regresyon yöntemi ilk olarak belirli bir oranda temettü dağıtma zorunluluğu bulunan firma grubuna yönelik uygulanmıştır. Şekil 33'te ağaç sayısına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 33: CB değişken önem dereceleri (2. Grup)

Şekil 33'te görüldüğü üzere ağaç sayısında meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 28'de en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 28: XGB optimum model parametreleri (1. Grup)

Parametre	Optimum Değer
<i>gamma</i>	0
<i>learning_rate</i>	0.1
<i>max_depth</i>	3
<i>n_estimators</i>	150

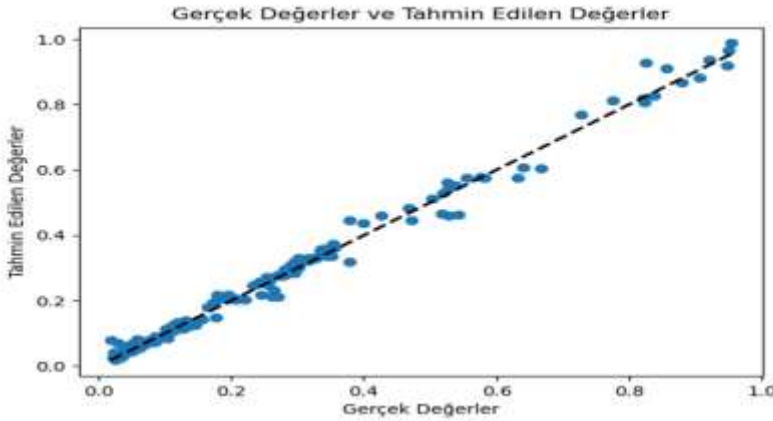
Tablo 28'e XGBoost Regresyon modeli için optimum parametreler gösterilmektedir. Bu değerlere göre model oluşturulmuş ve optimum model elde edilmiştir.

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile XGBoost Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 29'da modele ilişkin performans ölçütleri yer almaktadır.

Tablo 29: XGB performans ölçütleri (1. Grup)

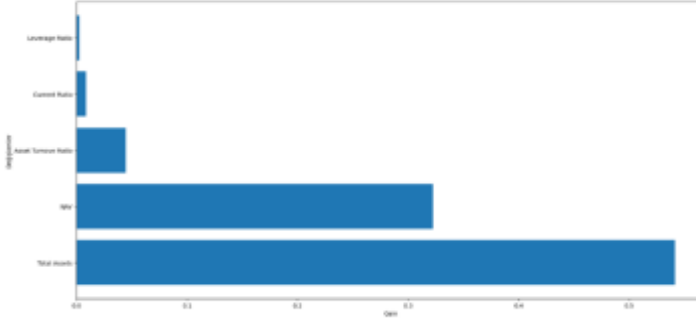
Ölçüt	Değer
R^2	0.991
MSE	0.0201
MAE	0.0124

Tablo 29'a göre, R2 değeri %99 olarak, hata değerleri ise 0'a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluşturduğunu göstermektedir. Şekil 34'te modele ait grafik gösterilmektedir.



Şekil 34: XGB model grafiği (1. Grup)

XGBoost orman regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, XGBoost regresyon yöntemine göre çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 35'te, modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.

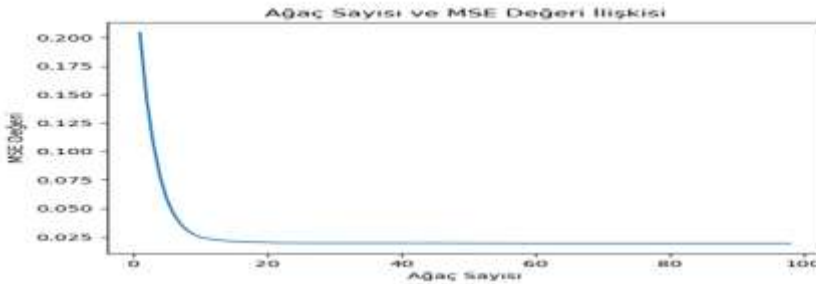


Şekil 35: XGB değişken önem dereceleri (1. Grup)

Şekil 35'e göre temettü dağıtma zorunluluğu bulunan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem derecelerine göre gösterilmektedir. Buna göre kârlılık üzerinde en önemli etkiye sahip değişken firma büyüklüğüdür (TA). Net varlık değeri (NAV) de kârlılık üzerinde önemli etkiye sahip diğer değişkendir. Cari oran ve kaldıraç oranı değişkenlerinin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır. Diğer bölümde, XGBoost Regresyon uygulaması, 2. Grup verileri ile gerçekleştirilmiştir.

5.10.2. XGBoost Regresyon Uygulamaları(2. Grup)

Bu bölümde, XGBoost regresyon yöntemi belirli bir oranda temettü dağıtma zorunluluğu bulunmayan firma grubuna (2. Grup) yönelik uygulanmıştır. İlk olarak parametre ayarlaması yapılmaksızın model oluşturulmuştur. Şekil 36'da ağaç sayısına göre MSE değerinde meydana gelen değişim gösterilmektedir.



Şekil 36: XGB ağaç sayısı – MSE ilişkisi (2. Grup)

Şekil 36'da görüldüğü üzere ağaç sayısında meydana gelen değişim model performansını etkilemektedir. Bu sebeple, en uygun modeli oluşturmak amacı ile ızgara araması (grid search) yöntemi ile hiper parametre optimizasyonu gerçekleştirilmiştir. Tablo 30'da en iyi model için gerekli olan, ızgara araması sonucunda elde edilen parametreler yer almaktadır.

Tablo 30: XGB optimum model parametreleri (2. Grup)

Parametre	Optimum Değer
<i>gamma</i>	0
<i>learning_rate</i>	0.1
<i>max_depth</i>	5
<i>n_estimators</i>	150

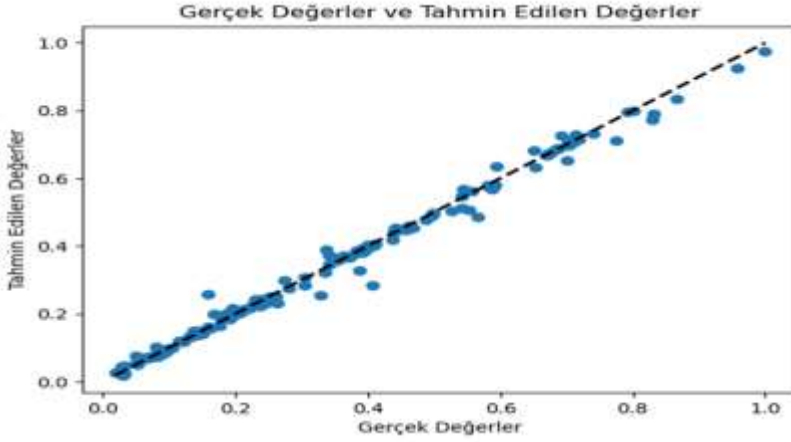
Tablo 30'da XGBoost Regresyon modeli için optimum parametreler gösterilmektedir. Bu değerlere göre model oluşturulmuş ve optimum model elde edilmiştir.

Izgara araması (grid search) yöntemine göre elde edilen hiperparametre değerleri ile XGBoost Regresyon modeli 1. Grup için oluşturulmuştur. Tablo 31'de modele ilişkin performans ölçütleri yer almaktadır.

Tablo 31: XGB performans ölçütleri (2. Grup)

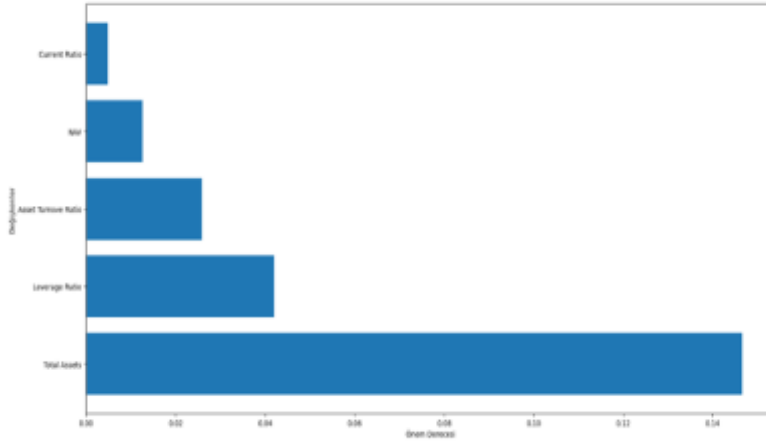
Ölçüt	Değer
R^2	0.992
<i>MSE</i>	0.0191
<i>MAE</i>	0.0097

Tablo 31'e göre, R^2 değeri %99 olarak, hata değerleri ise 0'a yakın tespit edilmiştir. Bu değerler, kârlılığın belirleyicilerinin tespiti için güçlü bir model oluşturduğunu göstermektedir. Şekil 37'de modele ait grafik gösterilmektedir.



Şekil 37: XGB model grafiği (2. Grup)

XGBoost regresyon yöntemine göre kurulan model başarılı bir performans göstermiştir. Daha sonra, XGBoost Regresyon yöntemine göre çalışmada yer alan değişkenlerin önem derecesi tespit edilmiştir. Şekil 38’de, modele göre bağımsız değişkenlerin önem dereceleri gösterilmektedir.



Şekil 38: XGB değişken önem dereceleri (2. Grup)

Şekil 38’e göre temettü dağıtma zorunluluğu bulunan GYO firmalarının, kârlılıkları üzerinde etkili olan faktörler önem dereceleri gösterilmektedir.

Buna göre kârlılık üzerinde en önemli etkiye sahip değişken firma büyüklüğüdür (TA). Net varlık değeri ve cari oran değişkenlerinin ise kârlılık üzerindeki etkisi diğer değişkenlere göre daha azdır.

Bu bölümde, her iki gruba ait veriler ile XGBoost Regresyon uygulaması gerçekleştirilmiştir. Yöntem ile her iki grup için başarılı sonuçlar elde edilmiştir. Firma büyüklüğü (TA) her iki grupta da kârlılığın önemli bir belirleyicisi olarak öne çıkmaktadır. Sonraki bölümde, üç farklı yöntem tarafından elde edilen sonuçlar karşılaştırılmaktadır.

5.11. Tartışma

Bu bölümde çalışmada kullanılan yöntemlere ait performansların karşılaştırılması yapılacaktır. Bu sayede, GYO firmalarına ait veriler üzerinde en başarılı tahmini yapan yöntem tespit edilecektir.

İlk olarak, 1.gruba ait veriler ile gerçekleştirilen uygulamalar karşılaştırılacaktır. Tablo 32’de bu gruba ait verilere uygulanan yöntemlerin performans ölçütleri gösterilmektedir.

Tablo 32: Yöntem performanslarının karşılaştırılması (1. Grup)

Yöntem	Önem Sırası	Performans Ölçütleri		
		R^2	MSE	MAE
RF	3	0.887	0.0215	0.0122
CB	1	0.994	0.0179	0.0104
XGB	2	0.991	0.0201	0.0124

Tablo 32’ye göre tüm yöntemler başarılı modeller üretmişlerdir. Bu yöntemler arasında en başarılı yöntemin CatBoost Regresyon yöntemi olduğu görülmektedir. Diğer yöntemlere göre daha yüksek R2 değerine ve daha düşük hata oranlarına sahip olması yöntemin diğer yöntemlere göre daha başarılı olduğunu göstermektedir. Bu yöntemi sırası ile XGBoost ve Rastgele Orman Regresyon yöntemleri takip etmektedir. Tablo 33’te 2. gruba ait verilere uygulanan yöntemlerin performans ölçütleri gösterilmektedir.

Tablo 33: Yöntem performanslarının karşılaştırılması (2. Grup)

Yöntem	Önem Sırası	Performans Ölçütleri		
		R^2	MSE	MAE
RF	3	0.9823	0.0295	0.0120
CB	2	0.9896	0.0226	0.0090
XGB	1	0.9920	0.0191	0.0097

Tablo 33'e göre tüm yöntemler başarılı modeller üretmişlerdir. Bu yöntemler arasında en başarılı yöntemin XGBoost Regresyon yöntemi olduğu görülmektedir. Diğer yöntemlere göre daha yüksek R2 değerine ve daha düşük hata oranlarına sahip olması yöntemin diğer yöntemlere göre daha başarılı olduğunu göstermektedir. Bu yöntemi sırası ile CatBoost ve Rassal Orman Regresyon yöntemleri takip etmektedir.

Her iki gruba ait veriler üzerinden uygulamalar yapılmıştır. Uygulamalar sonucunda yöntemler performans ölçütlerine göre karşılaştırılmıştır. Elde edilen bulgular, çalışmada kullanılan dört yöntemde başarılı sonuçlar elde ettiğini göstermektedir. XGBoost ve CatBoost Regresyon yöntemlerinin, Rassal Orman Regresyon yöntemine göre daha başarılı sonuçlar elde ettiği tespit edilmiştir.

Tablo 34'te her iki grup ve dört yöntemle göre kârlılık üzerinde en önemli etkiye sahip değişkenler yer almaktadır.

Tablo 34: Değişken önem sıraları

Grup	Yöntem	Değişken Önem Sırası				
		1.	2.	3.	4.	5.
1. Grup	RF	Net Varlık Değeri	Toplam Aktifler	Kaldıraç Oranı	Duran Varlık Devir Hızı	Cari Oran
	CB	Toplam Aktifler	Net Varlık Değeri	Duran Varlık Devir Hızı	Kaldıraç Oranı	Cari Oran
	XGB	Toplam Aktifler	Net Varlık Değeri	Varlık Devir Hızı	Cari Oran	Kaldıraç Oranı
2. Grup	RF	Toplam Aktifler	Net Varlık Değeri	Kaldıraç Oranı	Duran Varlık Devir Hızı	Cari Oran
	CB	Toplam Aktifler	Net Varlık Değeri	Duran Varlık Devir Hızı	Kaldıraç Oranı	Cari Oran
	XGB	Toplam Aktifler	Kaldıraç Oranı	Duran Varlık Devir Hızı	Net Varlık Değeri	Cari Oran

Tablo 34'te görüldüğü üzere, firma büyüklüğü değişkeni 1. Grup GYO firmaları için Rassal Orman Regresyon yöntemi hariç diğer tüm yöntemlerde en önemli değişken olarak tespit edilmiştir. 2. Grup GYO firmaları için uygulanan tüm yöntemlerde ise firma büyüklüğü göstergesi en önemli değişken olarak tespit edilmiştir. Bu sonuç, GYO firmaları kârlılıkları için varlık

oranlarının önemli bir etkiye sahip olduğunu göstermektedir. Bu tespitin, GYO firmalarının özellikle duran varlık ağırlıklı bir varlık portföyüne sahip olmalarından kaynaklandığı tahmin edilmektedir. Ayrıca, Cari Oran değişkeninin her iki grup için diğer değişkenlere göre daha az öneme sahip olduğu görülmektedir. Bu sonuç, GYO firmalarının kısa vadeli borçlarını ödeme gücünün kârlılık üzerinde diğer değişkenlere göre daha az öneme sahip olduğunu göstermektedir.

En iyi model, 1. Grup için CatBoost yöntemi ile oluşturulur iken 2. Grup için XGBoost yöntemi ile oluşturulmuştur. Her iki yöntemle elde edilen sonuçlar karşılaştırıldığında finansal oranların önem sırası 2 ve 4. Sıralarda değişmektedir. Temettü dağıtma zorunluluğu bulunmayan firmalarda 2. Önem sırasında Net Varlık Değeri yer almaktadır. Temettü dağıtma zorunluluğu bulunmayan firmalarda ise Kaldıraç Oranı yer almaktadır. Bu durum temettü dağıtma zorunluluğu bulunmayan firmaların elde ettikleri fonlar ile borçlarını ödemelerinden kaynaklanabilir.

BÖLÜM 6

SONUÇ VE ÖNERİLER

Çalışmada, konut GYO firmalarının kârlılığını etkileyen finansal göstergelerin önem sıralamalarını belirlemek ve bu amaç doğrultusunda kullanılan üç farklı veri madenciliği yöntemini karşılaştırmak amaçlanmıştır. Bu bağlamda, Rassel Orman Regresyonu, XGBoost ve CatBoost gibi ağaç tabanlı topluluk öğrenme yöntemleri kullanılmış ve karşılaştırılmıştır. Her üç yöntem de topluluk yaklaşımını benimsemekte olup, karar ağaçları üzerine kurulu bir yapıya sahiptir. Bununla birlikte, her bir yöntemin kendine özgü özellikleri ve avantajları bulunmaktadır. Bulgular, konut GYO firmalarının kârlılığını artırmak ve yönetim stratejilerini geliştirmek amacıyla uygulanabilir öneriler sunma potansiyeline sahiptir. Sonuçlar, kullanılan yöntemler arasında XGBoost ve CatBoost yöntemlerinin Rassel Orman Regresyonu yöntemine göre daha başarılı sonuçlar elde ettiğini göstermiştir. Bu tespitler, Sadaf (2023)'ın zararlı web sitelerini tespit etmeyi amaçladığı çalışma ve Kanaparthi (2023)'nin finansal kurumlarda kredi riskini tahmin etmeyi amaçladığı çalışmanın bulguları ile uyumaktadır. Bu çalışmalarda da XGBoost ve CatBoost yöntemleri diğer yöntemlere göre daha başarılı sonuçlar elde etmiştir. Sonuç olarak, XGBoost ve CatBoost, konut GYO firmalarının kârlılığını etkileyen faktörleri daha iyi tahmin etme yeteneğine sahiptir. Bu durum, veri madenciliği tekniklerinin kârlılık analizinde etkin bir şekilde kullanılabileceğini göstermektedir.

Analiz sonuçları, konut GYO firmalarının kârlılığı üzerinde varlıkların önemli belirleyici bir faktör olduğunu göstermektedir. Bu bulgular, yöneticilerin etkili bir varlık yönetimine odaklanmasının önemini vurgulamaktadır. Adnan vd. (2021), tarafından yapılan çalışma, GYO firmalarının etkili varlık yönetiminin gelirleri en üst düzeye çıkarma, giderleri en aza indirme ve mülk değerlerini artırma konularında önemli bir etkiye sahip olduğunu vurgulamaktadır. Bu etkili yönetim pratiği, işletme performansını iyileştirerek firma kârlılığında artış sağlama potansiyeline sahiptir. Bu bulgular, elde ettiğimiz sonuçla uyumlu bir şekilde varlıkların kârlılık üzerindeki etkisini vurgulamaktadır. Lubyanya vd. (2016) tarafından yapılan çalışma, duran varlıkların kârlılık üzerinde önemli bir etkiye sahip olduğunu ve yöneticilerin

duran varlık yönetimine önem vermesi gerektiğini vurgulanmaktadır. GYO firmalarının duran varlık ağırlıklı olduğu düşünüldüğünde, bu bulgu önemli bir tespittir. Ayrıca, farklı sektörler için yapılan çalışmalarda da (Banamtuan vd., 2020; Carstina vd., 2015) varlık yönetiminin kârlılık için önemli olduğu vurgulanmaktadır. Bu bağlamda, elde ettiğimiz sonuçlar ve literatürdeki diğer çalışmalar, GYO firmalarının varlık yönetiminin önemini vurgulamaktadır. Varlıkların etkin bir şekilde yönetilmesi, gelirlerin maksimize edilmesi, giderlerin minimize edilmesi ve mülk değerlerinin artırılmasıyla birlikte firma kârlılığında olumlu etkilerin gözlemlenebileceği sonucuna varılmıştır.

Firma Yöneticileri İçin Öneriler

Bu bölümde, konut GYO firmalarının kârlılığını artırmayı hedefleyen yöneticilere rehberlik etmek ve varlık yönetimi stratejilerini geliştirmelerine yardımcı olmak amacıyla 2 alanda öneriler sunulmaktadır. Bu öneriler aşağıda maddeler halinde yer almaktadır.

1. Varlık Yönetimi: Konut GYO firmalarının kârlılığını artırmayı hedefleyen yöneticiler, literatürde de yer alan tespitlere göre (Adnan vd., 2021; Banamtuan vd., 2020; Carstina vd., 2015; Lubyanyaya vd., 2016) varlık yönetimine odaklanmaları önem taşımaktadır. Etkin varlık yönetimi ile gelirlerin maksimize edilmesi, giderlerin minimize edilmesi ve mülk değerlerinin artırılmasıyla birlikte kârlılığın artırılması beklenmektedir. Yöneticiler, firma varlıklarını stratejik olarak yönetmeli ve potansiyel büyüme fırsatlarını değerlendirerek büyüklüklerini artırmayı göz önünde bulundurmalıdır.

2. Veri Madenciliği Yöntemlerinin Kullanımı: Çalışmanın bulguları, veri madenciliği yöntemlerinin konut GYO firmalarının kârlılığını analiz etmede etkili bir araç olduğunu göstermektedir. Yöneticilerin, varlık portföylerinin etkin bir şekilde yönetilmesi ve kârlılık potansiyelinin maksimize edilmesi için veri madenciliği yöntemlerinin kullanımını teşvik etmesi beklenmektedir. Özellikle XGBoost ve CatBoost gibi yöntemlerin kârlılığı tahmin etmede daha başarılı olduğu tespit edilmiştir. Literatürde de bu yöntemlerin diğer yöntemler ile kıyaslandığında başarılı sonuçlar elde ettiğini gösteren çalışmalar (Bentéjac vd., 2021; Emami vd., 2023; Kanaparthi, 2023; Lee, 2014; Luo vd., 2021; Rezapour ve Hansen, 2022; Sadaf, 2023; Tiwari vd., 2021). Ayrıca, Latha ve Bommi (2023) ve (Shobana Bai, 2023) tarafından

yapılan çalışmalarda CatBoost yönteminin başarısı gösterilmektedir. Gumus ve Kiran, (2017) yapılan çalışma da ise XGBoost yönteminin başarısı gösterilmektedir. Yöneticiler, bu yöntemleri uygulamak ve sonuçları analiz etmek için gerekli kaynakları ve uzmanlığı sağlamalıdır.

Yasa Yapıcılar İçin Öneriler

Çalışmanın sonuçları, yasa yapıcılara önemli bir perspektif sunmaktadır. Konut GYO firmalarının kârlılığını desteklemek ve sektörün sürdürülebilir büyümesini sağlamak için düzenlemeler ve politikalar gözden geçirilebilir. Ayrıca, veri madenciliği yöntemlerinin kullanımının teşvik edilmesi ve bu yöntemlerin doğru kullanımını sağlayacak eğitim programlarının geliştirilmesi önerilmektedir. Bu şekilde, sektörde daha bilinçli kararlar alınabilir, daha etkili stratejiler uygulanabilir ve kârlılık artırılabilir. Yasa yapıcıların dikkate almaları gereken bazı öneriler şunlardır:

1. Veri Madenciliği Yöntemlerinin Teşviki: Yasa yapıcılar, konut gayrimenkul yatırım ortaklıklarının kârlılığını artırmak için veri madenciliği yöntemlerinin kullanımını teşvik edebilirler. Bu yöntemler, sektörde daha bilinçli ve veriye dayalı kararlar alınmasını sağlayabilir. Yasa yapıcılar, veri madenciliği tekniklerinin kullanımını teşvik etmek için eğitim programları düzenleyebilir veya teşvikler sağlayabilirler.

2. Gayrimenkul Yatırım Ortaklıklarının Desteklenmesi: Yasa yapıcılar, konut gayrimenkul yatırım ortaklıklarının varlık yönetimi konusunda daha fazla desteklenmesini sağlayabilirler. Bu destek, sektörde daha etkin varlık yönetimi stratejilerinin geliştirilmesine ve uygulanmasına yardımcı olabilir. Yasa yapıcılar, sektörün sürdürülebilirliğini sağlamak için düzenlemelerde esneklik ve teşvikler sunabilirler. 11. Kalkınma Planı (T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı, 2019) Madde 685 ve 686 da konut ihtiyacı ifade edilmektedir. Madde 685 “Dar gelirliler başta olmak üzere, herkesin yeterli, yaşanabilir, dayanıklı, güvenli, kapsayıcı, ekonomik olarak karşılanabilir, sürdürülebilir, iklim değişikliğine dirençli, temel altyapı hizmetlerine sahip konuta erişiminin sağlanması temel amaçtır.” ifadesiyle konut erişiminin önemini belirtmektedir. Madde 686 ise “Kentleşme, nüfus artışı, yenileme ve afetten kaynaklanan konut ihtiyacı arz-talep dengesi gözetilerek karşılanacaktır.” konut ihtiyacının arz ve talep dengesine göre ele alınması gerektiğini ifade etmektedir. Bu bağlamda, konut gayrimenkul yatırım

ortaklıklarının desteklenmesi, konut ihtiyacının karşılanması ve 11. Kalkınma Planı'nın hedeflerine ulaşılması için önemli bir adım olabilir.

3. Yatırımcı Güveninin Güçlendirilmesi: Yasa yapıcılar, yatırımcı güvenini güçlendirecek politikaları gözden geçirebilirler. Şeffaflık, hesap verilebilirlik ve denetim mekanizmalarının güçlendirilmesi, yatırımcıların sektöre olan güvenini artırabilir. Bu da konut gayrimenkul yatırım ortaklıklarının sermaye çekme ve büyüme potansiyelini artırabilir.

4. Sektörün Ekonomik Büyüme Katkısı: Yasa yapıcılar, konut gayrimenkul yatırım ortaklıklarının sektörün ekonomik büyümesine katkıda bulunmasını teşvik edebilirler. Özellikle uzun vadeli yatırımları destekleyen düzenlemeler ve politikalar oluşturarak, sektörün istikrarlı ve sürdürülebilir bir büyüme elde etmesine sağlayabilirler.

Yatırımcılar İçin Öneriler

Çalışma sonucunda elde edilen tespitlere göre yatırımcılar için birtakım önerilerde bulunulabilir. Bu öneriler aşağıda maddeler halinde yer almaktadır.

1. Risk Değerlendirmesi: Yüksek bir net varlık değeri oranı, genellikle daha az finansal risk anlamına gelmektedir. Yatırımcılar, daha düşük finansal risk taşıyan konut GYO'ları tercih edebilirler.

2. Uzun Vadeli Yatırım: Yatırımcılar, büyük konut GYO'ları uzun vadeli yatırım stratejilerine dahil edebilirler. Büyük bir varlık tabanı genellikle daha istikrarlı bir gelir akışına ve daha fazla değer artışı potansiyeline işaret eder. Yatırımcılar, bu tür konut GYO'ları seçerek uzun vadeli karlılık fırsatları yakalayabilirler.

Araştırmacılar İçin Öneriler

Bu bölümde, araştırmacılara farklı bir bakış açısı kazandırmak amacı ile öneriler sunulmaktadır. Bu öneriler aşağıda maddeler halinde yer almaktadır.

1. Veri Madenciliği Yöntemlerinin Kullanımı: Veri madenciliği yöntemlerinin, konut GYO firmalarının kârlılığını analiz etmek için etkili bir yaklaşım olduğu bu çalışma tarafından gösterilmiştir. Araştırmacılara, farklı sektörlerde ve şirketlerde de benzer analizleri gerçekleştirmeleri ve sonuçları karşılaştırmaları önerilmektedir. Bu şekilde, veri madenciliği yöntemlerinin genel bir uygulanabilirliği ve etkinliği değerlendirilebilir. Örneğin farklı alanlarda yapılan çalışmalarda (Almaskati, 2022; H. Chen, 2023; Gumus ve

Kiran, 2017; Kanaparthi, 2023; Muhasshanah vd., 2023; Sari vd., 2023) da veri madenciliği yöntemlerin başarılı sonuçlar elde ettiği görülmektedir.

2. Ağaç Tabanlı ve Topluluk Öğrenme Yöntemlerinin Kullanımı: Ağaç tabanlı topluluk öğrenme yöntemlerinin kullanımı, konut GYO firmalarının kârlılığını etkileyen faktörleri tahmin etmek için başarılı sonuçlar vermiştir. Araştırmacılara, farklı topluluk öğrenme yöntemlerini, farklı ağaç tabanlı yöntemleri deneyerek sonuçları karşılaştırmaları ve en iyi performans gösteren yöntemi belirlemeleri önerilir. Örneğin, Demir ve Sahin (2022) çalışmalarında geleneksel Rassal orman yöntemi ile modern ağaç tabanlı yöntemleri karşılaştırmışlardır. Ayrıca, diğer veri madenciliği karşılaştırmalı çalışmalar yaparak sonuçların güvenilirliğini doğrulamak da önemlidir. Örneğin, Abdi vd. (2023); Bentéjac vd. (2021); Emami vd. (2023); Jhaveri vd. (2019) ve Luo vd. (2021) tarafından yapılan çalışmalarda bu şekilde karşılaştırmalar yer almaktadır.

KAYNAKÇA

- Abdi, Y., Momeni, E., & Armaghani, D. J. (2023). Elastic modulus estimation of weak rock samples using random forest technique. *Bulletin of Engineering Geology and the Environment*, 82(5), 176. <https://doi.org/10.1007/s10064-023-03154-y>
- Acharya, S. (2021, Mayıs 14). What are RMSE and MAE? <https://towardsdatascience.com/what-are-rmse-and-mae-e405ce230383>.
- Adnan, Y. Mohd., Lamin, N., Razali, M. N., Jalil, R. A., & Esha, Z. (2021). Real estate investment trusts' (REITs) asset management strategies within global REIT portfolios. *Real Estate Management and Valuation*, 29(1), 72-86. <https://doi.org/10.2478/remav-2021-0007>
- Ahmad Tarmizi, S. S., Mutalib, S., Abdul Hamid, N. H., & Abdul Rahman, S. (2019). A review on student attrition in higher education using big data analytics and data mining techniques. *International Journal of Modern Education and Computer Science*, 11(8), 1-14. <https://doi.org/10.5815/ijmeecs.2019.08.01>
- Akpınar, H. (2014). *Data: Veri madenciliği veri analizi*. Papatya Yayıncılık.
- Aktaş, M., & Darwish, Z. (2020). Türkiye'de gayrimenkul yatırım ortaklıkları ile girişim sermayesi yatırım ortaklıklarının finansal oranlar açısından karşılaştırılması. *International Journal of Social Science Research*, 9(1), 63-78. <https://dergipark.org.tr/en/pub/ijssresearch/issue/54863/727905>
- Alawi, S. J. S. Al, Jamil, J. M., & Shahanee, I. N. M. (2022). Predicting student performance using data mining approach: A case study in Oman. *Mathematical Statistician and Engineering Applications*, 71(4), 1389-1398.
- Albayrak, A. S., & Akbulut, R. (2008). Sermaye yapısını belirleyen faktörler: İMKB sanayi ve hizmet sektörlerinde işlem gören işletmeler üzerine bir inceleme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 22, 425-446.

- Allibhai, E. (2018, Kasım 3). Hold-out vs. Cross-validation in Machine Learning. <https://medium.com/@ejaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f>.
- Almaskati, N. (2022). The determinants of bank profitability and risk: A random forest approach. *Cogent Economics & Finance*, 10(1). <https://doi.org/10.1080/23322039.2021.2021479>
- Amatriain, X., Jaimes*, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. *Çinde Recommender Systems Handbook* (ss. 39-71). Springer US. https://doi.org/10.1007/978-0-387-85820-3_2
- Ambrose, B., & Linneman, P. (2001). REIT organizational structure and operating characteristics. *Journal of Real Estate Research*, 21(3), 141-162. <https://doi.org/10.1080/10835547.2001.12091049>
- Ambrose, B. W., Fuerst, F., Mansley, N., & Wang, Z. (2019). Size effects and economies of scale in European real estate companies. *Global Finance Journal*, 42. <https://doi.org/10.1016/J.GFJ.2019.04.004>
- Ambrose, B. W., Highfield, M. J., & Linneman, P. D. (2005). Real estate and economies of scale: The case of REITs. *Real Estate Economics*, 33(2), 323-350. <https://doi.org/10.1111/J.1540-6229.2005.00121.X>
- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The 'k' in k-fold cross validation. *20th European Symposium on Artificial Neural Networks*, 441-446.
- APREA. (2023). About - APREA Asia. <https://www.aprea.asia/about>
- Argüden, Y., & Erşahin, B. (2008). Veri madenciliği: veriden bilgiye, masraftan değere. ARGE Danışmanlık.
- Ashwini, B. P., Sumathi, R., & Sudhira, H. S. (2022). A dynamic model for bus arrival time estimation based on spatial patterns using machine learning.

International Journal of Engineering Trends and Technology, 70(9), 185-193. <https://doi.org/10.14445/22315381/IJETT-V70I9P219>

Aydoğan, E. K., Gencer, C., & Akbulut, S. (2008). Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesi. *Sigma : Mühendislik ve Fen Bilimleri Dergisi*, 26(1), 43-57.

Aytekin, S., & Kahraman, E. (2015). BIST gayrimenkul yatırım ortaklıkları endeksindeki (XGMYO) şirketlerin finansal etkinliklerinin veri zarflama analizi yöntemi ile değerlendirilmesi. *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 8(1), 289-301. <https://dergipark.org.tr/en/pub/niguiibfd/issue/19757/211539>

Baasith, A. (2021, Eylül 3). MSE vs MAE in machine learning. <https://baasith-shiyam1.medium.com/mse-vs-mae-in-machine-learning-714f586f968a>.

Badaoui, F., Amar, A., Ait Hassou, L., Zoglat, A., & Okou, C. G. (2017). Dimensionality reduction and class prediction algorithm with application to microarray Big Data. *Journal of Big Data*, 4(1), 32. <https://doi.org/10.1186/s40537-017-0093-4>

Bahadır, Y., & Haznedaroğlu, F. (2011). Konut finansmanı sistemleri ve türkiye uygulamasının geliştirilmesi. 6. İnşaat Yönetimi Kongresi.

Baker, M., Nagel, S., & Wurgler, J. (2006). The effect of dividends on consumption. <http://www.nber.org/papers/w12288>

Banamtuan, O., Zuhroh, D., & Sihwahjoeni, S. (2020). Asset management and capital ownership on firm value: Through profitability. *AFRE Accounting and Financial Review*, 3(1), 83-92.

Barclay, M. J., Smith, Jr, C. W., & Morellec, E. (2006). On the debt capacity of growth options. *The Journal of Business*, 79(1), 37-60.

Bartok, J., Habala, O., Bednar, P., Gazak, M., & Hluchý, L. (2010). Data mining and integration for predicting significant meteorological

- phenomena. *Procedia Computer Science*, 1(1), 37-46.
<https://doi.org/10.1016/J.PROCS.2010.04.006>
- Bayraktaroğlu, E. (2023). EPRA Global REIT Survey 2022.
- Ben-Shahar, D., Sulganik, E., & Tsang, D. (2020). Funds from operations versus net income: Examining the dividend relevance of reit performance measures. *Journal of Real Estate Research*, 33(3), 415-441.
<https://doi.org/10.1080/10835547.2011.12091308>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021a). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937-1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021b). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937-1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Bernard, S., Heutte, L., & Adam, S. (2009). On the selection of decision trees in Random Forests. 2009 International Joint Conference on Neural Networks, 302-307. <https://doi.org/10.1109/IJCNN.2009.5178693>
- Berry, M. W., Mohamed, A., & Yap, B. W. (2020). Supervised and unsupervised learning for data science (M. W. Berry, A. Mohamed, & B. W. Yap, Ed.). Springer International Publishing.
<https://doi.org/10.1007/978-3-030-22475-2>
- BI, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies*, 27(2), 87-101.
<https://doi.org/10.1111/j.1745-459X.2012.00370.x>
- Bilbao, I., & Bilbao, J. (2017). Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 173-177.
<https://doi.org/10.1109/INTELCIS.2017.8260032>

- Binti Mohamad, N. E. A., & bin Zolkifli, I. A. (2014). The determinant factors of real estate investment trust (REIT)'s performance: Evidence from Asian REITs. *Indonesian Capital Market Review*, 6(1), 53-59. <https://doi.org/10.21002/ICMR.V6I1.2987>
- Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14(1), 1-16. <https://doi.org/10.1186/1471-2105-14-106/FIGURES/7>
- Bose, I., & Mahapatra, R. K. (2001). Business data mining — a machine learning perspective. *Information & Management*, 39(3), 211-225. [https://doi.org/10.1016/S0378-7206\(01\)00091-X](https://doi.org/10.1016/S0378-7206(01)00091-X)
- Bossman, A., Umar, Z., & Teplova, T. (2022). Modelling the asymmetric effect of COVID-19 on REIT returns: A quantile-on-quantile regression analysis. *Journal of Economic Asymmetries*, 26. <https://doi.org/10.1016/J.JECA.2022.E00257>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. <https://doi.org/10.1007/BF00058655>
- Brighton, H., & Gigerenzer, G. (2012). Homo Heuristicus and the Bias–Variance Dilemma. *Action, Perception and the Brain*, 68-91. https://doi.org/10.1057/9780230360792_4
- Burnham, K. P., & Anderson, D. R. (2004). *Model selection and multimodel inference* (K. P. Burnham & D. R. Anderson, Ed.). Springer New York. <https://doi.org/10.1007/b97636>
- Buschjager, S., & Morik, K. (2018). Decision tree and random forest implementations for fast filtering of sensor data. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(1), 209-222. <https://doi.org/10.1109/TCSI.2017.2710627>
- Case, B. (2008) REITs Take Up the Melody. *IPE Real Estate* September, 76.

- Carstina, S., Siminica, M., Circiumaru, D., & Tanasie, A. (2015). Correlation analysis of the indicators of asset management and profitability. *International Journal in Economics and Business Administration*, 3(2), 3-21.
- CATBOOST. (2023). Overview - How training is performed | CatBoost. <https://catboost.ai/en/docs/concepts/algorithm-main-stages>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chang, K.-L., Chen, N.-K., Charles, S., Leung, K. Y., Chang, K.-L., Chen, N.-K., & Leung, C. K. Y. (2011). Monetary policy, term structure and asset return: comparing REIT, housing and stock. *J Real Estate Finan Econ*, 43, 221-257. <https://doi.org/10.1007/s11146-010-9241-8>
- Chapman, P., Khabaza, T., Clinton, J., Kerber, R., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide .
- Chatpatanasiri, R. (2005). Bayesian evidence framework for decision tree learning. *AIP Conference Proceedings*, 88-95. <https://doi.org/10.1063/1.2149783>
- Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. *Çinde Data Mining and Knowledge Discovery Handbook* (ss. 875-886). Springer US. https://doi.org/10.1007/978-0-387-09823-4_45
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/JAIR.953>
- Chehreh Chelgani, S., Nasiri, H., Tohry, A., & Heidari, H. R. (2023). Modeling industrial hydrocyclone operational variables by SHAP-CatBoost - A

- “conscious lab” approach. *Powder Technology*, 420, 118416. <https://doi.org/10.1016/j.powtec.2023.118416>
- Chen, H. (2023). Enterprise marketing strategy using big data mining technology combined with XGBoost model in the new economic era. *PLOS ONE*, 18(6), e0285506. <https://doi.org/10.1371/journal.pone.0285506>
- Chen, J. (2022, Mayıs 30). What Is the National Association of Real Estate Investment Trusts (Nareit)? <https://www.investopedia.com/terms/n/nareit.asp>.
- Chen, L.-P. (2021). Practical statistics for data scientists: 50+ essential concepts using R and Python. *Technometrics*, 63(2), 272-273. <https://doi.org/10.1080/00401706.2021.1904738>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Ching, N., David, Y., Chang, Y. W., & Low, S. C. (2021). Corporate Governance Mechanisms with Firm Performance: A Study on Malaysia’s and Hong Kong’s Real Estate Investment Trust (REITs) Public Listed Companies. *Journal of Governance Risk Management Compliance and Sustainability*, 1(1), 62-75. <https://doi.org/10.31098/JGRCS.V1I1.511>
- Chugh, A. (2020). MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better? <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>.

- Clements, J., Jones, T., & Yates, A. (2021). Worldwide Real Estate Investment Trust (REIT) Regimes.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. bs). Lawrence Erlbaum.
- Colestock, S. (2021). Three Types of REITs (and How to Invest in Them).
- Conklin, M., Powaga, K., & Lipovetsky, S. (2004). Customer satisfaction analysis: Identification of key drivers. *European Journal of Operational Research*, 154(3), 819-827. [https://doi.org/10.1016/S0377-2217\(02\)00877-9](https://doi.org/10.1016/S0377-2217(02)00877-9)
- Crocker, D. C. (1972). Some Interpretations of the Multiple Correlation Coefficient. *The American Statistician*, 26(2), 31-33. <https://doi.org/10.1080/00031305.1972.10477345>
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. *Çinde Ensemble Machine Learning* (ss. 157-175). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_5
- Czarnowski, I. (2022). Weighted Ensemble with one-class Classification and Over-sampling and Instance selection (WECOI): An approach for learning from imbalanced data streams. *Journal of Computational Science*, 61, 101614. <https://doi.org/10.1016/j.jocs.2022.101614>
- Çelik, E., & Arslanlı, K. Y. (2020). Gayrimenkul yatırım ortaklığı firmalarının piyasa değeri ve aktif karlılığını etkileyen finansal oranların panel veri analizi yöntemiyle belirlenmesi. *The Journal of Accounting and Finance*, 88, 255-274. <https://doi.org/10.25095/MUFAD.801491>
- Çelik, Ş., & Manan, T. (2018). Gayrimenkul yatırım ortaklıklarının risk ile performans ilişkisi. *Muhasebe ve Finans İncelemeleri Dergisi*, 1(1), 60-79. <https://doi.org/10.32951/MUFIDER.385054>
- Dasi, H., Ying, Z., & Yang, B. (2023). Predicting the consumed heating energy at residential buildings using a combination of categorical boosting

- (CatBoost) and Meta heuristics algorithms. *Journal of Building Engineering*, 71, 106584. <https://doi.org/10.1016/j.jobbe.2023.106584>
- Debreceňy, R. S., & Gray, G. L. (2010). Data mining journal entries for fraud detection: An exploratory study. *International Journal of Accounting Information Systems*, 11(3), 157-181. <https://doi.org/10.1016/j.accinf.2010.08.001>
- Demir, S., & Sahin, E. K. (2022). Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data. *Soil Dynamics and Earthquake Engineering*, 154, 107130. <https://doi.org/10.1016/j.soildyn.2021.107130>
- Demirdöven, M. (2009). Türkiye’de gayrimenkul sektörünün gelişimi ve gayrimenkul finansmanı [Yüksek Lisans]. Sosyal Bilimler Enstitüsü.
- Deran, A., Sarıay, M. A. İbrahim, & Savaş, İ. (2013). İMKB’ de işlem gören menkul kıymet yatırım ortaklıkları ve gayrimenkul yatırım ortaklıklarının finansal performanslarının karşılaştırılması. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 23(2), 195-203. <https://dergipark.org.tr/tr/pub/firatsbed/issue/45176/565597>
- Dev, V. A., & Eden, M. R. (2019). Gradient Boosted Decision Trees for Lithology Classification (ss. 113-118). <https://doi.org/10.1016/B978-0-12-818597-1.50019-9>
- Dietterich, T. G. (2002). Ensemble learning. *The handbook of brain theory and neural networks*, 2(1), 110-125.
- Dogan, Y. Y., Ghosh, C., & Petrova, M. (2019). On the Determinants of REIT Capital Structure: Evidence from around the World. *Journal of Real Estate Finance and Economics*, 59(2), 295-328. <https://doi.org/10.1007/S11146-018-9687-7/TABLES/12>

- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87. <https://doi.org/10.1145/2347736.2347755>
- Dong, G., Zhang, K., Huynh, K., & Zou, W. (2022). 2021 - 2022 Asia Reit Market Insight.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241-258. <https://doi.org/10.1007/s11704-019-8208-z>
- Durer, S., & Berrak Köten, A. (2021). Gayrimenkul yatırım ortaklıklarının kârlılık oranlarının borsa performansına etkisi.
- Egbert, J., & Plonsky, L. (2020). Bootstrapping Techniques. İçinde *A Practical Handbook of Corpus Linguistics* (ss. 593-610). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_24
- Elavarasan, D., & Vincent, D. R. (2020). Reinforced XGBoost machine learning model for sustainable intelligent agrarian applications. *Journal of Intelligent & Fuzzy Systems*, 39(5), 7605-7620. <https://doi.org/10.3233/JIFS-200862>
- Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data*, 4(1), 19. <https://doi.org/10.1186/s40537-017-0079-2>
- Emami, S., Nasiri, H., & SeyyedMomeni, A. (2023). A novel approach based on CatBoost and explainable artificial intelligence for diagnosis of COVID-19 cases using patients' symptoms. *Journal of Modeling in Engineering*.
- Enayati, M., Bozorg-Haddad, O., Pourgholam-Amiji, M., Zolghadr-Asli, B., & Tahmasebi Nasab, M. (2022). Decision Tree (DT): A Valuable Tool for Water Resources Engineering (ss. 201-223). https://doi.org/10.1007/978-981-19-2519-1_10

- EPRA. (2018). Building a Stronger Europe Stock Exchange Listed Property Companies.
- EPRA Global REIT Survey 2022. (2022). https://www.epra.com/application/files/9516/6203/9094/EPRA_Global_REIT_Survey_2022_ONLINE_v1.7_FINAL_VERSION.pdf
- EY. (2021). Economic contribution of REITs in the United States Prepared for Nareit.
- Fama, E. F., & French, K. R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *Review of financial studies*, 1-33.
- Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers & Industrial Engineering*, 101, 554-564. <https://doi.org/10.1016/J.CIE.2016.09.011>
- Farrell, S. (2023). EPRA Global REIT Survey 2022.
- Fatnassi, I., Slim, C., Ftiti, Z., & Maatoug, A. ben. (2014). Effects of monetary policy on the REIT returns: Evidence from the United Kingdom. *Research in International Business and Finance*, 32, 15-26. <https://doi.org/10.1016/j.ribaf.2014.01.001>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases (C. 17). www.ffly.com/
- Félix, D., Branco, J. M., & Feio, A. (2013). Temporary housing after disasters: A state of the art survey. *Habitat International*, 40, 136-141. <https://doi.org/10.1016/j.habitatint.2013.03.006>
- Feng, Z., Ghosh, C., & Sirmans, C. F. (2007). On the Capital Structure of Real Estate Investment Trusts (REITs). *The Journal of Real Estate Finance and Economics* 2007 34:1, 34(1), 81-105. <https://doi.org/10.1007/S11146-007-9005-2>

- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from Imbalanced Data Sets*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-98074-4>
- Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, 61, 863-905. <https://doi.org/10.1613/JAIR.1.11192>
- Francois, C., & JJ, A. (2017). *Deep Learning with R (MEAP Edition)*. Manning Publications. www.manning.com
- Gana, K., & Broc, G. (2019). *Structural Equation Modeling with lavaan*. John Wiley & Sons.
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2021). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115. <https://doi.org/10.1016/j.engappai.2022.105151>
- Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, 2(4), 42-47.
- Gao, T., & Liu, J. (2021). Application of improved random forest algorithm and fuzzy mathematics in physical fitness of athletes. *Journal of Intelligent & Fuzzy Systems*, 40(2), 2041-2053. <https://doi.org/10.3233/JIFS-189206>
- George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493-1507. <https://doi.org/10.5465/amj.2016.4005>
- Georgiev, G. (2002). The benefits of real estate investment.
- Geurts, P. (2009). Bias vs variance decomposition for regression and classification. *Çinde Data Mining and Knowledge Discovery Handbook*

- (ss. 733-746). Springer US. https://doi.org/10.1007/978-0-387-09823-4_37
- Gheorghe, M., & Petre, R. (2014). Integrating data mining techniques into telemedicine systems. *Informatica Economica*, 18(1/2014), 120-130. <https://doi.org/10.12948/issn14531305/18.1.2014.11>
- GITHUB.(2020).XGBoost.https://vatsalparsiya.github.io/ML_Knowledge/XGBoost/Readme.html#xgboost.
- Glascok, J. L., & Lu-Andrews, R. (2015). The Price behavior of REITS surrounding extreme market-related events. *Journal of Real Estate Finance and Economics*, 51(4), 441-479. <https://doi.org/10.1007/S11146-015-9495-2/METRICS>
- Glascok, J., & Lu-Andrews, R. (2016). An examination of macroeconomic effects on the liquidity of REITs. *Energy Economics*, 60, 168-175. <https://doi.org/10.1007/s11146-013-9406-3>
- Goebel, P. R., Harrison, D. M., Mercer, J. M., & Whitby, R. J. (2013). REIT momentum and characteristic-related REIT returns. *J Real Estate Finan Econ*, 47, 564-581. <https://doi.org/10.1007/s11146-012-9371-2>
- Graham, C., & Knight, J. (2020). Cash flows vs. earnings in the valuation of equity REITs. <https://doi.org/10.1080/10835547.2000.12089590>, 6(1), 17-25. <https://doi.org/10.1080/10835547.2000.12089590>
- Grupe, M. R., & DiRocco, C. J. (1999). The NAREIT index of REIT industry performance. *Real Estate Finance*, 16, 21-51.
- Gu, Q., Tian, J., Li, X., & Jiang, S. (2022). A novel Random Forest integrated model for imbalanced data classification problem. *Knowledge-Based Systems*, 250, 109050. <https://doi.org/10.1016/j.knosys.2022.109050>
- Gumus, M., & Kiran, M. S. (2017). Crude oil price forecasting using XGBoost. 2017 International Conference on Computer Science and Engineering (UBMK), 1100-1103. <https://doi.org/10.1109/UBMK.2017.8093500>

- Guo, F., Wang, G., Su, Z., Liang, H., Wang, W., Lin, F., & Liu, A. (2016). What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. *International Journal of Wildland Fire*, 25(5), 505. <https://doi.org/10.1071/WF15121>
- Gupta, S. (2015). A regression modeling technique on data mining. *International Journal of Computer Applications*, 116(9), 27-29. <https://doi.org/10.5120/20365-2570>
- Gülyüz, A. B. E. (2019). Gayrimenkul yatırım ortaklıkları ve borsa İstanbul A.Ş.'de işlem gören GYO'ların performanslarını etkileyen faktörlerin incelenmesi [Yüksek Lisans Tezi]. Bankacılık Ve Sigortacılık Enstitüsü.
- Gümüş, U. T., & Can Öziç, H. (2018). BIST'te gayrimenkul yatırım ortaklıkları sektöründe faaliyet gösteren işletmelerin veri zarflama analizi ile etkinliklerinin ölçülmesi. *Akademik Sosyal Araştırmalar Dergisi*, 6, 352-362.
- Günay, B., & Timur, E. (2019). Ekonomik katma değer ile borsa performans ölçütleri arasındaki ilişkinin incelenmesi: BIST gayrimenkul yatırım ortaklıkları endeksi (XGMYO) örneği. 6th International Congress on Accounting and Finance Research (ICAFR'19), 28-38.
- GYODER. (2022). Türkiye Gayrimenkul Sektörü 2022 3.Çeyrek Raporu.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining. concepts and techniques*, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems).
- Hancock, J., & Khoshgoftaar, T. M. (2020). Medicare fraud detection using CatBoost. 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 97-103. <https://doi.org/10.1109/IRI49571.2020.00022>
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7(1), 94. <https://doi.org/10.1186/s40537-020-00369-8>

- Hand, D. (2001). Data Mining. İçinde Encyclopedia of Environmetrics. Wiley. <https://doi.org/10.1002/9780470057339.vad002>
- Harine Rajashree, R., & Hariharan, M. (2021). A study on ensemble methods for classification (ss. 127-136). https://doi.org/10.1007/978-981-16-0289-4_10
- Harrison, D. M., Panasian, C. A., & Seiler, M. J. (2011). Further evidence on the capital structure of REITs. *Real Estate Economics*, 39(1), 133-166. <https://doi.org/10.1111/J.1540-6229.2010.00289.X>
- Hastie, T., Friedman, J., & Tibshirani, R. (2001). The elements of statistical learning. Springer New York. <https://doi.org/10.1007/978-0-387-21606-5>
- Hayta, Ö. (2009). Gayrimenkul yatırım ortaklıkları ve gayrimenkul yatırım ortaklıkları performans değerlendirmesi: İMKB`de bir uygulama [Yüksek Lisans]. Sosyal Bilimler.
- Hepsen, A., Berberoglu, M. G., & Aydın, O. (2017). Real estate investment trusts in Turkey: Structure, analysis, and strategy. *Pressacademia*, 6(2), 191-199. <https://doi.org/10.17261/Pressacademia.2017.508>
- Hepşen, A. (2009). Gayrimenkul sektöründe köpük oluşumları, ABD Kredi Krizinin Etkisiyle Gerileyen Konut Fiyatları ve “REIDIN.COM EMLAKENDEKS”.
- Hertzmann, A., Fleet, D., & Brubaker, M. (2015). Machine Learning and data mining lecture notes.
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481-5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hsieh, H. C., Claresta, V., & Ngoc Bui, T. M. (2020). Green building, cost of equity capital and corporate governance: Evidence from US real estate

- investment trusts. *Sustainability* 2020, Vol. 12, Page 3680, 12(9), 3680. <https://doi.org/10.3390/SU12093680>
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, 79, 403-408. <https://doi.org/10.1016/j.procir.2019.02.106>
- Hussain, S., Mustafa, Mohd. W., Jumani, T. A., Baloch, S. K., Alotaibi, H., Khan, I., & Khan, A. (2021). A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. *Energy Reports*, 7, 4425-4436. <https://doi.org/10.1016/j.egy.2021.07.008>
- Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 70, 163-172.
- Jadwal, P. K., Jain, S., Pathak, S., & Agarwal, B. (2022). Improved resampling algorithm through a modified oversampling approach based on spectral clustering and SMOTE. *Microsystem Technologies*, 28(12), 2669-2677. <https://doi.org/10.1007/s00542-022-05287-8>
- Jaiswal, S., & Gupta, P. (2022). Ensemble Approach: XGBoost, CATBoost, and LightGBM for Diabetes Mellitus Risk Prediction. 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), 1-6. <https://doi.org/10.1109/ICCSEA54677.2022.9936130>
- Jakpar, S., Tinggi, M., Tak, A. H., & Ruzlan, N. A. (2018a). Determinant factors of profitability in Malaysia's real estate investment trusts (M-REITS). *UNIMAS Review of Accounting and Finance*, 2(1). <https://doi.org/10.33736/URAF.1209.2018>
- Jamal, E. (2013). The Effectiveness and Efficiency of REITs Performance in Malaysia. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2273937>

- Jhaveri, S., Khedkar, I., Kantharia, Y., & Jaswal, S. (2019). Success prediction using random forest, catboost, xgboost and adaboost for kickstarter campaigns. *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 1170-1173. <https://doi.org/10.1109/ICCMC.2019.8819828>
- Jin, M., Wang, Y., & Zeng, Y. (2018). Application of Data Mining technology in financial risk analysis. *Wireless Personal Communications*, 102(4), 3699-3713. <https://doi.org/10.1007/S11277-018-5402-5/TABLES/4>
- Jurion, G. (2021). Worldwide Real Estate Investment Trust (REIT) Regimes.
- Kamath, C. (2001). On mining scientific datasets (ss. 1-21). https://doi.org/10.1007/978-1-4615-1733-7_1
- Kanaparthi, V. (2023). Credit risk prediction using ensemble machine learning algorithms. *2023 International Conference on Inventive Computation Technologies (ICICT)*, 41-47. <https://doi.org/10.1109/ICICT57646.2023.10134486>
- Karakuş, R., & Öksüz, S. (2021). BİST gayrimenkul yatırım ortaklıkları endeksi ile konut fiyat endeksi, faiz oranı ve enflasyon ilişkisi: ARDL sınır testi yaklaşımı. *Business & Management Studies: An International Journal*, 9(2), 751-764. <https://doi.org/10.15295/BMIJ.V9I2.1825>
- Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609-629. <https://doi.org/10.1016/j.ins.2021.11.036>
- Kavuncuoğlu, E., Uzunhisarcıklı, E., Barshan, B., & Özdemir, A. T. (2022). Investigating the performance of wearable motion sensors on recognizing falls and daily activities via machine learning. *Digital Signal Processing*, 126, 103365. <https://doi.org/10.1016/j.dsp.2021.103365>
- Kavzoglu, T. (2009). Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software*, 24(7), 850-858. <https://doi.org/10.1016/j.envsoft.2008.11.012>

- Kazak, J., Van Hoof, J., Świąder, M., & Szewrański, S. (2017). Real estate for the ageing society - The perspective of a new market. *Real Estate Management and Valuation*, 25(4), 13-24. <https://doi.org/10.1515/REMAV-2017-0026>
- Ke, L., Li, C., Zhong, T., Cai, Z., Wen, J., Wang, R., Qian, F., & Tang, H. (2021). Loan repayment behavior prediction of provident fund users using a stacking-based model. 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 37-43. <https://doi.org/10.1109/ICCCBDA51879.2021.9442613>
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics*.
- Khairulanuwar, A. J., & Chuweni, N. N. (2021). The significance and performance analysis of Malaysian real estate investment trusts. *International Journal of Law and Management*, 63(4), 417-430. <https://doi.org/10.1108/IJLMA-01-2020-0022>
- Khare, S., Gourisaria, M. K., Harshvardhan, G., Joardar, S., & Singh, V. (2021). Real estate cost estimation through data mining techniques. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012053. <https://doi.org/10.1088/1757-899X/1099/1/012053>
- Kim, B., Lee, D.-E., Hu, G., Natarajan, Y., Preethaa, S., & Rathinakumar, A. P. (2022). Ensemble machine learning-based approach for predicting of FRP–Concrete interfacial bonding. *Mathematics*, 10(2), 231. <https://doi.org/10.3390/math10020231>
- Kim, H., Mattila, A. S., & Gu, Z. (2002). Performance of hotel real estate investment trusts: a comparative analysis of Jensen indexes. *International Journal of Hospitality Management*, 21(1), 85-97. [https://doi.org/10.1016/S0278-4319\(01\)00026-3](https://doi.org/10.1016/S0278-4319(01)00026-3)
- Kim, P. (2017). *MATLAB deep learning*. MATLAB Deep Learning. <https://doi.org/10.1007/978-1-4842-2845-6>

- Koçak, A., & Ergün, M. A. (2023). Sağlıkta veri kalitesi ve veri madenciliği uygulamaları. *Journal of Interdisciplinary Innovation Studies*, 3(1), 23-30. <https://doi.org/10.56723/DYAD.1161993>
- Kohavi, R. (1998). Glossary of terms. *Machine learning*, 30, 271-274.
- Ksieniewicz, P., Krawczyk, B., & Woźniak, M. (2018). Ensemble of extreme learning machines with trained classifier combination and statistical features for hyperspectral data. *Neurocomputing*, 271, 28-37. <https://doi.org/10.1016/j.neucom.2016.04.076>
- Kumar, G., Yadav, S. S., Yogita, & Pal, V. (2022). Machine Learning-Based Framework to Predict Finger Movement for Prosthetic Hand. *IEEE Sensors Letters*, 6(6), 1-4. <https://doi.org/10.1109/LESENS.2022.3147518>
- Kumar, N. (2019, Mart 19). Advantages of XGBoost Algorithm in Machine Learning. <http://theprofessionalspoint.blogspot.com/2019/03/advantages-of-xgboost-algorithm-in.html>.
- Latha, R., & Bommi, R. M. (2023). Hybrid CatBoost Regression model based intrusion detection system in iot-enabled networks. 2023 9th International Conference on Electrical Energy Systems (ICEES), 264-269. <https://doi.org/10.1109/ICEES57979.2023.10110148>
- Lee, J. K., Williams, P. D., & Cheon, S. (2008). Data Mining in Genomics. *Clinics in Laboratory Medicine*, 28(1), 145-166. <https://doi.org/10.1016/j.cll.2007.10.010>
- Lee, S. (2014). The relationship between growth and profit: evidence from firm-level panel data. *Structural Change and Economic Dynamics*, 28, 1-11. <https://doi.org/10.1016/J.STRUECO.2013.08.002>
- Lee, S., Nguyen, N., Karamanli, A., Lee, J., & Vo, T. P. (2022). Super learner machine-learning algorithms for compressive strength prediction of high

- performance concrete. *Structural Concrete*, 24(2), 2208-2228. <https://doi.org/10.1002/suco.202200424>
- Lever, J., Krzywinski, M., & Altman, N. (2016). Points of Significance: Model selection and overfitting. *Nature Methods*, 13(9), 703-705. <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=15487091&v=2.1&it=r&id=GALE%7CA461963379&sid=googleScholar&linkaccess=fulltext>
- Lhessani, S. (2019, Eylül 28). What is the difference between training and test dataset? <https://lhessani-sajid.medium.com/what-is-the-difference-between-training-and-test-dataset-91308080a4e8>.
- Li, S., & Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. *Neural Computing and Applications*, 32(7), 1971-1979. <https://doi.org/10.1007/s00521-019-04378-4>
- Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., Lu, S., Fan, H., Yang, L., & Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, 8, 1087-1095. <https://doi.org/10.1016/j.egy.2022.02.251>
- Li, Y., & Chen, W. (2020). A comparative performance assessment of ensemble learning for credit scoring. *Mathematics*, 8(10), 1756. <https://doi.org/10.3390/math8101756>
- Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. *R news*, 2(3), 18-22. https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest
- Lipovetsky, S., & Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, 17(4), 319-330. <https://doi.org/10.1002/asmb.446>

- Liu, Y., Lin, H., Ding, L., Wang, W., & Liao, S. (2018). Fast Cross-Validation. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2497-2503. <https://doi.org/10.24963/ijcai.2018/346>
- Loo, W. K., Anuar, M. A., & Ramakrishnan, S. (2016). Integration between the Asian REIT markets and macroeconomic variables. *Journal of Property Investment and Finance*, 34(1), 68-82. <https://doi.org/10.1108/JPIF-12-2014-0070/FULL/PDF>
- Lu, C., & So, R. W. (2001). The relationship between reits returns and inflation: a vector error correction approach. *Çinde Review of Quantitative Finance and Accounting* (C. 16). Kluwer Academic Publishers.
- Lubyayaya, A. V., Izmailov, A. M., Nikulina, E. Y., & Shaposhnikov, V. A. (2016). Evaluation of the effect of non-current fixed assets on profitability and asset management efficiency. *International Journal of Environmental & Science Education*, 11(15), 7745-7753.
- Luna, Z. (2021, Temmuz 21). Understanding CRISP-DM and its importance in Data Science projects. <https://medium.com/analytics-vidhya/understanding-crisp-dm-and-its-importance-in-data-science-projects-91c8742c9f9b>.
- Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., & Sun, Y. (2021). Combination of feature selection and CatBoost for prediction: the first application to the estimation of aboveground biomass. *Forests*, 12(2), 216. <https://doi.org/10.3390/f12020216>
- Mach-Król, M., & Hadasik, B. (2021). On a certain research gap in big data mining for customer insights. *Applied Sciences*, 11(15), 6993. <https://doi.org/10.3390/app11156993>
- Maciejewski, T., & Stefanowski, J. (2011). Local neighbourhood extension of SMOTE for mining imbalanced data. 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 104-111. <https://doi.org/10.1109/CIDM.2011.5949434>

- Ma'in, M., Arifin, N. A. M., Hatta, M. F. M., Ibrahim, R. M., & Isa, S. S. M. (2018). Malaysian real estate investment trust performance: A panel data analysis - ProQuest. *Journal of Economic & Management Perspectives*, 12(1), 527-532. <https://www.proquest.com/openview/6f9be32f54d818b4e25ab2b97cd17ec6/1?cbl=51667&pq-origsite=gscholar>
- Ma'In, M., Arifin, N. A. M., Hatta, M. F. M., Hashim, M. H., & Isa, S. S. M. (2016). Determinants of Islamic real estate investment trust performance. *Advanced Science Letters*, 22(12), 4321-4325. <https://doi.org/10.1166/ASL.2016.8137>
- Maimon, O., & Rokach, L. (2005). Introduction to knowledge discovery in databases. *Çinde Data Mining and Knowledge Discovery Handbook* (ss. 1-17). Springer-Verlag. https://doi.org/10.1007/0-387-25465-X_1
- Maimon, O., & Rokach, L. (2009). Introduction to Knowledge discovery and data mining. *Data Mining and Knowledge Discovery Handbook*, 1-15. https://doi.org/10.1007/978-0-387-09823-4_1
- Mak, M. K., Ho, G. T., & Ting, S. L. (2011). A financial data mining model for extracting customer behavior. *International Journal of Engineering Business Management*, 59-72.
- Malek, N. H. A., Yaacob, W. F. W., Wah, Y. B., Md Nasir, S. A., Shaadan, N., & Indratno, S. W. (2022). Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(1), 598. <https://doi.org/10.11591/ijeecs.v29.i1.pp598-608>
- Marcot, B. G., & Hanea, A. M. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*, 36(3), 2009-2031. <https://doi.org/10.1007/s00180-020-00999-9>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The*

- Knowledge Engineering Review, 25(2), 137-166.
<https://doi.org/10.1017/S0269888910000032>
- Mavrogiorgou, A., Kiourtis, A., Manias, G., & Kyriazis, D. (2021). An Optimized KDD Process for Collecting and Processing Ingested and Streaming Healthcare Data. 2021 12th International Conference on Information and Communication Systems (ICICS), 49-56.
<https://doi.org/10.1109/ICICS52457.2021.9464551>
- Mendes-Moreira, J., Soares, C., Jorge, A. M., & De Sousa, J. F. (2012). Ensemble approaches for regression. ACM Computing Surveys (CSUR), 45(1). <https://doi.org/10.1145/2379776.2379786>
- Mert, B. S. (2012). Gayrimenkul yatırım ortaklıklarının yapısı, faaliyetleri ve gayrimenkul sektöründeki risklerin incelenmesi [Yüksek Lisan]. Sosyal Bilimler Enstitüsü.
- Micci-Barreca, D. (2001). A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. ACM SIGKDD Explorations Newsletter, 3(1), 27-32.
<https://doi.org/10.1145/507533.507538>
- Miles, J. (2014). R Squared, Adjusted R Squared. İçinde Wiley StatsRef: Statistics Reference Online. Wiley.
<https://doi.org/10.1002/9781118445112.stat06627>
- Ming-Syan Chen, Jiawei Han, & Yu, P. S. (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866-883. <https://doi.org/10.1109/69.553155>
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, Model Tuning, and Evaluation of Prediction Performance. İçinde Multivariate Statistical Machine Learning Methods for Genomic Prediction (ss. 109-139). Springer International Publishing.
https://doi.org/10.1007/978-3-030-89010-0_4

- Morri, G., Anconetani, R., & Benfari, L. (2021). Greenness and financial performance of European REITs. *Journal of European Real Estate Research*, 14(1), 40-61. <https://doi.org/10.1108/JERER-05-2020-0030>
- Morri, G., & Parri, E. (2017). US REITs capital structure determinants and financial economic crisis effects. *Journal of Property Investment and Finance*, 35(6), 556-574. <https://doi.org/10.1108/JPIF-07-2016-0055/FULL/PDF>
- Muhasshanah, M., Tohir, M., Ningsih, D. A., Susanti, N. Y., Umiyah, A., & Fitria, L. (2023). Comparison of the Performance Results of C4.5 and Random Forest Algorithm in Data Mining to Predict Childbirth Process. *CommIT (Communication and Information Technology) Journal*, 17(1), 51-59. <https://doi.org/10.21512/commit.v17i1.8236>
- Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, 78(3), 691. <https://doi.org/10.2307/2337038>
- Najm, S. M., Trzepieciński, T., & Kowalik, M. (2023). Modelling and parameter identification of coefficient of friction for deep-drawing quality steel sheets using the CatBoost machine learning algorithm and neural networks. *The International Journal of Advanced Manufacturing Technology*, 124(7-8), 2229-2259. <https://doi.org/10.1007/s00170-022-10544-1>
- NAREIT. (2022). REIT Industry Timeline | Nareit. <https://www.reit.com/investing/reit-basics/reit-industry-timeline>
- Nazlioglu, S., Gormus, N. A., & Soytaş, U. (2016). Oil prices and real estate investment trusts (REITs): Gradual-shift causality and volatility transmission analysis. *Energy Economics*, 60, 168-175. <https://doi.org/10.1016/J.ENECO.2016.09.009>
- Neelamegam, S., & Ramaraj, E. (2013). Classification algorithm in data mining: An overview . *International Journal of P2P Network Trends and Technology (IJPTT)*, 4(8), 369-374.

- Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, 34(21), 3711-3718. <https://doi.org/10.1093/bioinformatics/bty373>
- Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, 510, 1-14. <https://doi.org/10.1016/j.neucom.2022.08.055>
- Nieminen, P. (2022). Application of Standardized Regression Coefficient in Meta-Analysis. *BioMedInformatics*, 2(3), 434-458. <https://doi.org/10.3390/biomedinformatics2030028>
- Nishigaki, H. (2007). An analysis of the relationship between US REIT returns. *Economics Bulletin*, 13(1), 1-7.
- Njagi, B. N. (2021). Effect of Macroeconomic Variables on Performance of Real Estate Investment Trusts in the Nairobi Securities Exchange. University Of Nairobi.
- Nurhayati, Soekarno, I., Hadihardaja, I. K., & Cahyono, M. (2014). A study of hold-out and k-fold cross validation for accuracy of groundwater modeling in tidal lowland reclamation using extreme learning machine. 2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment, 228-233. <https://doi.org/10.1109/TIME-E.2014.7011623>
- Obenshain, M. K. (2004). Application of Data Mining Techniques to Healthcare Data. *Infection Control & Hospital Epidemiology*, 25(8), 690-695. <https://doi.org/10.1086/502460>
- Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-76917-0>
- Öndeş, T., & Barakalı, O. C. (2023). Faiz Oranlarının Gayrimenkul Yatırım Ortaklığı Karlılık Oranları Üzerinde Etkisi. *Muhasebe ve Finansman Dergisi*, 97, 49-62. <https://doi.org/10.25095/mufad.1184115>

- Özcan, N., & Gürol, B. (2020). Gayrimenkul Yatırım Ortaklıklarının Performanslarının İncelenmesi: Türkiye'deki Gayrimenkul Yatırım Ortaklıkları Üzerine Bir Araştırma. *Celal Bayar University Journal of Social Sciences*, 18(Özel Sayı), 1-16. <https://doi.org/10.18026/cbayarsos.468166>
- Palmer, B. (2022). Key Reasons to Invest in Real Estate.
- Pandey, A. K. (2022). Regression algorithms. <https://medium.com/@arunp77/regression-algorithms-29f112797724>.
- Pickell, K. (2021, Kasım 1). What Is Validation Data, And What Is It Used For? <https://medium.com/artificialis/what-is-validation-data-and-what-is-it-used-for-158d685fb921>.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21-44. <https://doi.org/10.1109/MCAS.2006.1688199>
- Polikar, R. (2012). Ensemble Learning. İçinde *Ensemble Machine Learning* (ss. 1-34). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_1
- Pradipta, G. A., Wardoyo, R., Musdholifah, A., Sanjaya, I. N. H., & Ismail, M. (2021). SMOTE for Handling Imbalanced Data Problem : A Review. 2021 Sixth International Conference on Informatics and Computing (ICIC), 1-8. <https://doi.org/10.1109/ICIC54025.2021.9632912>
- Prajwala, T. R. (2015). A Comparative Study on Decision Tree and Random Forest Using R Tool. *IJARCCCE*, 196-199. <https://doi.org/10.17148/IJARCCCE.2015.4142>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Andrey Gulin. (2018). CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, 31, 6638-6648.
- PWC. (2022). Structural Overview of the Real Estate Investment Industry.

- Radivojac, P., Chawla, N. V., Dunker, A. K., & Obradovic, Z. (2004). Classification and knowledge discovery in protein databases. *Journal of Biomedical Informatics*, 37(4), 224-239. <https://doi.org/10.1016/j.jbi.2004.07.008>
- Rahmawati, Y., Sari, A. F., & Utomo, C. (2023). The Effect of Consequences in Utilizing Real Estate Investment Trust (REIT) on Property Development. *International Journal on Advanced Science, Engineering & Information Technology*, 13(1), 173-179.
- Rajan, V., & Krishnan, V. G. (2022). A study on the use of machine learning and complex hierarchical structures to visualise text categorization. *Technoarete Transactions on Advances in Computer Applications (TTACA)*, 1(2), 9-16.
- Ratner, B. (2017). *Statistical and machine-learning data mining : techniques for better predictive modeling and analysis of big data (2. bs)*. CRC Press.
- Reddy, W., & Wong, W. W. (2018). Australian interest rate movements and A-REITs performance: An analysis by industry sector. *Pacific Rim Property Research Journal*, 24(1), 85-103. <https://doi.org/10.1080/14445921.2018.1436323>
- Rençber, Ö. F. (2021). Veri Bilimine Giriş ve Makine Öğrenme. İçinde Ö. F. Rençber (Ed.), *Veri Madenciliğinde Kullanılan regresyon modelleri ve R ile Uygulamalı Örnekler* (ss. 1-10). Nobel Akademik Yayımcılık.
- Rezapour, M., & Hansen, L. (2022). A machine learning analysis of COVID-19 mental health data. *Scientific Reports*, 12(1), 14965. <https://doi.org/10.1038/s41598-022-19314-1>
- Rice, L., Wong, E., & Kolter, Z. (2020). Overfitting in adversarially robust deep learning. *International Conference on Machine Learning* , 8093-8104.
- Rocks, J. W., & Mehta, P. (2022). Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models. *Physical*

- Review Research, 4(1), 013201.
<https://doi.org/10.1103/physrevresearch.4.013201/figures/6/medium>
- Rossetti, L. (2022, Kasım 5). Knowledge Discovery in Databases (KDD) and Data Mining. <https://medium.com/mlearning-ai/knowledge-discovery-in-databases-kdd-and-data-mining-b9350e5a191a>.
- Sadaf, K. (2023). Phishing Website Detection using XGBoost and Catboost Classifiers. 2023 International Conference on Smart Computing and Application (ICSCA), 1-6.
<https://doi.org/10.1109/ICSCA57840.2023.10087829>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4).
<https://doi.org/10.1002/widm.1249>
- Saltz, J. S. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. 2021 IEEE International Conference on Big Data (Big Data), 2337-2344.
<https://doi.org/10.1109/BigData52589.2021.9671634>
- Santos, M., & Azevedo, C. (2005). *Data Mining – Descoberta de Conhecimento em Bases de Dados*. FCA Publisher.
- Sarı, S. S., & Başakın, E. E. (2020). Estimating the index of real estate investment trusts with tree based regression models calibration of land surface models view project. International Marmara Social Sciences Congress, 310-317.
<https://www.researchgate.net/publication/346649127>
- Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*, 14(1), 155-162.
<https://doi.org/10.35970/infotekmesin.v14i1.1751>

- Sarkar, D., & Natarajan, V. (2019). *Ensemble Machine Learning Cookbook: Over 35 practical recipes to explore ensemble machine learning techniques using Python*. Packt Publishing Ltd.
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197-227. <https://doi.org/10.1007/BF00116037>
- Schapire, R. E. (2003). The Boosting Approach to Machine Learning: An Overview. *Çinde Nonlinear Estimation and Classification. Lecture Notes in Statistics* (ss. 149-171). https://doi.org/10.1007/978-0-387-21579-2_9
- Schneider, J. (1997, Şubat 7). Cross Validation. <http://www.cs.cmu.edu/~schneide/tut5/node42.html>.
<http://www.cs.cmu.edu/~schneide/tut5/node42.html>
- Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kuehnel, S., Badwitz, W., Dann, D., Kloker, S., Alekozai, E. M., & Lanquillon, C. (2020). Introducing DASC-PM: A Data Science Process Model. *ACIS 2020 Proceedings*. 45.
- Scikitlearn.(2023).<https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html>.<https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html>.
- Seçkin, N. (1998). Gayrimenkul Yatırım Kararlarında Fizibilite Analizi. *Fen Bilimleri Enstitüsü*.
- Sha, T. L. (2017). Effects of Price Earnings Ratio, Earnings Per Share, Book to Market Ratio and Gross Domestic Product on Stock Prices of Property and Real Estate Companies in Indonesia Stock Exchange - ProQuest. "Effects of Price Earnings Ratio, Earnings per Share, Book to Market Ratio And Gross Domestic Product on Stock Prices of Property And Real

- Estate Companies In Indonesia Stock Exchange, 11(1), 1743-1754.
<https://www.proquest.com/openview/d2295546bf23b6c69c78df2b471a12dd/1?pq-origsite=gscholar&cbl=51667>
- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- Shah, R. M., Butt, M. A., & Baba, M. Z. (2017). Predictive analytic modeling: A walkthrough. *International Journal of Advanced Research in Computer Science and Software Engineering*, 7(6), 421-426.
<https://doi.org/10.23956/ijarcsse/V7I6/0305>
- Sharma, A., Bhuriya, D., & Singh, U. (2017). Survey of stock market prediction using machine learning approach. 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 506-509. <https://doi.org/10.1109/ICECA.2017.8212715>
- Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127-137. [https://doi.org/10.1016/S0167-9236\(00\)00123-8](https://doi.org/10.1016/S0167-9236(00)00123-8)
- Shobana Bai, F. J. J. (2023). A Machine Learning Approach for Carbon di oxide and Other Emissions Characteristics Prediction in a Low Carbon Biofuel-Hydrogen Dual Fuel Engine. *Fuel*, 341, 127578.
<https://doi.org/10.1016/j.fuel.2023.127578>
- Sırma, İ. (2019). Effect of Real Estate Investment Trusts Portfolio Structure on the Market Performance. *alphanumeric journal*, 7(1), 25-36.
<https://doi.org/10.17093/alphanumeric.418874>
- Simpson, M. W., Ramchander, S., & Webb, J. R. (2007). The Asymmetric Response of Equity REIT Returns to Inflation. *The Journal of Real Estate Finance and Economics*, 34(4), 513-529.
<https://doi.org/10.1007/s11146-007-9023-0>

- Sklearn. (2023). sklearn.ensemble.RandomForestRegressor — scikit-learn 1.2.2 documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
- Skurichina, M., & Duin, R. P. W. (2002). Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Pattern Analysis & Applications*, 5(2), 121-135. <https://doi.org/10.1007/s100440200011>
- Smith, G. (2015). Simple Regression. *Çinde Essential Statistics, Regression, and Econometrics* (ss. 219-259). Elsevier. <https://doi.org/10.1016/B978-0-12-803459-0.00008-X>
- Sowmya, R., & Suneetha, K. R. (2017). Data Mining with Big Data. 2017 11th International Conference on Intelligent Systems and Control (ISCO), 246-250. <https://doi.org/10.1109/ISCO.2017.7855990>
- Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04), 687-719. <https://doi.org/10.1142/S0218001409007326>
- Schwartz-Driver, S. (2008). The Great Resi Rebound. *IPE Real Estate*, 2008, 74-5.
- Szczepanek, R. (2022). Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost. *Hydrology*, 9(12), 226. <https://doi.org/10.3390/hydrology9120226>
- Şahin, C. (2014). Firmaya özgü değişkenlerle gayrimenkul yatırım ortaklıkları (GYO) getirisi arasındaki ilişkiyi incelemeye yönelik bir uygulama. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 42, 11-18.
- Takagi, H., & Handler, A. (2021). Worldwide Real Estate Investment Trust (REIT) Regimes.

- Tang, J., Liu, F., Zou, Y., Zhang, W., & Wang, Y. (2017). An Improved Fuzzy Neural Network for Traffic Speed Prediction Considering Periodic Characteristic. *IEEE Transactions on Intelligent Transportation Systems*, 18(9), 2340-2350. <https://doi.org/10.1109/TITS.2016.2643005>
- Tarazi, R., & Hasan, M. Z. (2019). The effect of economic and fundamental factors on the Australian property performance. *Asian Academy of Management Journal of Accounting and Finance*, 15(2), 155-184. <https://doi.org/10.21315/aamjaf2019.15.2.7>
- T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı, S. ve B. B. (2019). On birinci kalkınma planının (2019-2023).
- Terko, A., Zunic, E., Donko, D., & Dzelihodzic, A. (2019). Credit Scoring Model Implementation in a Microfinance Context. 2019 XXVII International Conference on Information, Communication and Automation Technologies (ICAT), 1-6. <https://doi.org/10.1109/ICAT47117.2019.8939036>
- Tiwari, U., Sheoran, H., Jain, U., & Fatima, Z. (2021). Parkinson's disease prediction using xgboost, random forest and catboost. *International Journal for Research in Engineering and Emerging Trends ((IJREET)*, 5(1), 840-846.
- Tuysuzoglu, G., & Birant, D. (2020). Enhanced Bagging (eBagging): A Novel Approach for Ensemble Learning. *The International Arab Journal of Information Technology*, 17(4), 515. <https://doi.org/10.34028/iajit/17/4/10>
- TÜİK. (2022). Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları, 2022. <https://data.tuik.gov.tr/Bulten/Index?>
- United Nations. (1948). Universal declaration of human rights. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- United Nations. (1996). United Nations Conference on Human Settlements : Habitat II.

URL1: <https://finance.yahoo.com/screener/unsaved/a1879be4-e7df-4826-b290-c8bf51575c42>

URL2: <https://tr.investing.com/pro/watchlist/w-43542073.iwl/v-021d3af4>

Ünalı, M. O. (2018). BİST’de işlem gören gayrimenkul yatırım ortaklıklarının hisse performanslarını etkileyen firmaya özgü faktörlerin panel regresyon yöntemi ile analizi [Doktora Tezi]. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü .

Vangou, C., Griffin, K., Ng, K., & Howieson, A. (2021). Worldwide Real Estate Investment Trust (REIT) Regimes.

Vens, C., & Costa, F. (2011). Random Forest Based Feature Induction. 2011 IEEE 11th International Conference on Data Mining, 744-753. <https://doi.org/10.1109/ICDM.2011.121>

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. *Energy and Buildings*, 171, 11-25. <https://doi.org/10.1016/J.ENBUILD.2018.04.008>

Way, T. W., Sahiner, B., Hadjiiski, L. M., & Chan, H.-P. (2010). Effect of finite sample size on feature selection and classification: A simulation study. *Medical Physics*, 37(2), 907-920. <https://doi.org/10.1118/1.3284974>

Wei, X., Rao, C., Xiao, X., Chen, L., & Goh, M. (2023). Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Systems with Applications*, 219, 119648. <https://doi.org/10.1016/j.eswa.2023.119648>

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82. <https://doi.org/10.3354/cr030079>

- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. 4th international conference on the practical applications of knowledge discovery and data mining , 29-39.
- Xie, Y., & Peng, M. (2019). Forest fire forecasting using ensemble learning approaches. *Neural Computing and Applications*, 31(9), 4541-4550. <https://doi.org/10.1007/s00521-018-3515-0>
- Yadav, S., & Bhole, G. P. (2020). Handling Imbalanced Dataset Classification in Machine Learning. 2020 IEEE Pune Section International Conference (PuneCon), 38-43. <https://doi.org/10.1109/PuneCon50868.2020.9362471>
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. 2016 IEEE 6th International Conference on Advanced Computing (IACC), 78-83. <https://doi.org/10.1109/IACC.2016.25>
- Yat-Hung, C., Joinkey, S. C. K., & Bo-Sin, T. (2008). Time-varying performance of four Asia-Pacific REITs. *Journal of Property Investment and Finance*, 26(3), 210-231. <https://doi.org/10.1108/14635780810871605/FULL/PDF>
- Yavaş, M., Güran, A., & Uysal, M. (2020). Covid-19 Veri Kümesinin SMOTE Tabanlı Örnekleme Yöntemi Uygulanarak Sınıflandırılması. *European Journal of Science and Technology*, 258-264. <https://doi.org/10.31590/ejosat.779952>
- Yetgin, F. (2006). Dünya’da gayrimenkul yatırım ortaklıkları: ABD, Avrupa Birliği ve Malezya uygulamaları. *Marmara Üniversitesi Avrupa Topluluğu Enstitüsü Avrupa Araştırmaları Dergisi*, 14(2), 257-278. <https://doi.org/10.29228/mjes.207>
- Zaki, M. J., & Meira, Jr, W. (2014). *Data Mining and Analysis. İçinde Data Mining and Analysis.* Cambridge University Press. <https://doi.org/10.1017/CBO9780511810114>

- Zeleznikow, J., & Stranieri, A. (1997). Knowledge discovery in the split up project. In Proceedings of the 6th international conference on artificial intelligence and law , 89-97.
- Zhang, C., & Ma, Y. (2012). Ensemble Machine Learning Methods and Applications. Ensemble Machine Learning. <https://doi.org/10.1007/978-1-4419-9326-7/cover>
- Zhang, C.-X., & Zhang, J.-S. (2008). A local boosting algorithm for solving classification problems. Computational Statistics & Data Analysis, 52(4), 1928-1941. <https://doi.org/10.1016/j.csda.2007.06.015>
- Zhang, Y., Zhao, Z., & Zheng, J. (2020). CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. Journal of Hydrology, 588, 125087. <https://doi.org/10.1016/j.jhydrol.2020.125087>
- Zheng, Z., Lu, P., & Lantz, B. (2018). Commercial truck crash injury severity analysis using gradient boosting data mining model. Journal of Safety Research, 65, 115-124. <https://doi.org/10.1016/j.jsr.2018.03.002>
- Zhou, H., Geng, S., & Wang, L. Z. (2014). Relationship between Free Cash Flow and Corporate Performance Evidence from the Listed Real Estate Companies in China. Applied Mechanics and Materials, 556-562, 6445-6448. <https://doi.org/10.4028/www.scientific.net/amm.556-562.6445>
- Zuber, V., & Strimmer, K. (2011). High-Dimensional Regression and Variable Selection Using CAR Scores. Statistical Applications in Genetics and Molecular Biology, 10(1). <https://doi.org/10.2202/1544-6115.1730>
- Zügül, M., & Şahin, C. (2015). Faiz Oranı ve Enflasyonun Gayrimenkul Yatırım Ortaklığı Getirisi Üzerindeki Etkisine Yönelik Bir Uygulama. Akademik Bakış Dergisi, 49, 147-162. <https://dergipark.org.tr/en/pub/abuhsbd/issue/32942/366009>



ISBN: 978-625-367-723-7