



**AVRUPA FUTBOL KULÜPLERİNİN FİNANSAL
PERFORMANS TAHMİNİ: MAKİNE ÖĞRENME
MODEL TASARIMLARI**

Recep Tayyip BALIKÇILAR

Yüksek Lisans Tezi

İşletme

Ana Bilim Dalı

Prof. Dr. Hakan PABUÇCU

2025

(Her Hakkı Saklıdır)

T.C.
BAYBURT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İŞLETME ANA BİLİM DALI
İŞLETME TEZLİ YÜKSEK LİSANS PROGRAMI

AVRUPA FUTBOL KULÜPLERİNİN FİNANSAL PERFORMANS TAHMİNİ:
MAKİNE ÖĞRENME MODEL TASARIMLARI

YÜKSEK LİSANS TEZİ

Recep Tayyip BALIKÇILAR

Danışman: Prof. Dr. Hakan PABUÇCU

BAYBURT- 2025

KABUL VE ONAY

Prof. Dr. Hakan PABUÇCU danışmanlığında, Recep Tayyip BALIKÇILAR tarafından hazırlanan “Avrupa Futbol Kulüplerinin Finansal Performans Tahmini: Makine Öğrenme Model Tasarımları” başlıklı bu çalışma, 03.07.2025 tarihinde yapılan savunma sınavı sonucunda oy birliği ile başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Hakan PABUÇCU

Üye : Doç. Dr. Nurdan DEĞİRMENCİ

Üye : Dr. Öğr. Üyesi Eda AYVACIK

Bu tezin kabulü Lisansüstü Eğitim Enstitüsü Yönetim Kurulu’nun/...../20... tarih ve 20...../.....-..... sayılı kararıyla onaylanmıştır.

Doç. Dr. Bünyamin ALİM
Enstitü Müdürü

BEYANNAME

Bayburt Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre Prof. Dr. Hakan PABUÇCU danışmanlığında hazırlamış olduğum “Avrupa Futbol Kulüplerinin Finansal Performans Tahmini: Makine Öğrenme Model Tasarımları” başlıklı yüksek lisans tezimin bilimsel etik değerlere ve kurallara uygun, özgün bir çalışma olduğunu, aksinin tespit edilmesi halinde her türlü yasal yaptırımını kabul edeceğimi beyan ederim.

03.07.2025

Recep Tayyip BALIKÇILAR



ÖN SÖZ

Tez çalışmam boyunca bana rehberlik eden, bilgi ve tecrübesiyle katkı sunan değerli danışmanım Prof. Dr. Hakan PABUÇCU hocama en içten teşekkürlerimi sunarım. Kendisi, yalnızca akademik desteğiyle değil, aynı zamanda sabrı ve anlayışıyla da bu süreci daha verimli ve anlamlı kılmıştır.

Bu süreçte manevi desteğini her zaman yanımda hissettiğim aileme teşekkür ederim. Ailemin sevgisi ve sabrı, bu çalışmayı tamamlamamdaki en büyük güç kaynaklarımdan biri olmuştur. Bu çalışmanın ilgili alana katkı sağlamasını ve ileride yapılacak araştırmalara ışık tutmasını temenni ederim.

Recep Tayyip BALIKÇILAR



ÖZET

YÜKSEK LİSANS TEZİ

AVRUPA FUTBOL KULÜPLERİNİN FİNANSAL PERFORMANS TAHMİNİ: MAKİNE ÖĞRENME MODEL TASARIMLARI

Recep Tayyip BALIKÇILAR

Bayburt Üniversitesi

Lisansüstü Eğitim Enstitüsü

İşletme Ana Bilim Dalı

Tez Danışmanı: Prof. Dr. Hakan PABUÇCU

Bayburt-2025, Sayfa: 91

Profesyonel futbol kulüplerinin sürdürülebilir mali yapılar oluşturabilmeleri, yalnızca sportif başarıyla değil, aynı zamanda etkin finansal yönetimle de doğrudan ilişkilidir. Bu yönüyle değerlendirildiğinde, kulüplerin finansal performanslarının doğru şekilde analiz edilmesi ve geleceğe yönelik öngörülerin oluşturulması hem yönetsel karar alma süreçleri hem de dışsal finansal düzenlemelere uyum açısından kritik bir öneme sahiptir. Bu çalışmada, Avrupa'nın üst düzey futbol liglerinde faaliyet gösteren 26 profesyonel futbol kulübüne ait 1996-2022 yılları arasındaki finansal verileri kullanılarak, kulüplerin finansal performanslarının tahmin edilebilirliği araştırılmıştır. Bu durum göz önüne alındığında, özkaynak kârlılığı (ROE) ve varlık kârlılığı (ROA) olmak üzere iki temel performans göstergesi, 19 farklı finansal oran yardımıyla öngörülmeğe çalışılmıştır. Araştırmada hem istatistiksel hem de makine öğrenmesine dayalı toplam 10 farklı regresyon algoritması uygulanmış ve modellerin tahmin başarımları MAE, RMSE, R^2 gibi çeşitli istatistiksel ölçütler aracılığıyla karşılaştırmalı olarak değerlendirilmiştir. Elde edilen bulgular, finansal oranların zaman içinde yapısal değişkenlik gösterdiği ve doğrusal olmayan ilişkiler içerdiğini doğrulamış; analizler sonucunda, Additive Model Tree (AMT), SMOreg ve Gaussian Process gibi doğrusal olmayan modellerin ROE ve ROA tahmininde daha başarılı performans sergilediği görülmüştür. Buna karşılık, Elastic Net, Pace Regression ve Linear Regression gibi doğrusal modeller, özellikle karmaşık finansal yapılar karşısında sınırlı kalmış ve görece düşük tahmin başarısı göstermiştir. Bu sonuçlar, futbol kulüplerinin finansal performans analizinde doğrusal olmayan makine öğrenimi tekniklerinin önemini ve etkinliğini ortaya koymaktadır. Elde edilen bulgular, makine öğrenmesi tabanlı modellerin, futbol kulüplerinin finansal analizinde geleneksel yöntemlerin ötesine geçilerek etkili biçimde kullanılabileceğini göstermektedir. Ayrıca, kulüplerin finansal risklerinin erken teşhisi ve stratejik karar destek mekanizmalarının geliştirilmesi açısından bu tür modellerin önemli katkılar sağlayabileceği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Özkaynak kârlılığı, varlık kârlılığı, finansal performans, makine öğrenmesi, regresyon modelleri, finansal tahmin.

ABSTRACT

M. SC. THESIS

FINANCIAL PERFORMANCE FORECASTING OF EUROPEAN FOOTBALL CLUBS: MACHINE LEARNING MODEL DESIGNS

Recep Tayyip BALIKÇILAR

Bayburt University

Institute of Graduate Studies

Department of Business Administration

Thesis Advisor: Prof. Dr. Hakan PABUÇCU

Bayburt-2025, Pages: 91

The ability of professional football clubs to create sustainable financial structures is directly related not only to sporting success but also to effective financial management. In this context, correctly analyzing the financial performance of clubs and creating future forecasts is of critical importance in terms of both managerial decision-making processes and compliance with external financial regulations. In this study, the predictability of the financial performance of clubs was investigated using financial data from 1996 to 2022 belonging to 26 professional football clubs operating in Europe's top-level football leagues. Considering this situation, two basic performance indicators, return on equity (ROE) and return on assets (ROA), were tried to be predicted with the help of 19 different financial ratios. In the study, a total of 10 different regression algorithms based on both statistical and machine learning were applied and the prediction performances of the models were evaluated comparatively through various statistical measures such as MAE, RMSE, R^2 . The findings confirmed that financial ratios exhibit structural variability over time and contain nonlinear relationships; as a result of the analyses, it was observed that nonlinear models such as Additive Model Tree (AMT), SMOreg and Gaussian Process exhibited more successful performance in estimating ROE and ROA. In contrast, linear models such as Elastic Net, Pace Regression and Linear Regression were limited and showed relatively low prediction success, especially in the face of complex financial structures. These results reveal the importance and effectiveness of nonlinear machine learning techniques in the financial performance analysis of football clubs. The findings show that machine learning-based models can be used effectively in the financial analysis of football clubs, going beyond traditional methods. It has also been concluded that such models can make significant contributions in terms of early diagnosis of financial risks of clubs and development of strategic decision support mechanisms.

Keywords: Return on equity, return on assets, financial performance, machine learning, regression models, financial forecasting.

İÇİNDEKİLER

KABUL VE ONAY.....	II
BEYANNAME.....	III
ÖN SÖZ.....	IV
ÖZET.....	V
ABSTRACT.....	VI
İÇİNDEKİLER.....	VII
TABLolar.....	X
ŞEKİLLER.....	XI
KISALTMALAR.....	XII
SİMGELER.....	XII
GİRİŞ.....	1
Araştırmanın Konusu ve Problemi.....	4
Araştırmanın Amacı.....	4
Araştırmanın Önemi ve Gerekçesi.....	5
Araştırmanın Sınırlılıkları.....	5
Varsayımlar.....	6
BİRİNCİ BÖLÜM.....	7
1. KURAMSAL ÇERÇEVE.....	7
1.1. FUTBOLUN TARİHİ VE ENDÜSTRİ SÜRECİ.....	7
1.1.1. Futbol Tarihi.....	7
1.1.2. Futbolun Endüstrisi.....	8
1.2. FUTBOL KULÜPLERİNİN FİNANSAL YAPISI.....	10
1.2.1. Futbol Kulüplerinin Mali Gelir Kaynakları.....	11
1.2.2. Futbol Kulüplerinin Mali Gider Kaynakları.....	11
1.2.3. Başabaş Kuralının Hesaplamasının Sonuçları.....	12
1.3. LİTERATÜR ARAŞTIRMASI.....	12
1.4. MAKİNE ÖĞRENME.....	16
1.4.1. Denetimli Öğrenme.....	17
1.4.2. Denetimsiz Öğrenme.....	18
1.4.3. Yarı-Denetimli Öğrenme.....	19
1.4.4. Makine Öğrenme Sürecindeki Adımlar.....	20
1.5. MAKİNE ÖĞRENMESİ ALGORİTMALARI.....	20
1.5.1. YSA-Yapay Sinir Ağları.....	20
1.5.1.1. Biyolojik sinir ağları.....	21
1.5.1.2. Yapay sinir hücresi.....	22
1.5.1.3. Yapay sinir ağlarının yapısı.....	24
1.5.1.4. Yapay sinir ağının çalışması.....	25
1.5.1.5. Yapay sinir ağlarının eğitilmesi ve test edilmesi.....	25
1.5.2. DVM- Destek Vektör Makinaları.....	26
İKİNCİ BÖLÜM.....	29
2. YÖNTEM.....	29

2.1. WEKA YAZILIMI.....	29
2.1.1. Weka Arayüzü	29
2.1.2. Veri Setinin İşlenmesi	32
2.1.3. Regresyon Modeli	33
2.1.4. Seçilen Regresyon Modüllerinin Yüklenmesi.....	34
2.1.5. Regresyon Modülü Uygulaması	35
2.1.6. Modül Parametrelerinin Yapılandırılması.....	39
2.2. WEKA REGRESYON ALGORİTMALARI.....	41
2.2.1. MultilayerPerceptron (Çok Katmanlı Algılayıcı) Algoritması.....	41
2.2.2. Linear Regression Algoritması.....	41
2.2.3. SMOreg Algoritması	43
2.2.4. GaussianProcesses Algoritması.....	45
2.2.5. M5Rules Algoritması	46
2.2.6. Alternating Model Trees Algoritması	48
2.2.7. PaceRegression Algoritması.....	48
2.2.8. Isotonic Regression Algoritması	50
2.2.9. RepTree Algoritması	50
2.2.10. Elastic Net Algoritması	52
2.3. MODEL DEĞERLENDİRME KRİTERLERİ	53
2.3.1. R^2 -Determinasyon Katsayısı.....	53
2.3.2. Mean Absolute Error-MAE (Ortalama Mutlak Hata)	53
2.3.3. Root Mean Squared Error-RMSE (Hata Kareleri Ortalaması Karekökü).....	53
2.3.4. Mean Squared Error-MSE (Ortalama Kare Hatası)	54
2.3.5. Mean Absolute Percentage Error-MAPE (Ortalama Mutlak Yüzde Hatası)	54
2.3.6. Relative Absolute Error-RAE (Göreceli Mutlak Hata)	54
2.3.7. Root Relative Squared Error-RRSE (Kök Göreceli Kare Hata).....	54
ÜÇÜNCÜ BÖLÜM.....	55
3. BULGULAR	55
3.1. VERİ SETİ VE DEĞİŞKENLER	55
3.2. MODEL OLUŞTURMA.....	58
3.2.1. MultilayerPerceptron Modeli	58
3.2.2. Linear Regression Modeli	61
3.2.3. SMOreg Modeli.....	62
3.2.4. GaussianProcesses Modeli	64
3.2.5. M5Rules Modeli	66
3.2.6. Alternating Model Trees Modeli	67
3.2.7. PaceRegression Modeli	68
3.2.8. Isotonic Regression Modeli	70
3.2.9. RepTree Modeli.....	71
3.2.10. Elastic Net Modeli	72
3.3. MAKİNE ÖĞRENMESİ TABANLI MODELLERİN TAHMİN BAŞARIMININ KARŞILAŞTIRILMASI	75
SONUÇ, TARTIŞMA VE ÖNERİLER.....	78
KAYNAKÇA.....	83



TABLÖLAR

Tablo 1: MLP Modeli için Parametre Ayarları.....	40
Tablo 2: WEKA Regresyon Algoritmaları.....	41
Tablo 3: Avrupa Futbol Kulüpleri ve Dönemleri	56
Tablo 4: Çalışmada Kullanılan Değişkenler.....	57
Tablo 5: MLP-ROE için Parametre Ayarları.....	59
Tablo 6: MLP-ROA için Parametre Ayarları	59
Tablo 7: MLP için Sonuçlar	60
Tablo 8: Linear Regresyon-ROE ve ROA için Parametre Ayarları	61
Tablo 9: Linear Regresyon için Sonuçlar	61
Tablo 10: SMOreg- ROE için Parametre Ayarları	62
Tablo 11: SMOreg- ROA için Parametre Ayarları.....	63
Tablo 12: SMOreg için Sonuçlar	64
Tablo 13: GaussianProcess-ROE ve ROA için Parametre Ayarları.....	64
Tablo 14: GaussianProcess için Sonuçlar.....	65
Tablo 15: M5Rules-ROE ve ROA için Parametre Ayarları	66
Tablo 16: M5Rules için Sonuçlar	66
Tablo 17: AMT-ROE için Parametre Ayarları	67
Tablo 18: AMT-ROA için Parametre Ayarları.....	67
Tablo 19: AMT için Sonuçlar.....	68
Tablo 20: PaceRegression-ROE ve ROA için Parametre Ayarları	69
Tablo 21: PaceRegression için Sonuçlar	69
Tablo 22: Isotonic Regression için Sonuçlar	70
Tablo 23: RepTree-ROE ve ROA için Parametre Ayarları	71
Tablo 24: RepTree için Sonuçlar	72
Tablo 25: Elastic Net-ROE için Parametre Ayarları	73
Tablo 26: Elastic Net-ROA için Parametre Ayarları.....	74
Tablo 27: Elastic Net için Sonuçlar	74
Tablo 28: ROE için Model Sıralama	75
Tablo 29: ROA için Model Sıralaması	76

ŞEKİLLER

Şekil 1: Biyolojik Sinir Hücresi	21
Şekil 2: Yapay Sinir Hücresinin Yapısı	22
Şekil 3: Bir Yapay Sinir Ağı Örneği	25
Şekil 4: WEKA Görsel Arayüz Seçim Paneli	30
Şekil 5: WEKA Explorer Arayüzü.....	31
Şekil 6: Örnek Cvs (Virgülle Ayrılmış) Dosyası	32
Şekil 7: WEKA Explorer Arayüzünde, Cvs Yapısında Veri Seti	32
Şekil 8: WEKA Veri Seti Filtre Uygulama Penceresi	33
Şekil 9: WEKA Package Manager Seçimi.....	34
Şekil 10: WEKA Modül İndirme Arayüzü	35
Şekil 11: Weka Regresyon Uygulama Penceresi	36
Şekil 12: WEKA MLP Analiz Test Sonuçları	37
Şekil 13: Korelasyon Grafik Çıktısı.....	38
Şekil 14: Detaylı Korelasyon Grafik Çıktısı	39
Şekil 15: Modül Parametreleri Arayüzü	40
Şekil 16: Excel Formatlı ROA Veri Seti.....	58

KISALTMALAR

AMT	: Alternatif Ağaç
DL	: Derin Öğrenme
DVM	: Destek Vektör Makinaları
DVR	: Destek Vektör Regresyonu
FFP	: Finansal Fair Play
KNN	: K-En Yakın Komşu Algoritması
MLP	: Çok Katmanlı Algılayıcı
MLR	: Çoklu Doğrusal Regresyon
ML	: Makine Öğrenme
MAE	: Ortalama Mutlak Hata
MSE	: Ortalama Kare Hata
MAPE	: Ortalama Mutlak Yüzde Hatası
PCA	: Temel Bileşen Analizi
RAE	: Göreceli Mutlak Hata
RMSE	: Hata Kareleri Ortalaması Karekökü
RRSE	: Kök Göreceli Kare Hata
ROE	: Özkaynak Kârlılığı
ROA	: Varlık Kârlılığı
SMOreg	: Sıralı Minimal Optimizasyon
UEFA	: Avrupa Futbol Federasyonları Birliği
YSA	: Yapay Sinir Ağları

SİMGELER

C	: Ceza Katsayısı
€	: Ceza Boyutu
I	: Birim Matris
p	: Korelasyon Katsayısı
R²	: Determinasyon Katsayısı
P	: Korelasyon Katsayısı
ξ	: Hata Terimi
λ	: Düzenleme Parametresi
>	: Büyük
<	: Küçük
≥	: Büyük eşit
≤	: Küçük eşit
%	: Yüzde

GİRİŞ

Günümüz futbolunun sınırları aştığı ve milyonlarca insanın ilgisini çektiği bu zamanda, kulüplerin finansal istikrarı sadece sahadaki rekabet gücünü belirlemekle kalmamış, aynı zamanda küresel spor endüstrisindeki etkilerini de belirlemiştir.

Futbol kulüpleri, yalnızca sahadaki performanslarını değil, aynı zamanda uzun vadeli sürdürülebilirliklerini ve küresel nüfuzlarını da etkileyen çok yönlü mali yapılara sahip karmaşık varlıklar olarak faaliyet göstermektedir. Kayda değer transfer anlaşmalarından kazançlı yayın haklarına kadar futbolun mali ortamı, bu sporun büyüklüğünü yansıtan dinamik bir ekosistemdir. Sahada başarı arayışı saha dışı mali stratejilerle giderek daha fazla iç içe geçerken, futbol kulüplerinin finansal anatomisini anlamak hem yatırımcılar hem de taraftarlar için bir zorunluluk haline gelmektedir.

Kâr amacı gütmeyen futbol kulüplerinin temel hedeflerini, ticari kazanç amacıyla faaliyet gösteren profesyonel futbol kulüpleriyle karıştırmamak büyük önem taşımaktadır. Futbolun etkisini artıran ve onu her yerde bulunan bir olguya dönüştüren çok sayıda değişiklik meydana gelmiştir. Alınan karar sonrasında, sözleşmeli oyuncuların transfer edilebilmesi ve yabancı oyuncu sayısının artırılması, bazı eleştirmenler tarafından oyuncuların iş güvencesi ve sözleşme haklarının zedelenmesi bağlamında değerlendirilmiş; bu durum, profesyonel futbolculara yönelik yapısal bir istismar biçimi olarak nitelendirilmiştir. Bu aynı zamanda ücretlerde önemli bir artışa yol açmış ve önemli bir mali etkiye yol açmıştır (Andreff & Staudohar, 2000).

Futbol, önemli mali kaynakları harekete geçirebilen kazançlı bir sektördür. Ekonomik açıdan bakıldığında futbol ve diğer spor dallarında yaşanan profesyonelleşme süreci, futbolu bir oyun olmaktan çıkarıp, ekonomik düzeyde işleyen bir sektör haline getirmiştir. Bunun sonucunda kapitalist sistemin ilke ve inançları futbol sporu üzerinde egemenlik kurmaya başlamıştır. Futbolun evrimi, sporun daha geniş bir şekilde anlaşılmasına yol açmış, bu da futbol kulüpleri için finansal giriş ve çıkışların çeşitliliğinde ve büyüklüğünde bir artışa yol açmıştır. Yüksek bütçeli kulüplerin, özellikle dernek olarak faaliyet gösterenlerin, yönetim stratejilerini değiştirme ihtiyacı, onların daha yapılandırılmış ve iş odaklı bir çerçeveye sahip kurumsal yapılara dönüşmesini zorunlu kılmıştır.

Futbol, 21. yüzyılda sadece bir spor dalı olmaktan çıkarak küresel ölçekte milyarlarca dolarlık bir endüstri haline gelmiştir. Özellikle Avrupa'da medya gelirleri, sponsorluk anlaşmaları ve uluslararası turnuvalar aracılığıyla futbol kulüplerinin gelirleri son yirmi yılda

istikrarlı bir artış göstermiştir. UEFA verilerine göre, Avrupa kulüp futbolunun 20 yılı aşkın süredir yıllık ortalama %8,2 oranında gelir artışı kaydettiği belirtilmektedir. Bu büyüme, futbolun ekonomik gücünü ve uluslararası etkisini artırmıştır (Morrow, 2023). Futbolun küreselleşmesi, yerel ve uluslararası süreçlerin iç içe geçmesini beraberinde getirmiştir. Önde gelen futbol kulüpleri “glokal” (yerel ve küresel) transnasyonal şirketler olarak değerlendirilmektedir. Bu durum, futbolun sadece sportif bir etkinlik olmanın ötesine geçerek, ekonomik ve politik bir aktör haline geldiğini göstermektedir (Giulianotti & Robertson, 2004)

Türkiye’de futbol ekonomisi, özellikle 2000’li yıllardan itibaren önemli bir dönüşüm sürecine girmiştir. Süper Lig kulüplerinin artan yayın gelirleri, transfer harcamaları ve sponsorluk anlaşmaları, Türk futbolunun ekonomik dinamiklerinde ciddi değişikliklere neden olmuştur. Ancak bu büyüme, beraberinde finansal sürdürülebilirlik sorunlarını da gündeme getirmiştir. Kulüplerin gelir-gider dengesizlikleri, yüksek borçlanma düzeyleri ve UEFA’nın Finansal Fair-Play (FFP) düzenlemelerine uyum sağlama çabaları, finansal yönetim açısından önemli sorunlara yol açmıştır. Bu çerçevede, Türk futbol kulüplerinin mali yapıları, profesyonel yönetim anlayışı ve şeffaf finansal raporlama uygulamaları açısından yeniden değerlendirilmeye muhtaçtır. Süper Lig kulüplerinin yıllık finansal tablolarının analizi sonucunda, yapılan büyük yatırımlara rağmen, yüksek tutarlarda borçlu olduğu, büyük ölçüde likidite ve kârlılık sorunları yaşadığı ve sürekli olarak artan bir şekilde, finansal sorunlarla karşı karşıya olduğu sonucuna ulaşılmıştır (Uluyol, 2014).

Bu çerçevede futbol kulüplerinin finansal analizleri ve gelecek performans tahminleri, yalnızca sportif yönetim açısından değil; yatırım, regülasyon ve finansal sürdürülebilirlik açısından da stratejik öneme sahiptir. Özellikle Türkiye gibi finansal istikrarın sektörler arası farklılık gösterdiği ekonomilerde, kulüplerin mali yapılarını doğru analiz edebilmek ve bu yapılar üzerinden sağlam tahmin modelleri kurabilmek, karar alıcılar için hayati bir ihtiyaç haline gelmiştir.

Finansal performans ölçümü, şirketlerin likidite, finansal kompozisyon, operasyonel verimliliği ve karlılığı hakkında objektif verilerin elde edilmesini ve değerlendirilmesini içerir. Kârlılık göstergeleri, firmaların para üretme ve sürdürülebilir kârlılığı sürdürme yetenekleri hakkında fikir verir; mali yapı göstergeleri ise özsermaye ile borç arasındaki ilişkiyi gösterir. Her şirket öncelikle finansal performansından etkilenir. Geleneksel olarak finansal performansa ilişkin bu değerlendirmeler, belirli endüstrilere ve işletme türlerine odaklanılarak sınırlı bir şekilde yürütülmüştür.

Son zamanlarda, Avrupa futbol takımlarının önemli bir kısmı, mali performanslarındaki düşüşün ardından düşük performanstan kaynaklanan kayıplar bildirdikleri için mali zorluklarla karşı karşıya kalmıştır. UEFA, kulüpler için gelecekteki olası mali sorunları azaltmak amacıyla lisans düzenlemelerinin bir parçası olarak mali adil oyun (FFP) gerekliliklerini uygulamaya koymuştur. Bu yaklaşımın amacı, kulüplerin mali hesaplarını yakından izleyerek ve düzenleyerek borç temerrütlerini ve organik olmayan finansman gibi etik olmayan mali uygulamaları en aza indirmektir.

Finansal performansın öngörülmesine yönelik çalışmalar, genellikle geleneksel istatistiksel yöntemlere dayansa da son yıllarda yapay zekâ ve makine öğrenmesi teknikleri bu alanda güçlü alternatifler olarak öne çıkmaktadır. Literatür incelendiğinde, makine öğrenmesi algoritmaları, büyük veri setleri içerisindeki karmaşık örüntüleri tanıma, doğrusal olmayan ilişkileri modelleme ve hata düzeylerini minimize etme yetenekleri sayesinde finansal tahminlerde yüksek başarı oranları sunmaktadır. Özellikle kârlılık oranlarının modellenmesinde bu tür tekniklerin kullanımı, geleneksel yöntemlerin ötesine geçilerek daha bütüncül ve isabetli analizlerin yapılmasına olanak tanımaktadır.

Bu çalışmanın amacı, Avrupa'nın beş büyük futbol liginde faaliyet gösteren seçilmiş kulüplerin uzun dönemli finansal verilerini kullanarak, özkaynak kârlılığı (ROE) ve varlık kârlılığı (ROA) gibi temel performans göstergelerini farklı makine öğrenmesi algoritmalarıyla tahmin etmektir. Bu kapsamda 1996–2022 dönemine ait 26 kulübe ilişkin finansal oranlar analiz edilmiş ve 10 farklı regresyon tabanlı model kullanılarak bu oranların tahmin başarımı karşılaştırılmıştır. Çalışmada kullanılan modellerin performansları, ortalama mutlak hata (MAE), ortalama karekök hata (RMSE), açıklayıcılık katsayısı (R^2) gibi çeşitli istatistiksel ölçütler aracılığıyla değerlendirilmiştir.

Araştırmanın özgün yönü, futbol kulüpleri gibi finansal yapıları klasik şirketlerden önemli ölçüde farklılaşan organizasyonların, ileri seviye makine öğrenmesi algoritmaları aracılığıyla daha etkin analiz edilebileceğini göstermesidir. Ayrıca çalışma, zaman serisi verilerinin yapay zekâ modelleriyle nasıl bütünleştirilebileceğine ve bu entegrasyonun tahmin performansını nasıl artırabileceğine dair önemli bulgular sunmaktadır. Bu doğrultuda yapılan çalışmalarda elde edilen sonuçlar, yalnızca akademik literatüre katkı sağlamakla kalmayıp, aynı zamanda kulüplerin mali stratejilerinin oluşturulmasında ve finansal karar destek mekanizmalarının geliştirilmesinde de pratik faydalar sunmaktadır.

Araştırmanın Konusu ve Problemi

Spor kulüplerinin mali yapıları, yalnızca sportif başarılarla değil, aynı zamanda uzun vadeli sürdürülebilirliğe, yatırım planlamalarına ve düzenleyici kurumlarla olan finansal ilişkilerine de doğrudan etki etmektedir. Bu durum göz önüne alındığında, kulüplerin finansal durumlarının doğru şekilde analiz edilmesi ve gelecekteki performanslarının öngörülebilir hâle getirilmesi, etkili karar alma süreçleri açısından kritik bir gereklilik hâline gelmiştir.

Geleneksel finansal analiz yöntemleri çoğunlukla doğrusal modellere dayanmakta olup, karmaşık ve değişken yapıya sahip olan spor kulüplerinin mali göstergelerini açıklamada yetersiz kalabilmektedir. Özellikle finansal oranların zamanla yapısal değişkenlik göstermesi, doğrusal olmayan ilişkilerin ön plana çıkmasına neden olmakta ve bu durum klasik modelleme tekniklerinin tahmin gücünü sınırlamaktadır. Bu nedenle çalışmanın temel problemi, profesyonel futbol kulüplerinin finansal performanslarını temsil eden ROE ve ROA değişkenlerinin, makine öğrenmesi algoritmaları aracılığıyla doğruluk düzeyi yüksek bir biçimde tahmin edilip edilemeyeceğinin ortaya konulmasıdır.

Araştırmada, 1996-2022 yılları arasında 26 farklı kulübe ait 19 finansal oran kullanılarak, on farklı regresyon algoritmasının tahmin performansları karşılaştırmalı olarak analiz edilmiştir. Bu süreçte doğrusal ve doğrusal olmayan algoritmaların başarı düzeyleri çeşitli hata metrikleri ile değerlendirilmiş, böylece hangi modellerin futbol kulüplerinin finansal yapısını daha iyi temsil edebildiği ortaya konulmuştur. Çalışma, spor finansmanı literatürüne veri bilimi temelli bir yaklaşım kazandırmayı ve finansal karar destek sistemlerinin geliştirilmesine katkı sunmayı amaçlamaktadır.

Araştırmanın Amacı

Yüksek bütçeli futbol kulüpleri, endüstriyel gelişime adım atmak için profesyonelleşme ve şirketleşme yolunu benimsemektedir. Şirketleşme modelinin finansal kaynaklara daha kolay ulaşma imkânı sunması ve sermaye piyasalarına erişim sağlaması, düşük maliyetli fon oluşturarak futbol kulüplerini hızla kurumsallaştırmaktadır. Futbol kulüpleri, varlıklarını koruma ve finansal istikrarı sağlama konularında giderek daha fazla önem kazanmaktadır (Güngör & Kocamış, 2018).

Futbol kulüplerinin detaylı bir mali analizini yaparak, faaliyetlerini şekillendiren karmaşık ekonomik dinamikleri ortaya çıkarmaktır. Profesyonel futbolun sürekli gelişen ortamında kulüplerin mali sağlığı; rekabet etme, büyümeyi sürdürme ve spor endüstrisine özgü zorlukların üstesinden gelme becerilerinde çok önemli bir rol oynamaktadır. Futbol

kulüplerinin mali performansını, gelir yaratma, maliyet yönetimi ve karlılık gibi temel göstergeler merceğinden incelenmesi hedeflenmektedir.

Bu analiz, kulüpler tarafından kullanılan finansal stratejilerin verimliliği ve etkinliği hakkında fikir verecektir. Bu tez çalışmasının ana amacı, futbol kulüplerinin finansal performansını ölçmek ve tahmin etmek için çeşitli makine öğrenme yöntemlerini karşılaştırmaktır. Finansal performansın göstergeleri olarak ROE ve ROA gibi finansal oranların tahmini için en uygun modelin tasarlanması hedeflenmektedir.

Araştırmanın Önemi ve Gerekçesi

Futbol kulüplerinin mali analizini yürütmenin önemi ve mantığı, mali sağlığın bu spor kuruluşlarının sürdürülebilir başarısı ve rekabet gücünde oynadığı önemli rolün tanınmasında yatmaktadır. Futbol kulüpleri, yüksek gelir ve büyük bütçelerle faaliyet gösteren, aynı zamanda dünya çapında büyük bir ticari etkiye sahip olan organizasyonlardır. Bu kulüplerin finansal durumu, sadece sportif başarılarıyla değil, aynı zamanda kulübün sürdürülebilirliği ve uzun vadeli stratejik kararlarıyla da doğrudan ilişkilidir. Kulüplerin performansını anlamak ve finansal sağlığını ölçmek, kulüp yönetimleri, yatırımcılar ve sponsorluk anlaşmaları açısından son derece kritik bir öneme sahiptir. Finansal oranlar bu ölçümleri yapmak için yaygın olarak kullanılır. Finansal performans, genellikle karmaşık ve çok sayıda faktörü içerir. Bu nedenle, futbol kulüplerinin finansal performansını tahmin etmek ve analiz etmek için geleneksel istatistiksel yöntemler yetersiz kalabilir.

Son yıllarda, makine öğrenimi (ML) ve derin öğrenme (DL) gibi yapay zekâ yöntemlerinin finansal analizde kullanımının artması, daha doğru ve güvenilir tahminler yapmayı mümkün kılmaktadır. Özellikle yapay sinir ağları (YSA) gibi derin öğrenme yöntemleri, çok katmanlı ve karmaşık ilişkileri modelleyebilme yetenekleriyle bu tür tahminlerde etkili olabilir (Sezer vd., 2020).

Araştırmanın Sınırlılıkları

Çalışma kapsamı yalnızca 26 kulüp ile sınırlandırılmıştır. Bu durum, bulguların genellenebilirliğini kısıtlayarak Avrupa dışındaki futbol kulüplerine veya daha alt liglerde yer alan takımlara uygulanabilirliğini sınırlı hâle getirmektedir. Zaman aralığı olarak 1996-2022 yılları arasındaki dönem dikkate alınmış olup, daha güncel veya farklı dönemlere ait veriler ile yapılacak analizler farklı sonuçlar üretebilir. Finansal oran seçimi, literatürde sıkça kullanılan

ve veri erişimi mümkün olan göstergelerle sınırlı tutulmuş; bu nedenle başka önemli finansal oranlar çalışmaya dahil edilememiştir.

Kulüplerin mali yapılarını etkileyebilecek sportif başarı, ekonomik krizler, sponsorluk anlaşmaları, yönetim değişiklikleri veya düzenleyici kurumların müdahaleleri gibi dışsal faktörler modelleme sürecine dahil edilmemiştir. Bazı makine öğrenmesi algoritmalarının tahmin başarımı yüksek olsa da bu modellerin karar mekanizmalarının sınırlı ölçüde yorumlanabilir olması, sonuçların karar alıcılar açısından doğrudan uygulanabilirliğini zorlaştırabilmektedir.

Varsayımlar

Bu araştırmada, Avrupa'nın üst düzey futbol liglerinde faaliyet gösteren 26 kulübe ait 1996-2022 yılları arasındaki finansal verilerin doğru, güvenilir ve eksiksiz şekilde raporlandığı varsayılmıştır. Elde edilen verilerin ilgili kaynaklardan doğru biçimde temin edildiği kabul edilerek, analiz süreci bu veriler ışığında yürütülmüştür.

Araştırmada uygulanan regresyon algoritmalarının, veri setinin yapısal özelliklerini yeterli düzeyde yansıtacak biçimde tahminleme gücüne sahip olduğu varsayılmıştır. Özellikle makine öğrenmesi tabanlı modellerin, doğrusal olmayan ilişkileri yakalama kapasitesi doğrultusunda doğru öngörüler sunduğu değerlendirilmiştir. Bununla birlikte, yıllık verilerin kullanılması nedeniyle, geçmiş dönemlerin etkilerinin sınırlı düzeyde de olsa mevcut dönem performansına yansıdığı varsayılmıştır. Ayrıca farklı liglerde yer alan kulüplerin finansal göstergelerinin benzer ekonomik alanlarda analiz edilebileceği ve karşılaştırılabilir olduğu varsayılmıştır.

BİRİNCİ BÖLÜM

1. KURAMSAL ÇERÇEVE

Çalışmamızın bu bölümünde, amatör futboldan endüstriyel futbola geçişin tarihsel süreci, futbola dair önemli kilometre taşları ve endüstriyel futbolun oluşumuna etki eden faktörler ele alınmıştır. İlk alt başlıkta, futbolun ortaya çıkışı ve günümüze kadar olan evrimi saha içindeki değişimlere odaklanmıştır. İkinci alt başlık, futbolun endüstriyel bir boyut kazanmasına yol açan faktörleri çeşitli kaynaklardan derlenen çalışmalarla açıklanmıştır.

1.1. FUTBOLUN TARİHİ VE ENDÜSTRİ SÜRECİ

1.1.1. Futbol Tarihi

Futbol, günümüzde büyük bir popülerlik kazanmış olup, kökenleri tam olarak bilinmese de milattan önceki dönemlerde Çin, Japon, Türk ve Mısır gibi çeşitli toplumlarda oynandığına dair duvar resimleri ve antik yazılı kaynaklar tarafından gösterilmiştir. Ancak, coğrafyanın iklim ve inanç sistemlerinin etkisiyle, futbolun kuralları ve oynanış şekillerinde çeşitlilikler görülmüştür.

Antik Çin'in tarihî derinliklerinde, futbolun erken bir formu olarak kabul edilen Tsu Chu, MÖ 255 ile 206 yılları arasında Tsin Hanedanlığına kadar uzanır. Askerleri eğitmek amacıyla kullanılan beden eğitimi programının bir parçası olarak ortaya çıkan bu spor, zaman içinde evrim geçirerek Han Hanedanlığının sonraki dönemlerinde, MÖ 206 ile MS 220 yılları arasında yaygın bir şekilde oynanmıştır (History of football, 2018).

Japon kültüründe, günümüz futboluna karşılık gelen bir spor olan Kemari'nin kökenleri hakkında sınırlı sayıda belge bulunmaktadır. Kemari'nin bu olgunun kesin kökeni belirsizliğini korumasına rağmen Japon toplumunun eski yazılı kaynaklarından biri olan Nihongi' den elde edilen bilgiler, Kemari'nin M.Ö. 644'te Japon toplumunda oynandığını göstermektedir (Erdoğan, 2008).

Antik Yunan toplumunda var olan “Episkyros” adlı spor, çok az somut detayın hayatta kaldığı bir oyun türüydü. Buna karşın, Roma döneminde daha canlı bir şekilde varlığını sürdüren “Harpastum”, iki takım arasında daha küçük bir topun kullanıldığı bir spor olarak öne çıktı. Bu oyun, dikdörtgen bir alan üzerinde sınırlar ve bir merkez çizgisi ile belirlenmişti. Temel amaç, topu rakip takımın sınır çizgilerinden geçirmektir. Harpastum, 700-800 yıl boyunca büyük bir popülerlik kazandı. Ancak, Romalılar bu sporu Britanya'ya taşıdıklarında,

oyunun ayak kullanımının sınırlı olması nedeniyle etkileyici sonuçlar elde etmek zordu (FIFA, 2013).

Orta Asya'da, futbolun dokuz yüzyıl önce Kaşgarlı Mahmut tarafından “Tepük” olarak adlandırılarak oynandığı belirtilmiştir. Kaşgarlı Mahmut'un “Divân-ı Lügati't-Türk” adlı eserinde bu sporun varlığına dair bahsedilen bilgiler, zaman zaman düzenlenen yarışmalarda karşılaşmaların meydana geldiğini de ortaya koymaktadır (Abalı, 1974).

Günümüz modern futbolu 19. yüzyılın ortalarında İngiltere'de, farklı okullarda oynanan futbol türleri arasında bir standardizasyon ihtiyacı doğmuştu. 1863'te Football Association (Futbol Birliği) kurularak modern futbolun temel kuralları belirlendi (Gözcü, 2011).

1.1.2. Futbolun Endüstrisi

Modern Futbolun 19. yüzyılın ortalarında başlayan serüveni, sanayileşme ve modernleşme süreçleri ile, üretim ve organizasyon modelleri arasında paralellik göstermektedir. Futbolun endüstrileşme veya ticarileşme süreci hızlandıkça, futbol ekonomisi büyüme kaydetmiş, ancak bu büyümeyle birlikte futbol deneyimimizin boyutları da değişmiştir. Artık sadece oynamak ve haz almak değil, kazanmak ve başarmak ön plandadır, çünkü güçlü bir rekabet ortamı futbol sahalarında da varlık göstermektedir. Endüstriyel futbolun oluşumunda, sermayenin rahat dolaşımı sürecine benzer bir şekilde, futbolcuların serbest dolaşımının önemi ortaya çıkmıştır. Bosman kararları gibi önemli adımların atılması, futbolun küreselleşmesini hızlandırmıştır. Artık futbol, sadece bir oyun olmanın ötesinde, uluslararası bir arenada rekabet edilen bir endüstri haline gelmiştir. Bu süreçte, futbolun ekonomik büyümesi ve küresel etkileşim, sporun daha karmaşık ve çeşitli bir dinamik kazanmasına neden olmuştur (Talinçiler, 2008).

Futbolun profesyonelleşmesi, sporun ekonomik, kültürel ve toplumsal dinamiklerinde derin değişimlere neden olmuştur. Profesyonelleşme, oyuncuların daha fazla gelir elde etmelerine olanak tanırken, kulüplerin finansal sürdürülebilirlikleri ve rekabet avantajları üzerinde etkili olmuştur. Profesyonelleşme, futbolun geleneksel amatör statüsünden uzaklaşarak, oyuncuların rekabet yetenekleri karşılığında ücret almalarına ve kulüplerin kâr amaçlı organizasyonlara dönüşmelerine işaret eder. Bu süreç, özellikle 19. yüzyılın ortalarında İngiltere'de ortaya çıkmış olup, kademeli bir evrimle dünya genelinde yayılmıştır (Yılmaz, 2020).

Sanayi kapitalizmi, futbolda profesyonelleşme sürecini başlatırken aynı zamanda oluşturduğu aşırı sermaye birikimi ile futbolun küresel bir fenomen haline gelmesine olanak

tanımıştır. Bu dönemde, İngiltere'nin sanayileşme ve kapital birikimle sağladığı ekonomik avantajlar, devletin yayılcı ekonomik politikalarıyla birleşerek, dünyanın çeşitli bölgelerine sermaye akışını hızlandırmıştır (Yiğit, 2012).

Futbol müsabakaları, kulüpler ve sporcular, genel anlamda medya için büyük bir öneme sahiptir; özellikle de televizyon, bu önemli etkinlikleri öne çıkaran bir platform olarak öne çıkmaktadır. Toplumun futbola olan yoğun ilgisi, yüksek izlenme oranları, reklam ve sponsorluk anlaşmaları gibi faktörlerle birleşerek, müsabakaların yayın hakları, televizyon kanalları için en ilgi çekici içerikler arasında yer alır. Günümüzde, bu alandaki ulusal ve uluslararası rekabetin oldukça ciddi boyutlara ulaştığı görülmektedir (Talmiciler, 2010).

Sponsorluk çalışmaları, işletmelerin markalarını farklılaştırmalarına ve güçlü kurumsal mesajlar iletmelerine katkı sağlar. Bu çalışmalar, işletmelerin imajını güçlendirmeyi amaçlar hem kurumsal hem de marka boyutunda etkili olabilir. Özellikle sponsorluk çalışmaları, desteklenen olay veya kurumla bağlantılı olarak kurum kimliğinin oluşturulmasına ve güçlendirilmesine imkân tanır (Enginkaya, 2014).

Kulüpler, markalarla gerçekleştirdikleri sponsorluk anlaşmalarıyla sadece finansal destek sağlamakla kalmaz, aynı zamanda bu etkinlikler aracılığıyla geniş bir kitleye ulaşma ve marka bilinirliğini artırma fırsatını değerlendirirler. Bu çerçevede, futbolun sunduğu büyük izleyici kitlesi, markalar için etkili bir pazarlama platformu oluştururken, kulüpler de gelirlerini çeşitlendirme ve güçlendirme konusunda avantaj elde ederler.

Küreselleşme bağlamında, futbol kulüplerinin uluslararası turnuvalara ve etkinliklere katılımı, oyuncu transferleri ve medya hakları gibi faktörlerin, futbolun sadece bir yerel olgu olmaktan çıkarak küresel bir ticaret platformu haline gelmesine olanak tanıdığı belirtilir. Futbolun küresel pazarlama stratejileri konusundaki anlayışı zenginleştirir ve markaların bu dinamik pazarda başarılı olmaları için gerekli stratejileri belirlemede rehberlik eder (Saatçioğlu & Çakmak, 2019).

Oyuncu transferleri ve maaşlar, spor endüstrisinin önemli ekonomik unsurlarını oluşturur. Bu konu, kulüplerin rekabet avantajı sağlama, sportif başarı elde etme ve marka değerlerini artırma çabalarını şekillendiren karmaşık bir dinamiktir. Oyuncu transferleri, futbol kulüplerinin kadrolarını güçlendirme, taktiksel stratejilerini şekillendirme ve rekabet avantajı elde etme amacını taşır. Bu transferler sadece sportif başarıyı etkilemekle kalmaz, aynı zamanda kulüp kimliği ve taraftar bağlılığı gibi faktörlere de önemli bir etki yapar (Franceschi vd., 2023).

Stadyum yatırımlarının, futbol kulüplerinin gelirlerini artırmada, taraftar deneyimini geliştirmede ve marka değerini yükseltmede önemli bir rol oynadığını gösterilmektedir (Biscaia vd., 2016).

1.2. FUTBOL KULÜPLERİNİN FİNANSAL YAPISI

Yüksek bütçeli futbol kulüpleri, endüstriyel gelişime adım atmak için profesyonelleşme ve şirketleşme yolunu benimsemektedir. Şirketleşme modelinin finansal kaynaklara daha kolay ulaşma imkânı sunması ve sermaye piyasalarına erişim sağlaması, düşük maliyetli fon oluşturarak futbol kulüplerini hızla kurumsallaştırmaktadır. Futbol kulüpleri, varlıklarını koruma ve finansal istikrarı sağlama konularında giderek daha fazla önem kazanmaktadır (Güngör & Kocamış, 2018).

Futbol kulüplerinin finansal yapılarını anlamak için öncelikle gelir kaynaklarına odaklanmak gerekmektedir. Gelir kaynakları genellikle maç günü gelirleri, televizyon yayın hakları, sponsorluk anlaşmaları, oyuncu transferleri ve ticari faaliyetlerden oluşur. Bu kaynakların yönetimi ve çeşitlendirilmesi, kulüplerin mali sürdürülebilirliklerini etkileyen kritik faktörler arasında yer almaktadır. Öte yandan, harcamalar da finansal yapıyı etkileyen bir diğer önemli unsurdur. Oyuncu maaşları, transfer ücretleri, teknik ekip maliyetleri, stadyum bakımı gibi harcamalar, kulüplerin bütçe yönetimini belirleyen ana etkenlerdir. Bu harcamaların kontrol altında tutulması, kulüplerin finansal dengeyi sağlamalarına yardımcı olabilir. Borçlanma durumu da futbol kulüplerinin finansal yapısının önemli bir parçasını oluşturur. Kulüpler, genellikle büyük transferler veya altyapı projeleri için borçlanma yoluna gidebilirler. Ancak, bu borçların yönetimi ve geri ödenmesi, kulüp finansmanının sürdürülebilirliği açısından kritiktir. Futbol kulüplerinin finansal yapısı, gelir ve harcamaların dengesi, borçlanma durumu ve yönetim stratejileri gibi faktörleri içerir. Kulüplerin bu finansal unsurları başarılı bir şekilde yönetmeleri, uzun vadeli başarı ve sürdürülebilirlik açısından önemlidir.

Futbol, sadece saha içinde değil, aynı zamanda finansal yönetim açısından da büyük bir rekabetin ortamına sahiptir. FFP kuralları, futbol kulüplerinin ekonomik sürdürülebilirliklerini sağlamak ve finansal istikrarlarını korumak amacıyla UEFA tarafından 2010 yılında başlatılan bir inisiyatiftir. FFP kuralları, futbol kulüplerinin harcamalarını gelirleriyle dengelemelerini ve uzun vadeli mali sağlamlıklarını sürdürmelerini hedefler. Bu kurallar, kulüplerin sadece kendi finansal kaynaklarına dayanmalarını ve mali sorumluluklarını yerine getirmelerini amaçlar. Bu

doğrultuda, kulüplerin sürdürülebilir bir finansal modelle hareket etmeleri beklenir (Türkiye Futbol Federasyonu, 2016).

1.2.1. Futbol Kulüplerinin Mali Gelir Kaynakları

UEFA'nın "Kulüp Lisanslama ve FFP" dokümanının 58. maddesi ile Türkiye Futbol Federasyonu Kulüp Lisans ve FFP Talimatına göre, başabaş kuralının değerlendirilmesinde dikkate alınacak gelir unsurları aşağıdaki gibidir (Er, 2016):

- Bilet satışlarından elde edilen para (hem kombine hem de bireysel maçlar için).
- Reklamlardan ve sponsorluklardan kazanılan para.
- Yayın süresinin satışından elde edilen kazançlar.
- İşletmeler tarafından sağlanan hizmetlerin (kafeler ve restoranlar gibi) yanı sıra konferanslar ve diğer sınıflandırılmamış iş türlerinden elde edilen karlar.
- UEFA katkılarından ve ödülllerinden elde edilen kazançlar.
- Çeşitli operasyonel gelirler: sübvansiyonlar ve ödenekler gibi kamu kuruluşlarından elde edilen nakit; kira geliri, ödenen temettüler ve yukarıda gösterilen gelir bileşenlerine dahil olmayan kulüple ilgili futbol dışı faaliyetlerden kazanılan nakitler.
- Katılımcıların satışından elde edilen kazanç veya gelirler.
- Fiziksel varlıklar satıldığında kazanılan gelirler.
- Yurt dışında yapılan yatırımlardan ve paralardan elde edilen kazançlar.
- Para içermeyen krediler ve gelirler.
- Gerçeğe uygun değer olarak kabul edilen tutarın üzerinde gelir elde eden bağlantılı tarafları içeren işlemler.
- Bağlantısız üçüncü taraflarla gerçekleştirilen futbol dışı çalışmalardan elde edilen karlar.
- Lisansta adı geçen futbolcuya ilişkin kazançlar.

Başabaş noktası, alacaklı riskten korunma stratejilerinden kaynaklanan yükümlülükleri hafifletmek için verilen herhangi bir krediyi içermemelidir (Şaban & Demirci, 2016).

1.2.2. Futbol Kulüplerinin Mali Gider Kaynakları

UEFA'nın "Kulüp Lisanslama ve FFP" dokümanının 58. maddesi ile Türkiye Futbol Federasyonu Kulüp Lisans ve FFP Talimatına göre, başabaş kuralının değerlendirilmesinde dikkate alınacak gelir unsurları aşağıdaki gibidir (Er, 2016):

- Mal ve hizmetlerin fiyatları.

- Çalışanların maaşları.
- Ek işletme maliyetleri.
- Bir futbolcunun tescilinin alınması, satılması veya başka bir şekilde elden çıkarılmasıyla ilgili maliyetler ile amortisman veya değer düşüklüğüyle ilgili maliyetler.
- Ödenen temettüleri ve talep edilen mali harcamalar.

1.2.3. Başabaş Kuralının Hesaplamasının Sonuçları

Başabaş fazlası veya açığını bulmak, UEFA'nın başabaş hesaplamasında gelirlerden giderlerin çıkarılmasıyla yapılır. Gelir kalemleri giderleri aştığında başabaş fazlası, giderler gelir kalemlerinden az olduğunda ise başabaş açığı gösterilir. İzleme dönemini oluşturan tüm raporlama dönemleri için başabaş sonuçlarını topladığınızda, kümülatif başa baş sonucunu elde edersiniz. Toplam başabaş sonucunun olumlu olması durumunda, kulüp, izleme döneminde başabaş fazlası elde etmiştir. İncelenen dönem içerisinde, genel başabaş sonucunun olumsuz olması durumunda kulüp zarara uğrayacaktır (Şaban & Demirci, 2016).

1.3. LİTERATÜR ARAŞTIRMASI

Geleneksel finansal analiz yöntemlerinin sınırlılıkları, futbol kulüplerinin karmaşık finansal yapılarını değerlendirme ve tahmin etme süreçlerinde yeni yaklaşımlara duyulan ihtiyacı artırmıştır. Bu çerçevede, literatürde yapay sinir ağları gibi modern tekniklerin etkin bir araç olarak kullanıldığına dair pek çok çalışma yer almaktadır.

Gerrard (2005) tarafından 90'lı yılların sonuna doğru FA Premier League'de yarışan İngiliz profesyonel futbol kulüplerinin hem sportif hem de mali performanslarını inceleyen bir analiz yapılmıştır. Performans oranı analizi ve regresyon analizi tekniklerini kullanarak, profesyonel spor takımlarındaki oyunculuk yeteneği havuzunu etkileyen unsurları incelenmiştir. Profesyonel spor takımlarının performansta bir değişim yaşayıp yaşamadığını tespit etmeyi ve kulübün özel şahıslara sahiplik durumunun organizasyonel verimliliği etkileyip etkilemediğini tespit etmeyi amaçlanmıştır. Bir takımın sahiplik pozisyonunun takımın mali performansı üzerinde önemli bir etkiye sahip olduğunu öne sürmüştür.

Karadeniz ve İskenderoğlu (2022) çalışmasında, Avrupa futbol kulüplerinin mali tablolarını karşılaştırmayı ve COVID-19 salgınının bu sonuçları nasıl etkilediğini açıklamaktadır. Oran analizini kullanarak, UEFA'ya bağlı beş ulusal ligdeki 23 futbol kulübünün 2017-2021 dönemleri arasında mali tablolarını incelemiştir. Futbol kulüpleri son derece yüksek düzeyde likidite riski ve finansal riske sahip olduğunu, ancak son derece düşük

düzeyde kârlılık, faaliyet ve nakit akışı performansı var olduğunu saptamıştır. Futbol takımlarının faaliyetleri, karlılıkları ve nakit akışı sonuçları, COVID-19 salgınının ardından büyük bir düşüş yaşadığını ve buna karşılık olarak, yalnızca salgın öncesi ve sonrası dönemler aktivite oranları ve varlık karlılık oranları açısından karşılaştırıldığında istatistiksel olarak anlamlı bir fark bulunduğunu tespit etmiştir.

Sakınç vd. (2017), Avrupa borsalarında işlem gören 22 futbol kulübünün finansal performansını incelemek için futbol kulüplerini on finansal orana göre sıralamak için TOPSIS yaklaşımını kullanmıştır. Ardından sonuçları UEFA kulüp sıralamalarıyla ilişkisi için Spearman'ın sıra korelasyonunu kullanmıştır. Test sonucu finansal başarı ile sportif başarı arasında istatistiksel olarak anlamlı bir ilişkiye işaret etmediğini aktarmıştır.

Ecer ve Büyükaslan (2014), 2008'den 2012'ye kadar olan finansal oranları kullanarak dört Türk futbol kulübünün finansal performansını ölçmek için Kârlılık, yükümlülük ve likidite göstergelerine dayalı olarak on bir finansal parametre seçilip Gri İlişkisel Analiz yaklaşımını kullanmışlardır. Finansal başarıyı değerlendirmek için kullanılan finansal göstergelerin göreceli ağırlığını da bulmak istemişlerdir. Sonuçların, en iyi finansal performansın Fenerbahçe'de olduğunu ve en önemli finansal ölçütün borç göstergeleri olduğunu göstermiştir.

Güngör (2014) tarafından, 2007-2012 döneminde önemli bütçelere sahip dört Türk futbol kulübünün sportif başarıları ile finansal performansları arasındaki ilişkiyi incelemek amacıyla bir çalışma yürütülmüştür. Bu bağlantıyı değerlendirmek üzere oran analizi ve yapay sinir ağları modeli kullanılmıştır. Çalışma kapsamında, spor performansı ile profesyonel futbol faaliyetlerinden elde edilen toplam gelir arasında anlamlı bir ilişki olduğu belirlenmiştir. Reklam ve sponsorluk gelirleri arasında orta düzeyde bir ilişki saptanmıştır. Telif hakları, lisans gelirleri ve ticari gelir arasında ise yüksek düzeyde bir korelasyon tespit edilmiştir. Buna karşılık, maddi duran varlıklarla bir yıldan kısa vadeli tahvillere yapılan yatırımlar arasında zayıf bir korelasyon olduğu görülmüştür. Banka kredileri gibi kısa vadeli finansal yükümlülüklerle finansal kiralama borçları arasında güçlü bir ilişki olduğu ortaya konulmuştur.

Beyazgül ve Karadeniz (2019) tarafından gerçekleştirilen çalışmada, Türkiye ve Avrupa'daki futbol kulüplerinin 2013-2018 yılları arasındaki nakit akış profilleri karşılaştırmalı olarak analiz edilmiştir. Çalışmada, 29 futbol kulübüne ait nakit akış tabloları, "Nakit Akışlarının Sağlandığı Faaliyetler Yöntemi" temel alınarak değerlendirilmiştir. Bu yöntemde, işletmelerin nakit akışları; işletme, yatırım ve finansman faaliyetleri şeklinde sınıflandırılmış ve sekiz farklı nakit akış modeli tanımlanmıştır. Elde edilen bulgulara göre, Avrupa'daki futbol kulüplerinin büyük bir kısmının "başarılı işletme" (Model 2) profiline, Türkiye'deki kulüplerin

ise genellikle "hızlı büyüyen işletme" (Model 6) profiline sahip olduğu belirlenmiştir. Avrupa kulüplerinde işletme faaliyetlerinden pozitif nakit akışı sağlarken, Türkiye'deki kulüplerde nakit akışlarının ağırlıklı olarak finansman faaliyetlerinden kaynaklandığı tespit edilmiştir. Bu temel üzerinde, Avrupa kulüplerinin mali disiplin açısından daha sürdürülebilir bir yapıya sahip oldukları, Türkiye'deki kulüplerin ise borçlanma yoluyla büyümeye çalıştıkları sonucuna ulaşılmıştır. Çalışmada, futbol kulüplerinin finansal yönetim stratejilerinin ve nakit akış profillerinin, uzun vadeli başarı üzerinde belirleyici olduğu vurgulanmıştır.

Ika vd. (2021), İngiltere'nin saygın futbol kulüpleri Arsenal ve Manchester City'nin mali tabloları incelenmiştir. Eş zamanlı olarak bu futbol takımlarının mali performansları, 2015-2017 mali hesapları kullanılarak Zskoru Altman, Springate ve Zmijewski'nin iflas tahmin modelleriyle karşılaştırılarak değerlendirilmiştir. Üç finansal performans modeli, mali duruma ilişkin tutarlı bir şekilde doğru değerlendirmeler ortaya koymuştur; Arsenal ve Manchester City, iş raporlarında güçlü bir performans sergilediğini ayrıca Arsenal ve Manchester City'nin finansal performans analizi, Manchester City'nin Arsenal'e kıyasla daha üstün bir finansal konuma sahip olduğunu ortaya koymuştur.

Galariotis vd. (2018), başlangıçta kulüplerin mali ve ticari performans özelliklerine göre önceliklendirilmesi amacıyla Çok Kriterli Analiz Yöntemi uygulanmıştır. İkinci aşama, sıralamaları, lig durumunu ve spor performansını birleştirerek çeşitli parametreler arasındaki ilişkileri değerlendirmek için Kısmi En Küçük Kareler Yapısal Eşitlik Modelleme Yaklaşımını kullanmaktadır. İş performansı ile spor performansı arasında doğrudan ilişkili olduğunu gösteren pozitif bir korelasyon keşfedilmiştir. Ayrıca finansal performansın spor performansı üzerinde tek yönlü olumsuz etkisi olduğu da ortaya konmuştur. Artan gelir, sportif başarılar üzerinde olumlu bir etki yaratmıştır ve karşılıklı yarar sağlayan bir döngüde gelir üzerinde olumlu bir etkiye yol açtığı sonucuna ulaşmıştır.

Alaminos vd. (2020), 2016'dan 2018'e kadar ulusal liglerinin en üst iki seviyesinde yer alan 234 Avrupa profesyonel futbol kulübünden oluşan bir örnek kullanılmıştır. Bağımlı değişkenler, özsermaye hissedarlarının veya kurumsal özsermaye hissedarlarının parasının erişebildiği karı temsil eden RONW'dan ve faiz ve vergi öncesi kârın (EBIT) kullanılan sermayeye bölünmesini temsil eden ROCE'den oluşmaktadır. Kulüplerin mali performansı, çok katmanlı sinir ağları ile kuantum sinir ağları karşılaştırılarak analiz edilmiştir. Çok katmanlı sinir ağları tahminde iyi bir doğruluk sergilerken, kuantum sinir ağları en iyi doğruluk oranıyla daha iyi performans sergilemiştir. Analiz, finansal performansı açıklayan temel belirleyicilerin likidite, kaldıraç ve spor performansı ile ilgili özellikler olduğu belirlenmiştir.

Kayakuş vd. (2023), Polonya, Rusya ve Türkiye'den toplam 13 demir-çelik şirketini kapsayan çalışmada ROA ve ROE'yi yapay sinir ağları (YSA), destek vektör regresyonu (DVR) ve çoklu doğrusal regresyon (ÇDR) ile tahmin etmiştir. YSA yöntemi hem ROA ($R^2 = 0.864$) hem de ROE ($R^2 = 0.858$) tahmininde en başarılı model olarak öne çıkmıştır. SVR, orta düzeyde başarılı sonuçlar üretirken MLR en düşük doğruluk oranlarına sahip olmuştur. Bu sonuçlar, doğrusal olmayan ilişkilerin modellenmesinde makine öğrenimi tekniklerinin üstünlüğünü göstermektedir. Araştırmada kullanılan 2020–2022 dönemine ait çeyreklik finansal veriler, sürdürülebilir kârlılığın öngörülmesinde bu tür modellerin etkinliğini ortaya koymuştur.

Tutcu vd. (2024), Borsa İstanbul'da faaliyet gösteren 13 teknoloji firmasının 2000–2023 dönemindeki verilerini analiz etmiş, ROE ve ROA tahmininde yapay sinir ağları, karar ağaçları ve MLR modellerini karşılaştırmıştır. Yapay sinir ağları özellikle ROA'da daha yüksek tahmin başarısı gösterirken, ROE tahmininde ise karar ağaçları ve YSA benzer başarı seviyelerine ulaşmıştır. Bu bulgular, finansal performans göstergelerinin sektör özelinde farklı modellerle daha iyi tahmin edilebileceğini göstermektedir.

Yang ve Wang (2023), Çinli enerji şirketlerinin ROE ve ROA tahmininde, haber metinlerinden türetilen pazar duyarlılığı verilerini kullanarak BERT+PCA+ARIMA+XGBoost birleşik bir model önermiştir. Bu model, özellikle ROE tahmininde %95'e yakın R^2 değeri ile yüksek bir açıklayıcılık düzeyi sağlamıştır. ROA tahmininde ise R^2 değeri %91 civarındadır. Bu çalışma, geleneksel finansal oranların yanı sıra alternatif veri kaynaklarının ve metin madenciliği tekniklerinin tahmin performansına olan katkısını açıklamıştır.

Literatürdeki çalışmalar, profesyonel futbol kulüplerinin finansal performanslarının çeşitli yöntemlerle analiz edildiğini ve farklı modellerin etkinliklerinin araştırıldığını göstermektedir. Ayrıca finansal performans analizlerinde, geleneksel finansal oran analizlerinin yanı sıra yapay sinir ağları, destek vektör makineleri ve karmaşık çok katmanlı modeller gibi makine öğrenmesi tekniklerinin kullanımının giderek arttığı ve özellikle doğrusal olmayan ilişkilerin modellenmesinde bu yöntemlerin üstün performans gösterdiği literatürde vurgulanmaktadır.

Bu çalışma, mevcut literatüre paralel olarak, Avrupa'nın üst düzey futbol kulüplerine ait finansal verileri makine öğrenmesi ve istatistiksel regresyon algoritmalarıyla analiz ederek, finansal performans göstergelerinin tahmininde doğrusal olmayan modellerin etkinliğini ortaya koymayı hedeflemektedir. Özellikle, özkaynak kârlılığı (ROE) ve varlık kârlılığı (ROA) gibi temel finansal performans göstergelerinin tahmininde kullanılan modellerin karşılaştırmalı

performans analizleri, futbol kulüplerinin finansal sürdürülebilirliklerini anlamada ve yönetim stratejilerinin geliştirilmesinde önemli katkılar sunacaktır. Literatürde belirtilen bulgular ışığında, çalışmanın sonuçlarının hem akademik literatüre hem de futbol kulüplerinin finansal yönetim uygulamalarına değerli içgörüler kazandırması hedeflenmektedir.

1.4. MAKİNE ÖĞRENME

Makine öğrenimi, hesaplamalı sistemlerin açıkça programlanmış talimatlara ihtiyaç duymadan desenleri tanımlamasını ve verilere dayalı kararlar almasını sağlayan bir yapay zekâ dalıdır. 1950'lerde ortaya çıkan bu kavram, istatistik, optimizasyon ve bilişsel bilim gibi alanlardan teknikleri entegre ederek kademeli olarak gelişmiş ve veri merkezli modelleme metodolojilerini kapsayacak şekilde büyümüştür (Samuel, 1959).

Makine öğrenmesi algoritmaları, doğrudan programlama yoluyla belirli bir çıktıyı üretmek yerine, girdilerden öğrenerek belirli görevleri yerine getirebilen hesaplamalı sistemlerdir. Bu yönüyle, geleneksel “katı kodlamanın” aksine, “yumuşak kodlama” olarak tanımlanabilecek esnek yapılar geliştirirler. Bu algoritmalar, deneyim yoluyla yapılarını otomatik olarak güncelleyerek, zamanla istenen görevi daha etkin şekilde gerçekleştirebilir hâle gelirler. Bu adaptasyon süreci “öğrenme” olarak adlandırılır ve algoritma, yalnızca eğitim verileri üzerinde değil, aynı zamanda daha önce hiç karşılaşmadığı veriler üzerinde de doğru tahminler yapabilecek genel yetenekler geliştirir (Kotsiantis vd., 2007).

Makine öğrenme, istatistik, optimizasyon ve algoritma kuramlarını entegre ederek doğrusal ve doğrusal olmayan fonksiyonel ilişkiler üzerinden kestirim yapma olanağı sunar. Matematiksel olarak, bir makine öğrenme problemi genellikle bir fonksiyonun yaklaşık olarak bulunması problemi olarak ifade edilir:

$$f: X \rightarrow Y$$

Burada X giriş uzayı, Y ise çıktı uzayıdır (Mitchell, 1997).

Başarılı bir makine öğrenmesi algoritmasının iki temel avantajı bulunmaktadır. İlk olarak, tekrarlayan ve zaman alıcı insan faaliyetlerinin otomasyonu mümkün hale gelir. İkinci olarak, insan gözlemcilerinin çoğu zaman fark edemeyeceği karmaşık örüntüleri ve istatistiksel ilişkileri ortaya çıkarabilir (El Naqa & Murphy, 2015).

Makine öğrenmesi modellerinin değerlendirilmesinde kullanılan metrikler, probleme göre değişiklik gösterir. Sınıflandırma problemlerinde doğruluk (accuracy), kesinlik (precision), geri çağırma (recall) ve F1 skoru yaygın olarak kullanılırken, regresyon

problemlerinde ortalama kare hata (MSE) ve ortalama mutlak hata (MAE) gibi metrikler tercih edilir. Ayrıca çapraz doğrulama (cross-validation) gibi teknikler, modelin aşırı öğrenme veya eksik öğrenme eğilimlerini değerlendirmede önemli bir rol oynar (Berrar, 2019).

Büyük veri çağında, makine öğrenimi kritik bir önem haline gelmiştir ve karar destek sistemlerinin ve öngörücü analizlerin geliştirilmesinde önemli bir rol oynamıştır. Veri hacmindeki üstel büyüme, hesaplama gücündeki ilerlemelerle, makine öğrenimi tekniklerinin verimliliğini ve erişilebilirliğini artırmıştır. Çeşitli sektörlerde daha hızlı ve daha kesin karar almayı kolaylaştıran sistemler giderek daha fazla uygulanmaya başlanmıştır (Jordan ve Mitchell, 2015).

Makine öğrenmesinde en sık kullanılan algoritmalar arasında doğrusal regresyon, lojistik regresyon, destek vektör makineleri, karar ağaçları, rastgele ormanlar ve yapay sinir ağları yer almaktadır. Bu algoritmalar, doğruluk, hesaplama karmaşıklığı ve yorumlana bilirlik gibi çeşitli kriterlere göre değerlendirilir (Hastie vd., 2009).

1.4.1. Denetimli Öğrenme

Denetimli öğrenme, etiketli veri kümelerinin kullanılması yoluyla algoritmaların eğitilmesine olanak tanıyan bir makine öğrenmesi yaklaşımıdır. Bu yöntemde her giriş örneği, ona karşılık gelen doğru bir çıktı (etiket) ile sunulur. Model, bu veri çiftlerini kullanarak, verilen bir girdiye karşılık doğru çıktıyı tahmin etmeyi öğrenir.

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \subset X \times Y \quad (1)$$

Denklemden $x_i \in \mathbb{R}^d$ giriş vektörüdür, y_i ise sınıflandırma problemlerinde kategorik, regresyon problemlerinde sürekli hedef değeridir. Amaç, bu gözlemlerden öğrenerek $f: X \rightarrow Y$ fonksiyonunu tahmin etmektir. Bu fonksiyon genellikle bir model ailesi içinden seçilir ve öğrenme süreci sırasında parametreler güncellenerek tahmin doğruluğu artırılır.

Eğitim süreci boyunca model, tahmin hatasını minimize etmek amacıyla parametrelerini günceller. Bu ayarlama işlemi, genellikle gradyan inişi gibi optimizasyon algoritmaları aracılığıyla gerçekleştirilir. Modelin performansının sadece eğitim verisinde değil, aynı zamanda yeni ve görülmemiş veri üzerinde de iyi olmasını sağlamak için çapraz doğrulama gibi değerlendirme yöntemleri kullanılır. Bu sayede aşırı öğrenme ya da yetersiz öğrenme gibi yaygın sorunların önüne geçilir (Murphy, 2012, s. 2).

Bu öğrenme türünde yaygın olarak kullanılan bazı algoritmalar şunlardır:

- Yapay Sinir Ağları (Artificial Neural Networks): Karmaşık, doğrusal olmayan ilişkileri öğrenebilir.
- Naive Bayes: Olasılık temelli, basit ama etkili bir sınıflandırma algoritmasıdır.
- Doğrusal Regresyon (Linear Regression): Sürekli değişkenlerin tahmini için kullanılır.
- Lojistik Regresyon (Logistic Regression): İkili sınıflandırma problemleri için uygundur.
- Rastgele Orman (Random Forest): Karar ağaçlarının bir araya getirilerek oluşturulduğu toplu öğrenme yöntemidir.
- Destek Vektör Makineleri (DVM): Marjın maksimizasyonu esasına dayanarak veri ayırımını sağlar (Cortes & Vapnik, 1995).

1.4.2. Denetimsiz Öğrenme

Denetimsiz öğrenme, etiketlenmemiş veri kümelerini analiz etmek ve bu verilerdeki gizli yapıları keşfetmek amacıyla kullanılan bir öğrenme yaklaşımıdır. Bu yöntemde, veriler herhangi bir sınıf ya da kategori bilgisini içermediğinden, algoritmalar benzerlik veya yapısal özellikler temelinde otomatik gruplamalar (kümeleme) oluşturur. Bu gruplamalar genellikle “küme (cluster)” olarak adlandırılır. Veri içerisindeki örüntüleri, benzerlikleri ve farklılıkları keşfetme yeteneği sayesinde özellikle keşifsel veri analizi (exploratory data analysis), müşteri segmentasyonu, çapraz satış stratejilerinin geliştirilmesi, görüntü tanıma ve örüntü tanıma gibi pek çok alanda yaygın olarak kullanılmaktadır (Hastie vd., 2009).

Ayrıca, denetimsiz öğrenme algoritmaları boyut indirgeme amacıyla da önemli rol oynar. Bu sayede çok boyutlu veri setlerinde yer alan bilgilerin daha sade ve anlamlı alt temsilleri elde edilebilir. Yaygın olarak kullanılan bazı denetimsiz öğrenme algoritmaları şunlardır:

Kümeleme Algoritmaları: Belirli sayıda küme merkezi belirleyerek veriyi bu merkezlere en yakın olacak şekilde gruplar. K-Means Kümeleme;

$$\min_{C, \mu} \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (2)$$

Amaç; k adet küme merkezi $(\mu_1, \mu_2, \dots, \mu_k)$ belirleyerek her gözlemi en yakın merkezle eşleştirmek (MacQueen, 1967). Bu denklemde:

C_i : i . kümedeki gözlem setini,

μ_i : i . kümenin merkezini temsil eder.

Boyut Azaltma Teknikleri: Temel Bileşenler Analizi (Principal Component Analysis-PCA); yüksek boyutlu veri setlerinde yer alan varyansı en iyi şekilde açıklayan yeni bir eksen sistemi oluşturarak verinin boyutunu azaltan bir doğrusal dönüşüm yöntemidir. Bilgi kaybını en aza indirerek daha az boyutta veri temsili sağlamaktır.

$$\max_{\|w\|=1} Var(w^T X) \quad (3)$$

Amaç: Varyansı en yüksek doğrultular boyunca veriyi projekte ederek boyut sayısını azaltmak. Burada w veri kümesinin en yüksek varyans yönünü temsil eden özvektördür (Jolliffe & Cadima, 2016).

Olasılıksal Kümeleme Yöntemleri: Her bir gözlemin belirli bir kümeye ait olma olasılığı hesaplanır; örneğin Gaussian Mixture Model (GMM):

$$p(x) = \sum_{i=1}^k \pi_i \mathcal{N}(x|\mu_i, \Sigma_i) \quad (4)$$

Bu denklemde:

π_i : Küme i 'ye ait olma olasılığını

μ_i, Σ_i : Küme merkezleri ve kovaryanslarını temsil eder.

Gözlemler farklı normal dağılımların birleşiminden gelir varsayımı yapılır (Murphy, 2012, s. 341).

1.4.3. Yarı-Denetimli Öğrenme

Yarı-denetimli öğrenme, eğitim sürecinde hem etiketli hem de etiketsiz verileri birleştiren bir makine öğrenme yaklaşımıdır. Bu teknik, etiketli veri kıtlığı ancak etiketsiz veri bolluğu olduğunda özellikle yararlıdır. Yarı-denetimli öğrenmenin birincil amacı, modelin performansını, ilk öğrenme için etiketli verileri kullanarak geliştirmek ve aynı zamanda modelin doğruluğunu ve genelleme yeteneklerini geliştirmek için etiketsiz verileri dahil etmektir.

Makine öğrenimi, büyük veri bağlamında karar alma süreçlerini geliştiren ve otomatikleştiren dönüştürücü bir teknolojidir. Denetimli, denetimsiz ve yarı-denetimli öğrenme dahil olmak üzere çeşitli yaklaşımlar, çok çeşitli kullanım durumlarına uygulanabilir (LeCun vd., 2015).

1.4.4. Makine Öğrenme Sürecindeki Adımlar

Veri toplama ve hazırlama: Çalışmanın ilk aşamasını, araştırma problemine yönelik gerekli verilerin toplanması süreci oluşturur. Bu veriler genellikle farklı kaynaklardan elde edilir ve ham (işlenmemiş) biçimdedir. Bu nedenle, veri setinin analiz için uygun hale getirilmesi amacıyla kapsamlı bir veri ön işleme süreci gereklidir. Bu süreç; eksik verilerin giderilmesi, aykırı değerlerin tespiti ve veri türlerinin uygun biçimlere dönüştürülmesi gibi adımları içerir. Böylece, analizde kullanılacak veri kümesinin doğruluğu ve güvenilirliği sağlanmış olur (Çetin & Yıldız, 2022).

Keşifsel veri analizi: Bu aşamada, verinin temel yapısal özelliklerini ve örüntülerini ortaya koymak amacıyla kapsamlı ve sistematik bir keşifsel veri analizi gerçekleştirilir. Veri kümesindeki dağılımları, ilişkileri ve kalıpları belirlemek için istatistiksel yöntemler ve görsel teknikler uygulanır. Bu adım, uygun modelin tasarımına ve seçimine rehberlik eden değerli içgörüler sunar (Ghosh vd., 2018).

Model seçimi ve eğitimi: Bu aşamada, verilen sorun için en uygun makine öğrenimi algoritması seçilir. Bu, görevin niteliğine bağlı olarak denetlenen, denetlenmeyen veya yarı-denetimli öğrenme tekniklerini içerebilir. Seçilen algoritma daha sonra, tahmine dayalı bir model geliştirmek için hazırlanan eğitim verileri kullanılarak eğitilir (Chen vd., 2008).

Model değerlendirme ve iyileştirme: Eğitim tamamlandıktan sonra, modelin etkinliği ayrı bir veri kümesinde test edilerek değerlendirilir. Performansı doğruluk, kesinlik gibi ölçütler kullanılarak ölçülür. Bu değerlendirmelere dayanarak, model parametreleri performansını daha da artırmak için ince ayarlanabilir (Chicco & Jurman, 2020).

Model dağıtımı ve izlemesi: Son aşamada, eğitilen model gerçek dünya uygulamaları için dağıtılır. Performansı sürekli olarak izlenir ve gerektiğinde güncellenebilir veya yeniden eğitilebilir. Ek olarak, modelin tahminleri karar vermeyi desteklemek için iş süreçlerine veya otomatik sistemlere entegre edilir (Jordan & Mitchell, 2015).

1.5. MAKİNE ÖĞRENMESİ ALGORİTMALARI

1.5.1. YSA-Yapay Sinir Ağları

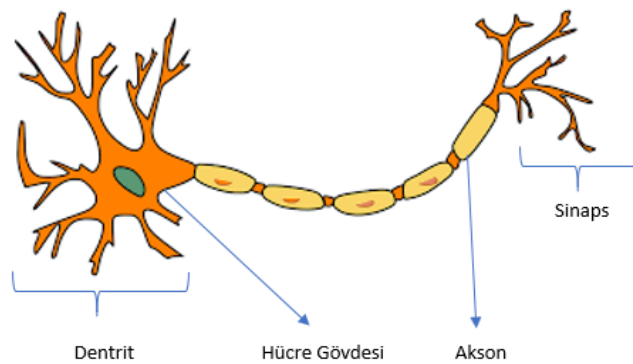
İnsan beyninin bilişsel süreçlerini taklit etmeyi amaçlayan Yapay Sinir Ağları, dışsal bir müdahaleye gerek duymaksızın öğrenme gerçekleştirebilir, veri yapılarındaki örüntüleri inşa edebilir ve yeni bilgiler türetebilir (Öztemel, 2012, s. 29).

Beynin öğrenme, ilişkilendirme, sınıflandırma ve genelleme gibi özelliklerinden ilham alan ilkeleri birleştiren Yapay Sinir Ağları, beyin sinir sisteminin temel bir modelini hesaplamalı platformlarda kopyalamaya çalışır. Beyindeki temel birim olarak hizmet veren nöron, bilgisayarlardaki işlevselliğinin basitleştirilmiş bir simülasyonunu oluşturmak için soyutlanmıştır (Anderson & McNeill, 1992).

YSA' lar ilgilenilen değişkenleri (bağımlı değişkenler, model çıktıları) bir veya daha fazla model girdisinin (bağımsız değişkenler, öngörücüler) bir fonksiyonu olarak tahmin eder. Bu, genellikle bir dizi bilinmeyen parametreye bağlı olan matematiksel bir ilişki (modelin işlevsel formu) aracılığıyla model girdilerini model çıktılarına dönüştürerek yapılır (Maier vd., 2023).

1.5.1.1. Biyolojik sinir ağları

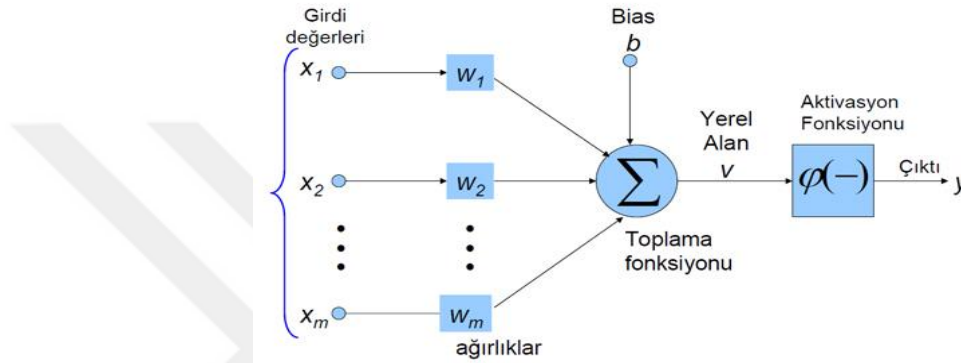
Bilim insanları, insan beyinde yaklaşık 10 milyar sinir hücresi bulunduğunu ve bu hücrelerin, aralarındaki iletişimi sağlayan yaklaşık 60 trilyon bağlantı ile birbirine bağlandığını ifade etmektedir. Nöronlar duyu organları aracılığıyla bilgi alırlar. Alınan bilgi daha sonra alıcı nöronlar tarafından işlenir ve bir sonraki hücreye iletilir. Bu yöntemin kullanılmasıyla bilgilerin merkezi sinir sistemine iletilmesi garanti altına alınır. Merkezi sinir sistemi gelen bilgiyi sinyal olarak yorumladıktan sonra bu sinyallere tepki olarak bir yanıt sinyali üretir. Dendritler, duyu organlarından veya birbirine bağlı diğer nöronlardan gelen uyarıların çekirdeğe iletilmesini kolaylaştırır. Dendrit, sinyalleri çekirdeğe iletir, daha sonra bunları toplayıp aksona iletir. Toplanan ve aksona aktarılan sinyaller, akson tarafından işlenir ve daha sonra dendritlere iletilir ve dendritler de bu sinyalleri sinapslara taşır (Yılmaz, 2020, s. 62).



Şekil 1: Biyolojik Sinir Hücresi

1.5.1.2. Yapay sinir hücresi

Yapay Sinir Hücresi, YSA'lar içindeki bilgi işlemede önemli bir bileşen olarak hizmet eder. Giriş sinyallerini alır, tanımlanmış matematiksel işlemleri kullanarak bunları işler ve bir çıkış sinyali üretir. Bu, biyolojik nöronların sinapslar aracılığıyla sinyalleri alma, işleme ve iletme şeklini taklit eder. Yapay nöronlar beş temel unsurdan oluşur: giriş düğümleri, sinaptik ağırlıklar, toplama süreçleri, aktivasyon fonksiyonları ve çıktı düğümleri; bu bileşenler, sinir ağı işlevselliği için gerekli olan çok yönlü hesaplamaları yürütmek üzere uyum içinde çalışır (Öztemel, 2012, s. 48).



Şekil 2: Yapay Sinir Hücresinin Yapısı

Girdiler

Yapay nöronlardaki girdiler, dışarıdan bilginin alındığı kanallar görevi görür. Sinir ağıları bağlamında bu girdiler, işlenmek üzere yapay nöronlara beslenen özellikleri veya sinyalleri temsil eder. Girdilerin doğasını ve önemini anlamak, sinir ağlarının dış verileri nasıl özümlediğini çözmede çok önemlidir (Öztemel, 2012, s. 52).

Ağırlıklar

Sinaptik ağırlıklar, bireysel girdilerin etkisinin modüle edilmesinde kritik bir rol oynar. Bu ağırlıklar, giriş düğümleri ile yapay nöron arasındaki bağlantının gücünü belirtir ve her bir girişin, nöronun genel hesaplaması üzerindeki etkisini belirler. Ağırlıkların atanması ve ayarlanması, sinir ağlarının belirli görevler için eğitiminin temel unsurlarıdır (Sağıroğlu vd., 2003).

Toplama fonksiyonu

Yapay nöronlar bağlamında toplama işlevi, giriş sinyallerinin ağırlıklı toplamını alır. Her giriş, giriş düğümü ile nöron arasındaki bağlantının gücünü yansıtan, karşılık gelen sinaptik

ağırlığıyla çarpılır. Bu ağırlıklı girişler daha sonra giriş sinyallerinin hem büyüklükleri hem de yönsel etkileri dahil edilerek toplanır (Abraham, 2005).

Aktivasyon fonksiyonu

Aktivasyon fonksiyonları, yapay nörona doğrusal olmayan bir özellik kazandırarak ağırlıklı verilerdeki karmaşık kalıpları ve ilişkileri öğrenmesini sağlar. Bu öge, birleştirilmiş girdilere, ağırlıklara ve nöronun iç durumuna göre nöronun çıktısını tanımlar. Yaygın aktivasyon fonksiyonları arasında sigmoid, hiperbolik tanjant ve düzeltilmiş doğrusal birim bulunur (Öztemel, 2012, s. 51).

Bazı aktivasyon fonksiyonları şöyledir:

- Lineer Fonksiyon:

Hücrenin çıkışları, gelen girişlerden gelen sinyaller olarak kabul edilir.

$$f(net) = net \quad (5)$$

- Step fonksiyonu:

Hücrenin çıkışı, gelen net giriş değerinin önceden tanımlanmış bir eşik altına düşüp düşmediğine veya onu aşmasına bağlı olarak 1 veya 0 şeklinde bir ikili durum varsayar.

$$f(net) = \begin{cases} 0 & \text{eğer } net \leq 0 \\ 1 & \text{eğer } net > 0 \end{cases} \quad (6)$$

- Sigmoid Fonksiyonu:

Sürekliliği, doğrusal olmaması ve türevlenebilirliği nedeniyle sigmoid aktivasyon fonksiyonu birçok uygulamada yaygın olarak kullanılmaktadır. Bu fonksiyon, her giriş için 0 ile 1 arasında sayısal bir değer hesapladığından yaygın olarak kullanılır.

$$f(net) = \frac{1}{1 + e^{-net}} \quad (7)$$

- Eşik Değer Fonksiyonu:

Çıkışı, gelen bilginin bir eşik aşım aşmadığına veya altına düşüp düşmediğine göre belirlenir ve 0 ile 1 arasında değerler elde edilir.

$$f(net) = \begin{cases} 0 & \text{eğer } net \leq 0 \\ net & \text{eğer } 0 < net < 1 \\ 1 & \text{eğer } net \geq 1 \end{cases} \quad (8)$$

- Hiperbolik Tanjant Fonksiyonu:

Hesaplama, tanjant fonksiyonunun gelen net giriş değerine uygulanmasını içerir.

$$f(net) = \frac{e^{net} - e^{-net}}{e^{net} + e^{-net}} \quad (9)$$

Hücrenin çıktısı

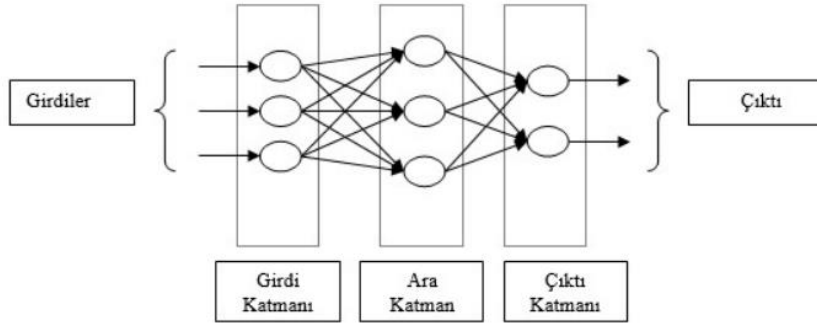
Çıkışlar, ağırlıklı toplama ve aktivasyon fonksiyonları yoluyla girdileri işledikten sonra yapay nöron tarafından üretilen nihai sonucu veya sinyali temsil eder. Bu elde edilen çıktı, genellikle dış dünyaya veya başka bir sinir hücresine iletilir. Aynı zamanda, hücre kendi ürettiği çıktıyı, bir sonraki adımda kendi girdisi olarak kullanabilir. Bu, sinir ağlarının karmaşıklığını artırarak bilgi işleme yeteneklerini geliştirir. Söz konusu yapı içerisinde, sinir hücresi hem çevresiyle etkileşime geçer hem de kendi içsel süreçlerini yönetir (Abraham, 2005).

1.5.1.3. Yapay sinir ağlarının yapısı

YSA'lar birden fazla katmandan oluşur ve her katman belirli bir işleve hizmet eder. Giriş katmanı ham verileri alır, gizli katmanlar bu bilgiyi işleyip dönüştürür ve çıkış katmanı nihai sonucu üretir. Derin sinir ağlarının birden fazla gizli katmanı vardır ve bu da "derin öğrenme" terimine katkıda bulunur (Patterson, 1998).

Giriş Katmanı, giriş veri kümesinin farklı özelliklerini veya özelliklerini temsil eden giriş Katmanındaki nöronlarla, harici verilerle etkileşime geçmek için ilk arayüz olarak işlev görür. Giriş ve çıkış Katmanları arasında yer alan gizli katmanlar, giriş verileri üzerinde karmaşık hesaplamalar gerçekleştirir. "Gizli" terimi girdiyle ilgilidir ve bu katmanların çıktıya göre algılanamayan özelliklerini ifade eder. Çıktı katmanı, nihai ağ tahminlerinin veya sınıflandırmalarının üretilmesinden sorumlu olan nihai katmandır. Çıktı Katmanındaki nöronlar, verilen görevin kesin gereksinimlerine göre özelleştirilmiş sonuçlar üretir. İstenilen sonuç ister ikili ister çok sınıflı kategorizasyon olsun, bu katmanın kurulumuna göre belirlenir.

Sinir ağlarının gizli katmanlarındaki doğrusal olmayan ve türevlenebilir aktivasyon fonksiyonlarının seçimi, karmaşık ilişkilerin yakalanması ve geri yayılım kullanılarak başarılı eğitimin sağlanması açısından çok önemlidir. Bu, bilginin entegrasyonunun ve ayrılmasının öğrenme sürecinin temel yönleri olduğu geri bildirim ağlarında özellikle önemlidir (Kecman, 2001, s. 255).



Şekil 3: Bir Yapay Sinir Ağı Örneği

1.5.1.4. Yapay sinir ağının çalışması

YSA'nın işleyiş prensibi, ağı sunulan girdilerin, çeşitli işlemlerden geçirilerek çıktı setinin oluşturulmasına dayanmaktadır. Ağı doğru çıktıyı üretebilmesi için eğitilmesi gerekmektedir, yani ağı örnek girdiler gösterilerek istenen çıktı elde edilmeye çalışılır. Bu süreçte, girdiler genellikle vektör formunda ağı sunulur ve ağı, bu girdi vektörleri üzerinden çıktı vektörlerini oluşturur. Yapay Sinir Ağlarının girdi vektörlerini nasıl çıktı vektörlerine dönüştürdüğü tam olarak anlaşılamamıştır. Bu nedenle, YSA'nın çalışma şekli genellikle bir "kara kutu" olarak nitelendirilir. Yani, iç mekanizmaları net bir şekilde bilinmediğinden dolayı, ağı olan güveni azaltabilir. Bu durum, YSA'nın tam olarak nasıl kararlar verdiğini anlamak konusunda belirsizlik yaratabilir (Öztemel, 2012, s. 54).

1.5.1.5. Yapay sinir ağlarının eğitilmesi ve test edilmesi

Öğrenme, ilk olarak rastgele atanan ağı başlangıç ağırlık değerlerini oluşturma görevini içerir. Ağı, karşılaştığı her numune için bu ağırlık değerlerini ayarlar. Prosedür, ağı doğru ağırlık değerlerine ulaşmasıyla sona erer; bu, örneklerle gösterilen olay hakkında genelleme yapma yeteneğini gösterir. Ağırlık değerlerinin değiştirilmesi, kullanılan öğrenme yaklaşımına bağlı olarak farklılık gösteren, öğrenme kuralları olarak bilinen belirli düzenlemelere göre gerçekleştirilir (Efe vd., 2000).

Yapay Sinir Ağlarında farklı öğrenme algoritmalarının kullanılmasının amacı, hatayı en aza indirmek ve gerçek çıktıya yakın bir yaklaşım elde etmektir. Bu işlem sırasında yapay sinir ağının (YSA) ağırlıkları her yinelemede güncellenir ve verilen hedefe ulaşmak için sürekli olarak değiştirilir. Ağı başarısı sıklıkla harici bir ölçüm kullanılarak değerlendirilir. Sağlanan giriş-çıkış çiftleri kullanılarak hedefe ulaşıldığında ağırlık değerleri Yapay Sinir Ağları için belgelenir. Yapay sinir ağlarında (YSA) öğrenme, istenilen sonuca ulaşılan kadar ağırlıkların ve sürenin sürekli olarak ayarlanması sürecini ifade eder. Yapay Sinir Ağı (YSA) öğrenme sürecini tamamladıktan sonra daha önce karşılaşılmamış yeni girdiler kullanılarak test edilir.

Bu testin amacı ağ tarafından üretilen çıktı ile istenen gerçek çıktı arasındaki benzerliği değerlendirmektir. Sınır ağına sağlanan örnek sayısı optimal değeri aşarsa, ağın ezberleme eğilimi gösterdiği kabul edilir. Genellikle ağ, mevcut verilerin %80'i kullanılarak eğitilir ve daha sonra performansı, kalan %20 kullanılarak değerlendirilir. Basitçe söylemek gerekirse, ağın genel fizibilitesi ve doğruluğu bu şekilde değerlendirilir (Öztemel, 2012, s. 55).

1.5.2. DVM- Destek Vektör Makinaları

DVM'yi diğer yöntemlerden ayıran temel özelliklerinden biri, veriler arasında maksimum marjın sağlayan bir hiper düzlem bulmasıdır. Bu hiper düzlem, sınıflar arasındaki mesafeyi mümkün olduğunca artırarak sınıflandırma hatalarını azaltır. Bu yöntem, sınıflar arasında birçok hiper düzlem mümkün olsa da en iyisinin en geniş marjınlı olan olduğunu varsayar. Böylece karar sınırının küçük varyasyonlara karşı duyarlılığı azaltılır. DVM'nin matematiksel temeli, bir kuadratik optimizasyon problemi çözülerek elde edilir. Lagrange çarpınları yöntemi ile çözülen bu problemde yalnızca destek vektörlerine ait veriler sınıflandırma fonksiyonuna katkıda bulunur (Burges, 1998).

Gerçek dünyadaki verilerin çoğu tam olarak ayrılabilir değildir. Bu nedenle DVM'de "soft margin" yaklaşımı geliştirilmiştir. Bu yöntemde, karar sınırının yanlış tarafında kalan veriler belirli bir toleransla kabul edilir ve bu tolerans slack değişkenleriyle modellenir. Bu toleransın büyüklüğü cezalandırılarak modelin genelleme başarısı korunur. Özellikle gürültülü verilerde, bu yaklaşım daha esnek ve dayanıklı sınıflandırmalar sağlar (Cortes & Vapnik, 1995).

DVM'nin doğrusal olmayan problemleri çözebilmesi ise kernel trick olarak bilinen yöntemle mümkün hale gelir. Bu yaklaşımda, doğrusal olarak ayrılamayan veriler daha yüksek boyutlu bir uzaya dönüştürülür ve bu yeni uzayda doğrusal bir hiper düzlem ile ayrılabilir hale gelir. Bu dönüşüm, veriler arasındaki iç çarpımın uygun bir çekirdek (kernel) fonksiyonu ile hesaplanmasıyla yapılır. En yaygın çekirdek fonksiyonları arasında polinom, Gaussian (RBF), sigmoid ve çok katmanlı algılayıcı (MLP) temelli fonksiyonlar yer alır. Doğru kernel seçimi, modelin başarısı üzerinde doğrudan etkilidir ve genellikle problem özelinde deneme-yanılma yoluyla belirlenir (Burges, 1998).

DVM'nin başarısını etkileyen bir diğer önemli konu modelin karmaşıklığıdır. C parametresi ile modelin öğrenme kapasitesi kontrol edilebilir. Düşük C değeri daha esnek, yüksek C değeri ise daha katı bir karar sınırı üretir. Ayrıca kernel parametreleri de modelin karmaşıklığını etkiler. Bu nedenle, parametre ayarlarının çoğu durumda çapraz doğrulama yöntemiyle belirlenmesi önerilir (Hsu vd., 2003).

DVM yalnızca sınıflandırma için değil, aynı zamanda regresyon problemleri için de kullanılabilir. Bu uygulama türü, literatürde Destek Vektör Regresyon (DVR) olarak adlandırılır. Regresyon uygulamalarında hata fonksiyonu uygun şekilde değiştirilerek hedef değişkenlerin tahmini yapılabilir. DVR’de amaç, tahmin edilen değerlerin gerçek değerden belirli bir tolerans aralığında kalmasını sağlamaktır. Yine doğrusal olmayan durumlarda çekirdek fonksiyonları yardımıyla yüksek boyutlu uzaylara geçilerek doğrusal regresyon uygulanabilir. DVM'nin en önemli güçlü yönleri arasında tek bir global optimuma sahip olması, aşırı öğrenmeye karşı dayanıklılığı ve yüksek boyutlu veri ile çalışma kapasitesi sayılabilir (Cortes & Vapnik, 1995).

Destek Vektör Regresyon (DVR), Destek Vektör Makineleri (DVM) algoritmasının regresyon problemlerine uyarlanmış hâlidir. Amaç, bir fonksiyon $f(x)$ bulmak ve bu fonksiyonun tahmin ettiği değerlerin, gerçek değerlerden belirli bir hata toleransı (ϵ) dahilinde sapmasını sağlamaktır. DVR, yalnızca “hata ϵ ’den büyükse” ceza uygular. Bu nedenle kullanılan kayıp fonksiyonu ϵ -insensitive loss function olarak adlandırılır.

Veriler bu şekilde kabul edildi: $(x_i, y_i) \in R^n \times R, i = 1, \dots, l$

Aşağıdaki regresyon fonksiyonunu öğrenmek:

$$f(x) = w * x + b \quad (10)$$

Optimizasyon Problemi:

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (11)$$

Koşuluyla:

$$\begin{cases} y_i - (w * x_i + b) \leq \epsilon + \xi_i \\ (w * x_i + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Burada ϵ , modelin tahminlerinin gerçek değerlerden ne kadar sapabileceğini belirleyen tolerans aralığıdır; yani tahmin edilen değer ile gerçek değer arasındaki fark ϵ ’den küçükse bu hata görmezden gelinir. Ancak bu tolerans sınırını aşan durumlar için slack değişkenleri (ξ_i, ξ_i^*) devreye girer; bu değişkenler, her bir veri noktası için toleransı aşan hatanın büyüklüğünü ifade eder. Bu hataların model üzerindeki etkisini sınırlamak amacıyla, modelin karmaşıklığını kontrol eden bir ceza parametresi olan C kullanılır. Ceza parametresi C , büyük hata yapan veri noktalarına uygulanacak cezanın ne kadar ağır olacağını belirler; dolayısıyla düşük C değerleri daha esnek modeller üretirken, yüksek C değerleri daha katı, düşük hatalı fakat genelleme yeteneği sınırlı modellerle sonuçlanabilir.

Dual Formülasyon ve Kernel Fonksiyonu: Eğer veriler doğrusal olarak uygun değilse, SVR de çekirdek (kernel) fonksiyonu kullanarak verileri yüksek boyutlu bir uzaya dönüştürebilir. Bu durumda regresyon fonksiyonu şu şekilde yazılır:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (12)$$

Lagrange çarpanları (α_i, α_i^*) , optimizasyon problemini çözmek için kullanılan parametrelerdir ve destek vektörlerini belirlemek için kritik öneme sahiptir. Modelde kullanılan kernel fonksiyonu $K(x_i, x)$, verilerin yüksek boyutlu uzaya dönüştürülmesini sağlayarak doğrusal olmayan verilerin doğrusal bir şekilde ayrılmasını mümkün kılar. Bu dönüşüm sayesinde veriler daha yüksek boyutlarda doğrusal olarak ayrılabilir hâle gelir. Ayrıca, bias terimi (b) regresyon fonksiyonunun karar sınırını kaydırarak modelin genelleme yeteneğini artırır (Smola & Schölkopf, 2004).

İKİNCİ BÖLÜM

2. YÖNTEM

Zaman serileri, bir değişkenin zaman içindeki ardışık gözlemlerinden oluşan veri setleridir ve özellikle finans ve ekonomi alanlarında yaygın olarak kullanılır. Bu tür verilerde trend (uzun vadeli eğilim), mevsimsellik (belirli periyotlarla tekrar eden dalgalanmalar), döngüsel hareketler (ekonomik döngüler) ve rastgele hareketler gibi bileşenler bulunur (Hamilton, 1994, s. 45). Zaman serisi analizi, geçmiş verilerden yola çıkarak gelecekteki değerleri tahmin etmeye yardımcı olur ve ARIMA gibi modeller sıklıkla tercih edilir (Box vd., 2015: 112). Bu analizler, örneğin bir futbol kulübünün yıllık gelirlerinin tahmin edilmesi gibi pratik uygulamalara sahiptir.

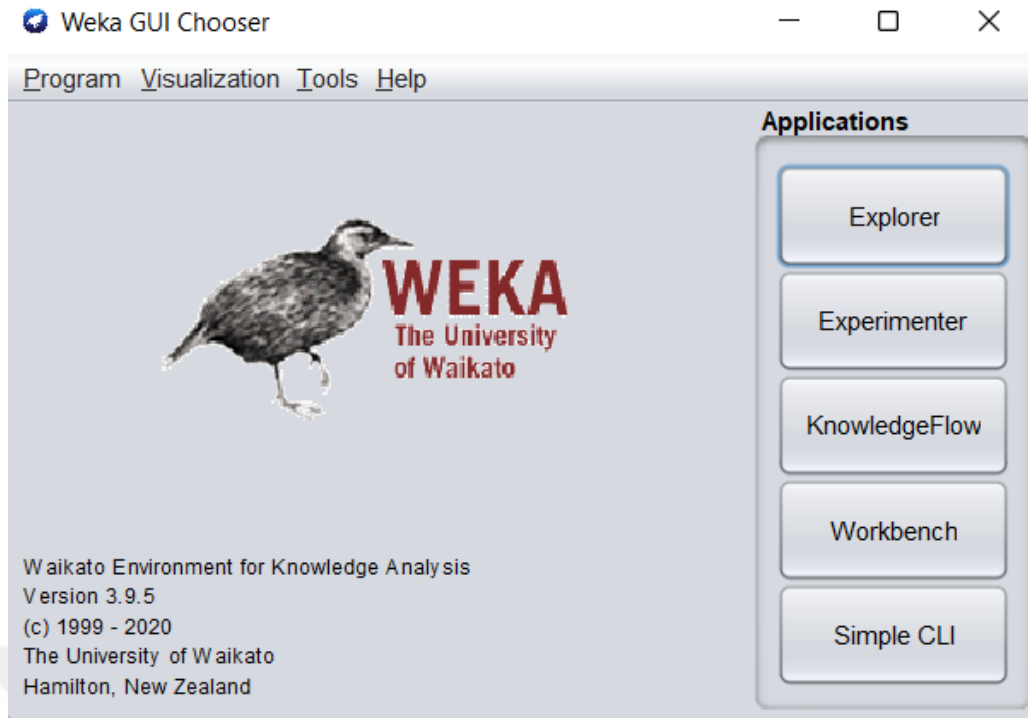
2.1. WEKA YAZILIMI

Weka (Waikato Bilgi Analizi Ortamı), veri madenciliği uygulamaları için tasarlanmış kapsamlı bir makine öğrenimi algoritmaları paketidir. Yeni Zelanda'daki Waikato Üniversitesi'nde geliştirilen bu yazılım ilk olarak 1997'de tanıtılmıştır. Weka'nın temel amacı, makine öğrenimi tekniklerini veri analizi görevlerine uygulamak isteyen araştırmacılar ve uygulayıcılar için sezgisel ve erişilebilir bir platform sağlamaktır. Kullanıcı dostu bir grafik arayüze sahiptir ve sınıflandırma, regresyon, kümeleme ve ilişki kuralı madenciliği dahil olmak üzere çok çeşitli veri madenciliği metodolojilerini destekler (Witten & Frank, 2002).

Java ile yazılan Weka, zamanla hem akademik hem de endüstriyel alanlarda yaygın olarak benimsenen bir araç haline gelmiştir. Yazılım, çok çeşitli algoritmaları kapsadığından, özellikle sınıflandırma ve regresyon analizi ile ilgili görevler ve ayrıca keşifsel veri analizi ve veri görselleştirme için etkilidir.

2.1.1. Weka Arayüzü

Çalışmamızda, kullanılacak olan WEKA 3.9.5 (1999-2020) versiyonu Şekil 4'te gösterildiği gibi kullanıcı arayüzü ekrana gelir.



Şekil 4: WEKA Görsel Arayüz Seçim Paneli

WEKA GUI seçim arayüzü, WEKA'nın grafiksel ortamını başlatmakta olup, bu ortamda Explorer, Experimenter, Knowledge Flow, Workbench ve SimpleCLI adında beş buton bulunmaktadır.

Explorer: WEKA ile veri keşfetme ortamı.

Preprocess Panel: Bu panel, veriyi bir veri tabanından, CSV dosyasından, ARFF formatından vb. alarak işleme sürecine sokar. Veriye yönelik filtreleme algoritmaları kullanılarak, verilerin bir formattan diğerine dönüştürülmesi sağlanabilir. Ayrıca, kullanıcı belirli kriterlere göre örnekleri ve öznitelikleri silebilir. Preprocess ekranında, belirli bir özniteliğe ait grafikler de görüntülenebilir.

Classify Panel: Bu panel, kullanıcıya sınıflandırma ve regresyon algoritmalarını veri setine uygulama imkânı sunar ve elde edilen modelin doğruluğunu tahmin eder. Sınıflandırma sonuçları, sınıflandırıcı çıktı alanında görüntülenebilir.

Cluster Panel: Bu panel, WEKA'nın kümeleme tekniklerine erişim sağlar.

Associate Panel: Bu panel, ilişkilendirme kurallarına erişim sağlar. İlişkilendirme kuralı için uygun parametreler seçildikten sonra, sonuçlar listesi görüntülenebilir veya kaydedilebilir.

Select Attributes Panel: Bu panel, veri setindeki tüm öznitelik kombinasyonları arasında arama yaparak, en uygun öznitelik alt kümesini seçmek ve bunun üzerinden tahmin yapma yeteneği sunar. Bu işlem, modelin doğruluğunu artırmaya yardımcı olabilir.

Visualize Panel: Bu panel, mevcut ilişkiyi 2D grafiklerle görselleştirir, böylece verilerin daha iyi anlaşılmasını sağlar. Grafikler, veri setinin görsel analizi için kullanılır.

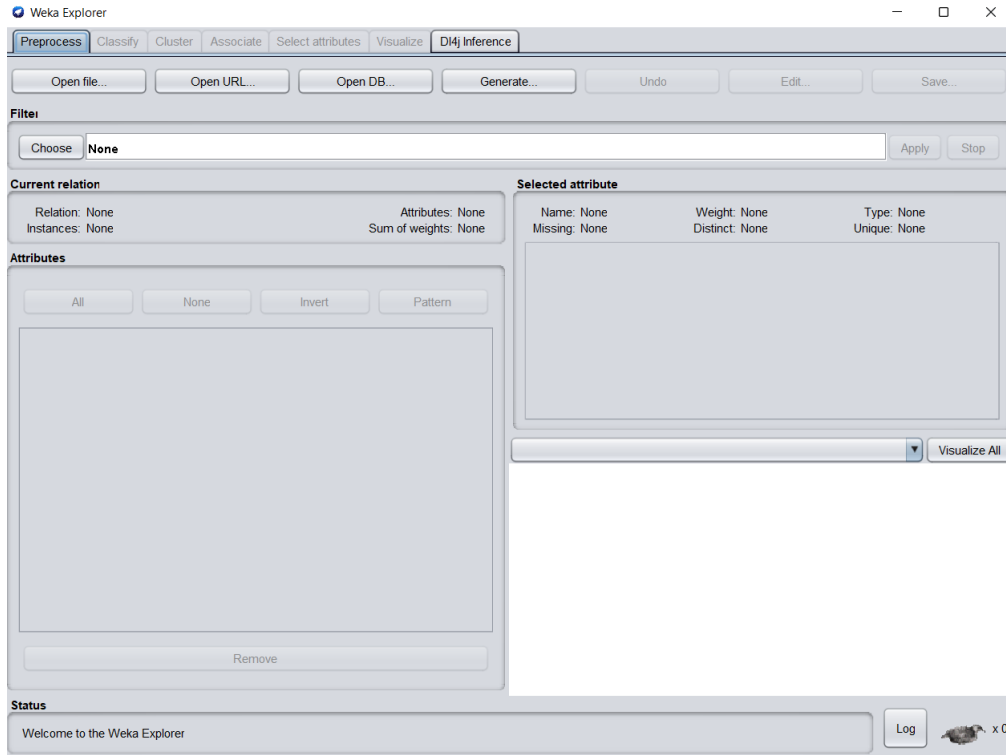
Experimenter: Farklı öğrenme algoritmalarını karşılaştırmak ve istatistiksel testler uygulamak amacıyla kullanılan bir deney ortamıdır.

KnowledgeFlow: Bu ortam, Explorer'ın sunduğu temel özellikleri aynı şekilde sunmakta olup, sürükle-bırak arayüzü ile kullanıcı etkileşimini kolaylaştırmaktadır. Artan öğrenme desteği ise önemli bir avantajdır.

Workbench: Kullanıcı tarafından seçilebilen 'perspektifler' içinde, diğer tüm uygulamaları birleştiren kapsamlı bir yazılımdır.

Simple CLI: İşletim sistemlerinde yerleşik komut satırı arayüzü bulunmayan durumlar için, WEKA komutlarını doğrudan çalıştırmayı mümkün kılan sade bir komut satırı arayüzü sunar.

Araştırmacı, kullanıcı arayüzünden çalışmasına yardımcı olacak butonu seçerek uygulamaya sürecine geçebilir. WEKA Explorer arayüzü Şekil 5'te gösterilmiştir.



Şekil 5: WEKA Explorer Arayüzü

Bu pencerede bulunan kullanılmayan butonlar, 'Open File' butonu seçilerek veri dosyalarının yüklenmesiyle aktif hale gelir.

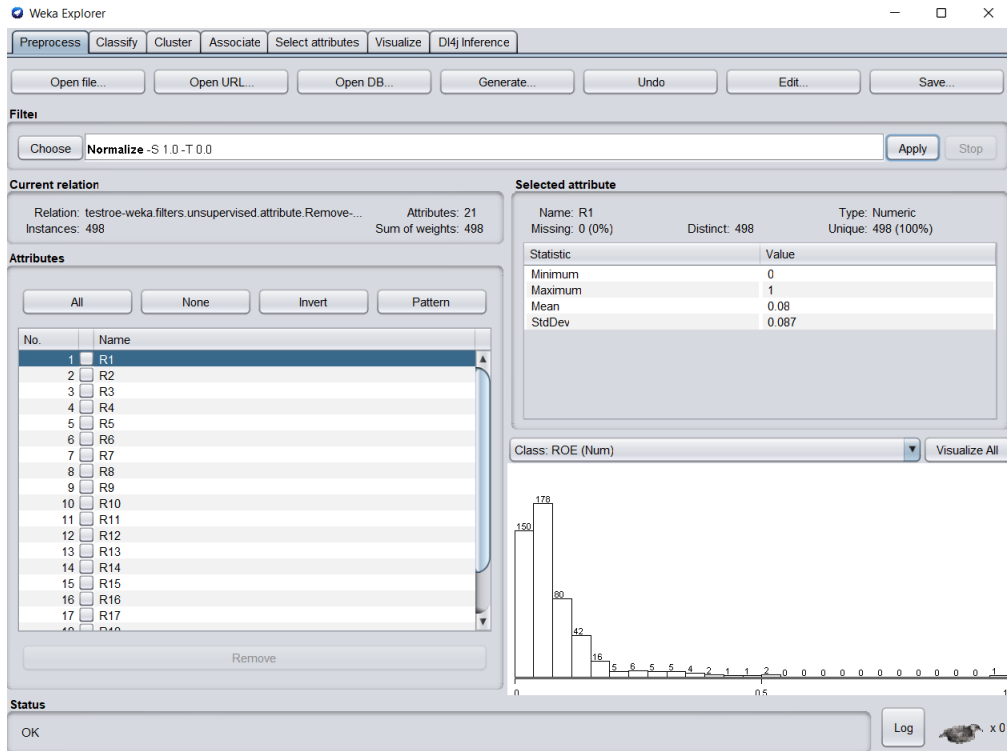
2.1.2. Veri Setinin İşlenmesi

Excel'de hazırlanan veri seti, 'Farklı Kaydet' özelliği aracılığıyla ve dosya formatı olarak 'CSV (virgülle ayrılmış)' tercihiyle kaydedilmelidir. Şekil 6'da çalışmamızda kullanılan örnek cvs dosyası gösterilmektedir.

```
Dosya   Düzenle   Görünüm
R1,R2,R3,R4,R5,R6,R7,R8,R9,R10,R11,R12,R13,R14,R15,R16,R17,R18,R19,ROA
0.445258983,1.08630342,3.008596868,0.511723297,-0.444487556,-0.008951832,0.36106646,1.362812483,-3.756531704,-1.132161956,-0.043391902,-2.997979798,-1.410542399,-0.024792754,47.07631318,89.19546907,31.84492272,-56.92736943,-0.032783958,0.226737968
1.386931635,1.071728724,4.259366831,-0.350750904,-0.47758249,-0.005407973,0.251616911,1.368801842,-3.711115514,-3.408658609,-0.037745195,-2.038383838,-0.950960967,-0.021492884,30.36175711,26.51111519,30.87996174,80.06594097,-0.023423116,0.199920745
1.010150137,1.030176949,3.957560169,-0.009182209,-0.528550933,-0.012912482,0.260306074,1.385987566,-3.591201983,-2.367781908,-0.051301115,-1.563636364,0.023791822,-0.049604997,-24.85436893,-23.09974569,5.584730358,32.88823979,-0.027018244,0.191821561
0.783383274,0.970712987,2.716021475,0.200767522,-0.47990305,-0.000807918,0.357402545,1.329478893,-4.035095785,-2.203678161,-0.034176245,-2.088080801,-0.609348659,-0.002260526,-28.71972318,19.2184605,65.11589128,83.71023381,-0.018310445,0.315708812
0.721448981,1.245289311,2.515748695,0.251959313,-0.663831916,-0.087043522,0.494997499,1.711605803,-2.405272408,-0.917047452,-0.02437024,-1.0.21201413,-0.35407147,-0.175846387,88.76551274,27.48505069,81.47980027,24.06086543,-0.079807177,0.338605741
1.408258029,1.136631866,1.613055615,-0.276743834,-0.767438342,-0.095929793,0.704645181,1.853381752,-2.171808511,-1.118617021,-0.027836879,-5.409893993,0.32429078,-0.136139145,31.41630901,-15.0492521,19.90203193,6.679066622,-0.100106496,0.271453901
2.052759459,0.827285922,0.831813207,-0.550253991,-0.634978229,-0.053519594,0.994557329,1.736393324,-2.357969943,-1.457009116,-0.035723085,-4.116607774,0.747228381,-0.053812477,-35.70640177,72.6070529,77.12082262,23.85340314,-0.089692982,0.287016507
4.994769874,1.533097686,1.50220403,-1.22718509,-0.750321337,-0.234254499,1.020565553,2.363753213,-1.733679001,-1.125382984,-0.022625501,-7.248,0.900070705,-0.229534005,41.45199063,2.74991912,-10.00578369,66.26044568,-0.169356529,0.427056328
4.204978038,1.104395604,1.235522485,-0.63302487,-0.151243493,-0.218623482,0.893869289,1.703007519,-2.422459893,-1.181406828,-0.043192102,-5.124,0.900452489,-0.244581042,1086.111111,-3.797074385,-36.88629312,-26.54731458,-0.233307148,0.526943645
1.467717718,0.468698668,0.799253035,-0.227413762,0.138346414,0.014601205,0.58642088,0.806534039,4.168867925,3.688679245,0.087735849,-0.432,-1.175471698,0.024898848,-280,30.2919708,39.59235669,145.7573853,0.015186916,-0.101886792
1.129169624,0.41477707,0.660178427,-0.046369427,0.234649682,0.012993631,0.628280255,0.688407643,2.209321341,1.300899428,0.075224857,0.24,-0.148814391,0.020681265,-93.80804954,138.030888,105.2824268,657.6190476,0.031326781,0.049059689
0.136363636,0.009937238,0.018339768,0.695606695,0.720188285,0.484309623,0.541841004,0.121861925,0.138773079,0.125074449,0.013103038,3.876,0.792138177,0.893822394,55.53772071,65.76,35.60283688,118.75,48.73684211,0.577129244
0.094674556,0.014184397,0.032,0.65106383,0.692198582,0.388652482,0.443262411,0.093617021,0.103286385,0.075117371,0.023474178,2.492,0.71830
```

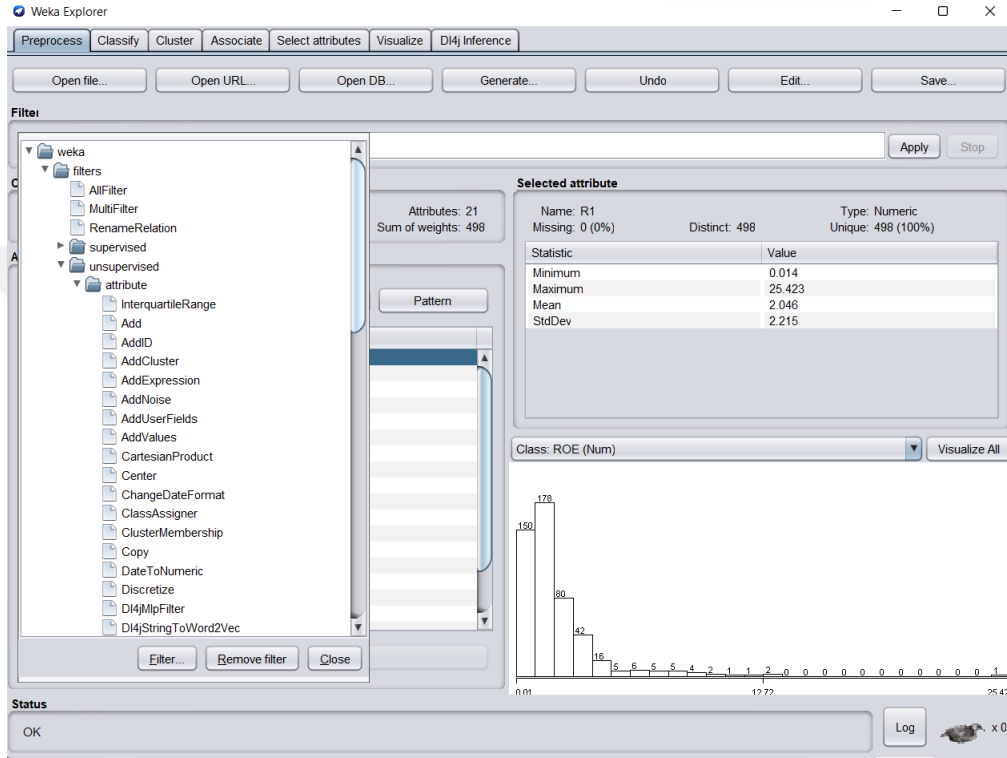
Şekil 6: Örnek Cvs (Virgülle Ayrılmış) Dosyası

WEKA Explorer arayüzünde 'Open File' penceresinden seçilen cvs dosya örneği Şekil 7'de gösterilmiştir.



Şekil 7: WEKA Explorer Arayüzünde, Cvs Yapısında Veri Seti

Bu pencerede çalışmamızda kullanılan veri seti örnek olarak gösterilmektedir. ‘Current relation’ kısmında; ‘Relation’, veri setinin ismini, ‘Attributes’, mevcut nitelik sayısını, ‘Instances’, toplam örnek sayısını ve ‘Sum of weights’, ağırlıkların toplamını belirtir. ‘Selected attribute’ bölümünde seçilmiş niteliğin bilgileri detaylı şekilde gösterilir ve bu kısmın alt bölümde, seçilen niteliğe ait görsel grafik sunulur. ‘Attributes’ alanında, veri setindeki tüm nitelikler seçici bir biçimde sıralanır ve bu bölümden nitelik seçimi yapılabilir ve işlenmeyecek nitelikler ‘Remove’ butonu ile silinebilir.



Şekil 8: WEKA Veri Seti Filtre Uygulama Penceresi

Şekil 8’ de ‘Filtre’ sekmesinden ‘Choose’ butonuna tıklayarak sırasıyla; ‘filters’, ‘unsupervised’, ‘attribute’, listesinden ‘normalize’ seçeneğiyle, Şekli 7’ de gösterildiği gibi veri setinin min-max değerlerini ‘0-1’ aralığında normalleştirme işlemi uygulanabilir.

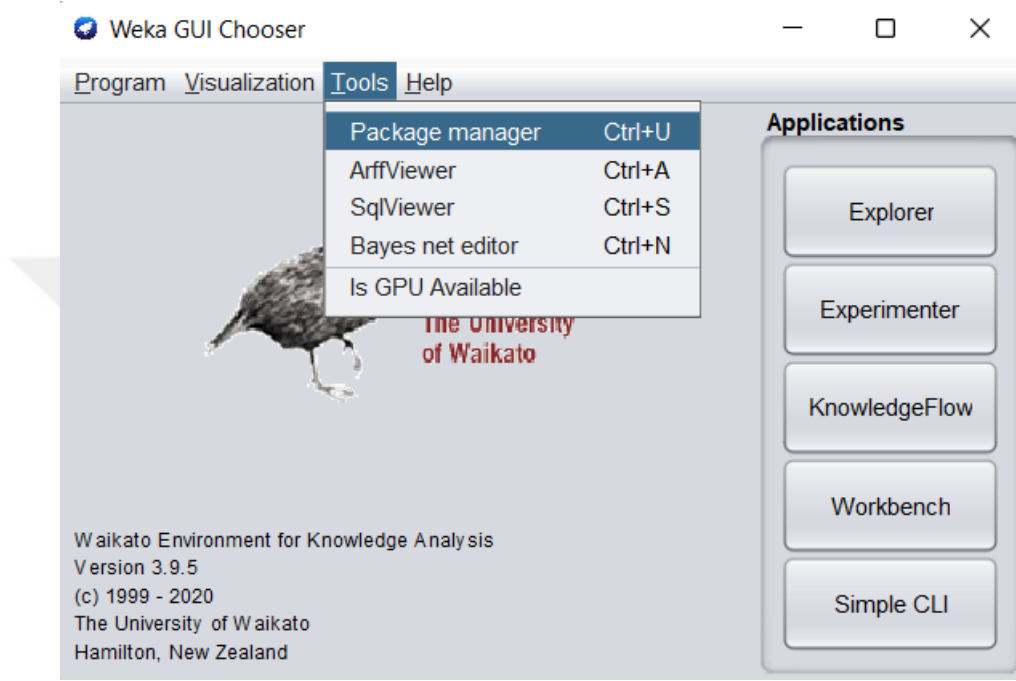
2.1.3. Regresyon Modeli

Regresyon, giriş verileri (özellikler) kullanılarak sürekli sayısal bir değer tahmin edilmesi işlemidir (Chatterjee & Hadi, 2015). WEKA’da regresyon, tahmin yapma amacıyla kullanılan çeşitli algoritmalar aracılığıyla gerçekleştirilir. Bu algoritmalar, sayısal değerlerin tahmin edilmesi için eğitim verilerini analiz ederek bir model oluşturur. Regresyon, giriş verileri ile bağımlı sayısal bir değişken arasında ilişki kurarak, gelecekteki veriler için tahminler

yapmayı hedefler. Bu süreç, farklı regresyon teknikleri kullanılarak, veriler üzerinde matematiksel bir modelleme yapılmasını sağlar (Frank vd., 2004).

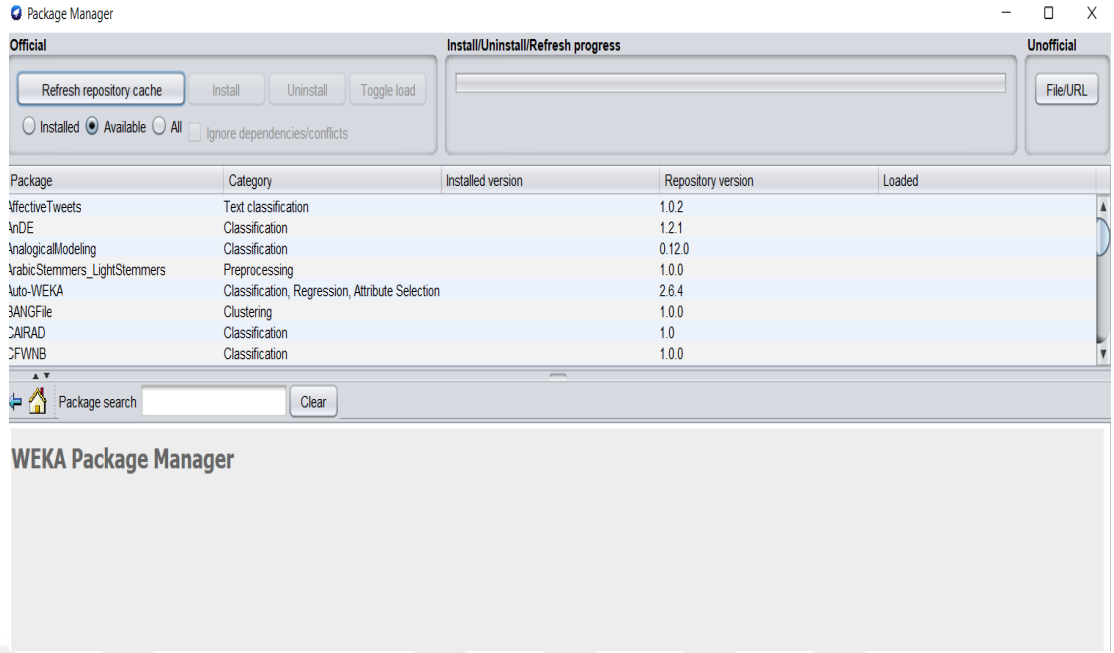
2.1.4. Seçilen Regresyon Modüllerinin Yüklenmesi

WEKA uygulamasında ‘Explorer’ arayüzünde bulunmayan regresyon modülleri, ‘Package Manager’ sekmesiyle karşıdan yüklemesi mümkündür. Şekil 9’daki adımlar takip edilir.



Şekil 9: WEKA Package Manager Seçimi

Uygulamayı açtığımızda karşımıza çıkan pencerenin üst kısmında yer alan ‘Tools’ listesinden ‘Package manager’ seçimiyle indirme arayüzü aktif hale gelecektir.

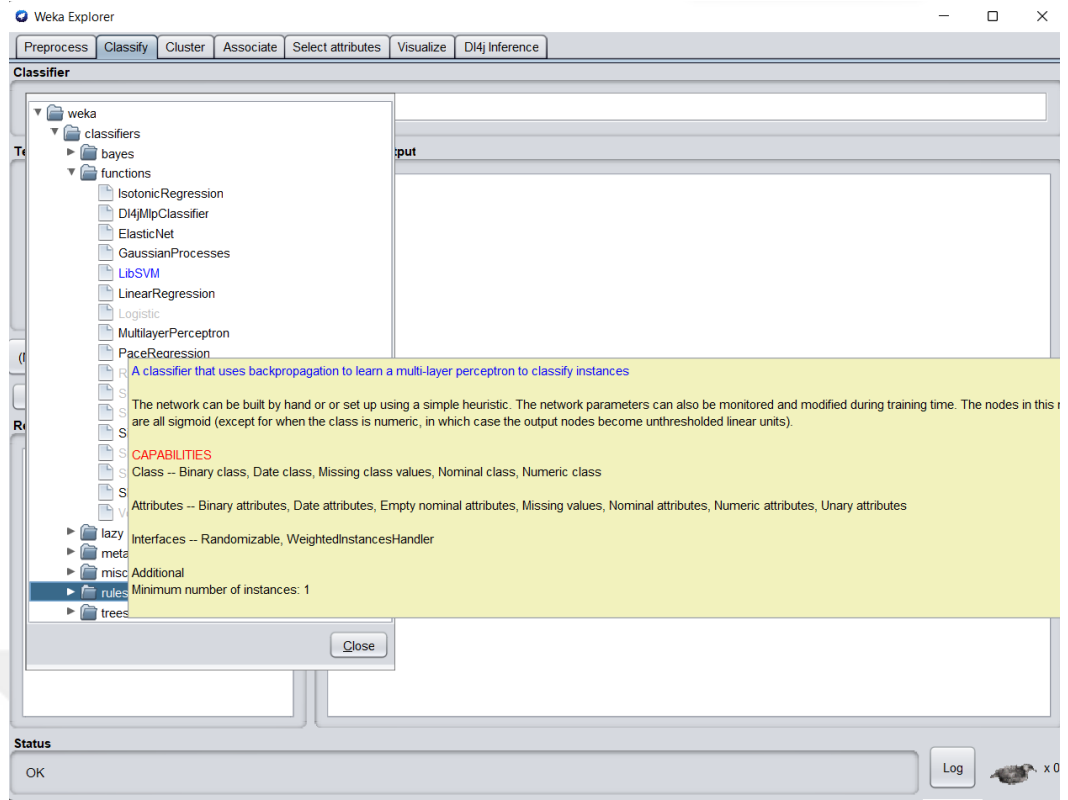


Şekil 10: WEKA Modül İndirme Arayüzü

Şekil 10'daki görselde 'Explorer' arayüzünde bulunmayan modüller listesi yer alır. Çalışmada kullanılacak herhangi bir modülü; 'Category' listesinden 'regression' özelliği taşıyan modül seçimi yapıldıktan sonra 'Install' seçeneği aktif hale gelecektir. Butona bastıktan sonra indirme işlemi başlamış olacak ve karşıdan yükleme işlemi tamamlanmış olacaktır. Daha sonrasında 'Explorer' arayüzünde indirmiş olduğumuz modül seçilerek çalışmanın analizine devam edilir.

2.1.5. Regresyon Modülü Uygulaması

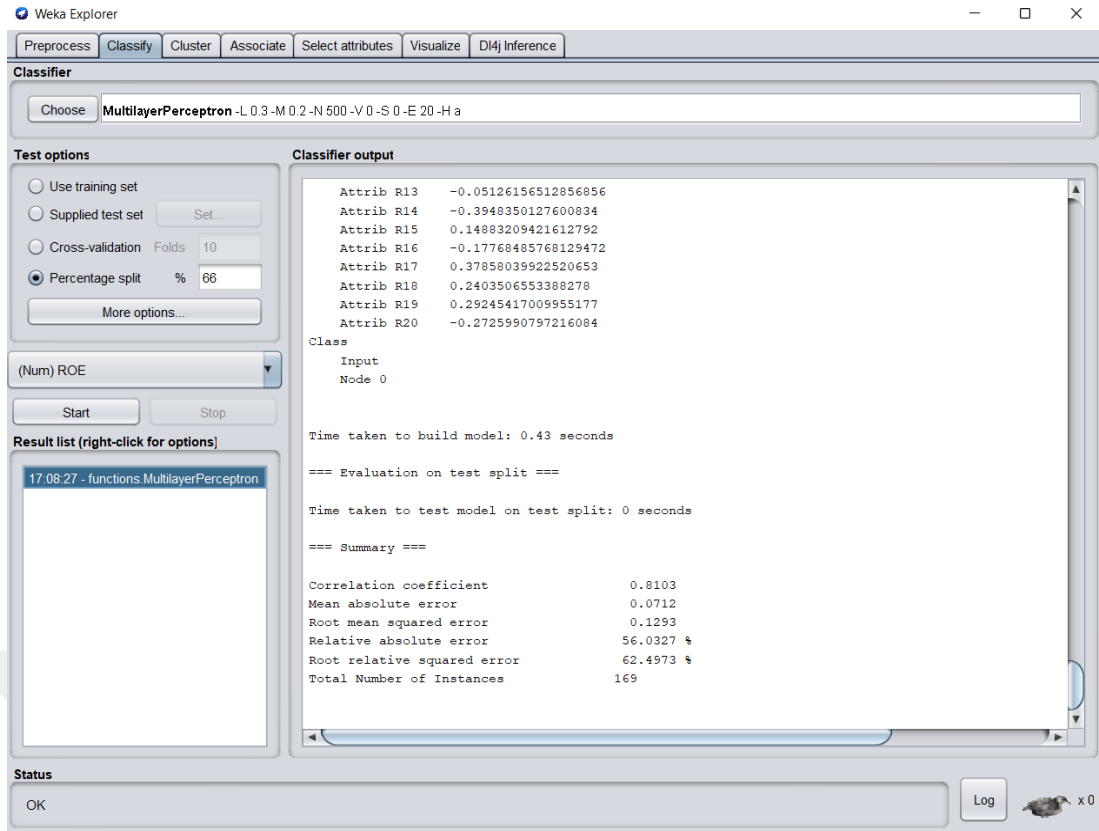
'Explorer' arayüzüne yüklenen veri setine öncelikle 'Preprocess' alanından filtreleme işlemi uygulanır, ardından veriler normalleştirilerek analiz için uygun hale getirilir. Bu işlemler tamamlandıktan sonra, 'Classify' alanında bulunan 'Classifier' modül seçme bölümünden 'Choose' butonuna basarak belirlenen regresyon modeli seçilir ve analiz süreci başlamaya hazır hale getirilir. Regresyon uygulama penceresi Şekil 11'de gösterilmiştir.



Şekil 11: Weka Regresyon Uygulama Penceresi

Modül seçilmeden önce, fare imleci modül isminin üzerine getirildiğinde, ilgili modül hakkında detaylı bilgilerin yer aldığı sarı renkte bir bilgi kutusu ortaya çıkar. Bu aşamada başlangıç örneği için belirlen regresyon modüllerinden ‘Multilayer Perceptron (Çok katmanlı algılayıcı)’ seçilmiştir.

Modül seçim işlemi tamamlandıktan sonra varsayılan ayarları değiştirmeden test süreci yapılmıştır.



Şekil 12: WEKA MLP Analiz Test Sonuçları

Analiz penceresinde yer alan ‘test options’ bölümü, bir makine öğrenimi modelinin performansını değerlendirmek için birden fazla değerlendirme stratejisi sunar. Bu bölüm, modelin doğruluğunu belirlemek ve yeni verilere genelleme yapma yeteneğini incelemek için farklı yöntemler sağlar:

Use training set (Eğitim verisini kullan): Modeli, aynı veri seti üzerinde test eder.

Supplied test set (Dış test verisi kullan): Modelin performansını değerlendirmek amacıyla, önceden tanımlanmış bir test veri setinin sisteme yüklenmesine olanak tanır.

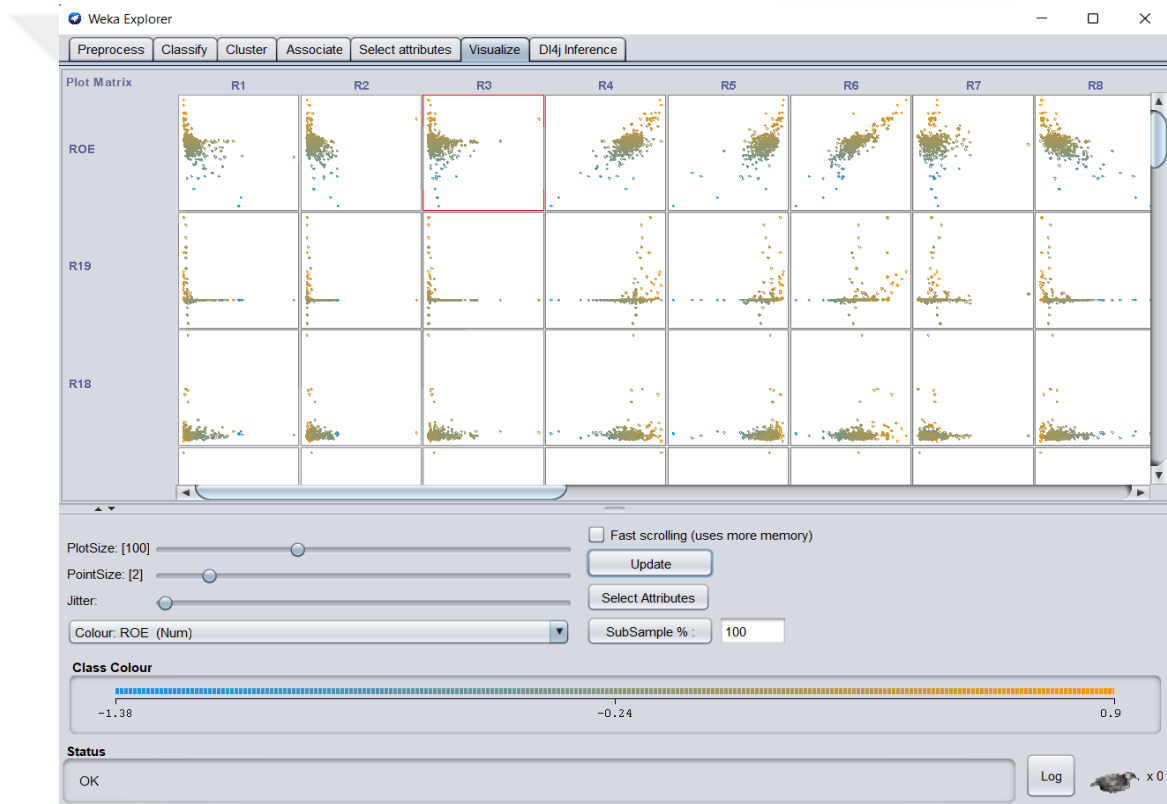
Cross-validation (Çapraz doğrulama): Veri, k katlı çapraz doğrulama tekniği ile belirli parçalara ayrılır. Varsayılan olarak 10 katlı çapraz doğrulama tercih edilse de bu değer kullanıcı tarafından değiştirilebilir. Model, rastgele oluşturulan alt veri kümeleri üzerinde değerlendirilerek performansı test edilir.

Percentage split (Yüzdelik ayırma): Veri setini belirli bir oran dahilinde eğitim ve test verisi olarak böler. Örneğin, %70 eğitim ve %30 test şeklinde bir ayırma yapıldığında, model verinin %70’lik kısmı ile öğrenir ve kalan %30’luk bölüm üzerinde değerlendirilir (Bouckaert vd., 2016).

Şekil 12’ de görüldüğü gibi ‘test options’ bölümünün hemen alt kısmında veri setine yüklenen bağımlı ve bağımsız değişkenlerin listesi bulunur (Bu çalışmada seçilmiş bağımlı

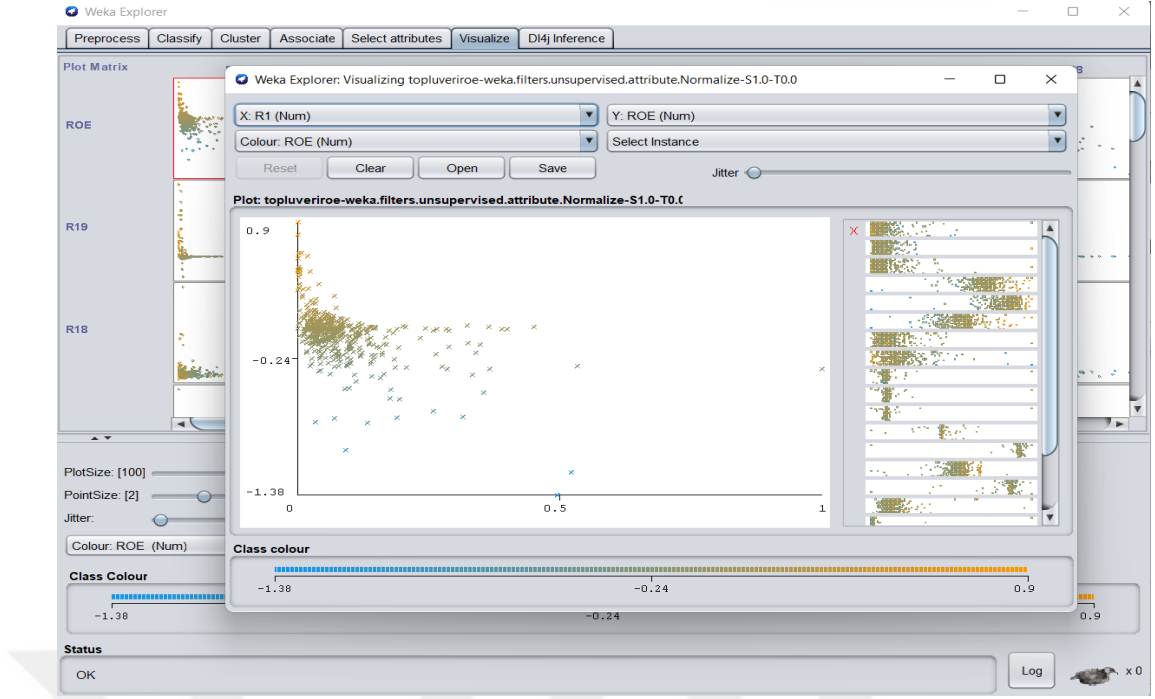
değişken ‘num (ROE)’ olarak gösterilmiştir. Test bölümü seçildikten sonra modülün varsayılan parametreleriyle ‘Start’ butonuna basılır ve analiz başlatılmış olur.

Modülün analiz sonuçları ‘Classifier Output’ bölümünde verilmektedir. Seçilen modülün başlangıç performansı, sağlanan hata metrikleri incelenerek değerlendirilebilir ve yorumlanabilir. Sonuçlar kısmında yer alan Correlation Coefficient; bağımlı ve bağımsız değişkenler arasındaki ilişkinin korelasyon katsayısını, Mean Absolute Error (MAE); hata değerlerinin mutlak ortalamasını, Mean Squared Error (MSE); hata değerlerinin karelerinin ortalamasını, Root Mean Squared Error (RMSE); hata kareleri ortalamasını, Mean Absolute Percentage Error (MAPE); hata değerlerinin yüzdesel mutlak ortalamasını gösterir. Değişkenler arasındaki korelasyon grafiklerine erişmek için ‘Explorer’ arayüzünde ‘Visualize’ sekmesine gidilmesi gerekmektedir.



Şekil 13: Korelasyon Grafik Çıktısı

‘Visualize’ sekmesinde Şekil 13’te gösterildiği tüm değişkenlerin aralarındaki korelasyon grafikleri yer alır. Herhangi bir grafik üzerine fare imleci yardımıyla basıldığında Şekil 14’ teki daha detaylı grafik bölümü açılır.

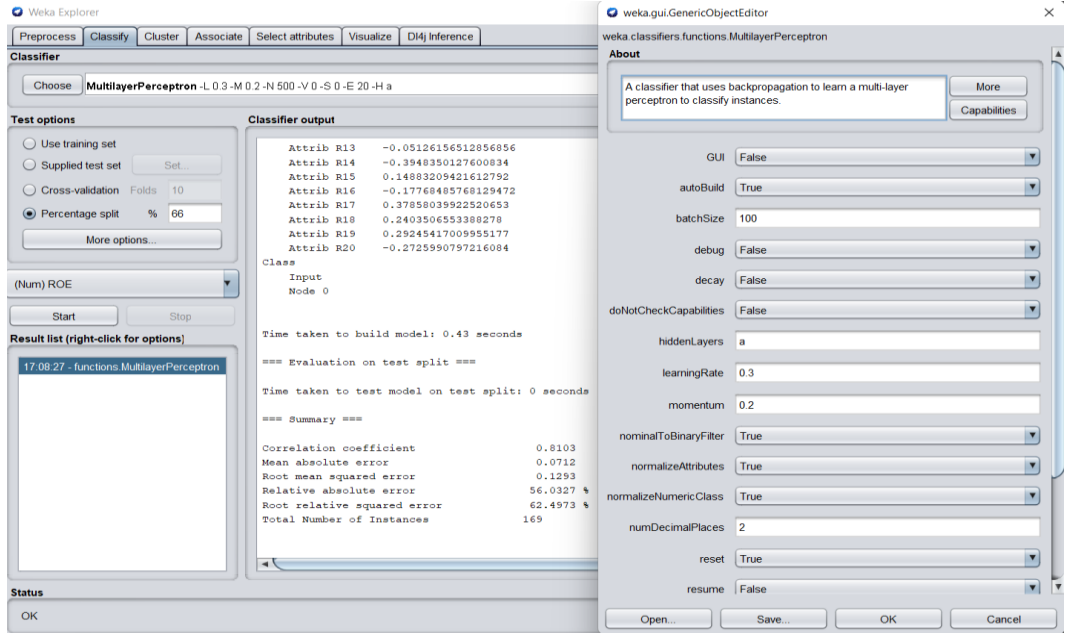


Şekil 14: Detaylı Korelasyon Grafik Çıktısı

Açılan bu grafik penceresinde ‘X’ ve ‘Y’ listelerinden değişkenler karşılıklı olarak korelasyon katsayıları karşılaştırılabilir.

2.1.6. Modül Parametrelerinin Yapılandırılması

Şekil 12’de uygulanan analizin ilk sonuçları verilmiştir. Sonuçlar analiz edildiğinde, modül parametreleri farklılaştırılarak daha iyi performansa ulaşılması hedeflenmektedir. Gerçekleştirilen tüm analizler sonucunda elde edilen en optimal çıktı, çalışmanın model oluşturma sürecini temsil edecektir.



Şekil 15: Modül Parametreleri Arayüzü

Şekil 13'te gösterildiği gibi, 'Explorer' penceresindeki 'Classifier' alanında seçilen modülün ismine fare imleci ile tıklanarak ilgili modülün parametre arayüzü açılabilir. Bu arayüzde bulunan batchSize, her eğitim adımında kullanılan veri alt kümesinin büyüklüğünü ifade eder; hidden layer, modeldeki gizli katmanların sayısını belirler; momentum, optimizasyon sürecinde ağırlık güncellemelerinin ivmesini kontrol eder; learning rate ise modelin öğrenme hızını düzenler. Bu parametreler için gerekli düzenlemeler yapıldıktan sonra 'ok' butonuna basılır ve yeni analiz için tekrar 'Start' butonuna basılır. Çalışmamızda ilerleyen bölümlerde modelin oluşturulmasında örnek teşkil etmesi amacıyla aşağıdaki Tablo 1'de ilgili parametreler belirtilmiştir.

Tablo 1: MLP Modeli için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Gizli Katman	1, 2, ...,10	3
Momentum Sabiti	0.1, 0.2, 0.3	0.2
Öğrenme Oranı	0.1, 0.2, 0.3	0.1

Tablo 1' de gösterilen seçilmiş seviyeler uygulandıktan sonra yeni bir analiz başlatılmıştır. Aşağıdaki Şekil 14'te de görüldüğü gibi korelasyon katsayısının arttığını ve hata metrik oranlarının düştüğü görülebilir. Ulaşılan en iyi performans sonucu yapılan analiz çalışmasının modelini oluşturur.

2.2. WEKA REGRESYON ALGORİTMALARI

Regresyon analizi, neden-sonuç ilişkisi gösteren iki veya daha fazla değişken arasındaki ilişkiyi incelemek ve bu ilişkilere dayalı tahminlerde bulunmak için yapılır. Analiz tek bir bağımsız değişkeni içerdiğinde buna tek değişkenli regresyon analizi denirken, birden fazla bağımsız değişkenin varlığı bunu çok değişkenli regresyon analizi olarak tanımlar. Doğrusal regresyon süreci, veri setinin yüklenmesi ve incelenmesiyle başlar. Daha sonra, veriler dilimleme yoluyla bölünür, ardından modeli oluşturmak için veri seti eğitilir ve bölünür. Son olarak, modelin güvenilirliğini sağlamak için modelin doğruluğu değerlendirilir. Regresyon analizinin birincil amacı, bir dizi bağımsız niteliğe dayalı olarak bağımlı değişkeni doğru bir şekilde tahmin eden etkili bir model geliştirmektir (Kadam vd., 2020).

Bu çalışmada karşılaştırmalı performans analizi yapılacak regresyon modelleri Tablo 2’de sunulmuştur.

Tablo 2: WEKA Regresyon Algoritmaları

No	Regresyon Algoritma İsmi		
1	MultilayerPerceptron	6	Alternating Model Trees
2	Linear Regression	7	Pace Regression
3	SMOreg	8	Isotonic regression
4	GaussianProcesses	9	RepTree
5	M5Rules	10	Elastic Net

2.2.1. MultilayerPerceptron (Çok Katmanlı Algılayıcı) Algoritması

Geri yayımlı sinir ağı, toplu olarak karmaşık çıktılar üreten birbirine bağlı işlem birimlerinden oluşur. Geri yayılım algoritması, sınıf etiketi tahminlerini geliştirmek için ağırlıkları yinelemeli olarak ayarlayarak çok katmanlı ileri beslemeli bir sinir ağında öğrenmeyi kolaylaştırır. Tipik bir ileri beslemeli sinir ağı bir giriş katmanı, bir veya daha fazla gizli katman ve bir çıktı katmanı içerir. Giriş katmanı, eğitim örneklerinden öznitelik değerleri alır, bunlar daha sonra ağırlıklandırılır ve çıktı katmanına ulaşmadan önce hesaplamaların gerçekleştiği gizli katmanlara geçirilir. Gizli katmanların sayısı değişebilse de çoğu durumda etkili öğrenme için tek bir gizli katman yeterlidir (Arora, 2012).

2.2.2. Linear Regression Algoritması

Linear Regression, öznitelik seçimini de dahil ederek çoklu doğrusal regresyon gerçekleştirmek için en küçük kareler yaklaşımını kullanır. Bu seçim süreci, açgözlü geriye

doğru eleme yöntemi kullanılarak veya başlangıçta tüm özniteliklere sahip tam bir model oluşturularak ve önceden tanımlanmış bir durdurma kriteri karşılanana kadar standart katsayılarına göre terimler sırayla kaldırılarak gerçekleştirilebilir. Her iki yaklaşım da sonlandırma koşulu olarak Akaike Bilgi Kriteri'nin (AIC) değiştirilmiş bir versiyonunu kullanır. Ek olarak, öznitelik seçimi gerekirse devre dışı bırakılabilir.

Uygulama iki önemli geliştirme içerir: devre dışı bırakılabilen eşdoğrusal öznitelikleri tespit etmek için sezgisel bir yöntem ve çoklu doğrusallık içeren durumları sabitleyen bir ridge parametresidir. Bu ridge parametresi ayrıca aşırı büyük katsayıları cezalandırarak aşırı uyumu azaltır. Teknik bir bakış açısından, Linear Regression algoritması, model genellemesini iyileştirmek için düzenlemeyi dahil ederek etkili bir şekilde ridge regresyon modeli olarak işlev görür (Bouckaert vd., 2016).

Genel anlamda, basit doğrusal regresyon modeli aşağıdaki gibi ifade edilir:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (13)$$

Burada:

- Y: Bağımlı değişken,
- X: Bağımsız değişken,
- β_0 : Sabit terim (intercept),
- β_1 : Bağımsız değişkenin katsayısı (eğim),
- ϵ : Hata terimi olup, modelin açıklayamadığı rastgele değişimleri temsil eder.

Çoklu doğrusal regresyon durumunda ise model şu şekilde genişletilir:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (14)$$

Bu modelde n adet bağımsız değişken bulunmaktadır.

Lineer regresyon modelinin katsayıları genellikle En Küçük Kareler (Ordinary Least Squares- OLS) yöntemi kullanılarak tahmin edilir. Bu yöntem, hata terimlerinin karelerinin toplamını en aza indiren parametreleri belirlemeyi hedefler.

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2 \quad (15)$$

Bu optimizasyon problemi, türev alınıp sıfıra eşitlenerek çözülür ve aşağıdaki normal denklemler elde edilir:

$$\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \quad (16)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (17)$$

Regresyon modelinin başarısını değerlendirmek için çeşitli istatistiksel ölçütler kullanılmaktadır:

- R-Kare (R^2) Katsayısı: Modelin bağımlı değişkendeki değişkenliği ne kadar açıkladığını gösterir.
- F Testi: Modelin genel anlamlılığını test eder.
- T Testi: Regresyon katsayılarının istatistiksel olarak anlamlı olup olmadığını değerlendirir.
- Durbin-Watson Testi: Hata terimlerinin otokorelasyona sahip olup olmadığını belirler.
- VIF (Variance Inflation Factor): Çoklu doğrusal bağlantı problemini değerlendirmek için kullanılır (Montgomery vd., 2021).

2.2.3. SMOreg Algoritması

SMOreg, yani Sequential Minimal Optimization, regresyon analizinde Destek Vektör Makineleri (DVM) kullanarak sayısal değerlerin tahmin edilmesini sağlayan bir tekniktir. Bu yöntem, DVM eğitim sürecinde önemli bir yer tutan kuadratik programlama optimizasyonu gibi karmaşık matematiksel problemleri ele alır. SMOreg, bu süreci daha etkin bir şekilde yönetmek amacıyla, büyük ölçekli optimizasyon problemini daha küçük alt problemlere bölerek her birini ayrı ayrı çözer ve böylece hesaplama sürecini kolaylaştırır. SMO algoritmasını optimize edilmiş bir şekilde uygular, böylece hesaplama maliyetini düşürür. Kernel fonksiyonları (çekirdek fonksiyonları) ile doğrusal olmayan regresyon modellerini destekler. Ceza parametresi ve epsilon hata toleransı gibi hiper parametreleri ayarlayarak modelin esnekliği artırılabilir. Özellik seçimi ve ölçeklendirme gerektirebilir, çünkü Destek Vektör Regresyonu, özellikle büyük ölçekli veri kümelerinde hassas ayarlamalar gerektirir (Nizar vd., 2024).

Destek Vektör Regresyonu (DVR), doğrusal ve doğrusal olmayan regresyon problemlerini çözmek için aşağıdaki optimizasyon problemini kullanır:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (18)$$

Kısıtlar:

$$y_i - (w_i, x_i) - b \leq \epsilon + \xi_i,$$

$$(w_i, x_i) + b - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, n$$

Burada;

- w ve b regresyon modelinin parametreleridir,
- ϵ hata payıdır,
- ξ_i ve ξ_i^* hata terimleridir,
- C cezalandırma parametresidir.

SMOreg algoritması, söz konusu optimizasyon problemini çözmek için Ardışık Minimal Optimizasyon (SMO) yönteminden yararlanmaktadır. Bu yöntem, klasik kuadratik programlama (QP) tekniklerine göre daha verimli olup, özellikle büyük ölçekli veri kümelerinde hesaplama maliyetini belirgin şekilde düşürmektedir.

SMOreg, doğrusal olmayan regresyon problemlerini çözmek için çekirdek dönüşümü kullanır. Çekirdek $K(x_i, x_j)$ fonksiyonu, giriş uzayını daha yüksek boyutlu bir özellik uzayına dönüştürerek doğrusal bir ayırma yüzeyi oluşturulmasını sağlar. Yaygın kullanılan çekirdek fonksiyonları şunlardır:

Doğrusal Çekirdek:

$$K(x_i, x_j) = x_i^T x_j \quad (19)$$

Polinom Çekirdek:

$$K(x_i, x_j) = (x_i^T x_j + c)^d \quad (20)$$

Radyal Tabanlı Fonksiyon (RBF) Çekirdek:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (21)$$

Sigmoid Çekirdek:

$$K(x_i, x_j) = \tanh(ax_i^T x_j + c) \quad (22)$$

SMOreg algoritması, her iterasyonda yalnızca iki Lagrange çarpanını güncelleyerek optimizasyon sürecini hızlandırır. Güncelleme adımları şu şekilde özetlenebilir:

- İki Lagrange çarpanı a_i ve a_j seçilir.
- Yeni a değerleri hesaplanır ve sınırlandırılır.
- Model parametreleri w ve b güncellenir.
- Durdurma kriteri sağlanana kadar süreç tekrar edilir.

SMOreg ile eğitilen modelin performansı, Ortalama Kare Hata (MSE), Kök Ortalama Kare Hata (RMSE), Ortalama Mutlak Hata (MAE) ve R-Kare (R^2) katsayısı gibi istatistiksel ölçütler kullanılarak değerlendirilir (Smola & Schölkopf, 2004).

2.2.4. GaussianProcesses Algoritması

GaussianProcesses, doğrusal olmayan regresyon problemleri için Bayesyen Gaussian süreçleri tekniğini uygular. Kullanıcılar, regresyon modelinin esnekliğini belirlemek amacıyla çekirdek (kernel) fonksiyonunu seçebilir ve uyum derecesini kontrol etmek için gürültü (noise) düzenleme parametresini ayarlayabilir. Model eğitilmeden önce, verilerin normalizasyon veya standartlaştırma işlemlerine tabi tutulması opsiyoneldir. Noktasal tahminler açısından bakıldığında, bu yöntem çekirdekli ridge regresyonu (kernel ridge regression) ile eşdeğer bir yapıdadır (Bouckaert vd., 2016).

Gaussian Process, bir fonksiyonun herhangi bir alt kümesi üzerinde ortak bir Gauss dağılımı oluşturacak şekilde tanımlanan bir dizi rastgele değişkenlerdir. Formülleştirildiğinde, bir GP şu şekilde ifade edilir (Williams & Rasmussen, 1995):

$$f(x) \sim GP(m(x), k(x, x')) \quad (23)$$

Burada:

- $f(x)$: Modellemek istediğimiz fonksiyon.
- $m(x)$: Ortalamayı temsil eden fonksiyon, genellikle sıfır olarak kabul edilir ($m(x) = 0$).
- $k(x, x')$: İki veri noktası arasındaki kovaryansı belirleyen çekirdek (kernel) fonksiyonudur.

Çekirdek fonksiyonu, GP's'nin temel bileşenlerinden biridir ve veriler arasındaki ilişkiyi tanımlar. Yaygın olarak kullanılan bazı çekirdek fonksiyonları şunlardır:

Radial Basis Function (RBF) veya Gaussian Kernel:

$$k(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \quad (24)$$

Burada σ^2 ölçek parametresi ve l uzunluk ölçeğini temsil eder.

Linear Kernel:

$$k(x, x') = x^T x' + c \quad (25)$$

Basit bir lineer ilişkiyi temsil eder.

Gaussian Process regresyonu, gözlemlenen verilerin (X, y) mevcut olduğu bir durumda, modelin hiperparametrelerini öğrenmek için maksimum olasılık tahmin veya Bayes optimizasyonu yöntemleri kullanılabilir. Verilerden elde edilen gözlemler, aşağıdaki dağılım ile tanımlanabilir:

$$y \sim N(m(X), K(X, X) + \sigma_n^2 I) \quad (26)$$

Burada $K(X, X)$ çekirdek matrisidir ve σ_n^2 gözlem gürültüsünün varyansını temsil eder.

Yeni bir veri noktası x_* için, tahmin edilen değer y_* ve belirsizlik (standart sapma) aşağıdaki şekilde hesaplanır:

- Kovaryans Hesaplama:

$$K_* = k(X, x_*) \quad (27)$$

$$K_{**} = k(x_*, x_*) \quad (28)$$

- Koşullu Dağılım: Koşullu dağılımın formülü kullanılarak tahmin yapılır:

$$p(y_* | X, y, x_*) = N(\mu_*, \sigma_*^2) \quad (29)$$

Burada ortalama ve varyans şu şekilde hesaplanır:

$$\mu_* = K_*^T (K + \sigma_n^2 I)^{-1} y \quad (30)$$

$$\sigma_*^2 = K_{**} - K_*^T (K + \sigma_n^2 I)^{-1} K_* \quad (31)$$

2.2.5. M5Rules Algoritması

M5Rules, M5' model ağaçlarını kullanarak regresyon kuralları elde eden bir algoritmadır (Onyelowe vd., 2025). Ridor algoritması, istisna durumlarını öğrenebilmek amacıyla öncelikle genel bir varsayılan kural oluşturur. Daha sonra, artan azaltılmış hata budama yöntemini kullanarak, en düşük hata oranına sahip istisnaları belirler. Süreç, her istisna için en uygun alt kuralların oluşturulmasıyla iteratif (yinelemeli) olarak devam eder (Janssen &

Fürnkranz, 2011). JRip, Ripper algoritmasını uygular ve kural kümesini optimize etmek için sezgisel küresel iyileştirme tekniklerini içerir. NNge, iç içe geçmemiş genelleştirilmiş örnekler kullanarak kurallar oluşturan bir en yakın komşu tabanlı öğrenme yöntemidir.

M5Rules, M5 algoritmasının kural tabanlı bir versiyonudur. Bu algoritma, karar ağaçlarının oluşturduğu kuralları doğrudan kullanarak, modelin yorumlana bilirliliğini artırmayı hedefler. M5Rules algoritmasının temel adımları aşağıdaki gibidir (Wang & Witten, 1996):

Ağaç Oluşturma: M5 algoritmasındaki gibi bir karar ağacı oluşturulur. Ancak burada, her bir düğüm için oluşturulan kurallar, daha sonra kullanıcı tarafından anlaşılır bir biçimde sunulmak üzere çıkarılır.

Kural Tabanlı Modelleme: Oluşturulan karar ağacı, her bir yaprak düğümde elde edilen lineer regresyon modelinin koşullarını temsil eden kurallara dönüştürülür. Bu kurallar, genellikle "Eğer... O zaman..." biçiminde ifade edilir.

Kural Seçimi ve Optimizasyonu: Oluşturulan kurallar arasında, en iyi tahmin performansını gösterenler seçilir. Bu seçim, doğruluk ve genel performans kriterlerine göre gerçekleştirilir.

Model Değerlendirme: Son olarak, modelin performansı test verileri üzerinde değerlendirilir. Modelin genel başarısını ölçmek için ortalama kare hatası (MSE) gibi metrikler kullanılır.

M5Rules algoritması hem istatistiksel öğrenme hem de kural tabanlı sistemler arasında köprü kurarak çalışır. Aşağıda, M5Rules'un temel matematiksel bileşenleri özetlenmiştir:

Regresyon Analizi: Modelleme, verilerin lineer ilişkilerini keşfetmek için regresyon analizi kullanır. Her bir alt küme için elde edilen regresyon denklemleri, şu şekilde ifade edilebilir (Quinlan, 1992):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (32)$$

Burada y , bağımlı değişken; x_i , bağımsız değişkenler; β_i , regresyon katsayıları ve ϵ , hata terimidir.

Kural Oluşturma: M5Rules'da kurallar, verilerdeki belirli eşik değerlerine dayanmaktadır. Örneğin, bir kural aşağıdaki gibi ifade edilebilir:

$$\text{Eğer } x_1 \leq a \text{ ve } x_2 > b \text{ ise } y \approx \hat{y} \quad (33)$$

Burada a ve b belirli eşik değerleridir ve $\hat{y}$ tahmin edilen değerdir.

2.2.6. Alternating Model Trees Algoritması

Alternatif ağaç yapısı, karar ağaçlarının genişletilmiş bir versiyonu olup, her düğümde birden fazla alternatif yol sunarak esnek ve daha güçlü bir modelleme kapasitesi sağlar. Geleneksel karar ağaçlarının aksine, alternatif ağaçlar belirli bir düğümde tek bir bölünme yerine, birden fazla alternatif karar yolu oluşturarak veri içindeki karmaşık ilişkileri daha iyi yakalamayı hedefler. Bu yapı, her bir düğümde farklı koşulları değerlendirerek, modelin daha esnek olmasına ve belirli bir örneğin farklı karar yolları üzerinden sınıflandırılmasına olanak tanır. Alternatif ağaçlar, kural tabanlı öğrenme ile ağaç tabanlı modellerin avantajlarını birleştirerek, genelleme yeteneğini artırır ve hata oranını minimize etmeye yardımcı olur (Frank vd., 2015).

2.2.7. PaceRegression Algoritması

Pace regresyonu, doğrusal modellerin oluşturulmasında kullanılan modern bir yaklaşımdır ve farklı modellerin rekabetini dikkate alarak regresyon analizini iyileştirmeyi amaçlar. Klasik en küçük kareler regresyonu, teorik olarak iyi tanımlanmış ve hesaplama açısından basit bir yöntem olsa da elde edilen modeller genellikle tatmin edici sonuçlar vermez. Pace regresyonu, her değişkenin modele katkısını değerlendirerek ve kümeleme analizi uygulayarak tahmin sürecinin istatistiksel temelini güçlendirir. Bu yaklaşım, değişkenlerin toplam regresyon üzerindeki etkilerini daha hassas bir şekilde belirlemeyi sağlar. Belirli düzenlilik koşulları altında, bağımsız değişken sayısı arttıkça Pace regresyonunun asimptotik olarak optimal olduğu kanıtlanmıştır. Wang ve Witten (1996) tarafından geliştirilen bu yöntem, özellikle yüksek boyutlu veri kümeleri üzerinde yapılan karşılaştırmalarda diğer regresyon modellerine kıyasla üstün performans sergilemiştir (Meshkin vd., 2009).

Pace Regression, klasik regresyon modellerinin genişletilmiş bir versiyonudur ve yüksek boyutlu verilerde modelin aşırı uyum (overfitting) riskini azaltmayı amaçlar. Bu yöntem, regresyon katsayılarını tahmin ederken, özellik seçiminde ve düzenlilemede (regularization) çeşitli teknikleri kullanır. Modelleme süreci, temel olarak aşağıdaki adımları içerir:

- Model Tanımı: Regresyon modeli, bağımlı değişken Y ve bağımsız değişkenler $X = (X_1, X_2, \dots, X_p)$ arasındaki ilişkiyi tanımlar:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (34)$$

Burada, β katsayıları modelin parametreleridir ve ϵ hata terimini temsil eder.

- **Düzenleme (Regularization):** Aşırı uyum riskini azaltmak için düzenleme terimleri eklenir. Bu doğrultuda, en yaygın olarak kullanılan yöntemler Lasso (L1) ve Ridge (L2) düzenleme teknikleridir. Düzenleme terimi, modelin karmaşıklığını kontrol ederek genel performansını artırır.

- **Katsayı Tahmini:** Düzenlenmiş model için katsayılar, aşağıdaki gibi bir kayıp fonksiyonu minimize edilerek tahmin edilir:

$$\min_{\beta} \left(\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \lambda R(\beta) \right) \quad (35)$$

Burada $R(\beta)$, düzenleme terimini temsil eder ve λ düzenleme parametresidir.

Pace Regression, yüksek boyutlu veri setlerinin analizinde kullanılan matematiksel yapılar içerir. Bu yapılar, regresyon katsayılarının tahmin edilmesi ve modelin değerlendirilmesinde önemli rol oynamaktadır (Wang & Leng, 2007).

- **Kayıp Fonksiyonu:** Kayıp fonksiyonu, gözlemlenen değerler ile modelin tahminleri arasındaki farkı ölçer. Bu fonksiyon, genellikle aşağıdaki gibi tanımlanır:

$$L(\beta) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (36)$$

- **Düzenleme Terimleri:** Lasso ve Ridge düzenlemesi, kayıp fonksiyonuna eklenir:
Lasso (L1) Düzenlemesi:

$$R(\beta) = \sum_{j=1}^p |\beta_j| \quad (37)$$

Ridge (L2) Düzenlemesi:

$$R(\beta) = \sum_{j=1}^p \beta_j^2 \quad (38)$$

Bu algoritmanın güçlü yönleri, verilerden elde edilen bilgiye dayalı karar verme yeteneği ve yüksek performanslı tahminler yapabilme becerisidir. Bununla birlikte, algoritmanın sınırlamaları arasında hesaplama maliyeti ve veri setindeki gürültünün etkileri bulunmaktadır. Uygulama sırasında dikkatli bir model değerlendirmesi ve optimizasyon süreci gerekmektedir (Tibshirani, 1996).

2.2.8. Isotonic Regression Algoritması

İsotonik regresyon, doğrusal regresyona benzer bir yaklaşım olmakla birlikte, veri setine bağlı olarak yukarı veya aşağı yönde değişkenlik gösterebilir ve doğrusal bir formda değildir. Bu yöntem, ortalama kare hatasını minimize etmeyi amaçlayan bir tekniktir. Seri halindeki gözlemler üzerinde serbest biçimli eğriler oluşturularak, verinin monotonluğunu koruyacak şekilde artmayan veya azalmayan bir doğrultuda ilerlemesini sağlar. İisotonik regresyon, gerçek gözlemlere mümkün olduğunca yakın bir uyum sergileyerek veri setindeki eğilimleri daha esnek bir şekilde modellemeye yardımcı olur (Kumar vd., 2022).

İsotonik Regresyon, verilen bir veri kümesi için bir fonksiyon tahmin ederken, bu fonksiyonun monoton olmasını sağlamak amacıyla kullanılan bir optimizasyon problemidir.

Verilen bir veri kümesi:

$$D = \{ (x_i, y_i) \mid i = 1, 2, \dots, n \} \quad (39)$$

Denklem 33 için, burada x_i bağımsız değişken ve y_i bağımlı değişkendir. Amaç, monoton artan bir tahmin fonksiyonu $f(x)$ bulmaktır, yani:

$$x_i \leq x_j \Rightarrow f(x_i) \leq f(x_j)$$

Eşitsizliğini sağlayan bir tahmin y_i vektörü elde etmek gerekir. İisotonik Regresyon problemi, aşağıdaki optimizasyon problemi olarak ifade edilir:

$$\min_{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (40)$$

Aşağıdaki kısıt altında:

$$\hat{y}_1 \leq \hat{y}_2 \leq \dots \leq \hat{y}_n$$

Bu problem, en küçük kareler tahminine benzer bir yapı gösterir, ancak tahmin edilen değerlerin monoton olması gerekmektedir (Leeuw vd., 2010).

2.2.9. RepTree Algoritması

RepTree, regresyon ağacı mantığını kullanarak farklı yinelenmelerde birden fazla ağaç oluşturur ve bu ağaçlar arasından en iyi olanı temsilci ağaç olarak seçer. Ağacın budama işleminde, yapılan tahminlerin ortalama kare hatası bir ölçüt olarak kullanılır. Reduced Error Pruning Tree (REP Tree), bilgi kazanımı veya varyans azaltma kriterine dayalı olarak karar ağaçları oluşturan hızlı bir öğrenme algoritmasıdır. REP Tree, bilgi kazanımını bölme ölçütü

olarak kullanarak karar/regresyon ağacı oluşturur ve azaltılmış hata budama (Reduced Error Pruning- REP) yöntemiyle budama işlemini gerçekleştirir. Sayısal değişkenler için sıralama işlemini yalnızca bir kez yaparak hesaplama maliyetini düşürür. Eksik veriler ise, C4.5 algoritmasının kesirli örnekler yöntemi kullanılarak işlenir. Bu algoritmanın etkinliği, UCI veri seti üzerinde uygulanarak test edilmiş ve altı farklı sınıf değerine sahip cinsiyet sınıflandırması (gender classification) için hata matrisleri oluşturularak değerlendirilmiştir (Srinivasan & Mekala, 2014).

RepTree algoritması, klasik karar ağacı yöntemleri ile benzer bir yaklaşım izler, ancak hata azaltmalı budama (REP) işlemi ile ağaç yapısının optimizasyonunu gerçekleştirir. RepTree, karar ağacını oluştururken aşağıdaki adımları takip eder:

Bölme Kriteri (Splitting Criterion): RepTree, en küçük varyans (minimum variance) ilkesine dayalı olarak veri kümesini böler. Bölme işlemi için, belirli bir düğümdeki veri kümesi D iki alt küme D_L ve D_R olarak ayrılır. Bölme işlemi, varyans azaltımı kullanılarak en uygun bölme noktası seçilerek yapılır (Witten vd., 2005, s. 281):

$$Varyans\ Azaltımı = Var(D) - \left(\frac{|D_L|}{|D|} Var(D_L) + \frac{|D_R|}{|D|} Var(D_R) \right) \quad (41)$$

Burada:

- $Var(D)$: Bölünmeden önceki veri kümesinin varyansı,
- $Var(D_L)$ ve $Var(D_R)$: Sol ve sağ alt düğümler için hesaplanan varyanslardır.

Yaprak Düğümlerin Oluşturulması: Eğer belirlenen bir durdurma kriteri sağlanırsa (örneğin, belirli bir düğümdeki örnek sayısı çok küçükse), bölme işlemi sonlandırılır ve düğüm bir yaprak düğüm haline gelir. Yaprak düğüm, veri kümesindeki hedef değişkenin ortalaması ile temsil edilir.

RepTree'nin en önemli aşamalarından biri hata azaltmalı budama (Reduced Error Pruning, REP) yöntemidir. Bu aşamada, ağaç büyütüldükten sonra aşırı uyumu engellemek için bazı düğümler kaldırılır.

Budama İşlemi:

- Ağaç, bölünmemiş bir doğrulama veri kümesi üzerinde test edilir.
- Eğer bir düğümün kaldırılması modelin hatasını artırmıyorsa, bu düğüm kaldırılır (budama yapılır).

- Budama işlemi, ağacın doğrulama veri kümesindeki hata oranı minimize edilene kadar devam eder.

Budama Kriteri:

- Tahmin doğruluğunun korunması: Eğer bir iç düğüm kaldırılırsa ve doğrulama veri kümesindeki hata önemli ölçüde artmıyorsa, düğüm silinir ve alt ağaç yerine bir yaprak düğüm konulur.
- Hesaplama maliyetinin azaltılması: Budama, ağaç büyüklüğünü küçülterek hem hesaplama süresini azaltır hem de modelin genelleştirme yeteneğini artırır.

2.2.10. Elastic Net Algoritması

L1 (Lasso) ve L2 (Ridge) norm cezalarını birleştirerek hem değişken seçimi hem de modelin kararlılığını sağlamayı amaçlayan bir düzenleme yöntemidir. Elastic Net yöntemi, iki aşamalı olarak tanımlanır. İlk olarak, "naif" tahmin (β^N), aşağıdaki düzenleme fonksiyonu kullanılarak elde edilir:

$$J(\beta) = \alpha \sum_{j=1}^p \beta_j^2 + (1 - \alpha) \sum_{j=1}^p |\beta_j| \quad (42)$$

Burada $\|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2$ ve $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ şeklinde tanımlanır. Düzenleme parametresi λ , iki negatif olmayan cezanın toplamıdır ($\lambda = \lambda_1 + \lambda_2$), ve $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ olarak ifade edilir. İkinci aşamada, son elastic net tahmini (β^E), naif tahminin yeniden ölçeklendirilmiş bir versiyonu olarak tanımlanır:

$$\beta^E = (1 + \lambda_2)\beta^N \quad (43)$$

Bu yeniden ölçeklendirme, naif tahminde gözlenen aşırı küçülmeyi azaltmak amacıyla önerilmiştir (Zou & Hastie, 2005).

Elastic Net, L1 normunun sağladığı değişken seçme avantajı ile L2 normunun sunduğu istikrarı bir araya getirerek Ridge regresyona benzer bir küçültme etkisi oluştururken, aynı zamanda Lasso gibi değişken seçimi yapma yeteneğine sahiptir.

Düzenleme temelli regresyon katsayı tahminleri, genellikle Bayesçi bir olasılık dağılımının modu olarak yorumlanabilir. Eğer olasılık bileşeni normal doğrusal regresyon modeli olarak ele alınırsa ($p(y | \beta) = N(y | X\beta, \sigma^2 I_n)$), Lasso tahmininin bir posterior dağılımın modu olarak yorumlanabilir. Bu çerçevede, Elastic Net hem Ridge hem de Lasso

cezalarının birleşimini içerdiği için, farklı veri yapılarıyla daha esnek ve dengeli bir model oluşturma potansiyeline sahiptir (Hans, 2011).

2.3. MODEL DEĞERLENDİRME KRİTERLERİ

2.3.1. R²-Determinasyon Katsayısı

Determinasyon katsayısı (R²), regresyon analizlerinde modelin açıklayıcılık gücünü nicel olarak değerlendirmek amacıyla kullanılan temel istatistiksel ölçütlerden biridir. Bu katsayı, bağımlı değişkenin toplam varyansının ne kadarının regresyon modeli tarafından açıklandığını gösterir ve 0 ile 1 arasında bir değer alır. R² değeri 1'e yaklaştıkça modelin veri üzerindeki açıklama gücü artmakta, 0'a yaklaştıkça ise modelin bağımlı değişkeni açıklama kapasitesinin zayıf olduğu anlaşılmaktadır. Aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (44)$$

- SS_{res} : Tahmin hatalarının karelerinin toplamı (residual sum of squares),
- SS_{tot} : Gözlemlerin ortalamadan sapmalarının karelerinin toplamını (total sum of squares) ifade eder (Chicco vd., 2021).

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

Test veri setindeki tüm örneklerde gerçek ve tahmin edilen değerler arasındaki mutlak farkların ortalamasını temsil eder. Her tahmin hatası, belirli bir örnek için gözlenen ve tahmin edilen değerler arasındaki mutlak sapmanın hesaplanmasıyla belirlenir. Bireysel bir model için MAE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$E_i = \frac{1}{n} \sum_{j=1}^n |p_i - \hat{p}_i| \quad (45)$$

2.3.3. Root Mean Squared Error-RMSE (Hata Kareleri Ortalaması Karekökü)

Kök ortalama kare hatası (RMSE), bir temel tahmin ediciye göre hatayı temsil eder. Göreceli kare hatasının karekökünü hesaplayarak, ortaya çıkan değer tahmin edilen değişkenle aynı birimde ifade edilir ve bu da yorumlamayı daha sezgisel hale getirir. Belirli bir model için RMSE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$E_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (p_i - \hat{p}_i)^2} \quad (46)$$

2.3.4. Mean Squared Error-MSE (Ortalama Kare Hatası)

MSE, hata karelerinin ortalamasıdır. RMSE'nin karesi olarak da düşünülebilir ve büyük hataları daha fazla cezalandırır. MSE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$E_i = \frac{1}{n} \sum_{j=1}^n (p_i - \hat{p}_i)^2 \quad (47)$$

2.3.5. Mean Absolute Percentage Error-MAPE (Ortalama Mutlak Yüzde Hatası)

MAPE, modelin tahmin hatalarının gerçek değerlere göre yüzdesel ortalamasını temsil eder ve hata oranını yüzde biriminde gösterir. MAPE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$E_i = \frac{100}{n} \sum_{j=1}^n \left| \frac{p_i - \hat{p}_i}{p_i} \right| \quad (48)$$

2.3.6. Relative Absolute Error-RAE (Göreceli Mutlak Hata)

RAE, modelin mutlak hata değerini, genellikle ortalama tahmine dayalı basit bir referans modelin hatasıyla kıyaslayan bir performans ölçütüdür. RAE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$E_i = \frac{\sum_{j=1}^n |p_i - \hat{p}_i|}{\sum_{j=1}^n |p_i - \bar{p}|} \times 100 \quad (49)$$

2.3.7. Root Relative Squared Error-RRSE (Kök Göreceli Kare Hata)

RMSE'nin, temel model (genellikle ortalama tahmin) ile kıyaslanarak değerlendirilen bir versiyonudur ve modelin hata seviyesini oranlı bir şekilde ifade eder. RRSE, aşağıdaki formül kullanılarak niceliksel olarak belirlenir:

$$\sqrt{\frac{\sum_{j=1}^n (p_i - \hat{p}_i)^2}{\sum_{j=1}^n (p_i - \bar{p})^2}} \times 100 \quad (50)$$

ÜÇÜNCÜ BÖLÜM

3. BULGULAR

3.1. VERİ SETİ VE DEĞİŞKENLER

Profesyonel futbolun sürekli gelişen ortamında kulüplerin mali sağlığı; rekabet etme, büyümeyi sürdürme ve spor endüstrisine özgü zorlukların üstesinden gelme becerilerinde çok önemli bir rol oynamaktadır. Bu çalışma, finansal analiz literatürüne makine öğrenme algoritmalarını entegre etmeyi amaçlayarak, en yüksek performansa sahip modeli tasarlamayı hedeflemektedir. Finansal performans analizinde yaygın olarak kullanılan iki oranı; ROE ve ROA bağımlı değişken olarak seçilmiştir.

ROE, bir şirketin hissedarları tarafından sağlanan özsermaye ile ne kadar kâr üretebildiğini ölçen bir finansal orandır. Şirketin hissedar sermayesini net gelir elde etmek için ne kadar verimli kullandığını gösterir. ROE, hissedarların yatırımlarından elde ettikleri getiriyi nicel olarak ifade ederek, şirketin özsermaye sahipleri için nasıl bir değer yarattığına dair önemli bilgiler sunar. Bu gösterge, yatırımcılar açısından büyük önem taşımakta olup, özsermaye yatırımlarının kârlılığı hakkında net bir perspektif sağlayarak gelecekteki büyüme potansiyeli ve temettü ödemeleri konusunda değerlendirme yapmalarına imkân tanır. ROE genellikle diğer finansal oranlarla birlikte kullanılarak, bir şirketin genel finansal performansına dair daha kapsamlı bir analiz yapılmasına yardımcı olur. ROE oranı, Net Gelir/Özsermaye formülü ile elde edilir (Damodaran, 2007).

ROA, bir şirketin toplam varlıklarını kâr elde etmek için ne kadar verimli kullandığını ölçen bir kârlılık oranıdır. Şirketin hem özkaynak hem de borç yoluyla sağladığı kaynakları ne derecede etkin bir şekilde gelir üretmek için kullandığını gösterir. ROA, özellikle yönetimin şirket varlıklarını nasıl değerlendirdiğini analiz etmek açısından önemli olup, yatırımcılar, analistler ve yöneticiler için kritik bir performans göstergesidir. Şirketin sahip olduğu her birim varlıktan ne kadar getiri sağladığını ortaya koyarak, operasyonel verimliliği değerlendirmeye yardımcı olur. Finansman kaynağından bağımsız olarak, şirketin yatırımlarını kâra dönüştürme kapasitesini ölçmek için yaygın olarak kullanılan bir metriktir. ROA oranı, Net Gelir/Toplam Varlıklar formülü ile elde edilir (Ikhwal, 2016).

Bu çalışma için örneklem büyüklüğü Avrupa'nın en üst liglerinden tespit edilen 26 futbol kulübünden oluşturmaktadır. Araştırmaya konu olan futbol kulüplerinin isimleri ve elde edilen veri dönemleri Tablo 3'te verilmiştir. Çalışmada kullanılan futbol kulüplerinin finansal tablolarına REFINITIV EIKON (<https://eikon.refinitiv.com/>) altyapısından ulaşılmıştır.

Tablo 3: Avrupa Futbol Kulüpleri ve Dönemleri

Ülke ve Lig	Futbol Kulüpleri	Dönem
İtalya (Serie A)	Juventus FC	2022-1998
	SS Lazio	2022-1998
	AS Roma	2021-2000
İngiltere (Premier League)	Manchester United P	2022-2009
	Leicester City	2001-1998
	Arsenal	2022-2003
Portekiz (Liga NOS)	FC Porto	2022-1999
	SL Benfica	2022-2004
	Sporting CP	2022-1999
	SC Braga	2022-2000
Danimarka (Superliga)	Brøndby IF	2022-1998
	Aalborg BK	2022-1997
	Silkeborg IF	2022-2001
Fransa (Ligue 1)	Olympique Lyon	2022-2003
Almanya (Bundesliga 1)	Borussia Dortmund	2022-2000
Hollanda (Erevedisie)	AFC Ajax	2022-1998
Türkiye (Spor Toto Süper Lig)	Galatasaray Sportif A.Ş.	2022-2001
	Fenerbahçe Futbol A.Ş.	2022-2003
	Beşiktaş Yat. A.Ş.	2022-2002
	Trabzonspor Sportif A. Ş	2021-2003
İskoçya (Scottish Premiership)	Aberdeen FC	2002-1999
	Celtic FC	2022-1999
	Heart Of Midlothian FC	2005-1996
	FC Kopenhag	2022-1997
	Rangers FC	2014-2013
İsveç (Allsvenskan)	AIK	2022-2003

Futbol kulüplerinin finansal tabloları Refinitiv altyapısından ilgili yıllara ait finansal tablolar indirildikten sonra Microsoft Excel yardımıyla bağımlı ve bağımsız değişkenler hesaplanmıştır. Hesaplanan finansal oranlar ve bağımsız değişkenler için atanan kodlar Tablo 4'te belirtilmiştir. Tabloda 19 adet bağımsız değişken, 2 adet bağımlı değişken gösterilmiştir.

Tablo 4: Çalışmada Kullanılan Değişkenler

Kod	Değişkenler	Kod	Değişkenler
ROE	Net Gelir/Özsermaye (Bağımlı Değişken)	R10	Kısa Vadeli Borçlar/Özkaynak
ROA	Net Gelir/Toplam Varlıklar (Bağımlı Değişken)	R11	Sabit Varlıklar/Özkaynak
R1	Cari Borçlar/Cari Varlıklar	R12	Net Kar/Hisse Sayısı
R2	Toplam Borç/Toplam Varlıklar	R13	Net Sermaye/Özkaynak
R3	Toplam Borç/Toplam Gelir	R14	FVÖK/Satış
R4	İşletme Sermayesi/Toplam Varlıklar	R15	Net gelir artışı
R5	Birikmiş Kazançlar/Toplam Varlıklar	R16	Net Satış Büyümesi
R6	FVÖK/Toplam Varlıklar	R17	Varlık Büyümesi
R7	Satışlar/Toplam Varlıklar	R18	Borç Büyümesi
R8	Toplam Borçlar/Toplam Aktifler	R19	Borç Karşılama Oranı
R9	Toplam Borçlar/Özkaynaklar		

Finansal oran hesaplamaları çalışmada kullanılan her kulüp ve dönemleri için ayrı ayrı hesaplanmıştır. Hesaplamalar bittikten sonra veri setinde eksik olan boşluklar ‘linear interpolation’ ile hesaplanarak tamamlanmıştır. Net Gelir Artışı, Net Satış Büyümesi, Varlık Büyümesi ve Borç Büyümesi değişkenleri yıllık oranlar olarak hesaplandığından, bir önceki yıla ait artış veya azalış verisine ihtiyaç duyulmaktadır. Bu nedenle, kulüplerin hesaplanan tablolarında başlangıç yıllarında bu oranlara ‘0’ sabit değeri atanmıştır.

Finansal oranlar, öncelikle ‘xlsx’ formatında Excel ortamında hazırlanmış olup, verilerin son hali ‘arff’ formatına dönüştürülmelidir. Çalışmada kullanılan veri setinin ‘arff’ formatına uygun hale getirilmesi için, ondalıklı sayısal değerlerde kullanılan ‘virgül’ ayracı, ‘nokta’ ile değiştirilmiştir. Bu değişiklik iki bağımlı değişken için iki ayrı dosya olarak uygulanmıştır. Şekil 16’da ROA için Excel formatlı dosya örneği gösterilmiştir.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	ROA
2	0.44526	1.0863	3.0085969	0.51172	-0.44449	-0.00895	0.36107	1.36281	-3.75653	-1.13216	-0.04339	-2.99798	-1.41054	-0.02479	47.0763	89.1955	31.8449	-56.9274	-0.03278	0.22674
3	1.38693	1.07173	4.2593668	-0.35075	-0.47758	-0.00541	0.25162	1.3688	-3.71112	-3.40866	-0.03775	-2.03838	0.95096	-0.02149	30.3618	26.5111	30.88	80.0659	-0.02342	0.19992
4	1.01015	1.03018	3.9575602	-0.00918	-0.52855	-0.01291	0.26031	1.38599	-3.5912	-2.36778	-0.0513	-1.56364	0.02379	-0.0496	-24.8544	-23.0997	5.58473	32.8882	-0.02702	0.19182
5	0.78338	0.97071	2.7160215	0.20077	-0.4799	-0.00081	0.3574	1.32948	-4.0351	-2.20368	-0.03418	-2.08081	-0.60935	-0.00226	-28.7197	19.2185	65.1159	83.7102	-0.01831	0.31571
6	0.72145	1.24529	2.5157487	0.25196	-0.66383	-0.08704	0.495	1.71161	-2.40527	-0.91705	-0.02437	-10.212	-0.35407	-0.17585	88.7655	27.4855	81.4798	24.0609	-0.07981	0.33861
7	1.40826	1.13663	1.6130556	-0.27674	-0.76744	-0.09593	0.70465	1.85338	-2.17181	-1.11862	-0.02784	-5.40989	0.32429	-0.13614	31.4163	-15.0493	19.902	6.67907	-0.10011	0.27145
8	2.05276	0.82729	0.8318132	-0.55025	-0.63498	-0.05352	0.99456	1.73639	-2.35797	-1.45701	-0.03572	-4.11661	0.74723	-0.05381	-35.7064	72.6071	77.1208	23.8534	-0.08969	0.28702
9	4.29477	1.5331	1.5021048	-1.22719	-0.75032	-0.24425	1.02057	2.36375	-1.73368	-1.12538	0.02263	-7.248	0.90007	-0.22853	41.452	2.74992	-10.0058	66.2604	-0.16936	0.42706
10	4.20498	1.1044	1.2355225	-0.63302	-0.15124	-0.21862	0.89387	1.70301	-2.42246	-1.18141	-0.04319	-5.124	0.90045	-0.24458	1086.11	-3.79707	-36.8863	-26.5473	-0.23331	0.52694
11	1.46772	0.4687	0.7992523	-0.22741	0.13835	0.0146	0.58642	0.80653	4.16887	3.68868	0.08774	-0.432	-1.17547	0.0249	-280	30.292	39.5924	145.757	0.01519	-0.10189
12	1.12917	0.41478	0.6601784	-0.04637	0.23465	0.01299	0.62828	0.68841	2.20932	1.3009	0.07522	0.24	-0.14881	0.02068	-93.808	138.031	105.282	657.619	0.03133	0.04906
13	0.13636	0.00994	0.0183398	0.69561	0.72019	0.48431	0.54184	0.12186	0.13877	0.12507	0.0131	3.876	0.79214	0.89382	55.5377	65.76	35.6028	118.75	48.7368	0.57713
14	0.09467	0.01418	0.032	0.65106	0.6922	0.38865	0.44326	0.09362	0.10329	0.07512	0.02347	2.492	0.71831	0.8768	-2.3511	10.424	6.65658	12.9412	27.4	0.48748
15	0.09169	5.64054	13.174558	0.63691	0.70651	0.36762	0.42814	0.06505	0.06958	0.06877	0.01052	2.552	0.68123	0.85866	10.9565	-1.04895	12.1289	70	0.06518	0.51618
16	0.06477	0.0017	0.0034965	0.61238	0.6955	0.43172	0.48516	0.04835	0.0508	0.04456	0.00802	2.3	0.64349	0.88986	13.4122	13.2673	13.1478	38.8889	254.5	0.51248
17	0.05873	0.00288	0.0059406	0.55374	0.67466	0.42131	0.48464	0.03647	0.03781	0.03582	0.01891	2.028	0.57413	0.86931	-2.68714	1.60966	4.0959	-23.4043	146.333	0.50448
18	0.08624	0.00599	0.0120724	0.4975	0.64136	0.43057	0.4965	0.05694	0.06038	0.04979	0.03178	2.084	0.52754	0.8672	-1.51229	1.42857	12.0941	193.75	71.8333	0.55191
19	0.02759	0.0045	0.0090543	0.63158	0.6439	0.50504	0.54871	0.01904	0.01941	0.01826	0.02283	2.116	0.64384	0.92041	204.023	117.778	80.0403	-15.7895	53.875	0.60388
20	0.11377	0.003	0.0060362	0.29839	0.38508	0.35081	0.45363	0.05522	0.04043	0.05319	0.696	0.31489	0.77333	-17500	144.565	2154.55	5.55556	35.9167	0.37021	
21	0.9	0.0015	0	0.09091	0.09091	0.09091	4.18182	0.86364	4.75	4.5	0.5	-1	0.5	0.02174	-8750	72.2826	1077.27	2.77778	17.9583	-0.25
22	3.49681	0.88977	2.2801177	-0.37261	-0.69247	-0.02759	0.39023	1.37206	-3.68777	-1.40261	-0.04153	-0.69907	1.0015	-0.0707	-13.1185	92.258	34.1224	173.975	-0.08963	0.33286
23	1.2954	0.98816	3.6298691	-0.05826	-0.76266	-0.14186	0.27223	1.33212	-4.011	-0.76921	-0.04677	-0.80463	0.17541	-0.52109	830.407	-31.2326	26.6975	-36.2258	-0.16101	0.57565
24	2.96065	0.8046	1.6042014	-0.3361	-0.72399	0.0286	0.50156	1.17822	-6.61089	-2.84767	-0.14044	-0.08648	1.88583	0.05702	-292.975	-2.9973	25.0645	-29.4099	0.02467	0.14607
25	3.53827	0.62002	0.958814	-0.64505	-0.87332	0.05292	0.64666	1.19048	-6.25	-4.72072	-0.04941	0.04481	3.38653	0.08183	-111.193	47.8559	38.0196	20.0614	0.02103	-0.08858
26	4.13472	0.62875	1.0416069	-0.78368	-1.2347	-0.06399	0.60364	1.29157	-4.4297	-3.54521	-0.04224	-0.40037	2.68779	-0.10601	-23.469	62.8505	36.5392	93.9339	-0.1685	0.71353
27	1.60504	0.64183	1.2681724	-0.27434	-1.40179	-0.14952	0.50611	1.79832	-2.25263	-0.91162	-0.01695	-13.0787	0.34365	-0.29543	254.454	-24.3073	7.51519	-20.5536	-0.30399	0.46494
28	2.7438	0.85492	1.1892317	-0.62594	-1.10807	0.0558	0.71889	1.45925	-3.17748	-2.14457	-0.03076	-3.68981	1.36297	0.07762	-9.84163	22.8337	-14.0898	21.8242	0.01107	0.24516
29	1.87332	0.62621	1.2454743	-0.32379	-0.69017	-0.08932	0.50279	1.13325	-8.50455	-5.2122	-0.0847	-4.09259	2.42987	-0.17765	-24.766	50.6545	18.4589	27.7455	-0.17771	0.8051
30	2.42162	0.49899	1.2621818	-0.37809	-0.68229	-0.20428	0.39534	1.02228	-45.8774	-28.9032	-0.50323	-5.43981	16.9677	-0.51673	18.0905	10.7531	6.37712	10.8636	-0.40738	7.58065
31	3.98128	0.34715	0.9142167	-0.46276	-0.54611	-0.09191	0.37972	1.10812	-10.2489	-5.7157	-0.17397	-17.15827	4.28006	-0.24205	177.159	5.37656	12.3347	67.2599	-0.26167	1.40736
32	4.3066	0.29531	0.7646797	-0.31867	-0.44305	0.05068	0.49962	1.43171	-3.31635	-0.9614	-0.04815	-12.8214	0.73816	-0.12102	-47.5912	2.68828	312.837	17.4526	0.04479	0.14286
33	2.52393	0.77092	0.8297771	-0.88085	-1.57447	-0.35745	0.55956	2.52766	-1.6546	-0.95497	-0.04085	-24.4643	0.5766	-0.06051	-21.6247	1.34414	20.2046	53.7369	-0.46366	0.31801

Şekil 16: Excel Formatlı ROA Veri Seti

Yapılan düzenlemelerin ardından, “xlsx” formatındaki Excel dosyası, WEKA yazılımının taniyabilmesi için dosyayı ‘farklı kaydet’ alanından ‘csv’ (Comma-Separated Values, virgülle ayrılmış değerler) formatı ile kaydedilmiştir. Şekil 6’da veri setinin ‘csv’ formatlı düzeni gösterilmiştir.

3.2. MODEL OLUŞTURMA

Bu bölümde, hazırlanan veri seti üzerine daha önceki bölümde belirlenen, WEKA kütüphanelerinde yer alan regresyon algoritmaları uygulanmıştır. Kullanılan her algoritma için toplanan veri seti, gözlemlerin %66’sıyla eğitim adımını ve gözlemlerin %33’üyle test adımını gerçekleştirmek için iki gruba ayrılmıştır. ROE ve ROA olmak üzere her iki bağımlı değişken için ayrı ayrı analizler gerçekleştirilmiş, elde edilen sonuçlar karşılaştırılarak en yüksek performansa sahip model belirlenmiştir.

3.2.1. MultilayerPerceptron Modeli

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 5 ve 6’da gösterildiği şekilde belirlenmiştir.

Batch Size, makine öğrenimi ve derin öğrenme modellerinin eğitimi sırasında, verisetinden aynı anda modele gönderilen örneklerin sayısını ifade eder. Eğitimin her adımında model, belirli sayıda veriyi alır, bunlar üzerinden ileri besleme ve geri yayılım işlemlerini yaparak ağırlıklarını günceller.

Gizli Katman, yapay sinir ağlarında girdi (input) ve çıktı (output) katmanları arasında yer alan ve ağırlık öğrenme kapasitesini artıran ara katmanlardır. Bu katmanlar, girdilerden gelen bilgileri işler ve daha karmaşık özellikleri öğrenmeye yardımcı olur. "Gizli" olarak

adlandırılmalarının nedeni, dışarıya doğrudan gözlemlenebilir bir çıktıları olmamasıdır; sadece iç hesaplamalarda rol oynarlar (LeCun vd., 2015).

Momentum sabiti, yapay sinir ağları gibi optimizasyon algoritmalarında kullanılan, öğrenme sürecini hızlandıran ve yerel minimumlardan kaçınmayı kolaylaştıran bir parametredir. Momentum, önceki adımların yönünü dikkate alarak ağırlık güncellemelerini daha dengeli ve hızlı hale getirir (Rumelhart vd., 1986).

Öğrenme oranı, makine öğrenimi ve derin öğrenme modellerinin eğitiminde, ağırlıkların güncellenme hızını belirleyen önemli bir hiper parametredir. Eğitim sırasında, her bir ağırlık, modelin hata fonksiyonunu minimize etmek amacıyla küçük adımlarla güncellenir. Öğrenme oranı, bu adımların büyüklüğünü belirler (Bengio, 2009).

Tablo 5: MLP-ROE için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Gizli Katman	1, 2, ...,10	3
Momentum Sabiti	0.1, 0.2, 0.3	0.2
Öğrenme Oranı	0.1, 0.2, 0.3	0.1

Tablo 5, Çok Katmanlı Algılayıcı (MLP) modeli kullanılarak ROE değişkeni için belirlenen parametre ayarlarını göstermektedir. Modelin performansını artırmak ve en uygun tahmin doğruluğunu sağlamak amacıyla farklı parametre seviyeleri değerlendirilmiş ve optimum değerler seçilmiştir. Batch size, modelin eğitim sürecinde aynı anda işlediği veri miktarını belirlemekte olup, 50 ile 100 arasında iki seviye test edilmiş ve 100 olarak seçilmiştir. Gizli katman sayısı, ağırlık öğrenme kapasitesini belirleyen kritik bir parametre olup, 1 ile 10 arasında değerlendirilmiş ve optimum değer 3 olarak belirlenmiştir. Momentum sabiti, ağırlık güncellemelerinde önceki adımların etkisini koruyarak öğrenme sürecini stabilize eden bir faktördür ve 0.1, 0.2 ve 0.3 seviyelerinde test edilerek 0.2 olarak seçilmiştir. Öğrenme oranı, modelin ağırlık güncellemelerini ne hızla gerçekleştireceğini belirleyen bir parametre olup, 0.1, 0.2 ve 0.3 seviyelerinde incelenmiş ve en uygun değer olarak 0.1 belirlenmiştir. Bu parametre seçimleri, modelin öğrenme kapasitesini optimize etmek ve aşırı öğrenmeyi önleyerek daha dengeli bir tahmin performansı elde etmek amacıyla yapılmıştır.

Tablo 6: MLP-ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100

Gizli Katman	1, 2, ...,10	7
Momentum Sabiti	0.1, 0.01, 0.001	0.001
Öğrenme Oranı	0.1, 0.2, 0.3	0.1

Tablo 6, MLP (Multilayer Perceptron) modelinin ROA için parametre ayarlarını ve bu parametreler için seçilen seviyeleri göstermektedir. Modelin optimizasyonu amacıyla, Batch Size parametresi için 50 ile 100 arasında iki seviye test edilmiş ve 100 değeri seçilmiştir. Bu, modelin daha geniş veri kümeleri ile daha etkili bir şekilde çalışmasını ve daha hızlı öğrenme süreci yaşamasını sağlamaktadır. Gizli katman parametresi, 1 ile 10 arasında değişen seviyelerde değerlendirilmiş ve 7 katmanlı bir yapı tercih edilmiştir. Bu seçim, modelin daha karmaşık yapıları ve ilişkileri öğrenebilmesi için yeterli derinliğe sahip olmasını sağlamaktadır. Momentum sabiti parametresi, 0.1 ile 0.001 arasında test edilen değerlerden 0.001 seçilmiştir. Bu düşük momentum değeri, ağırlık güncellemelerinin daha dikkatli ve istikrarlı bir şekilde yapılmasına olanak tanımıştır. Öğrenme oranı parametresi için 0.1, 0.2 ve 0.3 seviyeleri arasında 0.1 değeri seçilmiştir. Bu, modelin öğrenme sürecinde daha kontrollü bir hızla ilerlemesini sağlayarak, eğitim sürecinde daha doğru ve güvenilir sonuçlar elde edilmesini amaçlamaktadır. Bu parametre ayarları, MLP modelinin ROA tahminlerinde yüksek doğruluk ve verimli bir öğrenme süreci sağlamak için optimize edilmiştir.

Uygulama sonucunda MLP Modeli için elde edilen performans değerlendirme kriterleri Tablo 7’de sunulmuştur.

Tablo 7: MLP için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.069	0.113	54.501	54.483	0.013	0.756
ROA	0.809	2.641	90.450	72.507	6.977	0.653

Tablo 7, MLP (Multilayer Perceptron) modelinin ROE ve ROA için elde ettiği sonuçları sunmaktadır. ROE için R² değeri 0.756, ROA için ise 0.653 olarak hesaplanmıştır. Bu, modelin ROE'yi ROA'ya kıyasla daha iyi açıkladığını ve daha yüksek bir doğrulukla tahmin yaptığını göstermektedir. MAE ve RMSE değerleri, ROE için sırasıyla 0.069 ve 0.113, ROA için ise 0.809 ve 2.641 olarak belirlenmiştir. Bu, ROE için daha düşük hata seviyelerinin, ROA için ise daha yüksek tahmin hatalarının olduğunu işaret etmektedir. RAE ve RRSE metrikleri de benzer şekilde, ROE'nin tahmininde daha düşük hata oranları gösterirken, ROA için bu oranların daha yüksek olduğu gözlemlenmektedir. MSE değerleri, ROE için 0.013, ROA için ise 6.977 olarak hesaplanmış olup, ROA'nın tahmin hatalarının çok daha büyük olduğunu ve modelin ROE üzerinde daha başarılı olduğunu ortaya koymaktadır. Bu bulgular, MLP modelinin ROE için

daha doğru tahminler sağladığını, ancak ROA için daha yüksek hata paylarıyla çalıştığını göstermektedir.

3.2.2. Linear Regression Modeli

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 8’de gösterildiği şekilde belirlenmiştir.

Ridge parametresi, modelin genelleme gücünü artırmak ve aşırı uyum riskini azaltmak için kullanılan önemli bir parametredir. Alpha değeri, modelin ne kadar regülerize edileceğini belirler; yani katsayıların ne kadar küçültüleceğini ve modelin ne kadar basitleştirileceğini etkiler. Ridge parametresi büyüdükçe, modelin katsayıları küçülür, yani model daha basitleşir ve daha az aşırı uyum yapar (Tibshirani, 1996).

Tablo 8: Linear Regresyon-ROE ve ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Ridge	5, 10, ...,100	50

Tablo 8, Linear Regression modeli kullanılarak ROE ve ROA değişkeni için belirlenen parametre ayarlarını göstermektedir. Modelin performansını optimize edebilmek amacıyla farklı parametre seviyeleri değerlendirilmiş ve en uygun değerler seçilmiştir. Batch size, modelin eğitim sürecinde kullanılan veri alt kümelerinin büyüklüğünü ifade etmekte olup, 50 ile 100 arasında iki seviye test edilmiş ve optimum değer olarak 100 belirlenmiştir. Ridge parametresi, modelin aşırı öğrenme riskini azaltmak ve genelleştirme yeteneğini artırmak amacıyla 5 ile 50 arasında değerlendirilmiş ve en uygun değer 50 olarak seçilmiştir.

Uygulama sonucunda Linear Regresyon Modeli için elde edilen performans değerlendirme kriterleri Tablo 9’da sunulmuştur.

Tablo 9: Linear Regresyon için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.063	0.107	49.310	51.813	0.011	0.749
ROA	1.205	6.407	134.731	175.871	41.042	0.003

Tablo 9’da Doğrusal Regresyon (Linear Regression) modeli kullanılarak elde edilen tahmin performans sonuçları sunulmaktadır. ROE değişkeni için modelin performansı yüksek olup, MAE 0.063, RMSE 0.107 ve MSE 0.011 olarak hesaplanmıştır. Ayrıca, R² değeri 0.749 ile modelin ROE değişkenindeki varyansın yaklaşık %75’ini açıkladığını göstermektedir. Buna

karşılık, ROA değişkeni için modelin tahmin performansı oldukça zayıftır. MAE 1.205 ve RMSE 6.407 gibi yüksek hata değerlerine ek olarak, R^2 değeri yalnızca 0.003 olup modelin ROA değişkenini neredeyse hiç açıklayamadığını ortaya koymaktadır. Bu sonuçlar, Doğrusal Regresyon modelinin ROE için iyi bir tahmin başarısı gösterdiğini, ancak ROA değişkeni için yetersiz kaldığını göstermektedir.

3.2.3. SMOreg Modeli

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 10'da gösterildiği gibi belirlenen parametreler kullanılmıştır.

Kernel, doğrusal olmayan verileri daha yüksek boyutlu bir uzaya taşımak için kullanılan fonksiyonlardır ve genellikle DVM gibi algoritmalarda doğrusal olmayan verileri ayırtmada önemli bir rol oynar. Polynomial Kernel, doğrusal olmayan ayrımları modellemek için kullanılan bir çekirdek fonksiyonudur. Bu kernel, özellikle veriler arasında daha karmaşık, çok terimli ilişkiler olduğunda tercih edilir. Radial Basis Function (RBF) Kernel, en yaygın kullanılan çekirdek fonksiyonlarından biridir. Veri noktaları arasındaki uzaklığa göre benzerliği hesaplar. Verilerin küme benzeri yapılara sahip olduğu ve karmaşık karar sınırlarının gerektiği durumlarda etkilidir (Cortes & Vapnik, 1995). PUK Kernel, Pearson VII fonksiyonundan türetilmiş bir çekirdek fonksiyonudur ve hem RBF hem de Polynomial Kernel'lara esnek bir alternatif sunar ve şekil parametreleri ile eğrinin genişliğini ve sivrililiğini kontrol ederek farklı yapıları öğrenmeye uygundur. Bu özelliği sayesinde, farklı veri yapılarına esnek uyum sağlar (Üstün vd., 2007). RegOptimizer, regresyon ve sınıflandırma algoritmalarında, modelin öğrenme sürecinde kullanılan optimizasyon algoritmasını ifade eder. Bu parametre, modelin doğruluğunu artırmak amacıyla kullanılan teknikleri belirler. C parametresi, DVM ve regresyon modellerinde, modelin hata yapma toleransını belirleyen bir regularizasyon parametresidir (Platt, 1999).

Tablo 10: SMOreg- ROE için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Kernel	Poly, RBF, Puk	Poly
RegOptimizer	RegSMO, RegSMolmproved	RegSMolmproved
C	1, 2, 3, ...,50	1

Tablo 10, SMOREg algoritması için ROE tahmininde kullanılan parametre ayarlarını göstermektedir. Bu tabloda, modelin çeşitli hiper parametrelerinin aralıkları belirtilmiş olup, her parametre için yapılan seçimler de belirtilmiştir. Kernel parametresi, destek vektör makinelerinin çekirdek fonksiyonunu belirler; burada üç seçenek mevcuttur: Poly (polinom), RBF (radial basis function) ve Puk, ancak en uygun seçenek olarak Poly seçilmiştir. RegOptimizer, modelin optimizasyonu için kullanılan algoritmayı belirtir ve iki seçenekten RegSMO ile RegSMOImproved arasından RegSMOImproved tercih edilmiştir. Son olarak, C parametresi, modelin düzenliliğini kontrol eden ve modelin hata toleransını belirleyen bir parametre olup, aralık 1 ile 50 arasında olup, burada seçilen değer 1'dir.

Tablo 11: SMOREg- ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Kernel	Poly, RBF, Puk	RBF
Gamma	0.1, 0.01, 0.001	0.1
RegOptimizer	RegSMO, RegSMOImproved	RegSMOImproved
C	1, 2, 3, ..., 50	35

Tablo 11’de, ROA değişkenine yönelik olarak oluşturulan SMOREg regresyon modeli için gerçekleştirilen hiper parametre ayarları özetlenmektedir. Model eğitimi sürecinde, Batch Size parametresi 50 ve 100 olmak üzere iki farklı seviyede değerlendirilmiş ve modelin daha stabil ve verimli öğrenmesini sağladığı için 100 değeri tercih edilmiştir. Tahmin performansını etkileyen çekirdek (kernel) fonksiyonları arasında Polynomial, RBF (Radial Basis Function) ve Puk seçenekleri incelenmiş, doğrusal olmayan ilişkileri daha iyi yakalayabilme potansiyeli nedeniyle RBF çekirdeği seçilmiştir. RBF için önemli bir hiper parametre olan Gamma, modelin esnekliğini belirlemekte olup 0.1, 0.01 ve 0.001 seviyelerinde test edilmiş ve en iyi sonuç 0.1 değerinde elde edilmiştir. Optimizasyon algoritması olarak RegSMO ve RegSMOImproved alternatifleri değerlendirilmiş; iyileştirilmiş versiyon olan RegSMOImproved, daha düşük hata oranları sağladığı için tercih edilmiştir. Düzenleme parametresi olan C, 1’den 50’ye kadar artan değerlerle test edilmiş ve modelin genelleme yeteneği ile hata kontrolünü dengeleyen 35 değeri seçilmiştir. Bu parametre kombinasyonu, ROA değişkenine yönelik tahminleme başarısını maksimize edecek şekilde belirlenmiştir. Uygulama sonucunda SMOREg Modeli için elde edilen performans değerlendirme kriterleri Tablo 12’de sunulmuştur.

Tablo 12: SMOreg için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.057	0.099	44.897	47.943	0.010	0.777
ROA	0.683	2.222	76.367	61.008	4.937	0.794

Tablo 12, SMOreg modelinin ROE ve ROA değişkenleri için elde ettiği sonuçları sunmaktadır. ROE için elde edilen R² değeri 0.777 iken, ROA için bu değer 0.794 olarak belirlenmiştir, bu da modelin her iki değişkeni de iyi bir şekilde açıklayabildiğini göstermektedir. MAE ve RMSE değerleri, modelin tahminlerinde ortalama mutlak hata ve kök ortalama kare hata açısından ROE için sırasıyla 0.057 ve 0.099, ROA için ise sırasıyla 0.683 ve 2.222 olarak hesaplanmıştır. Bu sonuçlar, ROE için modelin daha düşük hata seviyeleriyle daha başarılı olduğunu, ancak ROA için tahmin hatalarının daha yüksek olduğunu göstermektedir. RAE ve RRSE metrikleri de benzer şekilde, ROE'nin tahmininde daha düşük hata oranları olduğunu ortaya koymaktadır. Öte yandan, MSE değerlerinin düşük olması, modelin genellikle düşük hata ile tahminler gerçekleştirdiğini göstermektedir.

3.2.4. GaussianProcesses Modeli

Modelin ölçüm hatalarını ve gürültüyü dikkate alarak daha güvenilir tahminler yapabilmek için ek parametreler içermektedir. Gürültü seviyesi, modelin veri içindeki rastgele hatalara duyarlılığını belirleyen önemli bir faktördür. Düşük gürültü seviyesi, modelin verilere daha sıkı uyum sağlamasına neden olurken, yüksek gürültü seviyesi ise modelin aşırı duyarlılığını azaltarak genelleştirme kapasitesini artırmaktadır. Tohum parametresi ise Gaussian Process modelindeki rastgele bileşenleri kontrol ederek aynı veri seti üzerinde farklı model çıktıların oluşmasını engeller ve tekrar edilebilir sonuçlar elde edilmesini sağlar. Bu parametrelerin uygun şekilde belirlenmesi, modelin doğruluğu ile genelleştirme kabiliyeti arasında optimum bir denge kurulmasına olanak tanımaktadır (Wolff vd., 2024).

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 13'te ROE ve ROA için gösterildiği gibi aynı parametreler kullanılmıştır.

Tablo 13: GaussianProcess-ROE ve ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Kernel	Poly, RBF, Puk	RBF
Gamma	0.1, 0.01, 0.001	0.1

Gürültü Seviyesi	0.1, 0.01, ..., 0.0001	0.1
Tohum	1, 2, ..., 10	1

Tablo 13, GaussianProcess modelinin ROE ve ROA için parametre ayarlarını ve seçilen seviyeleri sunmaktadır. Modelin performansını optimize etmek amacıyla çeşitli parametreler farklı seviyelerde test edilmiştir. Batch Size parametresi 50 ve 100 olmak üzere iki farklı seviyede değerlendirilmiş ve 100 değeri seçilmiştir. Kernel parametresi için Poly, RBF ve Puk seçenekleri arasında RBF (Radial Basis Function) seçilmiştir; bu, daha iyi genelleme ve doğruluk sağlamak için yaygın olarak tercih edilen bir çekirdek fonksiyonudur. Gamma parametresi için 0.1, 0.01 ve 0.001 değerleri arasından 0.1 seçilmiştir, bu da modelin daha yüksek etki alanı ile çalışmasını ve daha güçlü bir doğruluk sağlamasını amaçlamaktadır. Gürültü Seviyesi parametresi, 0.1, 0.01, ..., 0.0001 aralığında test edilmiş ve yine 0.1 değeri belirlenmiştir; bu, modelin daha az gürültü ile çalışarak verinin daha iyi bir şekilde modellenmesini sağlamak için seçilen bir seviyedir. Tohum parametresi için 1 ile 10 arasındaki değerler arasında en iyi sonuçları veren 1 değeri seçilmiştir, bu da modelin her çalıştırıldığında tutarlı sonuçlar vermesini sağlamaktadır. Bu parametre ayarları, GaussianProcess modelinin performansını en üst düzeye çıkarmak için dikkatlice optimize edilmiştir.

Tablo 14: GaussianProcess için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.059	0.101	46.195	48.776	0.010	0.774
ROA	0.820	2.592	91.652	71.159	6.717	0.580

Tablo 14, GaussianProcess modelinin ROE ve ROA için elde ettiği performans sonuçlarını sunmaktadır. ROE için R² değeri 0.774 olarak hesaplanırken, ROA için bu değer 0.580'e düşmektedir, bu da ROE'nin model tarafından daha iyi açıklandığını göstermektedir. ROE'nin tahmin hataları MAE ve RMSE açısından sırasıyla 0.059 ve 0.101 olarak hesaplanırken, ROA için bu değerler sırasıyla 0.820 ve 2.592'dir, bu da ROA için daha büyük hata oranlarına işaret etmektedir. RAE ve RRSE metrikleri de ROE'nin tahmininde daha düşük hata oranları gösterirken, ROA için bu oranların çok daha yüksek olduğu gözlemlenmektedir. MSE değerleri, ROE için 0.010, ROA için ise 6.717 olarak belirlenmiştir, bu da ROA'nın tahmin hatalarının çok daha yüksek olduğunu ve modelin ROE üzerinde daha başarılı olduğunu göstermektedir. Genel olarak, GaussianProcess modeli, ROE üzerinde daha yüksek doğruluk ve daha düşük hata ile tahminler yaparken, ROA üzerinde daha büyük hata paylarıyla çalışmaktadır.

3.2.5. M5Rules Modeli

Minimum örnek sayısı parametresi, özellikle karar ağaçları ve regresyon ağaçları gibi bölme öğrenme algoritmalarında kullanılan önemli bir hiper parametredir. Bu parametre, bir düğümün daha fazla bölünmesine izin verilebilmesi için gereken minimum gözlem (örnek) sayısını tanımlar (Witten vd., 2005).

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 15’te gösterildiği gibi ROE ve ROA için aynı parametreler kullanılmıştır.

Tablo 15: M5Rules-ROE ve ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Minimum Örnek Sayısı	1, 2, ...,10	1

Tablo 15’te, M5Rules algoritmasının ROE ve ROA değişkenleri için gerçekleştirilen hiper parametre ayarlarına yer verilmektedir. Modelin eğitiminde kullanılan Batch Size parametresi 50 ve 100 olmak üzere iki farklı seviyede değerlendirilmiş ve 100 değeri seçilmiştir. Ayrıca, Minimum Örnek Sayısı parametresi 1 ile 10 arasında denenmiş ve modelin esnekliğini artırmak amacıyla 1 olarak belirlenmiştir. Bu seçim, modelin her yaprakta çok az sayıda veriyle dahi kurallar türetebilmesine olanak tanıyarak ayrıntılı öğrenmeyi teşvik eder.

Tablo 16: M5Rules için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.054	0.104	42.613	50.080	0.011	0.767
ROA	1.242	6.368	138.828	174.799	40.539	0.004

Tablo 16’da M5Rules algoritması kullanılarak ROE ve ROA değişkenleri için elde edilen regresyon sonuçları özetlenmektedir. ROE değişkeni için modelin tahmin performansı oldukça güçlüdür; ortalama mutlak hata (MAE) 0.054, karekök ortalama hata (RMSE) 0.104 ve MSE 0.011 olarak gerçekleşmiştir. Özellikle R² katsayısının 0.767 olması, modelin ROE değişkenindeki varyansın yaklaşık %77’sini açıklayabildiğini göstermekte ve yüksek bir açıklayıcılık düzeyine işaret etmektedir. Buna karşın, ROA değişkeni için modelin performansı belirgin biçimde düşüktür; MAE 1.242, RMSE 6.368 ve R² değeri yalnızca 0.004 olarak hesaplanmıştır. Bu düşük R² değeri, modelin ROA değişkeni üzerindeki açıklayıcılığının neredeyse yok denecek kadar az olduğunu ve genelleme kabiliyetinin zayıf kaldığını ortaya koymaktadır.

3.2.6. Alternating Model Trees Modeli

Alternating Model Trees (AMT) algoritmasında iterasyon sayısı ve daralma parametreleri, modelin öğrenme sürecini kontrol eder. İterasyon sayısı, modelin eğitim sırasında kaç adımda güncelleneceğini belirler ve daha fazla iterasyon, modelin daha fazla öğrenmesini sağlar. Daralma ise her iterasyondaki öğrenme hızını kontrol eder, küçük bir değer seçildiğinde model daha yavaş öğrenir, bu da aşırı uyum riskini azaltarak daha genelleştirilmiş tahminler elde edilmesine yardımcı olur (Friedman, 2001). Bu parametrelerin dikkatli bir şekilde ayarlanması, modelin doğruluğunu ve genelleme yeteneğini artırabilir.

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 17 ve 18’de gösterildiği gibi belirlenen parametreler kullanılmıştır.

Tablo 17: AMT-ROE için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
İterasyon Sayısı	10, 20, ...,200	170
Daralma	0.1, 0.01, 0.001	0.01

Tablo 17’de, Additive Regression (AMT) algoritmasının ROE değişkeni için uygulanan parametre ayarları sunulmaktadır. Modelin eğitimi sırasında kullanılan Batch Size parametresi 50 ve 100 olmak üzere iki farklı seviyede değerlendirilmiş ve işlem verimliliği ile model kararlılığı göz önüne alınarak 100 değeri seçilmiştir. Bu parametre, algoritmanın her yinelemede işlediği örnek sayısını belirleyerek eğitim sürecinin hızını ve bellek kullanımını etkiler. İterasyon sayısı, yani modelin ardışık öğrenme adımlarının sayısı, 10 ile 200 arasında denenmiş ve optimum öğrenme düzeyine ulaşmak amacıyla 170 iterasyon kullanılmıştır. Bu değer, modelin hata oranlarını kademeli olarak azaltmasına olanak tanır. Daralma (shrinkage veya learning rate) parametresi ise 0.1 ile 0.001 aralığında test edilmesine karşın, daha hassas öğrenme sağlayabilmesi amacıyla düşük bir değer olan 0.01 tercih edilmiştir. Düşük daralma oranı, her bir adımda öğrenme etkisini sınırlandırarak aşırı öğrenme riskini azaltır ve daha istikrarlı sonuçlar üretir. Bu parametre kombinasyonu, modelin doğruluk ve genelleme kapasitesini dengeli biçimde geliştirmeyi hedeflemektedir.

Tablo 18: AMT-ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100

İterasyon Sayısı	10, 20, ...,200	150
Daralma	0.1, 0.01, 0.001	0.001

Tablo 18’de, Additive Regression (AMT) algoritmasının ROA değişkeni için uygulanan parametre ayarları detaylandırılmaktadır. Modelin eğitim sürecinde kullanılan batch size, işlem verimliliği ve bellek yönetimi açısından 100 olarak belirlenmiştir. İterasyon sayısı, yani modelin ardışık öğrenme adımlarının sayısı, 10 ile 200 arasında değerlendirilmiş ve optimum sonuç için 150 iterasyon tercih edilmiştir. Bu değer, modelin yeterli düzeyde öğrenmesine olanak sağlarken, aşırı öğrenme riskini sınırlamayı amaçlamaktadır. Daralma katsayısı (shrinkage/learning rate) ise modelin her adımda yaptığı öğrenmenin etkisini kontrol eder; burada 0.1’den 0.001’e kadar aralık test edilmiş ve daha yavaş ama istikrarlı bir öğrenme süreci sağlamak amacıyla 0.001 gibi düşük bir değer seçilmiştir. Bu seçim, modelin küçük adımlarla öğrenmesini sağlayarak genelleme kapasitesini artırmayı hedeflemektedir. Bu parametre kombinasyonu, ROA değişkeni için modelin uyum yeteneği ile genelleme gücü arasında dengeli bir yapı kurmayı amaçlamaktadır.

Tablo 19: AMT için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.057	0.107	44.960	54.858	0.011	0.800
ROA	0.503	2.045	56.281	51.744	41.833	0.703

Tablo 19’da Additive Regression (AMT) algoritması kullanılarak ROE ve değişkenleri için elde edilen regresyon performans metrikleri sunulmaktadır. ROE değişkeni açısından model iyi performans göstermiştir; ortalama mutlak hata (MAE) 0.057 ve karekök ortalama hata (RMSE) 0.107 olarak ölçülmüştür. Modelin açıklayıcılık gücünü gösteren R² katsayısı ise 0.800 olup, ROE değişkenindeki varyansın yaklaşık %80’inin model tarafından açıklandığını göstermektedir. Bu değer, yüksek düzeyde bir tahmin doğruluğuna işaret eder. ROA değişkeni için MAE 0.503, RMSE 2.045 ve MSE 41.833 olarak gerçekleşmiş, R² değeri 0.703 ile önemli düzeyde açıklayıcılık sağlanmıştır. Özellikle ROA değişkeni özelinde diğer algoritmalarla karşılaştırıldığında AMT’nin bu yüksek R² değeriyle öne çıktığı görülmektedir. Sonuçlar genel olarak, AMT algoritmasının her iki bağımlı değişken üzerinde de yüksek tahmin başarısı gösterdiğini ve genelleme yeteneğinin güçlü olduğunu ortaya koymaktadır.

3.2.7. PaceRegression Modeli

Bu model özellikle çoklu doğrusal regresyon modellerinde, değişken seçimi ve parametre tahmini süreçlerini optimize etmek için kullanılan bir yöntemdir. Bu yöntemde eşik

değeri, modelin hangi katsayıları sıfıra indirerek sadeleştirme yapacağına karar veren kritik bir hiper parametredir.

Eşik değeri, gürültüyü filtrelemek, aşırı öğrenmeyi önlemek ve modele daha iyi genelleme kabiliyeti kazandırmak amacıyla kullanılır. Eğer bir katsayının büyüklüğü belirlenen eşik değerinin altındaysa, model bu katsayıyı önemsiz olarak değerlendirip sıfıra eşitleyebilir. Böylece, yalnızca anlamlı değişkenlerin modele dahil edilmesi sağlanır (Wang, 2000).

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 20’de gösterildiği gibi aynı parametreler kullanılmıştır.

Tablo 20: PaceRegression-ROE ve ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Eşik Değer	1, 2, ...,10	2

Tablo 20, Pace Regression modeli için ROE ve ROA bağımlı değişkenlerine yönelik kullanılan parametre ayarlarını içermektedir. Modelin performansını optimize etmek amacıyla, Batch Size 100 olarak belirlenmiştir. Ayrıca, Pace Regression yönteminde önemli bir hiper parametre olan eşik değeri 1 ile 10 arasında değişen bir aralıkta değerlendirilmiş ve model için optimal değer olarak 2 seçilmiştir. Eşik değeri, modelin katsayı tahmin sürecinde belirli değişkenleri göz ardı etme veya dahil etme kriterini belirleyerek aşırı uyumun önlenmesine katkıda bulunabilir. Bu parametre ayarları, modelin genelleştirme kabiliyetini artırmak amacıyla optimize edilmiştir.

Tablo 21: PaceRegression için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.057	0.099	44.838	47.781	0.010	0.781
ROA	1.241	7.048	138.821	193.480	496.743	0.001

Tablo 21’de, Pace Regression algoritmasının ROE ve ROA değişkenleri üzerindeki regresyon performansı değerlendirilmektedir. ROE değişkeni için elde edilen sonuçlar modelin güçlü bir tahmin performansına sahip olduğunu göstermektedir; ortalama mutlak hata (MAE) 0.057, karekök ortalama hata (RMSE) 0.099 ve R² değeri 0.781 olarak hesaplanmıştır. Bu durum, modelin ROE değişkenindeki varyansın yaklaşık %78’ini başarıyla açıklayabildiğini ve iyi bir genelleme yeteneğine sahip olduğunu göstermektedir. Ancak ROA değişkeni için modelin başarısı oldukça sınırlıdır. MAE 1.241 ve RMSE 7.048 gibi yüksek hata değerleri ile, MSE'nin 496.743 gibi çok yüksek bir seviyede olması ve R² değerinin yalnızca 0.001 olması,

modelin ROA değişkeni üzerinde neredeyse hiçbir açıklayıcılık sağlayamadığını ortaya koymaktadır. Sonuç olarak, Pace Regression modeli ROE için yüksek doğrulukta tahminler üretebilirken, ROA üzerinde yeterli performansı gösterememektedir.

3.2.8. Isotonic Regression Modeli

İsotonic Regresyon algoritmasında, “hata ayıklama” ve “hata ayıklama yetenekleri kontrol etmeyin” parametreleri, modelin eğitim ve analiz süreçlerini etkileyen önemli ayarlardır.

Hata ayıklama = yanlış: Bu ayar, modelin eğitim ve tahmin süreçleri hakkında yalnızca temel çıktılar verir ve algoritmanın iç işleyişi hakkında ayrıntılı bilgi sağlamaz. Bu, kullanıcıların algoritmanın nasıl çalıştığını ayrıntılı olarak incelemek istemediği durumlar için uygundur ve genellikle daha temiz bir çıktı sağlar.

Hata ayıklama yetenekleri kontrol etmeyin = yanlış: Bu ayar, modelin veri kümesinin uygunluğunu kontrol etmesini sağlar. Yani, model veri setinin gereksinimlerini kontrol eder ve veri setinde herhangi bir uyumsuzluk tespit ederse, kullanıcıyı uyarır. Bu, modelin doğru çalıştığından emin olmak için önemlidir ve veri setindeki potansiyel hataları önceden tespit etmeye yardımcı olur.

Bu iki parametre, varsayılan olarak ayarlanmış durumdadır. İlgili parametreler seçildikten sonra gerçekleştirilen analiz sonuçları, Tablo 22’de sunulmuştur.

Tablo 22: Isotonic Regression için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.066	0.119	52.293	57.262	0.014	0.678
ROA	0.922	3.331	103.089	91.435	11.097	0.168

Tablo 22’de, Isotonic Regression algoritmasının ROE ve ROA değişkenleri üzerindeki tahmin performansına ilişkin sonuçlar sunulmaktadır. ROE değişkeni için modelin doğruluk düzeyi, diğer algoritmalara kıyasla daha düşüktür; ortalama mutlak hata (MAE) 0.066 ve RMSE 0.119 olarak hesaplanmış, açıklayıcılık gücünü ifade eden R² değeri ise 0.678 olarak gerçekleşmiştir. Bu sonuç, modelin ROE değişkenindeki varyansın yaklaşık %68’ini açıklayabildiğini göstermekte, orta düzeyde bir tahmin başarısını işaret etmektedir. ROA değişkeni için modelin performansı da benzer şekilde sınırlı kalmıştır; MAE 0.922, RMSE 3.331 ve R² değeri 0.168 olup, bu değişkendeki varyansın yalnızca yaklaşık %17’si açıklanabilmiştir. Genel olarak, Isotonic Regression algoritması her iki değişken için de kısıtlı bir tahmin kapasitesi sergilemiş ve özellikle ROA değişkeni üzerinde düşük açıklayıcılık

düzeyiyle dikkat çekmiştir. Bu bulgular, modelin doğrusal olmayan yapılarla sınırlı ölçüde baş edebildiğini göstermektedir.

3.2.9. RepTree Modeli

Modelin performansını ve genelleme yeteneğini etkileyen çeşitli hiper parametreler bulunmaktadır. Bu parametrelerden biri olan maksimum derinlik, ağacın en fazla hangi derinliğe kadar büyüebileceğini sınırlandırarak aşırı öğrenmeyi önlemeyi amaçlarken; minimum örnek sayısı, bir düğümün bölünebilmesi için gerekli en az gözlem sayısını tanımlar ve bu değerin düşük tutulması daha ayrıntılı ağaç yapılarının oluşmasına neden olabilir. Minimum varyans oranı parametresi ise regresyon problemlerinde, bir düğümün bölünebilmesi için gereken varyans oranını belirler. Ayrıca, kat sayısı, budama işlemi sırasında kullanılan k-katlı çapraz doğrulama sayısını ifade eder; bu parametrenin artırılması modelin daha istikrarlı hale gelmesini sağlarken işlem süresini de uzatabilir. Başlangıç sayısı, modelin öğrenmeye başlaması için gerekli minimum gözlem sayısını kontrol ederken; tohum parametresi, rastgelelik içeren işlemlerde tekrarlanabilir sonuçlar elde edilmesini sağlar. Bu hiper parametrelerin uygun şekilde ayarlanması, modelin hem öğrenme hem de genelleme performansını optimize etmek açısından kritik öneme sahiptir (Witten vd., 2005).

Analiz uygulamasında, algoritmanın varsayılan parametre ayarları çeşitli denemeler sonucunda optimize edilerek Tablo 23'te gösterildiği gibi aynı parametreler kullanılmıştır.

Tablo 23: RepTree-ROE ve ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Başlangıç Sayısı	1, 2, ...,10	1
Maks. Derinlik	1, 2, ...,20	13
Min. Örnek Sayısı	1, 2, ...,50	1
Min. Varyans Oranı	0.1, 0.01, ...,0.0001	0.0001
Kat Sayısı	1, 2, ...,20	13
Tohum	1, 2, ...,20	12

Tablo 23'te, RepTree algoritmasının ROE ve ROA değişkenleri için uygulanan hiper parametre ayarları detaylı biçimde sunulmaktadır. Model eğitimi sürecinde Batch Size parametresi 50 ve 100 olmak üzere iki farklı seviyede değerlendirilmiş ve 100 tercih edilmiştir. Başlangıç sayısı, modelin eğitim sürecindeki ilk çözümleme denemelerinin sayısını belirler ve 1 olarak belirlenmiştir. Modelin genel yapısını kontrol eden maksimum derinlik, yani karar

ağacının büyüme sınırı, 1 ile 20 arasında test edilerek optimum dengeyi sağlayan 13 değeri seçilmiştir. Minimum örnek sayısı, bir yaprak düğümde bulunması gereken asgari gözlem sayısını ifade eder ve modelin genelleme yeteneğini artırmak amacıyla 1 olarak belirlenmiştir. Minimum varyans oranı, dal oluşturulmadan önce düğümdeki varyansın ne kadar azaltılması gerektiğini belirleyen kritik bir parametredir; burada oldukça hassas ayrımlar yapılabilmesi için 0.0001 gibi düşük bir eşik kullanılmıştır. Ayrıca, kat sayısı, yani tekrar edilen modelleme adımlarının sayısı 13 olarak seçilerek modelin istikrarı artırılmış; tohum parametresi ise rastlantısal bileşenlerin kontrolü amacıyla 12 olarak belirlenmiştir. Bu parametrik yapı, RepTree algoritmasının hem esneklik hem de doğruluk dengesini sağlamaya yönelik olarak optimize edilmiştir.

Tablo 24: RepTree için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.053	0.100	42.035	48.486	0.010	0.775
ROA	0.845	3.241	94.476	88.983	10.509	0.211

Tablo 24’te sunulan sonuçlar, RepTree algoritmasının ROE ve ROA değişkenleri üzerindeki tahmin performansını ortaya koymaktadır. ROE değişkeni açısından modelin performansı iyi düzeydedir; ortalama mutlak hata (MAE) 0.053, karekök ortalama hata (RMSE) 0.100 ve açıklayıcılık oranını gösteren R² değeri 0.775 olarak elde edilmiştir. Bu sonuçlar, modelin ROE’deki varyansın yaklaşık %78’ini açıklayabildiğini ve dengeli bir tahmin başarısı sunduğunu göstermektedir. ROA değişkeni için ise modelin performansı daha zayıf kalmıştır; MAE 0.845, RMSE 3.241 ve R² 0.211 olup, bu değişkendeki varyansın yalnızca %21’lik kısmı model tarafından açıklanabilmektedir. Özellikle RAE ve RRSE değerlerinin yüksek olması, modelin göreceli hata düzeyinin arttığını göstermektedir. RepTree algoritması ROE tahminlerinde başarılı olurken, ROA üzerinde düşük açıklayıcılık gücüyle sınırlı bir performans sergilemiştir.

3.2.10. Elastic Net Modeli

Elastic Net algoritması, Lasso (L1) ve Ridge (L2) regresyonlarının birleşimini kullanarak modelin genelleme yeteneğini artırmayı amaçlar. Bu algoritmanın başarısını doğrudan etkileyen bazı önemli parametreler vardır. Alpha, Lasso ve Ridge arasındaki dengeyi belirler; Alpha = 1 Lasso’yu, Alpha = 0 ise Ridge’i ifade eder. Düşük bir alpha değeri, daha güçlü Ridge regularizasyonuna yol açarken, yüksek bir değer Lasso regularizasyonunun etkisini artırır (Zou & Hastie, 2005). Epsilon, modelin yakınsama eşiğini belirler ve çözümleme

sürecinde modelin doğruluğunu ayarlamak için kullanılır. Daha küçük epsilon değerleri daha hassas bir model öğrenmesine imkân tanırken, daha büyük değerler öğrenme sürecini hızlandırır (Hastie vd., 2009). Maksimum tekrarlar, algoritmanın çalışma süresi boyunca gerçekleştirdiği iterasyon sayısını belirler. Daha fazla iterasyon, modelin daha doğru sonuçlar vermesini sağlayabilir ancak aşırı uzun iterasyonlar zaman alıcı olabilir. Eşik parametresi, modelin tahmin ettiği değerlerin sıfırdan küçük olup olmayacağını belirler ve bu sayede modelin daha hassas hale gelmesini sağlar ve sayısal model parametresi, model seçimi sürecinde çeşitli konfigürasyonların denenmesini sağlayan bir hiper parametredir. (Zou & Hastie, 2005). Sayısal katmanlar, veri kümesinin kaç alt gruba (fold) bölüneceğini belirleyen hiper parametredir. K-katlı çapraz doğrulamada (k-fold cross-validation), model k sayıda alt kümede eğitilir ve test edilir. Her iterasyonda, bu alt kümelerden biri doğrulama için ayrılırken geri kalanı eğitim için kullanılır. Bu süreç, modelin genelleme kabiliyetini artırır ve overfitting riskini azaltır (Kohavi, 1995).

Tablo 25: Elastic Net-ROE için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Batch Size	50, 100	100
Alpha	0.1, 0.01, 0.001	0.1
Epsilon	0.1, 0.01, ...,0.000001	0.000001
Maksimum Tekrarlar	100, 200, ...,1000	1000
Eşik	0.1, 0.01, ...,0.000001	0.000001
Sayısal Katmanlar	1, 2, ...,50	10
Sayısal Modeller	1, 2, ...,50	30

Tablo 25'te yer alan parametre ayarları, ROE değişkeni için Elastic Net algoritmasının en verimli şekilde çalışmasını sağlayacak şekilde optimize edilmiştir. Batch Size 100 olarak belirlenerek her iterasyonda işlenecek veri sayısı sınırlandırılmış, bu da daha hızlı ve istikrarlı bir öğrenme süreci sağlamıştır. Alpha değeri 0.1 seçilerek Lasso regularizasyonunun baskın olması sağlanmış ve böylece modelde önemli olmayan değişkenlerin elenmesi hedeflenmiştir. Epsilon parametresinin 0.000001 gibi düşük bir seviyede belirlenmesi, modelin daha hassas bir çözümle yakınsama gerçekleştirmesine olanak tanımıştır. Maksimum iterasyon sayısı 1000 olarak belirlenmiş, bu da modelin yeterli sayıda güncelleme yapmasına imkân vererek öğrenme sürecini dengelemektedir. Eşik değeri 0.000001 ile oldukça düşük seviyede tutulmuş, böylece tahmin edilen değerlerin duyarlılığı artırılmıştır. Ayrıca, 10 sayısal katman ve 30 model yapılandırması kullanılarak modelin yapısal karmaşıklığı ve çeşitliliği artırılmış, en uygun

modelin seçilmesine zemin hazırlanmıştır. Bu parametre ayarları genel olarak hem modelin doğruluğunu hem de genelleme gücünü artırmaya yönelik ideal bir yapı sunmaktadır.

Tablo 26: Elastic Net-ROA için Parametre Ayarları

Parametre	Kullanılan Seviye Aralıkları	Seçilen Seviye
Yığın Boyutu	100, 200, ...,2000	100
Alpha	0.1, 0.01, 0.001	0.1
Epsilon	0.1, 0.01, 0.001	0.1
Maksimum Tekrarlar	100, 1000, ...,10000	1000
Eşik	0.1, 0.01, 0.001	0.01
Sayısal Katmanlar	1, 2, ...,20	10
Sayısal Modeller	1, 2, ...,50	50

Tablo 26’da Elastic Net algoritmasının ROA değişkeni için uygulanan parametre ayarları ayrıntılı şekilde sunulmaktadır. Modelin eğitim sürecinde veri işlem etkinliğini belirleyen batch size 100 olarak seçilmiştir. Regülerizasyonun dengeleyici bileşenlerinden biri olan alpha değeri, modelin L1 ve L2 normlarını birleştirme derecesini belirlerken; epsilon ise doğruluk eşliğini temsil etmekte ve her ikisi de 0.1 olarak belirlenerek modelin hem düzenliliği hem de hassasiyeti sağlanmıştır. Maksimum tekrar sayısı, modelin öğrenme sürecindeki iterasyon sınırını gösterir ve 1000 olarak belirlenmiştir. Eşik değeri ise katsayıların güncellenmesindeki toleransı kontrol eder ve 0.01 olarak optimize edilmiştir. Ayrıca, modelin yapısal karmaşıklığını ifade eden sayısal katman sayısı 10, birleştirilen sayısal modellerin sayısı ise 50 olarak belirlenmiştir.

Tablo 27: Elastic Net için Sonuçlar

Bağımlı Değişken	MAE	RMSE	RAE	RRSE	MSE	R ²
ROE	0.060	0.104	47.377	50.183	0.011	0.762
ROA	0.892	3.674	99.713	100.866	13.501	0.005

Tablo 27’de, Elastic Net regresyon modeli ile elde edilen ROE ve ROA tahmin performans sonuçları yer almaktadır. ROE değişkeni için modelin ortalama mutlak hata (MAE) değeri 0.060, kök ortalama kare hata (RMSE) değeri ise 0.104 olarak hesaplanmıştır. Bu değerler, modelin tahmin hatalarının düşük olduğunu ve ROE değişkenini iyi açıkladığını göstermektedir. Ayrıca, RAE 47.377 ve RRSE 50.183 oranlarının düşük olması başarılı olduğunu göstermektedir. MSE (0.011) ve R² (0.762) değerleri de modelin ROE üzerindeki açıklayıcılığının yüksek olduğunu desteklemektedir. Öte yandan ROA değişkeni için modelin performansı zayıftır. MAE 0.892, RMSE 3.674 ve MSE 13.501 olarak bulunmuştur. Özellikle

RAE ve RRSE değerlerinin yüksek olması, modelin görece hata düzeyinin arttığını göstermektedir. R^2 değerinin 0.005 gibi oldukça düşük bir seviyede kalması, Elastic Net modelinin ROA değişkenini açıklamada yetersiz kaldığını açıkça ortaya koymaktadır. Bu durum, modelin ROE tahmininde başarılı olmasına rağmen ROA için aynı etkiyi göstermediğini göstermektedir.

3.3. MAKİNE ÖĞRENMEŞİ TABANLI MODELLERİN TAHMİN BAŞARIMININ KARŞILAŞTIRILMASI

Çalışmada kullanılan farklı makine öğrenmesi algoritmalarının finansal performans göstergeleri olan ROE (Özsermaye Karlılığı) ve ROA (Aktif Karlılığı) üzerindeki tahmin başarılarını değerlendirmek amacıyla karşılaştırılmıştır. Bu bölümde, her bir modelin hata metrikleri (MAE, RMSE, RAE, RRSE, MSE) ile doğruluk ölçütü olan determinasyon katsayısı (R^2) dikkate alınarak karşılaştırmalı bir analiz yapılmıştır. Böylece, hangi modelin daha düşük hata oranlarıyla ve daha yüksek açıklayıcılarla en başarılı sonuçları verdiği belirlenmiş, modellerin güçlü ve zayıf yönleri ortaya konmuştur.

Tablo 28: ROE için Model Sıralama

Model	MAE	RMSE	RAE	RRSE	MSE	R^2
ATM	0.057	0.107	44.960	54.858	0.011	0.800
SMOreg	0.057	0.099	44.897	47.943	0.010	0.777
Pace	0.057	0.099	44.838	47.781	0.010	0.781
Reptree	0.053	0.100	42.035	48.486	0.010	0.775
Gaussian	0.059	0.101	46.195	48.776	0.010	0.774
ElasticNet	0.060	0.104	47.377	50.183	0.011	0.762
M5rules	0.054	0.104	42.613	50.080	0.011	0.767
Linear	0.063	0.107	49.310	51.813	0.011	0.749
MLP	0.069	0.113	54.501	54.483	0.013	0.756
Isotonic	0.066	0.119	52.293	57.262	0.014	0.678

Tablo 28, ROE için farklı modellerin performansını çeşitli hata metrikleri ile değerlendiren bir sıralamayı sunmaktadır. En yüksek doğruluğa sahip model ATM olup, MAE (0.057), RMSE (0.107) ve R^2 (0.800) değerleri ile en iyi sonuçları elde etmiştir. Bu, ATM modelinin en düşük hata oranlarına sahip olduğunu ve ROE'nin tahmininde oldukça güçlü bir performans sergilediğini göstermektedir. SMOreg ve Pace modelleri, sırasıyla 0.057 ve 0.057 MAE değerleri ile ATM'ye yakın doğruluk sunmuş, ancak R^2 değerleri sırasıyla 0.777 ve 0.781 ile ATM'nin gerisinde kalmıştır. Reptree, Gaussian ve M5rules modelleri de iyi sonuçlar elde

etmiş, ancak R^2 değerleri sırasıyla 0.775, 0.774 ve 0.767 ile daha düşük kalmıştır. ElasticNet ve Linear modelleri ise daha yüksek MAE ve RMSE değerlerine sahip olup, sırasıyla 0.762 ve 0.749 R^2 değerleri ile en düşük doğruluğa sahip modeller arasında yer almaktadır. MLP modeli ise 0.069 MAE ve 0.113 RMSE değerleriyle daha yüksek hata oranlarına sahip olup, R^2 değeri 0.756 ile orta sıralarda yer almaktadır. Son olarak, Isotonic modeli, en yüksek hata oranlarına sahip olan ve en düşük R^2 (0.678) değerine ulaşan modeldir. Bu sıralamalar, ATM modelinin ROE tahminlerinde en başarılı model olduğunu, diğer modellerin ise değişen doğruluk seviyelerinde performans gösterdiğini ortaya koymaktadır.

Tablo 29: ROA için Model Sıralaması

Model	MAE	RMSE	RAE	RRSE	MSE	R^2
SMOreg	0.683	2.222	76.367	61.008	4.937	0.794
ATM	0.503	2.045	56.281	56.147	41.833	0.703
MLP	0.809	2.641	90.450	72.507	6.977	0.653
Gaussian	0.820	2.592	91.652	71.159	6.717	0.580
Reptree	0.845	3.241	94.476	88.983	10.509	0.211
Isotonic	0.922	3.331	103.089	91.35	11.097	0.168
ElasticNet	0.892	3.674	99.713	100.866	13.501	0.005
M5rules	1.242	6.368	138.828	174.799	40.539	0.004
Linear	1.205	6.407	134.731	175.871	41.042	0.003
Pace	1.241	7.048	138.821	193.480	496.743	0.001

ROA tahmini için uygulanan çeşitli regresyon modellerinin karşılaştırmalı analizi, en yüksek başarıyı SMOreg modelinin sağladığını göstermektedir. SMOreg modeli, 0.683 MAE ve 2.222 RMSE gibi nispeten düşük hata değerlerine sahip olup, 0.794 R^2 değeri ile ROA değişkenini açıklama gücü açısından en yüksek performansı sergilemiştir. Hata oranı daha düşük olmasına rağmen, açıklayıcılığı SMOreg'e göre daha sınırlı olan ATM modeli ise 0.503 MAE, 2.045 RMSE ve 0.703 R^2 değeri ile ikinci sırada yer almıştır. Orta düzeyde başarı gösteren modeller arasında MLP ($R^2 = 0.653$) ve Gaussian ($R^2 = 0.580$) yer almakta olup, görece daha düşük açıklayıcılık katsayıları ile sınırlı bir tahmin performansı sunmuşlardır. Buna karşın Reptree, ElasticNet ve Isotonic gibi modeller hem hata oranları hem de açıklayıcılık güçleri bakımından zayıf sonuçlar vermiştir. Özellikle Reptree modelinin 0.211, Isotonic'in ise yalnızca 0.168 R^2 değeri ile düşük açıklama kapasitesine sahip olduğu görülmüştür. En başarısız üç model olan Pace, M5rules ve Linear, çok yüksek RMSE ve MSE değerlerine ve son derece düşük R^2 katsayılarına (sırasıyla 0.001, 0.004 ve 0.003) sahiptir. Bu durum, söz

konusu modellerin ROA deęiřkenini tahmin etmede yetersiz olduęunu açıkça ortaya koymaktadır. Analiz kapsamında en iyi performansı gösteren model SMOreg olurken, ATM modeli hata metriklerindeki başarısıyla güçlü performans olarak deęerlendirilmiřtir. Dięer modeller, özellikle karmařık ve doğrusal olmayan yapıya sahip finansal verilerde yetersiz kalmıřtır.



SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmada, Avrupa'nın üst düzey futbol liglerinde faaliyet gösteren 26 profesyonel futbol kulübünün 1996-2022 yılları arasındaki finansal verileri kullanılarak, kulüplerin finansal performansları tahmin edilmiştir. Çalışmada, finansal performansı temsilen ROE ve ROA olmak üzere iki temel bağımlı değişken seçilmiştir. Bu değişkenler, literatürde sıklıkla kullanılan ve şirketlerin kârlılık düzeylerini ölçmede etkin göstergeler arasında yer almaktadır (Altman, 1968; Courtis, 2003). Bu gerçeklerden hareketle, ROE ve ROA olmak üzere iki bağımlı değişkenin, bağımsız değişken olan toplam 19 finansal oran aracılığıyla tahmin edilebilirliği değerlendirilmiştir. Modelleme sürecinde, çok sayıda makine öğrenmesi ve istatistiksel öğrenme algoritması uygulanmış, kullanılan her algoritma için toplanan veri seti, gözlemlerin %66'sıyla eğitim adımını ve gözlemlerin %33'üyle test adımını gerçekleştirmek için iki gruba ayrılmıştır. Bu algoritmaların tahmin performansları çeşitli istatistiksel başarı ölçütleri kullanılarak karşılaştırılmıştır. Analiz süreci WEKA yazılımı aracılığıyla gerçekleştirilmiş ve 10 farklı regresyon algoritmasının çıktıları doğrultusunda sonuçlar yorumlanmıştır. Modellerin başarı düzeyleri MAE, RMSE, RAE, RRSE, MSE ve R^2 gibi temel istatistiksel ölçütler aracılığıyla karşılaştırılmıştır.

Finansal tahminleme, özellikle zaman serisi verileri kullanılarak geçmişteki gözlemlere dayanarak gelecekteki değerlerin öngörülmesini amaçlayan önemli bir istatistiksel uygulamadır. Zaman serileri, gözlemlerin zaman boyutuna göre sıralandığı ve genellikle belirli aralıklarla kaydedildiği veri yapılarını ifade eder. Finansal veriler doğası gereği ardışık ve bağımlı gözlemlerden oluştuğu için zaman serisi analizine uygundur (Box vd., 2015). Futbol kulüpleri gibi karmaşık gelir ve gider yapısına sahip kurumlar için finansal performansın modellenmesi, sadece statik analizlerle değil aynı zamanda zaman içerisindeki değişimlerin incelenmesini gerektirir. Kulüplerin gelirleri, yayın hakları, bilet satışları, sponsorluklar ve sportif başarıya bağlı ödüller gibi zamanla değişkenlik gösteren birçok etkene bağlı olarak dalgalanma gösterir. Literatürde de zaman serisi tabanlı finansal tahmin modellerinin, geleneksel yöntemlere kıyasla daha düşük hata oranları ve daha yüksek öngörü doğruluğu bulunduğu ortaya konulmuştur (Makridakis vd., 2018; Nelson, 1991)

ROE değişkeninin tahmininde kullanılan makine öğrenimi modelleri performans açısından karşılaştırıldığında, doğrusal ve doğrusal olmayan algoritmalar arasında anlamlı farklar gözlemlenmiştir. Özellikle ATM modeli, en yüksek determinasyon katsayısı ($R^2 = 0.800$) ile en başarılı tahmin performansını sergilemiştir. Bu model aynı zamanda düşük MAE (0.057) ve RMSE (0.107) değerleriyle hem hata düzeyini minimize etmiş hem de ROE

üzerindeki varyansı yüksek oranda açıklayabilmiştir. Bununla birlikte Pace ve SMOreg modelleri de benzer şekilde düşük hata oranları ve yüksek açıklayıcılık düzeyi ile öne çıkmaktadır (R^2 : 0.781 ve 0.777).

Reptree, Gaussian ve M5rules gibi algoritmalar da görece yüksek performans göstermiştir. Ancak bu modeller, Pace ve ATM'ye kıyasla biraz daha düşük R^2 değerlerine sahiptir. ElasticNet ve Linear Regression gibi parametrik modeller ise daha düşük açıklayıcılık düzeyine sahip olup, özellikle ROE gibi karmaşık yapıları tahmininde sınırlı kalmıştır.

Öte yandan, MLP ve Isotonic Regression modelleri, yüksek hata oranları ve düşük R^2 değerleriyle diğer modellere göre daha zayıf performans göstermiştir. Özellikle Isotonic Regression modeli, en düşük R^2 (0.678) ve en yüksek MSE (0.014) değeriyle, ROE tahmini açısından en düşük güvenilirliğe sahip model olmuştur.

Bu değerlendirmeler ışığında, ROE değişkeninin tahminlenmesinde ATM algoritması, sahip olduğu yüksek R^2 değeri ve düşük hata metrikleri ile en başarılı model olarak öne çıkmaktadır. Bu durum, ATM algoritmasının, verilerdeki karmaşık yapıyı diğer modellere kıyasla daha iyi temsil ettiğini ve ROE üzerinde etkili olan faktörleri daha güçlü bir şekilde açıklayabildiğini göstermektedir.

ROA değişkeninin tahmin performansına ilişkin model karşılaştırması, farklı makine öğrenimi algoritmalarının bu orana yönelik öngörü gücünü önemli ölçüde farklılaştırdığını göstermektedir. Elde edilen sonuçlara göre, SMOreg modeli, en yüksek açıklayıcılık katsayısı ($R^2 = 0.794$) ile ROA'yı en başarılı şekilde tahmin eden model olmuştur. Bu model aynı zamanda düşük RMSE (2.222) ve MAE (0.683) değerleriyle, tahmin hatasını minimize etmede de etkili bir performans sergilemiştir.

ATM modeli, SMOreg'e kıyasla daha düşük hata metrikleri (MAE = 0.503, RMSE = 2.045) göstermesine karşın, R^2 değeri 0.7031 ile görece olarak daha düşük kalmıştır. Bu durum, modelin tahminleri hedef değişkene daha yakın olsa da varyansın daha sınırlı kısmını açıkladığını göstermektedir.

Diğer yandan, Reptree, Gaussian ve MLP gibi modeller orta düzeyde bir tahmin başarısı sağlamış; özellikle Reptree modeli oldukça düşük R^2 (0.211) ile dikkat çekmiş ve yüksek hata değerleri nedeniyle istikrarsız bir görünüm sergilemiştir.

Özellikle Linear Regression, ElasticNet, M5rules, Pace ve Isotonic Regression gibi modellerin ROA üzerindeki tahmin performansı oldukça zayıf kalmıştır. Bu modellerin R^2 değerleri %1'in altında ya da çok düşük seviyelerde olup, hata oranları da oldukça yüksektir.

Bu durum, bu modellerin ROA gibi deęişken yapılı ve dıřsal etkilerden abuk etkilenen bir kârlılık oranını anlamlandırmada ve tahmin etmede yetersiz kaldığını ortaya koymaktadır. Bu sonuçlar, literatürde vurgulanan finansal oranların zaman içinde duraęan olmayan yapısını (Nelson & Plosser, 1982) ve heterojen daęılım özelliklerini desteklemektedir.

Elde edilen bulgular, futbol kulüplerine ait finansal verilerin doğrusal olmayan ilişkiler içerdığını ve bu nedenle doğrusal modellerin tahmin başarımında sınırlı kaldığını göstermektedir. Özellikle AMT ve SMOreg gibi doğrusal olmayan yapıyı modelleyebilen algoritmalar, deęişkenler açısından yüksek doğruluk oranlarına ulaşmıştır. Bu nedenle, söz konusu modellerin hem R^2 deęerleri hem de hata metrikleri açısından üstün performans sergilediğı görölmektedir.

Makine öğrenmesi teknikleri, destek vektör regresyonu, rastgele orman, XGBoost genellikle doğrusal modellere kıyasla daha yüksek tahmin başarıları elde etmiştir (Kayakuř vd., 2023; Tutcu vd., 2024). Demir-elik sektöründe yapılan bir alıřmada, YSA modeliyle ROE ve ROA tahmininde sırasıyla %85–86 civarında R^2 deęerleri elde edilmiştir (Kayakuř vd., 2023); bu oranlar, bu alıřmadaki YSA modelinin başarısı ile tutarlıdır. Borsa İstanbul örneklemlerinde de benzer biçimde YSA ve doğrusal modellerin özellikle ROA ve ROE tahmininde öne ıktığı gözlemlenmiştir (Tutcu vd., 2024).

Yang ve Wang (2023) tarafından önerilen BERT+PCA+ARIMA+XGBoost tabanlı yaklaşım gibi literatürde, pazar duyarlılığına dayalı metin verisi kullanılarak regresyon modellerinde %95 düzeyinde R^2 deęerine ulaşan yöntemler de bulunmaktadır. Yapılan analizlerde doğrudan metin verisi kullanılmamış olsa da benzer modellerin yüksek başarı sağladığı görölmüřtür. Bu da metin verisinin katkısı dışında makine öğrenimi modellerinin temel yapısal avantajlarının güçlü tahmin kabiliyeti sunduğunu ortaya koymaktadır. Çevresel, sosyal ve yönetim verileri kullanılarak gerekleştirilen bir alıřmada, uzun kısa vadeli bellek modeli, ROA ve ROE tahminlerinde en düşük RMSE ve MAE deęerlerine ulaşarak en yüksek R^2 deęerlerini elde etmiştir. Bu sonuçlar, derin öğrenme modellerinin doğrusal modellere kıyasla daha üstün performans sergilediğini göstermektedir (Hsu vd., 2025). Alaminos vd. (2020) tarafından Avrupa futbol kulüpleri üzerinde gerekleştirilen bir alıřmada, kuantum sinir aęları ve yapay sinir aęlarının geleneksel regresyon yöntemlerine kıyasla daha yüksek tahmin doğruluęu sunduęu tespit edilmiştir. Bu bulgu, özellikle karmařık finansal sistemlerde klasik modellerin yetersiz kaldığını ve geliřmiş yapay zekâ tabanlı yaklaşımların öne ıktığını göstermektedir. Ayrıca, kuantum sinir aęı modellerinin özellikle sportif performans

değişkenleri ile finansal göstergeler arasındaki çok boyutlu ilişkileri daha etkin şekilde öğrenebildiği vurgulanmaktadır.

ROA, net kârın toplam varlıklara bölünmesiyle elde edilir; bu nedenle varlıkların değerlemesi ve amortisman politikaları sonuca doğrudan etki eder. Örneğin, bir firmanın yeni edindiği bir varlığı doğrudan gider olarak muhasebeleştirilmesi veya amortismanına tabi tutması, net kârı (ve dolayısıyla ROA'yı) önemli ölçüde değiştirebilir. Ayrıca ROA, borçlanma etkisini dolaylı olarak yansıtır; bir firma borçlanarak varlıklarını büyüttüğünde ROA düşerken, eşdeğer bir firma özkaynak artırımıyla bunu yaptığında ROA daha az etkilenir (Öztürk & Şahin, 2013). Bu durum, farklı sermaye yapılarındaki firmaların ROA değerlerini birbirleriyle doğrudan karşılaştırmayı ve geleceğe dönük tahminini zorlaştırır. ROA hesaplaması, varlıkların bilanço değeri gibi muhasebe temelli “geçmiş” bir değere dayandığından, mevcut piyasa koşullarını anlık olarak yansıtmayabilir. Bu nedenlerle, ROA değerleri muhasebe politikaları ve varlık bileşenlerindeki farklılıklara bağlı olarak yüksek veri gürültüsü içerir.

Varlık yoğunluğu sektörler arasında çok farklıdır; ağır sanayi ve altyapı firmalarının varlıkları yüksek ve sabit maliyetli iken teknoloji veya hizmet sektörlerinde varlıklar nispeten daha düşük ve esnek olabilir. Bu farklılıklar, ROA'daki sapmaları ve trendleri sektöre özgü kılar. V4 ülkelerindeki sektörel ortalamalarla yapılan bir çalışmada, sektör bazında ortalama ROA'nın tahmin edilmesinin genellikle ROE'den daha güvenilir olduğu saptanmıştır (Kristóf ve Virág, 2022). Bu tespitler ışığında, futbol kulüpleri bağlamında ROA tahminlerinin zorluğu daha da belirgin hale gelmektedir. Futbol sektörü, geleneksel endüstriyel veya hizmet sektörlerinden oldukça farklı bir varlık yapısına sahiptir. Kulüplerin en değerli varlıkları genellikle fiziksel olmayan; futbolcu bonservisleri, marka değeri, yayın hakları gibi soyut unsurlardan oluşmaktadır. Ayrıca, oyuncu transferleri gibi yüksek maliyetli ve dönemsel işlemler, toplam varlıklar ve net gelir üzerinde ani dalgalanmalara yol açabilir (Szymanski, 2003). Bu durum, ROA oranının istikrarını azaltmakta ve öngörülebilirliğini sınırlamaktadır.

Muhasebe hesapları ve finansal raporlarda yer alan tek seferlik gelir-gider kalemleri, değer düşüklük karşılıkları veya işletme dışı hareketler gibi ögeler ROA'nın temsil gücünü zayıflatabilir. Bu gürültü, basit regresyon yöntemlerinin yetersiz kalmasına neden olabilir. Hsu vd. (2025), ROA tahmininde çoklu doğrusal regresyon modeli kullanmış ve elde edilen R^2 değeri negatif çıkmıştır; bu, modelin veri değişkenliğini yakalayamadığını göstermiştir. Derin öğrenme modelleri ise aynı çalışmada hem ROA hem ROE tahminlerinde en düşük RMSE'yi sağlayarak verideki karmaşıklığı yakalayabilmiştir.

Çalışmanın bulguları, futbol kulüplerinin finansal analizinde geleneksel yöntemlerin ötesine geçilerek, yapay zekâ temelli modellere başvurulmasının sağlıklı karar alma süreçlerine katkı sağlayabileceğini göstermektedir. Additive Model Tree gibi karar ağaçlarına dayalı, SMOreg gibi yorumlanabilir yapay zekâ modelleri; kulüplerin yalnızca mevcut durumlarını analiz etmekle kalmayıp, gelecekteki olası mali eğilimleri tahmin etmede de etkin biçimde kullanılabilir. Bu tür modeller, finansal risklerin erken teşhisi, bütçe planlaması ve yatırım stratejilerinin belirlenmesinde karar destek aracı olarak değerlendirilebilir. Ayrıca UEFA'nın uyguladığı finansal fair-play düzenlemeleri gibi dışsal finansal kısıtların da bu modeller aracılığıyla simüle edilerek kulüplerin uyum süreçleri analiz edilebilir.

Bu çalışma spor finansmanı alanına makine öğrenmesi uygulamaları açısından önemli bir katkı sunmaktadır. Elde edilen sonuçlar, futbol kulüplerinin finansal performansının yapay zekâ tabanlı algoritmalar ile yüksek doğrulukla tahmin edilebileceğini göstermekte, böylece geleneksel finansal analiz yaklaşımlarının dijitalleşme sürecinde yeniden düşünülmesi gerektiğini ortaya koymaktadır. Finansal analiz ve karar alma süreçlerinde veri bilimi temelli yaklaşımların benimsenmesi, spor kulüplerinin sürdürülebilir mali yapılar oluşturmaya katkı sağlayacaktır.

Mevcut çalışma yalnızca finansal oranlarla sınırlı kalmış olsa da futbol kulüplerinin dinamik yapısı göz önüne alındığında bilet satışları, taraftar etkileşimi, sosyal medya görünürlüğü ve taraftar sadakati ölçümleri gibi finans dışı değişkenlerin modellere entegre edilmesi büyük önem taşımaktadır. Bu tür çok boyutlu veri setleri ile özellikle ağ-tabanlı öğrenme yöntemleri gibi daha ileri düzey modellerin performansının artabileceği öngörülmektedir. Bu durum, kulüplerin finansal yapısını etkileyen çok katmanlı faktörlerin daha bütüncül şekilde analiz edilmesine olanak tanıyacaktır. Mevcut analiz futbol kulüplerinin verileri kısıtlı olduğundan ötürü yıllık verilerle yürütülmüş olup, bu durum özellikle kısa vadeli değişkenlerde öngörü gücünü sınırlandırmaktadır. Verilerin çeyrek dönem ya da aylık bazda toplanması, özellikle likidite, nakit oranı gibi kısa dönemli finansal oranların daha hassas biçimde tahmin edilmesine olanak tanır. Ayrıca daha yüksek frekanslı veriler, zaman serisi modellerinin öğrenme kapasitesini de artıracaktır.

KAYNAKÇA

- Abalı, A. (1974). *Cumhuriyetin 50. yılında gençlik ve spor*. Ankara: Kültür Bakanlığı.
- Abraham, A. (2005). Artificial neural networks. In P. Sydenham & R. Thorn (Eds.), *Handbook of measuring system design* (pp. 901–908). Hoboken: John Wiley & Sons.
- Alaminos, D., Esteban, I., & Fernández-Gámez, M. A. (2020). Financial performance analysis in European football clubs. *Entropy*, 22(9), 1056.
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Anderson, D., & McNeill, G. (1992). Artificial neural networks technology. *Kaman Sciences Corporation*, 258(6), 1-83.
- Andreff, W., & Staudohar, P. D. (2000). The evolving European model of professional sports finance. *Journal of Sports Economics*, 1(3), 257-276.
- Arora, R. (2012). Comparative analysis of classification algorithms on different datasets using WEKA. *International Journal of Computer Applications*, 54(13).
- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1-127.
- Berrar, D. (2019). Cross-validation. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Cambridge: Academic Press.
- Beyazgül, M., & Karadeniz, E. (2019). Futbol kulüplerinin nakit akış profillerinin analizi: Uluslararası bir karşılaştırma. *Journal of Accounting and Taxation Studies*, 12(3), 605-624.
- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2016). *WEKA manual for version 3-9-1* (pp. 1–341). Hamilton: University of Waikato.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. Hoboken: John Wiley & Sons.
- Burges, C. J. (1998). A Tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121-167.
- Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*. Hoboken: John Wiley & Sons.
- Chen, W. C., Tseng, S. S., & Hong, T. P. (2008). An efficient bit-based feature selection method. *Expert Systems with Applications*, 34(4), 2858-2869.

- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, 1-13.
- 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.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297.
- Courtis, P. G. (2003). Du Pont ratio: A comprehensive measure of business performance. *European Research Studies Journal*, 6, 21-34.
- Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2), 299-312.
- Damodaran, A. (2007, July). *Return on capital (ROC), return on invested capital (ROIC) and return on equity (ROE): Measurement and implications*. Retrieved from <https://ssrn.com/abstract=1105499>
- De Leeuw, J., Hornik, K., & Mair, P. (2010). Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32, 1-24.
- Ecer, F., & Büyükaslan, A. (2014). Measuring performances of football clubs using financial ratios: The gray relational analysis approach, *American Journal of Economics*, 4(1), 62-71.
- Efe, MO ve Kaynak, O. (2000). Kontrol mühendisliği uygulamalarında yapay sinir ağlarının öğrenme mekanizmalarının sabitlenmesi ve sağlamlaştırılması. *Uluslararası Akıllı Sistemler Dergisi*, 15 (5), 365-388.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, M. J. Murphy, & J. R. Hugo (Eds.), *Machine learning in radiation oncology: Theory and applications* (pp. 3-11). Cham: Springer International Publishing.
- Enginkaya, E. (2014). Futbol taraftarlarının sponsor markalara ilişkin tutum, satın alma niyeti ve takımla özdeşleşmeleri arasındaki ilişki. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 36(2), 145-158.
- Er, A. S. (2016). *UEFA fair play düzenlemeleri*. <https://www.erdem-erdem.av.tr/bilgi-bankasi/uefa-finansal-fair-play-duzenlemeleri> adresinden edinilmiştir.
- Erdoğan, İ. (2008). Futbol ve futbolu inceleme üzerine. *İletişim Kuram ve Araştırma Dergisi*, 26(2), 1-58.
- FIFA. (2013). *Origins of the game*. Retrieved from <http://www.fifa.com/classicfootball/history/the-game/origins.html>
- Franceschi, M., Brocard, J. F., Follert, F., & Gouguet, J. J. (2024). Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys*, 38(3), 577-600.

- Frank, E., Hall, M., Trigg, L., Holmes, G., & Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics*, 20(15), 2479-2481.
- Frank, E., Mayo, M., & Kramer, S. (2015, April). Alternating model trees. In *Proceedings of the 30th annual ACM symposium on applied computing* (pp. 871-878). New York: ACM.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Galariotis, E., Germain, C., & Zopounidis, C. (2018). A combined methodology for the concurrent evaluation of the business, financial and sports performance of football clubs: the case of France. *Annals of Operations Research*, 266(1), 589-612.
- Gerrard, B. (2005). A resource-utilization model of organizational efficiency in professional sports teams. *Journal of Sport Management*, 19(2), 143-169.
- Ghosh, A., Nashaat, M., Miller, J., Quader, S., & Marston, C. (2018). A comprehensive review of tools for exploratory analysis of tabular industrial datasets. *Visual Informatics*, 2(4), 235-253.
- Giulianotti, R., & Robertson, R. (2004). The globalization of football: a study in the glocalization of the ‘serious life’. *The British Journal of Sociology*, 55(4), 545-568.
- Gözcü, E. (2011). *Futbol endüstrisinde markalaşma ve bir ulusal futbol takımı analizi* (Yayınlanmamış yüksek lisans tezi). Marmara Üniversitesi, İstanbul, Türkiye.
- Güngör, A. (2014). Futbol endüstrisinde sportif başarı ile finansal performans arasındaki ilişkinin analizi ve Türkiye uygulaması. *Sosyal Bilimler Dergisi*, (1), 16-36.
- Güngör, A., & Kocamış, T. U. (2018). Halka açık futbol kulüplerinde finansal performansın TOPSIS yöntemi ile analizi: İngiltere uygulaması. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(3), 1846-1859.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*, 106(496), 1383-1393.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Boosting and additive trees. In *The elements of statistical learning: Data mining, inference, and prediction* (pp. 337–387). New York: Springer.
- History of Football. (2018). *Chinese and Tsu Chu*. <http://historyoffootballreally.blogspot.com/2009/09/chinese-and-tsu-chu.html> adresinden edinilmiştir.
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A practical guide to support vector classification* (Technical Report). Department of Computer Science and Information Engineering, National Taiwan University.

- Hsu, W. L., Lin, Y. L., Lai, J. P., Liu, Y. H., & Pai, P. F. (2025). Forecasting corporate financial performance using deep learning with environmental, social, and governance data. *Electronics*, 14(3), 417.
- Ika, S. R., Udin, K., Nugroho, J. P., & Koenti, I. J. (2021, March). Assessing the financial performance of English football clubs: Arsenal and Manchester City. In *The 3rd International Conference on Banking, Accounting, Management and Economics (ICOBAME 2020)* (pp. 57–62). Amsterdam: Atlantis Press.
- Ikhwal, N. (2016). Analisis ROA dan ROE terhadap profitabilitas bank di bursa efek Indonesia. *Al-Masraf: Jurnal Lembaga Keuangan dan Perbankan*, 1(2), 211-227.
- Janssen, F., & Fürnkranz, J. (2011, July). Heuristic rule-based regression via dynamic reduction to classification. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 22, No. 1, p. 1330). Darmstadt: TU Darmstadt.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kadam, V. S., Kanhere, S., & Mahindrakar, S. (2020). Regression techniques in machine learning & applications: a review. *Int. J. Res. Appl. Sci. Eng. Technol*, 8(10), 826-830.
- Karadeniz, E., & Iskenderoglu, O. (2022). Futbol kulüplerinin finansal performansı ve Covid-19 salgınının etkisi: Avrupa Futbol Federasyonları Birliği'ne bağlı beş ulusal lig karşılaştırması. *Business and Economics Research Journal*, 13(2), 299-317.
- Kayakus, M., Tutcu, B., Terzioglu, M., Talaş, H., & Ünal Uyar, G. F. (2023). ROA and ROE forecasting in iron and steel industry using machine learning techniques for sustainable profitability. *Sustainability*, 15(9), 7389.
- Kecman, V. (2001). *Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models*. Cambridge: MIT press.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 14, No. 2, pp. 1137–1145). California: Stanford University.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160(1), 3–24.
- Kristóf, T., & Virág, M. (2022). What drives financial competitiveness of industrial sectors in Visegrad Four countries? Evidence by use of machine learning techniques. *Journal of Competitiveness*, 14(4), 117-136.
- Kumar A.K., Reddy S. V. Sarojamma B. (2022). Data mining techniques for rainfall data using WEKA. *International Journal of Computer Sciences and Engineering*, 45-48.

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- MacQueen, J. (1967, January). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (Vol. 5, pp. 281-298). Berkeley: University of California Press.
- Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., ... & Humphrey, G. B. (2023). Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental Modelling & Software*, 167, 105776.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *Plos One*, 13(3), e0194889.
- Meshkin, A., Sadeghi, M., & Ghasem-Aghaee, N. (2009). Prediction of relative solvent accessibility using pace regression. *EXCLI Journal*, 8, 211–217.
- Mitchell, T. (1997). *Machine learning*. New York: McGraw-Hill Higher Education.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. Hoboken: John Wiley & Sons.
- Morrow, S. (2023). Football's ever-changing economics. In *The People's Game? Football, Finance and Society* (pp. 1-51). Cham: Springer International Publishing.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge: MIT Press.
- Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-162.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
- Nizar, N., Zainudin, A. D., Albada, A., & Shan, C. M. (2024). Forecasting short-term FTSE bursa Malaysia using WEKA. *Information Management and Business Review*, 16(2), 104-114.
- Onyelowe, K. C., Hanandeh, S., Kamchoom, V., Ebid, A. M., Reyes Silva, F. D., Allauca Palta, J. L., ... & Avudaiappan, S. (2025). Developing advanced datadriven framework to predict the bearing capacity of piles on rock. *Scientific Reports*, 15(1), 11051.
- Öztemel, P. D. (2012). *Yapay sinir ağları*. İstanbul: Papatya Yayıncılık Eğitim.
- Öztürk, M. B., & Şahin, E. E. (2013). İMKB'de işlem gören spor firmalarının sermaye yapısının belirleyicileri üzerine bir analiz. *Verimlilik Dergisi*, (1), 7-24.
- Patterson, D. W. (1998). *Artificial neural networks: theory and applications*. Upper Saddle River: Prentice Hall PTR.

- Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods: Support vector learning* (pp. 185–208). Cambridge: MIT Press.
- Biscaia, R., Ross, S. D., Yoshida, M., Correia, A., & Rosado, A. (2016). Investigating the role of fan club membership on perceptions of team brand equity in football. *Sport Management Review*, 19(2), 157–170.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Saatçioğlu, C., & Çakmak, U. D. (2019). Avrupa ve Türkiye’de futbol ekonomisi: Karşılaştırmalı bir inceleme. *Sakarya İktisat Dergisi*, 8(4), 331-350.
- Sağiroğlu, Ş., Erler, M., & Beşdok, E. (2003). *Mühendislikte yapay zekâ uygulamaları-I: Yapay sinir ağları*. Kayseri: Ufuk Kitabevi.
- Sakinc, I., Acikalin, S., & Soyguden, A. (2017). Evaluation of the relationship between financial performance and sport success in European football. *Journal of Physical Education and Sport*, 17, 16.
- Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229.
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199-222.
- Srinivasan, D. B., & Mekala, P. (2014). Mining social networking data for classification using reptree. *International Journal of Advance Research in Computer Science and Management Studies*, 2(10).
- Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, 41(4), 1137-1187.
- Şaban, M., & Demirci, F. (2016). Finansal fair play ve Türkiye'deki dört büyük futbol kulübünün uyum düzeyinin incelenmesi. *Mali Çözüm Dergisi / Financial Analysis*, (137), 25–49.
- Talimciler, A. (2008). Futbol değil iş: Endüstriyel futbol. *İletişim Kuram ve Araştırma Dergisi*, 26(2), 89-114.
- Talimciler, A. (2010). Y. Akın (Ed.), *Bir spor kulübünün gündelik hayatımızdaki yeri ve basınla olan ilişkileri içinde*. *Kafsinakaf* (ss. 147–178). İstanbul: İletişim Yayınları.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.

- Tutcu, B., Kayakuş, M., Terzioğlu, M., Ünal Uyar, G. F., Talaş, H., & Yetiz, F. (2024). Predicting financial performance in the IT industry with machine learning: ROA and ROE analysis. *Applied Sciences*, 14(17), 7459.
- Türkiye Futbol Federasyonu. (2016). *Kulüp lisans ve finansal fair play talimatı*. <https://www.tff.org/Resources/TFF/Documents/TALIMATLAR/Kulup-Lisans-ve-Finansal-Fair-Play-Talimati.pdf> adresinden edinilmiştir.
- Uluyol, O. (2014). Süper lig futbol kulüplerinin finansal performans analizi. *Yaşar Üniversitesi E-Dergisi*, 9(34), 5716-5731.
- Üstün, B., Melssen, W. J., & Buydens, L. M. C. (2007). Visualisation and interpretation of support vector regression models. *Analytica Chimica Acta*, 595(1-2), 299-309.
- Quinlan, J. R. (1992, November). Learning with continuous classes. In A. Adams & L. Sterling (Eds.), *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence* (pp. 343–348). Hobart: University of Tasmania.
- Wang, H., & Leng, C. (2007). Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association*, 102(479), 1039-1048.
- Wang, Y., & Witten, I. H. (1997). Induction of model trees for predicting continuous classes. In *Proceedings of the Poster Papers of the 9th European Conference on Machine Learning* (pp. 128–137). Prague: University of Economics.
- Wang, Y. (2000). A new approach to fitting linear models in high dimensional spaces (Doctoral dissertation, University of Waikato). Retrieved from <https://hdl.handle.net/10289/1183>
- Williams, C. K. I., & Rasmussen, C. E. (1995). Gaussian processes for regression. In D. S. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 514–520). Cambridge: MIT Press.
- Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 31(1), 76-77.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Amsterdam: Elsevier.
- Wolff, T. M., Lopez, V. G., & Müller, M. A. (2024). Gaussian processes with noisy regression inputs for dynamical systems. *arXiv preprint arXiv:2408.08834*. <https://arxiv.org/abs/2408.08834>
- Yang, Y., & Wang, Z. (2023). Prediction of return on equity of the energy industry based on equity characteristics. *Frontiers in Energy Research*, 11, 1136914.
- Yılmaz, A. (2020). *Yapay zekâ*. İstanbul: Kodlab Yayın Dağıtım.
- Yılmaz, A. H. (2020). Endüstriyel futbolda finansal istikrarın önemi. *Fiscaoeconomia*, 4(2), 498-523.

- Yiğit, E. (2012). *Kapitalizm sürecinde futbolun ticarileşmesi*. (Yayımlanmamış yüksek lisans tezi). İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, Türkiye.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301-320.



ÖZ GEÇMİŞ

Recep Tayyip BALIKÇILAR, ilköğrenimini ve lise eğitimini İstanbul'da tamamlamıştır. 2020 yılında Recep Tayyip Erdoğan Üniversitesi İktisadi ve İdari Bilimler Fakültesi İktisat Bölümünden lisans derecesi ile mezun olmuştur.

Akademik kariyerine yüksek lisans düzeyinde devam etmiş, Bayburt Üniversitesi Lisansüstü Eğitim Enstitüsü İşletme Anabilim Dalı'nda yüksek lisans eğitimini 2025 yılında başarıyla tamamlamıştır.

