ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE HİSSE SENEDİ SEÇİMİ VE FİNANSAL DEĞİŞKENLERİN ROLÜ

Taner ANT

İSTATİSTİK ANABİLİM DALI

ANKARA
2025

Her hakkı saklıdır

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE HİSSE SENEDİ SEÇİMİ VE FİNANSAL DEĞİŞKENLERİN ROLÜ

Taner ANT

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç. Dr. Esin KÖKSAL BABACAN

Bu çalışmada, makine öğrenmesi yöntemleri kullanılarak hisse senedi getirilerinin tahmin edilmesi ve finansal değişkenlerin bu tahminlerdeki rolünün incelenmesi amaçlanmıştır. Çalışma kapsamında Rastgele Orman, XGBoost, LightGBM ve SVR (Destek Vektör Regresyonu) modelleriyle çeyreklik getiri tahminleri yapılmış, pencere genişliği (expanding window) yöntemiyle zaman serisi analizi gerçekleştirilmiştir. 10 yıllık veriler kullanılarak hisse senetlerinin çeyreklik getirileri tahmin edilmiştir. Finansal değişkenler (fiyat/kazanç oranı, borç/özsermaye vb.) modellere girdi olarak eklenmiştir. İlk yıllardan başlayarak her çeyrekte model eğitilmiş ve bir sonraki çeyreğin getirisi tahmin edilmiştir. Her çeyrekte en yüksek getiri potansiyeli olan ilk 10 hisse senedi seçilmiş ve bu hisselerin performansı izlenmiştir. 2024-2025 dönemi için modellerin önerdiği ilk 10 hisse senedinden bir portföy oluşturulmuş ve bu portföyün performansı NASDAQ 100 endeksi ile karşılaştırılmıştır. Portföy getirileri, pasif bir yatırım stratejisine göre değerlendirilmiştir. Modellerin tahminlerinde en etkili olan finansal değişkenler belirlenmiştir. En etkili on değişken analiz edilerek, hangi faktörlerin hisse getirilerini daha fazla etkilediği yorumlanmıştır.

Temmuz 2025, 57 sayfa

Anahtar Kelimeler: Makine Öğrenmesi, Destek Vektör Makineleri, Karar Ağaçları, Portföy Optimizasyonu.

ABSTRACT

Master Thesis

STOCK SELECTION WITH MACHINE LEARNING METHODS AND THE ROLE OF FINANCIAL VARIABLES

Taner ANT

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Assoc. Prof. Esin KÖKSAL BABACAN

This study aims to predict stock returns using machine learning methods and examine the role of financial variables in these predictions. Within the scope of the study, quarterly return forecasts were made using Decision Trees, XGBoost, LightGBM, and SVR (Support Vector Regression), and time series analysis was conducted using the expanding window method. The quarterly returns of stocks were predicted using data from the last 10 years. Financial variables (e.g., price-to-earnings ratio, debt-to-equity ratio, etc.) were included as inputs in the models.. Beginning with the earliest years, the models were retrained in each quarter to forecast the return for the subsequent quarter. In each period, the top 10 stocks with the highest predicted return potential were selected, and the performance of these stocks was tracked. For the 2024–2025 period, a portfolio was constructed from the top 10 stocks recommended by the models, and its performance was compared to that of the NASDAQ 100 index. The portfolio returns were evaluated relative to a passive investment strategy. Furthermore, the most influential financial variables in the models' predictions were identified. The top ten variables were analyzed in detail to interpret which financial factors had the greatest impact on stock returns.

July 2025, 57 pages

Key Words: Machine Learning, Support Vector Machines, Decision Trees, Portfolio Optimization

ÖNSÖZ ve TEŞEKKÜRLER

Bu tez çalışması süresince her aşamada göstermiş olduğu emek, rehberlik, destek ve önerileri için danışmanım Sayın Doç. Dr. Esin KÖKSAL BABACAN'a teşekkür ederim. Bugüne kadar bana göstermiş oldukları emek ve destekleri her zaman hissettiğim aileme ve arkadaşlarıma teşekkür ederim.

Taner ANT
Ankara, Temmuz 2025

İÇİNDEKİLER

TEZ ONAY SAYFASI
ETİK i
ÖZET ii
ABSTRACT iii
ÖNSÖZ ve TEŞEKKÜRLER iv
SİMGELER DİZİNİ vii
ÇİZELGELER DİZİNİ x
1. GİRİŞ 1
1.1 Literatür Taraması 2
2. PORTFÖY YÖNETİMİ YAKLAŞIMLARI 6
2.1 Geleneksel Portföy Yönetimi 6
2.2 Modern Portföy Yönetimi (Markowitz Yaklaşımı) 6
2.3 Makine Öğrenmesi ile Portföy Yönetimi 7
3. DEĞİŞKEN YORUMLAMA YÖNTEMLERİ 9
3.1 Değişken Önemi (Feature Importance) Ölçümleri 9
3.1.1 Düğüm safsızlığına dayalı değişken önemi 9
3.1.2 Ortalama safsızlık azalımı (Mean Decrease in Impurity - MDI) 10
3.1.3 Permütasyon tabanlı değişken önemi (MDA) 10
3.2 SHAP Değerleri (SHapley Additive exPlanations) 11
4. FİNANSAL ORANLAR 13
4.1 Kârlılık Oranları 13
4.2 Likedite Oranları 14
4.3 Borçluluk Oranları 15
4.4 Piyasa Değerleme Çarpanları 16
5. ARAŞTIRMA YÖNTEMİ VE VERİ SETİ 17
5.1 Veri Toplama Süreci 17
5.2 Veri Ön İşleme Süreci 19
5.3 Veri Setinin Yapısal Özellikleri 19
5.4 Analiz Yöntemi 19
6. KULLANILAN MODELLER 22
6.1 Karar Ağaçları (Decision Trees) 22
6.2 Rastgele Orman Algoritması 25
6.3 XGBoost (Extreme Gradient Boosting) Algoritması 27
6.3.1 Temel prensipler 27
6.4 LightGBM (Light Gradient Boosting Machine) Algoritması 28
6.4.1 Temel prensipler 29
6.4.2 Yaprak odaklı (Leaf-wise) büyüme stratejisi 29
6.5 Destek Vektör Regresyonu (SVR) 30
7. PERFORMANS METRİKLERİ 33
7.1 Kök Ortalama Kare Hatası (RMSE) 33

7.2 Sharpe Oranı 34
7.3 Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error - MAPE).... 35
8. ARAŞTIRMA BULGULARI 36
8.1 Rastgele Orman Modeli Sonuçları 36
8.2 XGBoost Modeli Sonuçları 39
8.3 LightGBM Modeli Sonuçları 41
8.4 Destek Vektör Regresyonu Modeli Sonuçları 44
8.5 Karşılaştırmalı Değerlendirme 47
9. SONUÇ VE TARTIŞMA 50
KAYNAKLAR 52
EK 1 Modellerin Seçtiği NASDAQ 100 Şirketlerinin Listesi ve Sektör Bilgileri 56
ÖZGEÇMİŞ 57

SİMGELER DİZİNİ

K	Toplam ağaç sayısı
N	Gözlem sayısı
T	Yaprak sayısı
ρ	Korelasyon katsayısı
σ^2	Varyans
γ	Düzenleme parametresi
λ	L2 düzenleme katsayısı
ε	Hata toleransı
ξ	Gevşeme değişkeni (slack variable)
ϕ	SHAP değeri
$\Delta i(j)$	Safsızlık azalımı

Kısaltmalar

RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
SVR	Support Vector Regression
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
SHAP	SHapley Additive exPlanations
MDI	Mean Decrease in Impurity
MDA	Mean Decrease in Accuracy

ROE	Return on Equity
ROA	Return on Assets
P/E	Price-to-Earnings Ratio
P/B	Price-to-Book Ratio
D/E	Debt-to-Equity Ratio
FVÖK	Faiz ve vergi öncesi kazanç
FAVÖK	Faiz, Vergi, Amortisman Öncesi Kâr
NASDAQ	National Association of Securities Dealers Automated Quotations
SHAP	SHapley Additive exPlanations
BIST	Borsa İstanbul A.Ş.
PD	Piyasa Değeri
DD	Defter Değeri

ŞEKİLLER DİZİNİ

Şekil 5.1 Pencere Genişliği yöntemi çalışma prensibi 20
Şekil 5.2 Modelin çalışma akış şeması 21
Şekil 6.1 SVR temel çalışma mekanizması 31
Şekil 8.1 Rastgele Orman modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılaştırması 38
Şekil 8.2 XGBoost modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması. . 41
Şekil 8.3 LightGBM modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması. 44
Şekil 8.4 Destek Vektör Regresyonu modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması. 46

ÇİZELGELER DİZİNİ

Çizelge 5.1 Örnek veri seti 18
Çizelge 8.1 Rastgele Orman modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri 36
Çizelge 8.2 Rastgele Orman modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları 37
Çizelge 8.3 XGBoost modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri 39
Çizelge 8.4 XGBoost modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları 40
Çizelge 8.5 LightGBM modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri 42
Çizelge 8.6 LightGBM modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları 43
Çizelge 8.7 Destek Vektör Regresyonu modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları 45
Çizelge 8.8 Modeller tarafından önerilen portföylerin 2024 yılı getirisi 47
Çizelge 8.9 Modellerde en yüksek SHAP değerine sahip 10 finansal oran 49

1. GİRİŞ

Günümüz finans dünyasında hisse senedi getirilerinin tahmini, hem akademisyenler hem de yatırımcılar için giderek karmaşıklaşan bir problem haline gelmiştir. Geleneksel temel ve teknik analiz yöntemlerinin sınırlılıkları, yapay zeka ve makine öğrenmesi tekniklerinin bu alanda giderek daha fazla kullanılmasını zorunlu kılmıştır. Bu çalışma, makine öğrenmesi yöntemleriyle hisse senedi getirilerinin tahmin edilmesi ve finansal değişkenlerin bu tahminlerdeki rolünün sistematik olarak incelenmesini amaçlamaktadır.

Hisse senedi piyasaları, ekonomik göstergelerden şirket performanslarına, küresel gelişmelerden yatırımcı psikolojisine kadar pek çok faktörden etkilenen dinamik bir yapı sergilemektedir. Bu karmaşıklık, geleneksel istatistiksel yöntemlerin yetersiz kalmasına neden olurken, makine öğrenmesi modelleri büyük veri setlerindeki karmaşık ilişkileri öğrenme ve tahminleme konusunda önemli avantajlar sunmaktadır. Özellikle son yıllarda XGBoost ve LightGBM gibi gelişmiş algoritmaların finansal tahmin problemlerinde başarılı sonuçlar vermesi, bu alandaki çalışmaları hızlandırmıştır.

Bu çalışmada, Rastgele Orman, XGBoost, LightGBM ve Destek Vektör Regresyonu (SVR) gibi dört farklı makine öğrenmesi algoritması kullanılarak hisse senedi getirilerinin çeyreklik tahminleri yapılmıştır. pencere genişliği yöntemiyle gerçekleştirilen zaman serisi analizinde, son 10 yıllık veriler kullanılarak dinamik bir tahmin modeli oluşturulmuştur. Modelleme sürecinde, fiyat/kazanç oranı (P/E), borç/özsermaye oranı (D/E), özsermaye kârlılığı (ROE) ve piyasa değeri/defter değeri (P/B) gibi temel finansal oranlar sistematik olarak dahil edilmiştir.

Çalışmanın en önemli özgün katkılarından biri, her çeyrekte modelin yeniden eğitilerek bir sonraki çeyreğin getirisinin tahmin edilmesi ve en yüksek getiri potansiyeline sahip ilk 10 hisse senedinin belirlenmesidir. Bu yöntemle oluşturulan portföyün performansı, pasif yatırım stratejisinin temsilcisi olarak NASDAQ 100 endeksi (teknoloji odaklı şirketlerin hisselerinin işlem gördüğü bir Amerikan borsası) ile karşılaştırılmıştır. Ayrıca, SHAP değerleri ve değişken önemi analizleri yoluyla modellerin tahminlerinde en etkili olan finansal değişkenler belirlenmiştir.

Bu araştırmanın bulguları, yatırım fonları ve portföy yöneticileri için veriye dayalı hisse seçim stratejileri geliştirme konusunda önemli ipuçları sunmaktadır. Aynı zamanda, şirket değerleme süreçlerinde hangi finansal göstergelerin daha kritik olduğunu ortaya koyması açısından da değer taşımaktadır. Akademik literatüre katkısı ise makine öğrenmesi yöntemlerinin hisse senedi tahminlerindeki etkinliğini ampirik olarak kanıtlaması ve finansal değişkenlerin göreceli önemini sistematik bir şekilde ortaya koymasıdır.

Çalışmanın ilerleyen bölümlerinde önce literatür taraması ve kuramsal çerçeve sunulacak, ardından kullanılan yöntem ve veri seti detaylı bir şekilde açıklanacaktır. Daha sonra deneysel bulgular ve analiz sonuçları paylaşılacak, son olarak ise elde edilen bulguların teorik ve pratik çıkarımları tartışılarak gelecekteki çalışmalar için önerilerde bulunulacaktır. Bu giriş bölümü, araştırmanın genel çerçevesini çizmekte ve temel yaklaşımını özetlemektedir.

1.1 Literatür Taraması

Hisse senedi getirilerinin tahmini, uzun yıllardır finansal literatürde hem kuramsal hem de uygulamalı düzeyde araştırma konusu olmuştur. Başlangıçta lineer regresyon ve faktör modelleri gibi geleneksel yöntemler öne çıkarken, son yıllarda makine öğrenmesi tabanlı modellerin veri içerisindeki karmaşık örüntüleri daha başarılı şekilde yakalayabildiği gösterilmiştir. Bu doğrultuda yapılan çalışmalar, finansal değişkenlerin hisse getirileri üzerindeki etkisini daha esnek, veri odaklı ve güçlü algoritmalarla modellemeyi hedeflemiştir. Aşağıda yer verilen araştırmalar, bu dönüşüm sürecini yansıtan önemli katkılar sunmaktadır.

Krauss vd. (2017), S&P 500 hisse senetlerinde derin öğrenme modelleri, rastgele orman ve boosting algoritmalarını karşılaştırarak uzun vadeli yatırım performansını analiz etmiştir. Sonuçlar, rastgele orman ve boosting algoritmalarının lineer modellerden daha iyi sonuç verdiğini, LSTM gibi derin öğrenme modellerinin ise veri örüntülerini daha etkili yakalayarak tahmin gücünü artırdığını göstermiştir. Çalışma ile makine öğrenmesi

yöntemlerinin hisse seçiminde alternatif bir yöntem olarak uygulanabileceği gösterilmiştir.

Gu vd. (2020), ABD hisse senedi piyasasında makine öğrenmesi algoritmalarıyla faktör bazlı getiri tahmini yapmıştır. Çalışmada, Ridge, lasso, elastic net, rastgele orman ve neural network modelleri karşılaştırılmıştır. Bulgular, geleneksel faktör modellerine kıyasla makine öğrenmesi modellerinin daha yüksek doğrulukla hisse getirisi öngördüğünü ve faktör yapılarını daha esnek şekilde yakaladığını ortaya çıkarmıştır. Çalışma, nicel yatırım stratejilerinde makine öğrenmesinin artan önemini vurgulamaktadır.

Lee ve Lee (2020), Kore hisse senedi piyasasında LightGBM ve XGBoost algoritmalarını kullanarak hisse senedi getirilerini tahmin etmeye yönelik çok faktörlü bir model geliştirmiştir. Araştırmada, geleneksel faktör modelleri ile makine öğrenmesi tabanlı modellerin tahmin performansları karşılaştırılmıştır. LightGBM algoritmasının daha yüksek doğruluk ve daha düşük hata oranı sağladığı gözlenmiştir. Bulgular, Kore borsasında faktör bazlı hisse seçim stratejilerinde makine öğrenmesi tekniklerinin güçlü bir alternatif sunduğunu ortaya koymuştur.

Metsomäki (2020), Finlandiya piyasasında finansal oranların hisse senedi getirileri ile olan ilişkisini ve makine öğrenmesi yöntemleri ile hisse senedi getirisi tahminini incelemiştir. 2014-2018 dönemi için Nasdaq Helsinki'den elde edilen veriler kullanılarak yapılmış olan çalışmada, karlılık oranları ve piyasa değeri oranlarının hisse senedi getirileriyle en güçlü ilişkiye sahip olduğu bulunmuştur. Kullanılan modellerin hisse getirilerini doğru bir şekilde tahmin etmede başarılı olmadığını ve modellerin doğruluk farklarının marjinal düzeyde kaldığını göstermiştir.

Ouyang (2022) çalışmasında makine öğrenimi algoritmalarının borsa portföy stratejilerindeki potansiyelini araştırmıştır. New York Borsası (NYSE) verilerini kullanarak farklı algoritmaların portföy performansını tahmin etme gücünü karşılaştırmıştır. Destek vektör regresyonu ve sinir ağları gibi geleneksel yöntemlerin Dow Jones Endeksi'nden daha iyi performans gösterdiğini bulmuştur. Ancak,

çalışmasında Adaboost regresyonu ve Bayesian ridge regresyonu gibi yeni nesil algoritmaların daha zayıf sonuçlar verdiğini gözlemlemiştir. Çalışma, makine öğrenimi yöntemlerinin finans sektöründe umut vaat ettiğini, ancak parametre optimizasyonu ve gerçek dünya koşullarının dikkate alınması gerektiğini vurgulamaktadır.

Chaweewanchon ve Chaysiri (2022), Markowitz'in ortalama-varyans portföy optimizasyonunu (MV) geliştirmek için makine öğrenmesi yöntemlerini kullanmıştır. Araştırmacılar, hisse senedi seçiminde evrişimsel sinir ağları (CNN) ve çift yönlü uzun kısa süreli bellek (Bi-LSTM) modellerini birleştiren hibrit bir model önermişlerdir. Amaçları, geleneksel MV modelinin zayıflıklarını ele alarak, daha iyi getiri ve risk dengesi sağlayan optimal bir portföy oluşturma stratejisi geliştirmektir. Çalışmada, Tayland Borsası SET50 endeksinden alınan verilerle modelin performansı test edilmiş ve önerilen yaklaşımın Sharpe oranı, ortalama getiri ve risk açısından daha iyi sonuçlar verdiği gösterilmiştir.

Gao vd. (2022), Çin hisse senedi piyasasında makine öğrenimi yöntemlerini kullanarak çok faktörlü bir hisse senedi seçim stratejisi geliştirmeyi amaçlamıştır. Makale, destek vektör regresyonu (SVR), rastgele orman ve ridge regresyonu gibi çeşitli makine öğrenimi algoritmalarını geleneksel doğrusal regresyon yöntemleriyle karşılaştırmıştır. Amaç, hangi yöntemin hisse senedi getirilerini tahmin etmede ve istikrarlı bir başarı oranı sağlamada en iyi performansı gösterdiğini belirlemektir. Çalışma, Çin A-hisseleri piyasasına özgü özellikleri dikkate alarak, SVR algoritmasının yüksek dereceli polinomlara daha iyi uyum sağladığını, böylece getiri ve risk yönetimi açısından doğrusal regresyondan daha üstün olduğunu göstermiştir. Sonuçlar, makine öğrenimi yöntemlerinin finansal piyasalarda nicel yatırım stratejilerini geliştirmede potansiyel taşıdığını vurgulamaktadır.

Candar (2022), yaptığı tez çalışması ile hisse senedi portföylerinin optimizasyonu için pekiştirmeli öğrenme algoritmalarının kullanımını incelemiştir. Çalışma, gelişmekte olan bir piyasa olan Türkiye'deki BIST30 endeksini temel alarak, farklı pekiştirmeli öğrenme modellerinin portföy getirilerini ve risklerini nasıl etkilediğini değerlendirmeyi amaçlamıştır. Tezin amacı, yapay zeka ve makine öğrenmesi yöntemlerinin finansal

piyasalarda portföy yönetimi süreçlerini iyileştirmedeki potansiyelini araştırmaktır. Bu çalışma, hisse senedi portföy optimizasyonunda Pekiştirmeli Öğrenme (RL) yöntemlerinin başarılı sonuçlar verdiğini ve piyasa değeri ağırlıklı endekslerden daha iyi performans sağlanabileceğini göstermiştir.

Zhang vd. (2022), Çin hisse senedi piyasasına SHAP analizini entegre ederek makine öğrenmesi tabanlı bir hisse seçimi yaklaşımı önermiştir. Çalışmada XGBoost ve LightGBM modelleri kullanılarak getiriler tahmin edilmiş, SHAP değerleri ile hangi finansal faktörlerin tahmini en fazla etkilediği açıklanmıştır. Sonuçlar, değerleme çarpanlarının ve kârlılık oranlarının getiriler üzerinde belirleyici etkisi olduğunu, SHAP analizi sayesinde modelin yorumlanabilirliğinin arttığını göstermiştir. Bu yaklaşım, yatırımcıların "kara kutu" ML modellerine duyduğu güvensizliği azaltmayı hedeflemektedir.

Hu vd. (2022), Çin borsasında SHAP yöntemiyle XGBoost modelinin yorumlanabilirliğini değerlendirmiştir. Çalışmada, finansal oranlar ve teknik göstergeler kullanılarak hisse getirileri tahmin edilmiş, SHAP analizi ile en etkili faktörler belirlenmiştir. Bulgular, özsermaye kârlılığı ve piyasa değeri gibi faktörlerin tahminlerde belirleyici olduğunu, SHAP'in modelin güvenilirliğini artırarak yatırımcıların karar destek süreçlerine katkı sağladığını ortaya çıkarmıştır.

Literatürde bu alanda benzer temalarda çalışan başka araştırmacılar da yer almaktadır. Örneğin, Li ve Ma (2019) Çin piyasasında makine öğrenmesi algoritmalarıyla hisse getirilerini tahmin etmiş; Kim ve Lee (2021) SHAP yöntemini derin öğrenme modelleriyle birleştirmiş; Patel vd. (2015), SVR ve Rastgele Orman algoritmalarıyla Borsa performans tahmini yapmıştır.

2. PORTFÖY YÖNETİMİ YAKLAŞIMLARI

Portföy yönetimi, yatırımcıların risk ve getiri hedeflerine uygun olarak varlık seçimi yapmasını sağlayan stratejiler bütünüdür. Finans literatüründe portföy yönetimi, zaman içinde farklı yaklaşımlar çerçevesinde gelişmiştir. Genel olarak bu yaklaşımlar Geleneksel Portföy Yönetimi, Modern Portföy Yönetimi (Markowitz Yaklaşımı) ve Makine Öğrenmesi ile Portföy Yönetimi olarak üç ana başlık altında incelenebilir.

2.1 Geleneksel Portföy Yönetimi

Geleneksel portföy yönetimi, yatırımcıların tekil varlıkların analizi yoluyla karar vermesine dayanır. Bu yaklaşımda yatırımcılar, şirketlerin finansal tabloları, temel analiz göstergeleri (örneğin, F/K oranı, temettü verimi, bilanço kalitesi) ve teknik analiz sinyalleri gibi kriterleri değerlendirerek yatırım yaparlar (Graham ve Dodd, 1934).

Bu yöntemin temel varsayımları şunlardır:

- Piyasalar tamamen rasyonel değildir; yatırımcılar piyasa fırsatlarını kendi bilgileriyle değerlendirebilir.
- Risk genellikle tekil varlık düzeyinde ele alınır; portföy seviyesinde risk çeşitlendirme yerine varlık seçiminde yoğunlaşılır.
- Yatırım kararları subjektif olup, yatırımcı deneyimine ve piyasa öngörüsüne bağlıdır.

Geleneksel yaklaşım, hisse senedi seçiminde bireysel hisse analizine büyük önem verirken, portföyün genel optimizasyonu konusunda sınırlı bir çerçeve sunmaktadır.

2.2 Modern Portföy Yönetimi (Markowitz Yaklaşımı)

Modern portföy yönetimi (MPT), Harry Markowitz'in 1952 tarihli çalışmasıyla finans literatürüne kazandırılmıştır. Markowitz, yatırımcıların yalnızca beklenen getiriyi değil, portföyün riskini de dikkate alarak karar vermesi gerektiğini matematiksel olarak ortaya koymuştur (Markowitz, 1952).

Bu yaklaşımın temel ilkeleri:

- **Risk ve getiri dengesi:** Yatırımcılar getirilerini artırmak isterken, aynı zamanda riski minimize etmeye çalışmaktadırlar.
- **Çeşitlendirme (Diversification):** Portföy içindeki varlıkların birbirleriyle olan korelasyonu dikkate alınarak, riski azaltacak şekilde varlıklar bir araya getirilmelidir.
- **Etkin Sınır (Efficient Frontier):** Belirli bir risk seviyesinde en yüksek getiriyi sağlayan portföyler, etkin sınır üzerinde yer alır.
- **Varyans-Kovaryans Matrisi:** Portföy riskini hesaplamak için varlıklar arasındaki kovaryans dikkate alınır.

Markowitz yaklaşımı, yatırımcıların sadece tek tek varlıklara odaklanmak yerine portföy seviyesinde optimizasyon yapmasına olanak tanıyan ilk sistematik teoridir. Bu teori, daha sonra **Sermaye Varlık Fiyatlama Modeli (CAPM)** ve **Arbitraj Fiyatlama Teorisi (APT)** gibi finansal modellerin temelini oluşturmuştur (Sharpe, 1964; Ross, 1976).

2.3 Makine Öğrenmesi ile Portföy Yönetimi

Günümüzde artan veri erişimi ve hesaplama gücü sayesinde, portföy yönetiminde makine öğrenmesi ve yapay zeka tabanlı yaklaşımlar önemli bir yer edinmiştir. Makine öğrenmesi, özellikle büyük veri analizi, faktör bazlı yatırım stratejileri ve dinamik portföy optimizasyonu gibi alanlarda modern portföy teorisini tamamlayıcı bir rol oynamaktadır (De Prado, 2020).

Makine öğrenmesi tabanlı portföy yönetimi yöntemleri:

- **Zaman Serisi Modelleri:** GARCH, LSTM gibi modeller ile hisse senedi getirileri ve volatilitesi tahmin edilebilir (Bollerslev, 1986).
- **Kümeleme Algoritmaları:** Hisse senetleri arasındaki benzerlikleri belirleyerek optimal portföy grupları oluşturulabilir (Gu vd., 2020).

- **Takviyeli Öğrenme (Reinforcement Learning):** Algoritmalar, piyasa koşullarına adapte olarak alım-satım stratejileri geliştirebilir (Fischer vd., 2019).
- **Genetik Algoritmalar ve Optimizasyon:** Markowitz'in optimizasyon yaklaşımı, genetik algoritmalar ile geliştirilerek dinamik portföy seçim stratejileri oluşturulabilir (Chen ve Yeh, 2000).

Portföy yönetimi yaklaşımları tarihsel olarak geleneksel yatırım stratejilerinden, Markowitz'in modern portföy teorisine ve günümüzde makine öğrenmesi temelli portföy optimizasyonuna doğru evrilmiştir. Geleneksel yaklaşım bireysel varlık seçiminde yatırımcı öngörüsüne dayanırken, modern portföy teorisi portföy seviyesinde optimizasyon sağlamıştır. Makine öğrenmesi ise veri odaklı ve dinamik yaklaşımlar ile portföy yönetimini bir üst seviyeye taşımaktadır.

3. DEĞİŞKEN YORUMLAMA YÖNTEMLERİ

Bu bölümde makine öğrenmesi yöntemlerinde kullanılan değişken yorumlama yöntemleri anlatılacaktır.

3.1 Değişken Önemi (Feature Importance) Ölçümleri

Karar ağaçları ve bu ağaçlara dayalı topluluk yöntemleri (örneğin, Rastgele Orman, XGBoost, LightGBM), modelin tahmin sürecinde hangi özelliklerin ne ölçüde etkili olduğunu belirlemek için değişken önemi (feature importance) kavramını kullanır. Bu ölçüm, modelin yorumlanabilirliğini artırır ve özellikle yüksek boyutlu veri setlerinde önemli değişkenlerin seçilmesine yardımcı olur (Breiman, 2011).

3.1.1 Düğüm safsızlığına dayalı değişken önemi

Karar ağaçlarında, bir düğümdeki safsızlık (impurity) ölçüsü, örneğin Gini katsayısı veya bilgi kazancı (information gain), bir özelliğin ne kadar bilgi sağladığını gösterir. Bir özelliğin önemi, bu özelliğin kullanıldığı tüm düğümlerdeki safsızlık azalımının toplamı olarak hesaplanabilir (Louppe, 2014).

Özellik x için önem değeri şu şekilde tanımlanır:

$$\text{Importance}(x) = \sum_{j \in Sx} p(j) \cdot \Delta i(j) \tag{3.1}$$

Burada:

- $p(j)$: Düğüm j 'ye ulaşan örneklerin oranı,
- $\Delta i(j)$: Düğüm j 'deki safsızlık azalımı,
- Sx : x değişkeni üzerinde yapılan tüm bölünme (split) düğümlerinin kümesi.

Eşitlik (3.1)'de sözü geçen hesaplama, her bir özellik için tüm ağaçlarda yapılan bölünmelerdeki safsızlık azalımlarının ağırlıklı toplamını ifade eder.

3.1.2 Ortalama safsızlık azalımı (Mean Decrease in Impurity - MDI)

Rastgele Orman gibi topluluk yöntemlerinde, her bir özellik için yukarıda tanımlanan önem değerleri tüm ağaçlar üzerinde ortalanarak hesaplanır. Bu yöntem, **Ortalama Safsızlık Azalımı (MDI)** olarak bilinir.

Özellik x için MDI şu şekilde ifade edilir:

$$\mathrm{MDI}(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j \in St,x} p_t(j) \cdot \Delta i_t(j) \tag{3.2}$$

Burada:

- T: Toplam ağaç sayısı,
- $p_t(j)$: Ağaç t 'deki düğüm j 'ye ulaşan örneklerin oranı,
- $\Delta i_t(j)$: Ağaç t 'deki düğüm j 'deki safsızlık azalımı,
- *St,x*: Ağaç t içinde, x değişkeni üzerinde yapılan tüm bölünme (split) düğümlerinin kümesi.

MDI, modelin tüm ağaçlarında bir özelliğin ne kadar bilgi sağladığını ölçer. Ancak bu yöntem, özellikle kategorik değişkenlerin sınıf sayısı fazla olduğunda, öznitelikler lehine **yanlı (biased)** sonuçlar üretebilir (Strobl vd., 2007).

3.1.3 Permütasyon tabanlı değişken önemi (MDA)

Permütasyon tabanlı yöntemler, bir özelliğin değerlerini rastgele karıştırarak modelin doğruluğundaki düşüşü ölçer. Bu yaklaşım, modelin genel performansı üzerinde bir özelliğin etkisini değerlendirir (Fisher vd., 2019). Bu yöntem test verisi üzerinden uygulandığında genellikle daha güvenilir sonuçlar sunar (Molnar, 2022).

3.2 SHAP Değerleri (SHapley Additive exPlanations)

Makine öğrenmesi modelleri günümüzde oldukça karmaşık yapılar içerebilmekte; bu durum modelin karar verme sürecinin açıklanmasını zorlaştırmaktadır. **Model yorumlanabilirliği (model interpretability)**, özellikle finans, sağlık, hukuk gibi yüksek riskli uygulamalarda büyük önem taşımaktadır. Bu ihtiyaca yanıt olarak geliştirilen yöntemlerden biri **SHAP (SHapley Additive exPlanations)** değerleridir (Lundberg ve Lee, 2017).

SHAP, **kooperatif oyun teorisine** dayanan ve **her bir özelliğin** modele yaptığı katkıyı adil bir şekilde dağıtmayı amaçlayan bir açıklama yöntemidir. Temelinde, her bir değişkenin model tahminine katkısını ölçmek için **Shapley değerleri** kullanılır. Bu değerler, Nobel ödüllü ekonomist **Lloyd Shapley** tarafından tanımlanmıştır (Shapley, 1953). Bir modelin çıktısını, aşağıdaki gibi her bir özelliğin katkısının toplamı olarak ifade etmek mümkündür:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{3.3}$$

Burada:

- $f(x)$: modelin tahmini,
- ϕ_0: tüm veri setinin ortalama tahmini (baseline),
- ϕ_i: i -inci özelliğin katkısıdır (SHAP değeri),
- M: toplam özellik sayısıdır.
-

Shapley değeri ϕ_i şu şekilde tanımlanır:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(M - |S| - 1)!}{M!} \left[f_{S \cup \{i\}}(x) - f_S(x) \right] \tag{3.4}$$

Bu formül, özelliğin her alt kümede (S) eklenmesi durumunda modele olan katkısının ortalamasını alır. Bu da özelliğin tahmin üzerindeki **marjinal etkisinin** adil bir biçimde hesaplanmasını sağlar.

SHAP, diğer açıklama yöntemlerinden farklı olarak aşağıdaki **istenen özellikleri** garanti eder:

- **Yerindelik (Local Accuracy)**: SHAP değerlerinin toplamı model tahminine eşittir.
- **Simetri (Symmetry)**: İki değişken aynı katkıyı yapıyorsa aynı SHAP değerini alır.
- **Eksiklik (Missingness)**: Modelde yer almayan bir özelliğin katkısı sıfırdır.
- **Eklenti (Additivity)**: SHAP, toplamsal modelleme yaklaşımıyla açıklama yapar.

SHAP değerleri; ağaç tabanlı modeller (XGBoost, LightGBM,Rastgele Orman), derin öğrenme modelleri ve lineer modeller gibi birçok farklı yapay öğrenme modeliyle kullanılabilir.

SHAP görselleştirmeleri için,

- **Force Plot**: Bir gözleme özel katkıları görselleştirir.
- **Summary Plot**: Özelliklerin önem sırasını ve etkilerini genel düzeyde gösterir.
- **Dependence Plot**: Özelliğin değeri ile SHAP değeri arasındaki ilişkiyi gösterir.

grafikleri kullanılır. Bu grafikler, özellikle finansal karar destek sistemlerinde değişken etkilerini açıklamak için oldukça kullanışlıdır.

SHAP değerleri, finansal modellemelerde değişkenlerin etkilerini şeffaf biçimde analiz etmek için sıklıkla kullanılır. Örneğin, hisse senedi seçim modellerinde değişkenlerin getiri üzerindeki etkisini ölçmek, kredi skorlama modellerinde karar sürecini açıklamak, piyasa risk modellerinde özelliklerin riske katkısını ayırt etmek gibi alanlarda etkin biçimde kullanılmaktadır (Molnar, 2022; Lundberg vd., 2020).

4. FİNANSAL ORANLAR

Şirketlerin mali yapılarının, kârlılıklarının, likiditelerinin ve piyasa değerlemelerinin analizinde kullanılan temel finansal oranlar aşağıda açıklanmıştır. Bu oranlar, yatırımcılar, analistler ve akademisyenler tarafından yaygın biçimde kullanılmaktadır (White vd., 2003).

4.1 Kârlılık Oranları

Bu oranlar, şirketin kârlılık düzeyini ve faaliyetlerinin verimliliğini ölçmek amacıyla kullanılır.

Özsermaye Karlılığı (Return on Equity - ROE): Özsermaye karlılığı, şirketin özkaynaklarını kullanarak elde ettiği net kârı ölçer. Yatırımcının şirkete koyduğu sermaye karşılığında ne kadar kâr elde edildiğini gösterir.

$$\text{Özsermaye Karlılığı} = \frac{\text{Net Dönem Kârı}}{\text{Toplam Özsermaye}} \tag{4.1}$$

ile hesaplanır. Yüksek ROE, şirketin özsermaye kullanımında etkin olduğunu gösterir.(Penman, 2013).

Aktif Kârlılığı (Return on Assets - ROA): Aktif kârlılığı, şirketin tüm varlıklarını kullanarak elde ettiği net kârı gösterir. Varlıkların kârlılık açısından verimli kullanılıp kullanılmadığını ölçer.

$$\text{Aktif Kârlılığı} = \frac{\text{Net Dönem Kârı}}{\text{Toplam Aktifler}} \tag{4.2}$$

ile hesaplanır. ROA, genellikle sektörler arası kıyaslamada tercih edilen bir ölçüttür (White vd., 2003).

Net Kâr Marjı (Profit Margin): Şirketin satışlarından elde ettiği net kâr oranını gösterir.

$$Net\ Kâr\ Marjı = \frac{Net\ Kâr}{Net\ Satışlar} \tag{4.3}$$

ile hesaplanır. Yüksek kâr marjı, maliyetlerin iyi kontrol edildiğini ve satışların kârlı olduğunu gösterir.

Faaliyet Kâr Marjı (Operating Margin): Şirketin faaliyet gelirleri içindeki operasyonel kârını gösterir.

$$Faaliyet\ Kâr\ Marjı = \frac{FaaliyetKârı(\ FVÖK)}{Net\ Satışlar} \tag{4.4}$$

ile hesaplanır. Şirketin ana faaliyetlerinden ne kadar verimli gelir elde ettiğini gösterir (Damodaran, 2012).

FAVÖK Marjı: Vergi, faiz ve amortisman öncesi kazancın satışlara oranını ifade eder.

$$FAVÖK\ Marjı = \frac{FAVÖK}{Net\ Satışlar} \tag{4.5}$$

ile hesaplanır. Özellikle sektörler arası kıyaslama için yaygın biçimde kullanılır. Amortisman gibi nakit dışı giderleri dışladığı için nakit yaratma kapasitesini daha iyi yansıtır.

4.2 Likedite Oranları

Bu oranlar, şirketin kısa vadeli yükümlülüklerini karşılama yeteneğini gösterir.

Cari Oran (Current Ratio): Şirketin kısa vadeli yükümlülüklerini karşılayıp karşılayamayacağını gösteren temel likidite oranıdır.

$$Cari\ Oran = \frac{Dönen\ Varlıklar}{Kısa\ Vadeli\ Yükümlülükler} \tag{4.6}$$

ile hesaplanır. Genel olarak 2'nin üzerindeki değerler sağlıklı kabul edilir, ancak sektör farklılıkları dikkate alınmalıdır (Fridson ve Alvarez, 2011).

Likidite Oranı (Quick Ratio / Asit-Test Oranı): Stoklar hariç en likit varlıklarla kısa vadeli borçların karşılanma oranıdır.

$$Likidite\ Oranı = \frac{Dönen\ Varlıklar\ -\ Stoklar}{Kısa\ Vadeli\ Yükümlülükler} \tag{4.7}$$

ile hesaplanır. Likidite riskinin daha net değerlendirilmesini sağlar.

4.3 Borçluluk Oranları

Şirketin finansal yapısı ve borçlanma düzeyi hakkında bilgi verir.

Borç / Özsermaye Oranı (Debt to Equity Ratio):Şirketin özkaynaklarına oranla ne kadar borçla finanse edildiğini gösterir.

$$Borç/Özsermaye\ Oranı = \frac{Toplam\ Borç}{Toplam\ Özsermaye} \tag{4.8}$$

ile hesaplanır. Finansal kaldıraç düzeyini ve sermaye yapısını yansıtır.

Borç / Aktif Oranı (Debt to Assets Ratio): Toplam varlıkların ne kadarının borçla finanse edildiğini gösterir.

$$Borç/Aktif\ Oranı = \frac{Toplam\ Borç}{Toplam\ Varlıklar} \tag{4.9}$$

ile hesaplanır. Yüksek oran, yüksek risk anlamına gelebilir.

4.4 Piyasa Değerleme Çarpanları

Piyasa değerleme çarpanları, piyasa yatırımcılarının şirketi nasıl değerlediğini gösterir; hisse fiyatlamasında önemlidir.

Hisse Fiyatı (Stock Price): Şirketin piyasadaki güncel hisse fiyatıdır. Yatırımcıların şirket değerlemesi üzerindeki algısını yansıtır.

Fiyat / Kazanç Oranı (Price to Earnings - P/E Ratio): Yatırımcıların her 1 TL'lik kâr için ne kadar ödeme yaptığını gösterir.

$$F/K = \frac{Hisse\ Fiyatı}{Hisse\ Başına\ Düşen\ Kâr} \tag{4.10}$$

ile hesaplanır. Yüksek P/E, büyüme beklentisinin yüksek olduğuna işaret edebilir.

Fiyat / Satış Oranı (Price to Sales - P/S Ratio): Şirketin hisse fiyatının satışlarına oranıdır.

$$F/S = \frac{Piyasa\ Değeri}{Net\ Satışlar} \tag{4.11}$$

ile hesaplanır. Zarar eden firmalar için P/E yerine kullanılabilir (Damodaran, 2012).

Piyasa Değeri / Defter Değeri Oranı (Price to Book - P/B Ratio): Şirketin piyasa değerinin muhasebe defter değerine oranıdır.

$$PD/DD = \frac{Piyasa\ Değeri}{Toplam\ Özsermaye} \tag{4.12}$$

ile hesaplanır. 1'in üzerindeki değerler, piyasanın şirkete defter değerinden fazla bir değer biçtiğini gösterir.

5. ARAŞTIRMA YÖNTEMİ VE VERİ SETİ

5.1 Veri Toplama Süreci

Bu çalışmada seçilen 12 finansal oran, şirketlerin kârlılık, likidite, borçluluk ve piyasa değerleme yapılarını temsil eden temel göstergeler arasında yer almaktadır. Söz konusu oranlar, yatırım kararlarında en yaygın kullanılan ve literatürde finansal performans analizlerinde sıkça tercih edilen metriklerdir. Özellikle fiyat/kazanç, özsermaye karlılığı, fiyat/satış, borç/özsermaye ve cari oran gibi göstergeler, hem kurumsal yatırımcıların hem de akademik modellerin temel değişkenleri arasında yer almaktadır (Penman, 2013; Damodaran, 2012). Ayrıca bu oranlar, makine öğrenmesi tabanlı tahmin modellerinde sıkça kullanılmış ve yatırım getirisi üzerindeki etkileri ampirik olarak test edilmiştir (Gu vd., 2020; Krauss vd., 2017). Bu nedenle, çalışmada yer alan oranlar hem teorik hem de uygulamalı finans literatüründe güçlü bir temele sahiptir.

Kullanılan finansal veriler, ABD Menkul Kıymetler ve Borsa Komisyonu'nun (SEC) resmi veri tabanından[1] temin edilmiştir. Araştırma kapsamında aşağıdaki 12 temel finansal oran seçilmiştir:

1. Karlılık Oranları:
    - Özsermaye Kârlılığı (ROE)
    - Aktif Kârlılığı (ROA)
    - Net Kâr Marjı (Profit Margin)
    - Faaliyet Kâr Marjı (Operating Margin)
    - FAVÖK Marjı (EBITDA Margin)
2. Likidite Oranları:
    - Cari Oran (Current Ratio)
    - Asit-Test Oranı (Quick Ratio)
3. Kaldıraç Oranları:

[1] ABD Menkul Kıymetler ve Borsa Komisyonu (SEC) Veri Tabanı: *https://www.sec.gov*

- Borç/Özsermaye Oranı (Debt to Equity)
- Borç/Varlık Oranı (Debt to Assets)

4. Değerleme Oranları:
    - Fiyat/Kazanç Oranı (PE Ratio)
    - Fiyat/Satış Oranı (PS Ratio)
    - Fiyat/Defter Değeri Oranı (PB Ratio)

Örnek teşkil etmesi açısından, Apple Inc.(AAPL) şirketine ait 2023 yılı çeyreklik finansal oranları **Çizelge 5.1**'de özetlenmiştir. Bu tablo, çalışmada kullanılan 12 temel finansal oranın zaman içindeki dalgalanmasına dair genel bir görünüm sağlamaktadır.

Çizelge 5.1 Örnek veri seti

Tarih	**2023-03-31**	**2023-06-30**	**2023-09-30**	**2023-12-31**
Hisse	AAPL	AAPL	AAPL	AAPL
Çeyreklik Getiri(%)	15.9	-8.6	-25.2	31.8
Özsermaye Karlılığı (ROE)	%38.9	%33.0	%36.9	%45.8
Aktif Karlılığı (ROA)	%7.3	%5.9	%6.5	%9.6
Net Kar Marjı	%25.5	%24.3	%25.6	%28.4
Faaliyet Marjı	%29.9	%28.1	%30.1	%33.8
FAVÖK Marjı	%32.9	%31.8	%33.1	%36.1
Cari Oran	0.94	0.982	0.988	1.073
Likidite Oranı	0.878	0.923	0.944	1.024
Borç/Özsermaye	1.763	1.813	1.994	1.458
Borç/Varlık	0.33	0.326	0.351	0.306
Hisse Fiyatı($)	163.04	192.047	169.742	191.13
Fiyat/Kazanç Oranı	106.941	152.384	115.885	87.781
Fiyat/Satış Oranı	27.244	37.037	29.724	24.898
PP/DD	41.567	50.263	42.807	40.178

5.2 Veri Ön İşleme Süreci

Elde edilen ham veriler üzerinde aşağıdaki ön işleme adımları uygulanmıştır:

1. **Veri Birleştirme**: Her şirket için ayrı ayrı indirilen finansal oran verileri tek bir veri setinde birleştirilmiştir.
2. **Zaman Aralığı Sınırlaması**: Veri seti son 10 yıllık dönemle (2014-2024) sınırlandırılmıştır.
3. **Sayısal Format Standardizasyonu**: Tüm değişkenler 0.000 basamak formatına dönüştürülerek veri tutarlılığı sağlanmıştır.
4. **Eksik Veri Doldurma**: Eksik gözlemler için bir önceki dönem değerleriyle doldurma yöntemi kullanılmıştır. Bu yöntem, finansal verilerin zaman içindeki sürekliliği göz önüne alınarak tercih edilmiştir.

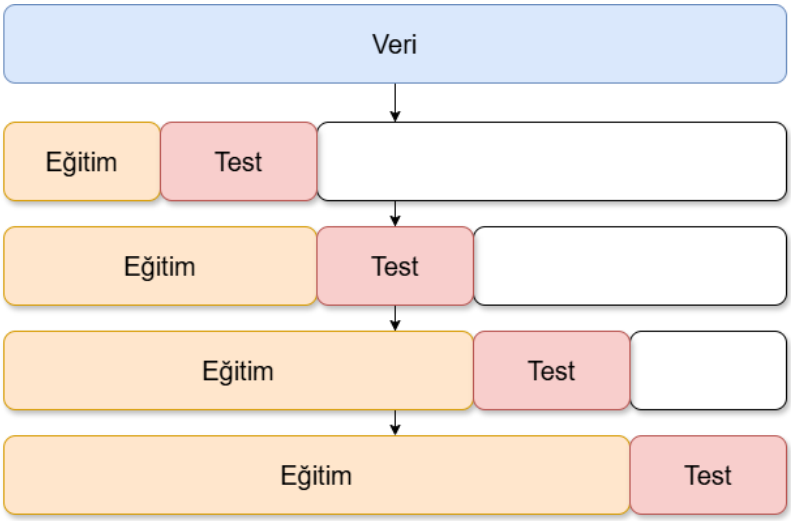
5.3 Veri Setinin Yapısal Özellikleri

Oluşturulan nihai veri seti aşağıdaki özelliklere sahiptir:

- **Zaman Aralığı**: 2014-2024 (çeyreklik veriler)
- **Gözlem Sayısı**: 40 çeyrek (10 yıl)
- **Şirket Sayısı**: 100 adet (çalışmada analiz edilen şirket sayısı)
- **Değişken Sayısı**: 12 finansal oran + hisse getirisi

5.4 Analiz Yöntemi

Zaman serisi analizinde kullanılan pencere genişliği yöntemiyle, ilk yıldan başlayarak her çeyrekte model yeniden eğitilmiş ve bir sonraki çeyreğin getirisi tahmin edilmiştir. Bu yöntemin çalışma prensibi görsel olarak Şekil 5.1 ile verilmiştir.

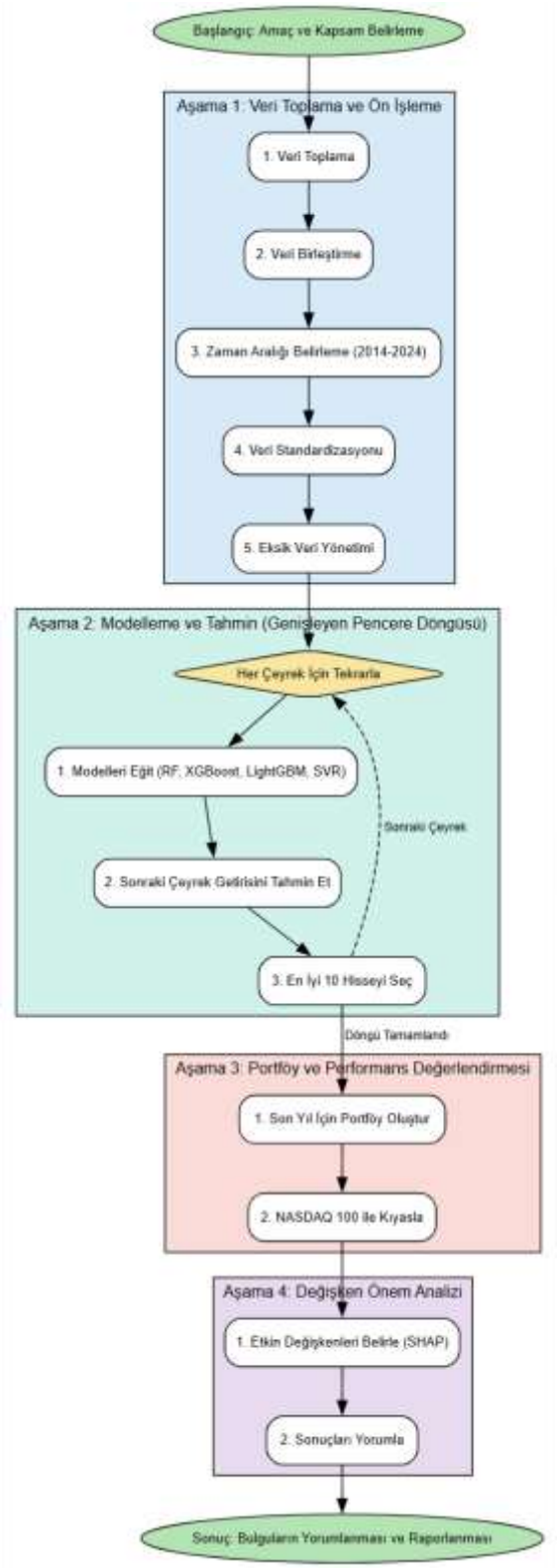


Şekil 5.1 Pencere Genişliği yöntemi çalışma prensibi

Tez çalışmasında Rastgele Orman, XGBoost, LightGBM, Destek Vektör Regresyonu (SVR) Makine Öğrenmesi Modelleri kullanılmıştır. Model Değerlendirme Metrikleri olarak Kök Ortalama Kare Hata (RMSE), Ortalama Mutlak Yüzde Hata (MAPE) ve Portföy getirisi karşılaştırmaları kullanılmıştır.

Her çeyrekte modeller tarafından en yüksek getiri potansiyeli olduğu tahmin edilen ilk 10 hisse senedi seçilmiş ve bu hisselerin performansı izlenmiştir. Son yılda ise bu stratejiyle oluşturulan portföyün performansı, pasif yatırım stratejisi temsilcisi olarak NASDAQ 100 endeksiyle karşılaştırılmıştır.

Analizler Python programlama dili kullanılarak gerçekleştirilmiştir. Kullanılan temel kütüphaneler ise Pandas, NumPy (veri işleme), Scikit-learn (makine öğrenmesi modelleri) ve XGBoost, LightGBM (gradient boosting algoritmaları), Matplotlib, Seaborn (görselleştirme) kütüphaneleridir. Şekil 5. 2 ile modelin çalışma akış şeması verilmiştir.



Şekil 5.2 Modelin çalışma akış şeması

6. KULLANILAN MODELLER

Tezin bu bölümünde analizlerde kullanılan modellerin kısa açıklamaları verilmiştir.

6.1 Karar Ağaçları (Decision Trees)

Karar ağaçları, hem sınıflandırma hem de regresyon problemleri için kullanılan, kural tabanlı ve sezgisel olarak kolay anlaşılabilen bir makine öğrenmesi yöntemidir. Model, veri kümesini özniteliklere göre bölerek bir ağaç yapısı oluşturur. İç düğümler (nodes) bir özelliğe dayalı karar noktalarını, yaprak düğümler (leaves) ise sınıf etiketlerini (sınıflandırmada) veya tahmin edilen değerleri (regresyonda) temsil eder (Breiman vd., 1984).

Karar ağacı algoritması, her bölme adımında veri setini bir özniteliğe göre iki veya daha fazla alt gruba ayırır. Amaç, her alt grubun mümkün olduğunca saf (homojen) olmasını sağlamaktır. Saflığı ölçmek için çeşitli kriterler kullanılır.

Bir düğümdeki alt gruba ayırma ölçütleri şunlardır:

Gini Katsayısı (Gini Impurity) : Gini katsayısı, rastgele seçilen iki örneğin farklı sınıflara ait olma olasılığını ölçer. Düğüm t için Gini katsayısı şöyle tanımlanır:

$$Gini(t) = 1 - \sum_{i=1}^{C} p(i|t)^2 \tag{6.1}$$

Burada: C, sınıf sayısını, $p(i|t)^2$, düğüm t 'deki örneklerden sınıf i 'ye ait olanların oranını ifade eder. Daha düşük Gini değeri, daha yüksek saflık anlamına gelir (Breiman vd., 1984).

Entropi (Entropy) ve Bilgi Kazancı (Information Gain): Entropi, düğümdeki sınıf dağılımının düzensizliğini ölçer ve şu şekilde tanımlanır:

$$Entropy(t) = -\sum_{i=1}^{C} p(i|t)\log_2\big(p(i|t)\big) \tag{6.2}$$

Bilgi kazancı ise, bir bölmenin sağladığı düzensizlik azalmasıdır:

$$InformationGain = Entropy(parent) - \sum_{k=1}^{K} \frac{N_k}{N} Entropy(child_k) \tag{6.3}$$

Burada:

- $parent$, bölünmeden önceki düğüm,
- $child_k$, bölünme sonrası oluşan alt düğümler,
- N_k, alt düğümdeki örnek sayısı,
- N, ana düğümdeki örnek sayısıdır.

Bilgi kazancı, ilk olarak ID3 algoritmasında kullanılmaya başlanmıştır (Quinlan, 1986).

Sınıflandırma ve Regresyon İçin Karar Ağaçları

Sınıflandırma Karar Ağaçları: Gini katsayısı veya entropi ölçütleri kullanılarak, sınıflar arasında ayrım yapılır.

Regresyon Karar Ağaçları: Bölünmeler, düğümdeki hedef değişkenin varyansını azaltmaya yönelik yapılır. Bir düğümdeki varyans şöyle tanımlanır:

$$Var(t) = \frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - \overline{y_t})^2 \tag{6.4}$$

Burada $\overline{y_t}$, düğümdeki hedef değişkenin ortalama değeridir.

Bölünme sonrası varyans azalımı şu şekilde hesaplanır:

$$ReductioninVariance = Var(parent) - \sum_{k=1}^{K} \frac{N_k}{N} Var(child_k) \qquad (6.5)$$

Karar Ağaçları Algoritmaları

Karar ağaçları, kullanılan bölme kriterlerine ve büyüme stratejilerine göre farklı algoritmalarla uygulanabilir. En yaygın karar ağacı algoritmaları şunlardır:

ID3 (Iterative Dichotomiser 3) Algoritması: ID3 algoritması, J. R. Quinlan tarafından geliştirilmiştir (Quinlan, 1986). Bölme kriteri olarak **bilgi kazancı (information gain)** kullanır. Her adımda, veri setini en yüksek bilgi kazancı sağlayacak özellik üzerinden böler. Dezavantajı, bilgi kazancının çok fazla farklı değer alabilen öznitelikleri tercih etmeye meyilli olmasıdır.

C4.5 Algoritması: C4.5 algoritması, ID3'ün geliştirilmiş bir versiyonudur (Quinlan, 1993). Özellik seçiminde bilgi kazancı oranı (**gain ratio**) kullanılır. Gain ratio, bilgi kazancını normalleştirerek fazla değerli özniteliklere yönelme eğilimini azaltır. Ayrıca, C4.5 hem kategorik hem de sürekli verilerle çalışabilir ve eksik değerleri de işleyebilir.

CART (Classification and Regression Trees) Algoritması: CART algoritması, Breiman ve arkadaşları tarafından geliştirilmiştir (Breiman vd., 1984). Hem sınıflandırma hem de regresyon problemleri için uygundur.

- Sınıflandırma problemlerinde, **Gini katsayısı** kullanılarak veri bölünür.
- Regresyon problemlerinde ise, düğümler **en küçük kareler yöntemi** kullanılarak bölünür.

CART ağaçları ikili ağaçlar üretir; her düğüm sadece ikiye ayrılır.

CHAID (Chi-squared Automatic Interaction Detector) Algoritması: CHAID algoritması, özellikle çok değerli kategorik öznitelikler ile çalışmak için tasarlanmıştır. Bölme kararlarını **ki-kare (χ^2)** testi ile verir. İstatistiksel anlamlılık temel alınarak, en uygun bölme seçilir (Kass, 1980).

Karar ağacı algoritmaları, veriyi tam ayırana kadar büyüyebilir. Ancak, bu durumda model veriye aşırı uyum (overfitting) gösterir. Bu problemi önlemek için iki ana yaklaşım vardır:

Ön Budama: Ağacın büyümesini erken durdurarak kontrol etmek (örneğin, maksimum derinlik belirlemek, minimum örnek sayısı şartı koymak).

Son Budama: Ağaç tamamen büyütüldükten sonra, doğrulama setinde performansı artırmayan dalların kesilmesi (Quinlan, 1993).

Karar ağaçları, kredi riski skorlama (Yeh ve Lien, 2009), tıbbi tanı sistemleri, hisse senedi fiyat tahmini (Chen vd., 2021) gibi birçok farklı alanda yaygın biçimde kullanılmaktadır.

6.2 Rastgele Orman Algoritması

Karar ağaçları yüksek önyargı veya yüksek varyans ile sonuçlanabilen tekil modellerdir. Bu sorunun üstesinden gelmek ve daha güçlü genelleme yeteneği elde etmek amacıyla, bir topluluk (ensemble) öğrenme yöntemi olan Rastgele Orman algoritması geliştirilmiştir (Breiman, 2001). Rastgele Orman, birçok karar ağacının birleşiminden oluşur ve bu ağaçların çıktılarının birleştirilmesiyle daha sağlam ve güvenilir tahminler yapılmasını sağlar.

Rastgele Orman algoritmasının temel ilkesi, her bir karar ağacının eğitiminde hem bootstrap örnekleme (veri alt örneklemesi) hem de özellik alt kümesi seçimi gibi rastgeleliklerin dahil edilmesidir. Böylece modeller arası korelasyon azaltılır ve aşırı öğrenme (overfitting) riski düşürülür.

Modelin Yapısı ve İşleyişi: Rastgele Orman algoritması, aşağıdaki temel adımları izleyerek çalışır:

1. Eğitim verisinden **B** kez rastgele örnekleme (bootstrap sampling) yapılır.

2. Her bir örneklem için, sınırlı sayıda özellikten (√p sınıflamada, p/3 regresyonda önerilir) rastgele seçim yapılarak bir karar ağacı oluşturulur.
3. Tüm ağaçlar eğitildikten sonra, sınıflama probleminde **çoğunluk oyu** (majority voting), regresyonda ise **ağaçların ortalaması** alınarak nihai tahmin yapılır.

Rastgele Orman ve Varyans Azaltımı: Rastgele Orman algoritmasının başarısının temelinde, varyansı azaltarak daha düşük genelleme hatası elde etmesi yer alır. B bireysel karar ağacının tahmin çıktısı $\widehat{f_b}(x)$ ile gösterilir ise, rastgele ormanın nihai çıktısı:

$$\widehat{f_{RF}}(x) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f_b}(x) \tag{6.6}$$

şeklinde tanımlanır.

Topluluk modellerinin toplam varyansı aşağıdaki gibi ifade edilir:

$$Var\left(\widehat{f_{RF}}\right) = \rho \cdot \sigma^2 + \frac{1-\rho}{B} \cdot \sigma^2 \tag{6.7}$$

Burada:

- ρ: Ağaçlar arasındaki korelasyon katsayısıdır.
- σ^2: Tekil bir karar ağacının varyansıdır.
- B: Ağaç sayısıdır.

Bu formül, ağaçlar arasındaki korelasyonun düşürülmesi ve çok sayıda ağaç kullanımı ile varyansın azaltılabileceğini göstermektedir (Hastie vd., 2009).

Rasgele Orman modelinin avantajları,

- Aşırı öğrenmeye karşı dayanıklıdır.
- Hem sınıflama hem regresyon için uygundur.
- Özellik önem derecelerini hesaplama imkânı sunar.

- Geniş veri setlerinde yüksek doğruluk sağlar.

olarak verilebilir. Dezavantajları ise,

- Karar mekanizması tekil ağaçlara kıyasla daha az yorumlanabilirdir.
- Hesaplama maliyeti yüksek olabilir (özellikle çok sayıda ağaç varsa).
- Çok büyük veri setlerinde bellekte fazla yer kaplayabilir.

biçiminde verilir.

6.3 XGBoost (Extreme Gradient Boosting) Algoritması

Son yıllarda birçok makine öğrenmesi yarışmasında yüksek başarı gösteren XGBoost (Extreme Gradient Boosting), gradyan artırmalı ağaç temelli modellerin optimize edilmiş bir sürümüdür (Chen ve Guestrin, 2016). XGBoost, özellikle büyük veri setlerinde ve karmaşık ilişkilere sahip problemlerde yüksek doğruluk ve hız sağlaması nedeniyle popülerleşmiştir.

6.3.1 Temel prensipler

XGBoost, boosting yaklaşımına dayalı bir topluluk (ensemble) yöntemidir. Boosting, zayıf öğrenicileri (weak learners), ardışık şekilde eğitip hataları azaltarak güçlü bir model oluşturmayı amaçlar (Friedman, 2001). Bu bağlamda, her yeni ağaç, önceki ağaçların yapamadığı tahmin hatalarını düzeltmek üzere eğitilir.

Modelin çıktısı şu şekilde ifade edilir:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{6.8}$$

Burada:

- $\hat{y}_i$: i. gözlemin model tarafından yapılan tahmini,
- K: toplam ağaç sayısı,

- f_k: k. karar ağacı (regresyon ağacı),
- $\mathcal{F}$: tüm olası ağaç fonksiyonları kümesidir.

Amaç, toplam kayıp fonksiyonunu (loss) ve ağaçların karmaşıklığını minimize etmektir:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6.9}$$

Burada:

- $l(y_i, \hat{y}_i)$: Gerçek değer ile tahmin arasındaki kayıp (ör. kare hatalar),
- $\Omega(f) = \gamma T + \frac{1}{2}\lambda|w|^2$: Model karmaşıklığı cezası (T: yaprak sayısı, w: yaprak skorları).

Bu yapı sayesinde XGBoost, aşırı öğrenmeyi kontrol eden düzenleme (regularization) mekanizmaları içerir.

XGBoost'un öne çıkan avantajları şunlardır:

- **Aşırı öğrenmeye karşı dirençlidir:** Düzenleme parametreleri (γ/λ) sayesinde model karmaşıklığı kontrol edilir.
- **Paralel işlemeye uygundur:** Ağaç yapısı, bölünmeler arasında paralel hesaplamaya olanak tanır.
- **Kayıp fonksiyonu esnektir:** Sınıflama ve regresyon dahil olmak üzere farklı problemlerde kullanılabilir.
- **Eksik veriye dayanıklıdır:** Kaybolan verileri otomatik olarak yönlendirir.

6.4 LightGBM (Light Gradient Boosting Machine) Algoritması

LightGBM (Light Gradient Boosting Machine), gradyan artırma (gradient boosting) yöntemine dayanan ve Microsoft tarafından geliştirilen bir makine öğrenmesi algoritmasıdır (Ke vd., 2017). Özellikle büyük veri setlerinde, yüksek boyutlu özelliklerde ve düşük gecikme süresi gerektiren uygulamalarda daha verimli çalışması

amacıyla geliştirilmiştir. LightGBM, XGBoost gibi ağaç temelli boosting algoritmalarına alternatif olarak geliştirilmiş; daha hızlı, daha düşük bellek kullanımı ve daha iyi doğruluk oranları sunabilen bir çözüm haline gelmiştir.

6.4.1 Temel prensipler

LightGBM, diğer gradyan artırmalı yöntemler gibi, art arda zayıf öğreniciler (karar ağaçları) inşa ederek her adımda hatayı azaltmaya çalışır. Ancak, XGBoost'tan farklı olarak LightGBM şu iki yenilikçi tekniği kullanır:

Gradient-based One Side Sampling (GOSS): Gradyan değeri düşük olan örnekleri rastgele örnekleme yerine eler ve sadece büyük gradyanlara sahip örnekleri tutarak bilgi kaybını minimize eder. Bu sayede, veri miktarı azalmasına rağmen model doğruluğu korunur.

Exclusive Feature Bundling (EFB): Seyrek ve birbirini dışlayan (mutually exclusive) özellikler birleştirilerek boyut azaltımı sağlanır. Bu, özellikle yüksek boyutlu veri setlerinde bellek kullanımını ve hesaplama süresini azaltır.

6.4.2 Yaprak odaklı (Leaf-wise) büyüme stratejisi

LightGBM’in en belirgin farkı, geleneksel ağaçlarda kullanılan **seviye odaklı (level-wise)** büyüme yerine **yaprak odaklı (leaf-wise)** stratejiyi kullanmasıdır. Bu yöntem, en yüksek bilgi kazancını sağlayan yaprağı derinleştirmeyi hedefler. Bu sayede daha hızlı kayıp azalımı sağlanır.

Geleneksel seviye odaklı yaklaşımda tüm yapraklar aynı seviyede büyütülürken, LightGBM’de yalnızca en verimli yaprak genişletilir:

- **Level-wise (örneğin XGBoost):** Daha dengeli ama daha yavaş.
- **Leaf-wise (LightGBM):** Daha hızlı öğrenir ama aşırı öğrenmeye (overfitting) daha açık olabilir.

LightGBM algoritmasının avantajları,

- Büyük veri setlerinde yüksek hız ve düşük bellek kullanımı sağlar.
- Gelişmiş örnekleme ve özellik paketleme teknikleri ile verimlidir.
- Hem sınıflama hem regresyon görevlerinde yüksek doğruluk sağlar.
- GPU desteği ile paralel hesaplama imkânı sunar.

ile verilebilir. Dezavantajları ile

- Yaprak odaklı büyüme aşırı öğrenmeye yol açabilir, dikkatli parametre ayarı gerektirir.
- XGBoost kadar esnek özelleştirme olanakları sunmaz.

biçiminde verilir.

6.5 Destek Vektör Regresyonu (SVR)

Destek Vektör Regresyonu (SVR), Support Vector Machine (SVM) algoritmasının regresyon problemlerine uyarlanmış bir versiyonudur (Vapnik, 1995). SVR, veriler arasındaki ilişkileri öğrenerek gelecekteki değerleri tahmin etmeye çalışır ve özellikle yüksek boyutlu, karmaşık veri yapılarında başarılı performans göstermesiyle bilinmektedir (Smola ve Schölkopf, 2004).

SVR, hedef değişken ile girdiler arasındaki ilişkiyi öyle bir şekilde öğrenmek ister ki, tahmin hatası belirli bir tolerans (ε) içinde kalsın. Model, bu toleransın dışında kalan hataları cezalandırır ve genel olarak margin of tolerance kavramına dayanılır.

SVR, aşağıdaki doğrusal regresyon fonksiyonunu kullanarak model kurar:

$$f(x) = \langle w, x \rangle + b \tag{6.10}$$

Burada:

- w: ağırlık vektörü,

- x: giriş vektörü,

- b: bias terimidir.

Modelin optimizasyon problemi şu şekilde tanımlanır:

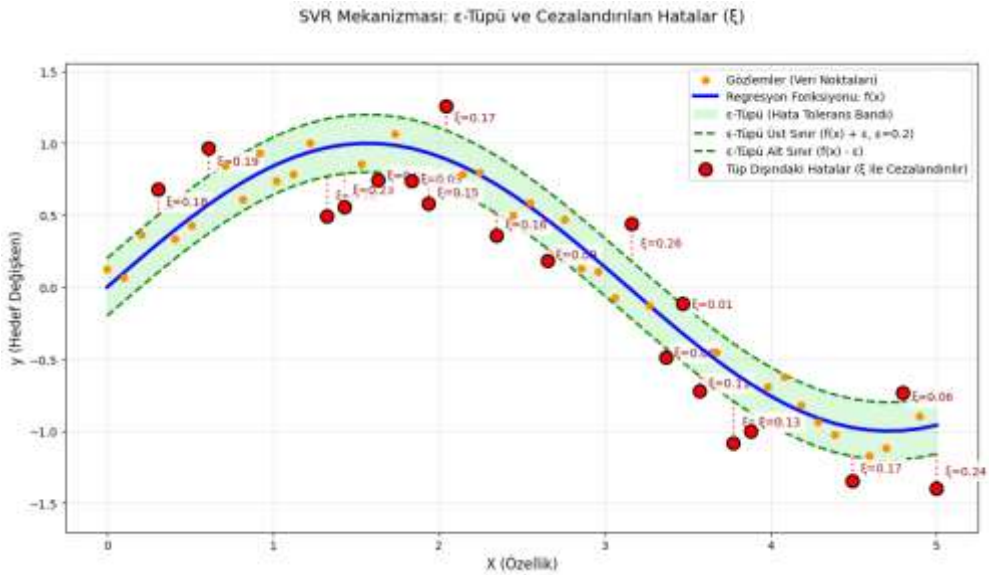
$$\min_{w,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{6.11}$$

Modelin kısıtları,

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{6.12}$$

biçimindedir. Burada, ε, hata toleransı, ξ_i, ξ_i^* tolerans dışındaki hataları temsil eden gevşeme değişkenleri, C, hata toleransına karşı duyarlılığı belirleyen düzenleme (regularization) parametresidir.

Bu yapı, hem hata hem de model karmaşıklığını minimize etmeye çalışır.



Şekil 6.1 SVR temel çalışma mekanizması

Şekil 6.1'de SVR'nin temel mekanizması gösterilmektedir. Ortadaki çizgi regresyon fonksiyonunu (f(x)) ifade ederken, çizginin üstü ve altındaki iki yatay çizgi ε-tüpünü

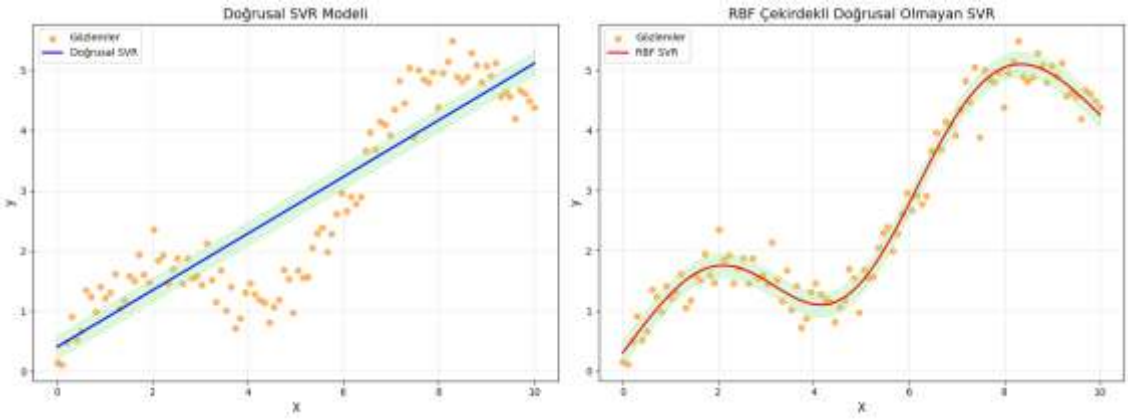
temsil eder. Bu tüp içine düşen tahmin hataları cezalandırılmazken, dışına çıkan gözlemler ξ değerleriyle cezalandırılır.

Çekirdek Fonksiyonu (Kernel Trick): SVR doğrusal olmayan ilişkileri öğrenmek için çekirdek (kernel) fonksiyonları kullanabilir. En yaygın kullanılan çekirdek fonksiyonları:

- Doğrusal: $K(x_i, x_j) = x_i^T x_j$
- RBF (Radial Basis Function):

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{6.14}$$

RBF, finansal veriler gibi karmaşık yapıları modellemek için özellikle uygundur.



Şekil 6.2 Doğrusal ve doğrusal olmayan SVR modelleri

Şekil 6.2'de doğrusal ve doğrusal olmayan SVR modelleri karşılaştırılmaktadır. Soldaki grafik doğrusal SVR modelini, sağdaki grafik ise RBF çekirdeği kullanan doğrusal olmayan SVR modelini temsil eder. RBF çekirdeği, karmaşık eğilimleri daha başarılı bir şekilde yakalayabilmektedir.

SVR, finansal piyasalarda hisse senedi fiyat tahmini, volatilite öngörüsü ve portföy optimizasyonu gibi alanlarda sıklıkla kullanılmaktadır (Kim, 2003; Cao ve Tay, 2003). Özellikle küçük ve gürültülü veri kümelerinde SVR, yüksek genelleme kapasitesi sayesinde etkili tahminler sunmaktadır.

7. PERFORMANS METRİKLERİ

Bu bölümde makine öğrenmesi yöntemlerinde yaygın olarak kullanılan performans metrikleri verilecektir.

7.1 Kök Ortalama Kare Hatası (RMSE)

Makine öğrenmesi modellerinin performansını değerlendirmek için çeşitli hata metrikleri kullanılmaktadır. Regresyon tabanlı modellerde tahmin edilen ve gerçek değerler arasındaki farkın ölçülmesi kritik bir adımdır (Hyndman ve Koehler, 2006). Bu çalışmada, model performansını değerlendirmek için **Kök Ortalama Kare Hatası (Root Mean Squared Error - RMSE)** metriği kullanılacaktır.

RMSE, tahmin edilen ($\hat{y}_i$) ve gerçek (y_i) değeri arasındaki farkın karesinin ortalaması alınarak hesaplanır ve ardından karekökü alınarak hata ölçüsü elde edilir. RMSE'nin matematiksel formülü şu şekildedir:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7.1}$$

Burada, y_i gerçek değerleri; $\hat{y}_i$, model tarafından tahmin edilen değerleri; N, toplam gözlem sayısını ifade eder.

Finansal tahminleme çalışmalarında RMSE, modelin tahmin doğruluğunu belirlemek için sıkça kullanılmaktadır. Büyük RMSE değerleri, modelin tahminlerinde yüksek sapma olduğunu gösterirken, düşük RMSE değerleri modelin daha başarılı tahminler ürettiğini gösterir (Makridakis vd., 1998).

Hisse senedi getirilerinin tahmininde RMSE'nin düşük olması, modelin piyasadaki getirileri daha doğru tahmin ettiğini gösterir. Ancak finansal zaman serilerinin doğası gereği yüksek volatilite içermesi nedeniyle, tek bir hata metriğine dayanarak model

performansını değerlendirmek yeterli olmayabilir (Gu vd., 2020). Bu nedenle, RMSE ile birlikte diğer hata metrikleri de göz önünde bulundurulmalıdır.

Bu çalışmada, oluşturulan modellerin tahmin performansını değerlendirmek için RMSE metriği kullanılacak ve sonuçlar piyasa endeksi ile karşılaştırılarak analiz edilecektir.

7.2 Sharpe Oranı

Sharpe Oranı, yatırımın birim risk başına ne kadar fazla getiri sağladığını ölçen önemli bir performans göstergesidir. İlk olarak William F. Sharpe tarafından geliştirilmiş ve 1966 yılında yayımlanan çalışmasıyla literatüre kazandırılmıştır (Sharpe, 1966). Finansal performans analizinde sıklıkla kullanılan bu oran, özellikle portföylerin risk-düzeltilmiş getirilerini kıyaslamada önemli bir araç olarak kabul edilmektedir.

Sharpe Oranı şu şekilde tanımlanır:

$$\text{Sharpe Oranı} = \frac{R_p - R_f}{\sigma_p} \tag{7.2}$$

Burada, R_p, Portföyün ortalama getirisi; R_f, risksiz getiri oranı (örneğin devlet tahvili faizi); σ_p, portföyün getirisinin standart sapması (yani riski) şeklinde ifade edilir.

Bu metrik, yatırımcının portföyü ile risksiz bir yatırım arasında ne kadar fazla getiri sağladığını ve bu fazladan getirinin ne kadar volatilite karşılığında elde edildiğini gösterir. Sharpe Oranı'nın yüksek olması, yatırımın riskine göre daha fazla getiri sunduğu anlamına gelir. Oranın sıfırdan büyük olması pozitif performansa, sıfırdan küçük olması ise riskin getiriden fazla olduğuna işaret eder (Bodie vd., 2014).

Sharpe Oranı 1'in üzerinde olan portföyler genellikle "iyi", 2'nin üzerindekiler "çok iyi" ve 3'ün üzerindekiler ise "mükemmel" performans olarak değerlendirilir (Francis ve Ibbotson, 2002). Ancak bu eşikler görecelidir ve analiz edilen döneme, varlık sınıfına ya da piyasa koşullarına bağlı olarak değişiklik gösterebilir.

7.3 Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error - MAPE)

Regresyon modellerinin performansını değerlendirmek için kullanılan en yaygın hata metriklerinden biri **Ortalama Mutlak Yüzde Hata (MAPE)**'dır. Bu ölçüt, modelin tahminlerinin gerçek değerlere göre ortalama ne kadar yüzde sapma gösterdiğini ifade eder. MAPE, modelin doğruluğunu yüzdesel bir ölçümle sunduğundan yorumlanması oldukça sezgiseldir (Hyndman ve Koehler, 2006).

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7.3}$$

Burada, y_i, i. gözlemdeki gerçek değer; $\hat{y}_i$, i. gözlemdeki model tarafından tahmin edilen değer; n, toplam gözlem sayısı.

MAPE, tüm örneklerdeki tahmin hatasının gerçek değere oranının ortalamasını alarak yüzdesel hata verir. Bu nedenle modelin "gerçeğe ne kadar yakın" tahmin yaptığını, % cinsinden yorumlamak mümkündür.

- Yorumlaması sezgiseldir ve farklı veri setleri arasında karşılaştırma yapmaya olanak tanır.
- Birimler arası bağımsızdır, çünkü oransal hatayı ifade eder.
- Uygulamada özellikle işletme, finans ve enerji tahminlerinde sıkça tercih edilir.
- Gerçek değerin sıfıra yakın olduğu durumlarda hata oranı sonsuza yaklaşır veya tanımsız hale gelir.
- Küçük gerçek değerler, oransal olarak büyük hatalara yol açtığı için MAPE bu tür veri setlerinde aşırı duyarlıdır.
- Negatif ya da sıfıra yakın değerlerin bulunduğu modellerde uygun değildir.
- Bu nedenlerle, MAPE sıklıkla RMSE (Root Mean Squared Error) veya MAE (Mean Absolute Error) gibi diğer hata metrikleriyle birlikte kullanılır.

8. ARAŞTIRMA BULGULARI

Bu bölümde, çalışmada kullanılan makine öğrenmesi modellerinin tahmin başarıları, değişken önem düzeyleri ve SHAP analizlerine dair bulgular sunulmaktadır. Modellerin performansı ve değişkenlerin tahmin gücü üzerindeki etkileri ayrı ayrı değerlendirilmiştir.

8.1 Rastgele Orman Modeli Sonuçları

Aşağıda, Rastgele Orman modeli ile yapılan analiz sonucunda, finansal değişkenlerin ortalama önem düzeyleri ve SHAP etkileri verilmiştir. Modelin hangi değişkenlere daha fazla ağırlık verdiği ve bu değişkenlerin tahmin gücü üzerindeki katkısı değerlendirilmiştir.

Çizelge 8.1 Rastgele Orman modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri

Değişken Adı	Önemi(Yüzde%)	SHAP Değeri
Cari Oran	0.173	0.009
Aktif Karlılığı	0.172	0.032
Likidite Oranı	0.166	0.008
Fiyat/Satış	0.11	0.050
Faaliyet Kâr Marjı	0.10	0.019
Net Kâr Marjı	0.07	0.016
EBITDA Marjı	0.06	0.011
PD/DD	0.04	0.008
Özsermaye Kârlılığı	0.035	0.008
Borç/Özsermaye	0.032	0.007
Fiyat/Kazanç	0.030	0.011
Borç/Aktif	0.017	0.004

Çizelge 8.1'deki Rastgele Orman modeline göre, **Cari Oran (%17.3)**, **Aktif Karlılık (%17.2)** ve **Likidite Oranı (%16.6)** en yüksek yapısal öneme sahip finansal oranlardır.

Ancak SHAP analizine göre, tahmin sonuçlarına en büyük katkıyı sağlayan değişken **Fiyat/Satış Oranı (0.050)** olmuştur. Bu bulgu, bir değişkenin model tarafından sık kullanılıyor olmasının, her zaman tahmin üzerinde yüksek etkiye sahip olduğu anlamına gelmediğini ortaya koymaktadır. Örneğin, **Cari Oran** yüksek sıklıkla kullanılmış olmasına rağmen tahmin değerine etkisi görece düşüktür ($SHAP = 0.009$). Buna karşılık, **Aktif Karlılık**, hem model tarafından sıkça kullanılmış hem de tahmin üzerinde anlamlı etki göstermiştir ($SHAP = 0.032$). Genel olarak, Rasgele Orman modeli kârlılık ve likidite oranlarına dayanırken, **Fiyat/Satış oranı** gibi değerleme çarpanlarının tahmin başarısında kritik bir rol oynadığı görülmektedir.

Çizelge 8.2 Rastgele Orman modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları

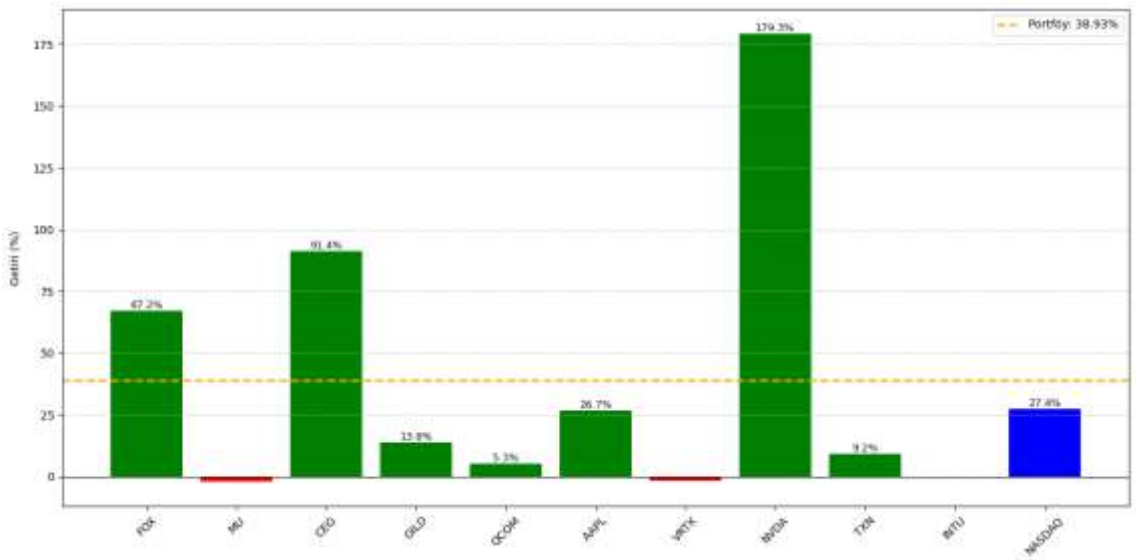
Hisse Kısaltması	Getiri(Yüzde%)	Seçilme Sayısı
FOX	67.23	21
MU	-2.14	19
CEG	91.39	15
GILD	13.84	11
QCOM	5.32	11
AAPL	26.66	9
VRTX	-1.61	9
NVDA	179.29	8
TXN	9.20	8
INTU	0.08	8

Çizelge 8.2 incelendiğinde, Rastgele Orman modeli tarafından en yüksek getiri potansiyeline sahip olduğu tahmin edilen ilk 10 hisse senedinin, çeşitli sektörlerden ve farklı risk profillerine sahip şirketlerden oluştuğu görülmektedir. Bu portföyde en dikkat çekici hisse, **%179.29** ile **NVDA (NVIDIA)** olup, en yüksek tahmini getiriyi sunmaktadır. Bunun yanı sıra, **CEG (%91.39)** ve **FOX (%67.23)** hisseleri de önemli getiri potansiyeline sahip diğer adaylar olarak öne çıkmaktadır.

Ancak portföyde **MU (-%2.14)** ve **VRTX (-%1.61)** gibi negatif getiri tahmini olan hisselerin de yer alması, modelin seçim kriterinin sadece geçmiş performans ve finansal oranlara dayalı olarak yüksek getiri potansiyeli öngördüğünü, kısa vadeli piyasa dalgalanmalarına tam duyarlı olamayabileceğini göstermektedir.

Toplamda portföyün tahmini ortalama getirisi, **NASDAQ 100** endeksinin aynı dönemdeki %27.42'lik ortalama getirisiyle karşılaştırıldığında, seçilen hisselerin bir kısmının endeksin üzerinde getiri sağlama potansiyeli taşıdığı anlaşılmaktadır. Bu durum, Rastgele Orman modelinin piyasa ortalamasının üzerinde getiriler hedefleyebilecek hisse gruplarını başarılı şekilde filtreleyebildiğine işaret ederken, aynı zamanda yüksek getiri beklentisinin yanında belirli ölçüde oynaklık riski de barındırdığını göstermektedir.

Bu nedenle yatırımcının risk tercihlerine bağlı olarak, model tarafından önerilen hisseler, aktif portföy yönetim stratejilerinde getiri potansiyelini artırmak amacıyla kullanılabilir; ancak model çıktıları, piyasa haberleri ve makroekonomik göstergeler gibi güncel bilgilerle desteklenmelidir (Gu vd., 2020; Zhang vd., 2022).



Şekil 8.1 Rastgele Orman modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılaştırması

8.2 XGBoost Modeli Sonuçları

XGBoost modeli, gradyan artırma yaklaşımına dayalı güçlü bir tahmin algoritmasıdır. Çizelge 8.3 ile modele göre hesaplanan değişken önem düzeyleri ve SHAP etkileri karşılaştırmalı olarak verilmiştir.

Çizelge 8.3 XGBoost modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri

Değişken Adı	Önemi(Yüzde%)	SHAP Değeri
Aktif Karlılığı	0.194	0.033
Cari Oran	0.192	0.025
Fiyat/Satış	0.155	0.052
FAVÖK Marjı	0.088	0.018
Faaliyet Kâr Marjı	0.082	0.029
Likidite Oranı	0.076	0.019
Net Kâr Marjı	0.048	0.024
Özsermaye Kârlılığı	0.046	0.024
Borç/Özsermaye	0.038	0.011
PD/DD	0.031	0.018
Fiyat/Kazanç	0.027	0.015
Borç/Aktif	0.022	0.013

XGBoost modelinin değişken önem analizine göre en yüksek öneme sahip finansal oranlar sırasıyla Aktif Karlılık (%19.4), Cari Oran (%19.2) ve Fiyat/Satış oranıdır (%15.5). Ancak SHAP analizine göre tahmin çıktısı üzerinde en yüksek etkiye sahip değişken Fiyat/Satış oranı (0.052) olup, bu değişkenin karar sürecine etkisi önem sırasından daha fazladır. Bu durum, bazı değişkenlerin modelde sık kullanılsa da etkisinin sınırlı olabileceğini ortaya koymaktadır. Genel olarak modelin, kârlılık ve likidite oranlarına daha fazla ağırlık verdiği; borçluluk oranlarının ise karar sürecinde daha az belirleyici olduğu görülmektedir.

Çizelge 8.4 XGBoost modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları

Hisse Kısaltması	Getiri(Yüzde%)	Seçilme Sayısı
FOX	67.23	15
TXN	9.20	14
MU	-2.14	12
CEG	91.39	11
QCOM	5.32	11
GILD	13.84	10
AMD	-18.17	9
EA	6.64	9
AAPL	26.66	8
INTU	0.08	8

Çizelge 8.4 incelendiğinde, XGBoost modeli tarafından en yüksek getiri potansiyeline sahip olduğu tahmin edilen ilk 10 hisse senedinde, teknoloji ve hizmet sektörlerinin ağırlıkta olduğu dikkat çekmektedir. Listenin en yüksek tahmini getiriye sahip hissesi, %91.39 ile CEG olurken, bunu %67.23 getiri ile FOX ve %26.66 ile AAPL (Apple) takip etmektedir. Bu durum, modelin yüksek büyüme potansiyeline sahip ve piyasa değeri yüksek şirketlere odaklandığını göstermektedir.

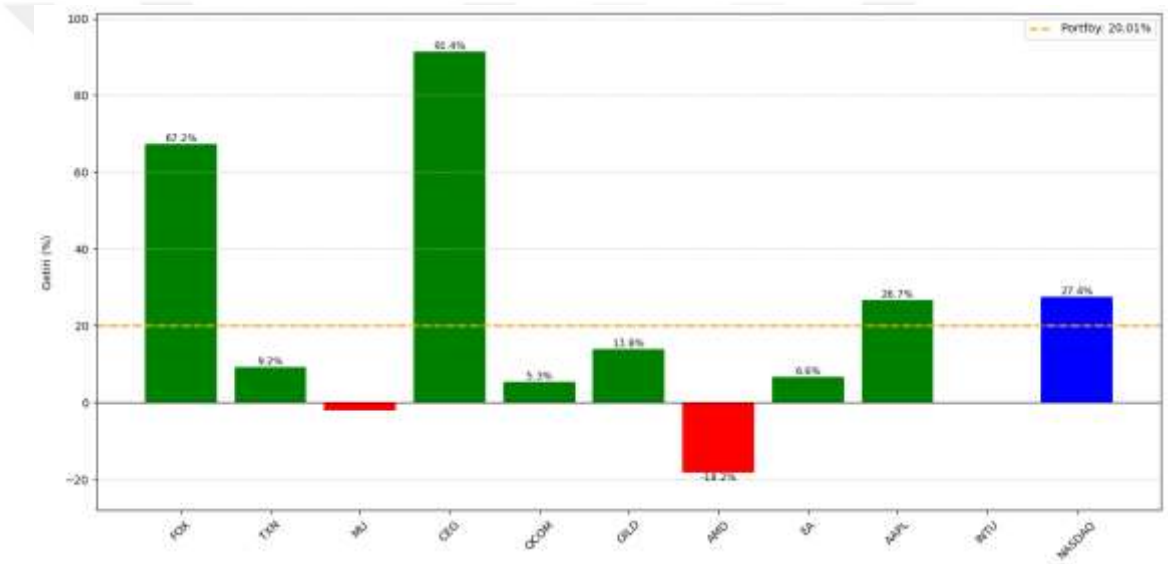
Diğer yandan, MU (-%2.14) ve AMD (-%18.17) gibi bazı hisseler için negatif getiri tahmini yapılmış olmasına rağmen bu hisselerin de ilk 10 içinde yer alması, XGBoost modelinin seçim kriterinin yalnızca tahmini getiriye değil, aynı zamanda değişkenler arası karmaşık ilişkilere ve modelin genel tahmin dengesine dayandığını göstermektedir (Chen ve Guestrin, 2016)

Seçilme sayılarının birbirine yakın dağılması (örneğin FOX: 15, TXN: 14, MU: 12) modelin bazı hisselerde tahmin istikrarını koruduğunu, yani çeyrekler boyunca benzer tahminler ürettiğini göstermektedir. Bu özellik, XGBoost algoritmasının ardışık hata azaltma ve ağırlıklı öğrenme mekanizması sayesinde elde edilen bir avantaj olarak değerlendirilebilir (Friedman, 2001).

Portföyün genel getirisi NASDAQ 100 endeksiyle kıyaslandığında, bazı hisselerin endeks ortalamasının üzerinde performans potansiyeli taşıdığı, ancak negatif getirili hisselerin bu

potansiyeli sınırlayabileceği anlaşılmaktadır. Bu bulgu, XGBoost modelinin yüksek getiriyi hedeflerken, veri içi varyasyonları da koruduğunu ve portföydeki riskin yönetilmesi gerektiğini işaret etmektedir (Lee ve Lee, 2020).

Özetle, XGBoost tarafından önerilen hisse portföyü, yüksek büyüme potansiyeline sahip seçilmiş şirketleri içerirken, yatırımcıların bu önerileri uygularken güncel piyasa haberleri ve risk yönetimi stratejileri ile destekleme gerekliliğini ortaya koymaktadır (Gao vd., 2022).



Şekil 8.2 XGBoost modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması

8.3 LightGBM Modeli Sonuçları

LightGBM modeli, büyük veri setlerinde daha hızlı ve etkin çalışabilen bir boosting algoritmasıdır. Çizelge 8.5 ile bu modele ilişkin değişken önemi ve SHAP çıktıları verilmiştir.

Çizelge 8.5 LightGBM modelinde finansal değişkenlerin ortalama önemi ve SHAP etkileri

Değişken Adı	Önemi(Yüzde%)	SHAP Değeri
Fiyat/Kazanç	0.45	0.019
Fiyat/Satış	0.41	0.038
Faaliyet Kâr Marjı	0.38	0.034
Özsermaye Kârlılığı	0.37	0.026
FAVÖK Marjı	0.34	0.021
Aktif Karlılığı	0.33	0.036
Likidite Oranı	0.20	0.027
PD/DD	0.04	0.013
Net Kâr Marjı	0.035	0.032
Cari Oran	0.032	0.027
Borç/Aktif	0.030	0.017
Borç/Özsermaye	0.017	0.011

LightGBM modelinde değişken önem analizi, **Fiyat/Kazanç (%45)**, **Fiyat/Satış (%41)** ve **Faaliyet Kâr Marjı (%38)** oranlarının karar sürecinde sıklıkla kullanıldığını göstermektedir. Ancak SHAP değerleri dikkate alındığında, modelin tahmin sonuçlarını en fazla etkileyen değişkenlerin **Fiyat/Satış Oranı (0.038)**, **Aktif Karlılık (0.036)** ve **Faaliyet Kâr Marjı (0.034)** olduğu görülmektedir. Özellikle Fiyat/Kazanç oranının sık kullanılmasına rağmen düşük SHAP değerine sahip olması, bu değişkenin karar yapısında yer almasına rağmen nihai tahmine etkisinin sınırlı olduğunu göstermektedir. Bu durum, bazı değişkenlerin modelin içinde yapısal olarak yer alsa bile, tahmin sonucuna etkisinin görece düşük olabileceğini ortaya koymaktadır. Genel olarak LightGBM modeli, **değerleme çarpanları ve kârlılık oranlarına** daha fazla ağırlık verirken, **borçluluk oranlarının** karar sürecine katkısı sınırlı kalmıştır.

Çizelge 8.6 LightGBM modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları

Hisse Kısaltması	Getiri(Yüzde%)	Seçilme Sayısı
FOX	67.23	21
MU	-2.14	13
TXN	9.67	12
EBAY	42.76	10
ROST	10.27	9
QCOM	5.32	9
BKNG	38.85	9
CEG	91.39	9
EA	6.64	9
GILD	13.84	8

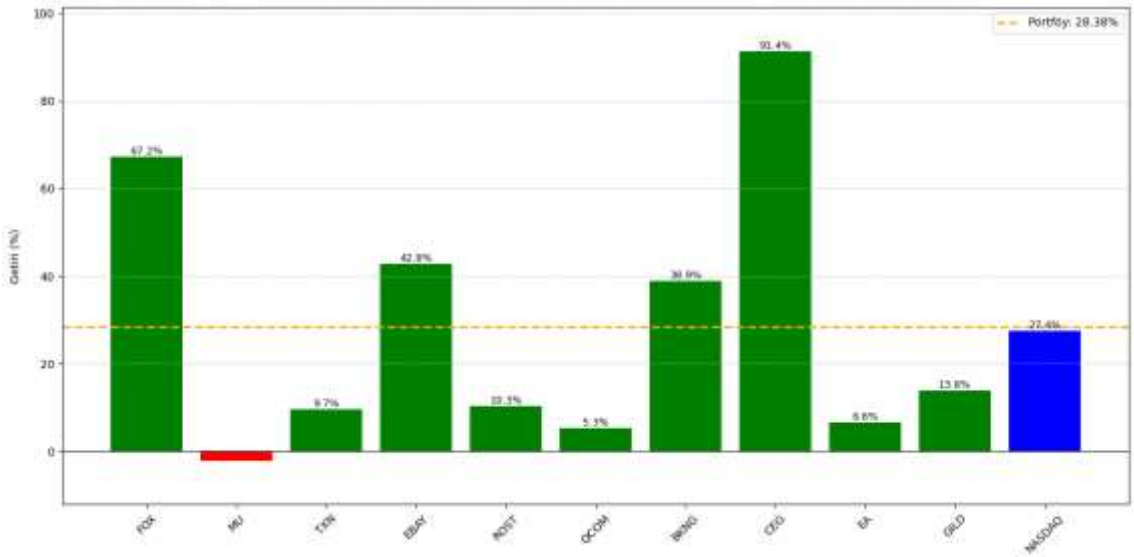
Çizelge 8.6 incelendiğinde, LightGBM modeli tarafından en yüksek getiri potansiyeline sahip olduğu tahmin edilen ilk 10 hisse senedinin, teknoloji, enerji ve tüketici hizmetleri sektörlerinden geldiği görülmektedir. Listenin en yüksek tahmini getirisine sahip hissesi %91.39 ile **CEG (Constellation Energy)** olurken, bunu %67.23 ile **FOX** ve %42.76 ile **EBAY** takip etmektedir. Bu dağılım, LightGBM modelinin yalnızca piyasa değeri yüksek değil, aynı zamanda dönemsel olarak aşırı performans gösterme potansiyeli taşıyan hisselere de yöneldiğini göstermektedir.

Modelin seçtiği hisseler arasında yer alan **MU (-%2.14)** gibi negatif getiri tahmini yapılmış hisselerin bulunması, LightGBM algoritmasının yalnızca pozitif getiri beklentisine göre değil, değişkenler arası kompleks ilişkileri gözeterek seçim yaptığını ortaya koymaktadır. Bu yaklaşım, modelin bazı durumlarda potansiyel geri dönüş sinyallerine odaklandığını ya da geçmiş dönem verilerine göre iyileşme olasılığı barındıran hisseleri de dikkate aldığını göstermektedir.

Seçilme sayıları açısından değerlendirildiğinde, **FOX** hissesi 21 kez seçilerek açık ara öne çıkarken, diğer hisselerin seçilme sayıları daha dengeli bir dağılım göstermektedir

(örneğin TXN: 12, EBAY: 10, BKNG: 9). Bu durum, LightGBM algoritmasının bazı hisselerde tahmin istikrarı sağladığını ve çeyrekler boyunca benzer tercihlerde bulunduğunu göstermektedir. Bu özellik, özellikle leaf-wise büyüme stratejisine dayanan model yapısının karar ağaçlarında belirli desenleri tekrar yakalama eğilimiyle uyumludur (Ke vd., 2017).

Sonuç olarak, LightGBM modelinin oluşturduğu portföyde yüksek getiri potansiyeli taşıyan hisseler yer almakta olup, bu hisselerin bir kısmı NASDAQ 100 endeksinin getiri ortalamasını aşabilecek düzeydedir. Bununla birlikte, negatif getirili hisselerin de listede yer alması, portföyün oynaklık içerebileceğini ve yatırım kararlarında ek filtreleme ya da risk kontrol mekanizmalarının gerekebileceğini göstermektedir.



Şekil 8.3 LightGBM modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması

8.4 Destek Vektör Regresyonu Modeli Sonuçları

Destek Vektör Regresyonu (SVR), doğrusal olmayan karmaşık yapıları modelleyebilmesiyle öne çıkar. Bu başlık altında, SVR modeli ile tahmin edilen getiri sonuçlarına ilişkin özet bulgular sunulmaktadır.

Çizelge 8.7 Destek Vektör Regresyonu modelinin en yüksek getiri potansiyeli tahmin ettiği 10 hisse senedi ve modeldeki seçilme sıklıkları

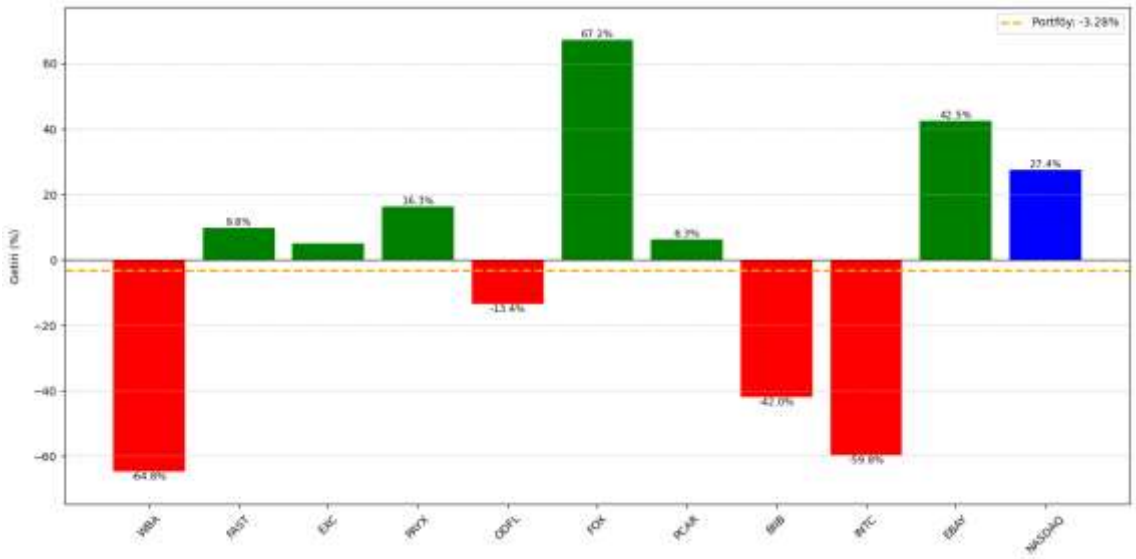
Hisse Kısaltması	Getiri(Yüzde%)	Seçilme Sayısı
WBA	-64.80	18
FAST	9.79	18
EXC	4.99	17
PAYX	16.28	16
ODFL	-13.39	16
FOX	67.23	15
PCAR	6.33	15
BIIB	-42.03	14
INTC	-59.76	13
EBAY	42.51	11

Çizelge 8.7 incelendiğinde, SVR modeli tarafından en yüksek getiri potansiyeline sahip olduğu tahmin edilen ilk 10 hisse senedinin oldukça heterojen bir sektör dağılımına sahip olduğu görülmektedir. Modelin önerdiği hisseler arasında hem pozitif hem de ciddi ölçüde negatif getiri tahminleri dikkat çekmektedir. Listenin en yüksek pozitif getiri tahmini, %67.23 ile **FOX** ve %42.51 ile **EBAY** hisselerine aitken, **WBA (-%64.80)**, **BIIB (-%42.03)** ve **INTC (-%59.76)** gibi hisselerde ciddi oranda negatif getiriler tahmin edilmiştir. Bu durum, SVR modelinin lineer olmayan karmaşık yapıları yakalama potansiyeline rağmen, bazı dönemlerde aşırı uç değerli tahminler üretebildiğini göstermektedir.

Modelin hem pozitif hem de negatif getiri tahmini yaptığı hisseleri ilk 10 portföye dahil etmesi, SVR algoritmasının yalnızca beklenen getiriyi değil, aynı zamanda veri içindeki kalıpları ve sapmaları da göz önünde bulundurarak seçim yaptığına işaret etmektedir (Smola ve Schölkopf, 2004). Bu yaklaşım, destek vektörlerinin özellikle marj dışında kalan gözlemlerden etkilenmeye açık olmasıyla da örtüşmektedir.

Seçilme sıklıklarına bakıldığında, **WBA** ve **FAST** hisseleri 18 kez seçilerek listenin başında yer almaktadır. Bu durum, SVR modelinin bazı hisseleri dönemler boyunca istikrarlı biçimde önermeye devam ettiğini göstermektedir. Bununla birlikte, modelin en çok seçtiği hisselerin bazıları (örneğin **WBA** ve **ODFL**) için negatif getiri tahmin edilmesi, SVR'nin portföy seçiminde potansiyel olarak yanlış sinyalleri de içerebildiğini göstermektedir.

Bu çerçevede değerlendirildiğinde, SVR modeli tarafından oluşturulan portföyde yüksek getirili hisseler yer almakla birlikte, negatif getirili tahminlerin fazlalığı modelin genel performansını sınırlayıcı bir etki yaratabilir. Özellikle **WBA, INTC ve BIIB** gibi hisselerin seçilme sayılarının yüksek olmasına karşın tahmini getirilerinin negatif olması, bu modelin portföy optimizasyonunda diğer modellere kıyasla daha yüksek bir oynaklık ve tahmin belirsizliği barındırabileceğini göstermektedir (Cao ve Tay, 2003).



Şekil 8.4 Destek Vektör Regresyonu modelinin seçtiği hisse senetleri ve NASDAQ 100 karşılatırması

8.5 Karşılaştırmalı Değerlendirme

Uygulanan modellerin tahmin sonuçları doğrultusunda oluşturulan portföylerin getirileri analiz edilmiş, elde edilen sonuçlar NASDAQ 100 endeksiyle karşılaştırılmıştır.

Çizelge 8.8 Modeller tarafından önerilen portföylerin 2024 yılı getirisi

Model	Sharpe Oranı (risksiz getiri %5 olarak)	Portföy Getirisi(%)	NASDAQ 100 Getirisi(%)	Fark (%)
Rastgele Orman	1.94	38.93	27.42	+11.51
XGBoost	0.80	20.01	27.42	-7.41
LightGBM	1.69	28.38	27.42	+0.96
Destek Vektör Reg.	-1.10	-3.29	27.42	-25.10

Rastgele Orman modeli, %38.93'lük portföy getirisi ve **1.94** Sharpe oranı ile en yüksek risk düzeltilmiş performansı sunmuş; NASDAQ 100 endeksinin yıllık %27.42'lik getirisini **11.51 puan aşarak** önemli bir üstünlük sağlamıştır. Bu sonuç, modelin hem yüksek getiri sağlama hem de göreli olarak düşük volatiliteyle çalıştığını göstermektedir. Sharpe oranının 1.5 üzeri olması, portföyün yatırım açısından oldukça güçlü bir performansa sahip olduğunu kanıtlar niteliktedir.

LightGBM modeli de oldukça başarılı sonuçlar vermiştir. %28.38'lik yıllık getirisiyle NASDAQ 100'ü sınırlı bir farkla da olsa **geride bırakmış (+0.96)** ve **1.69'luk Sharpe oranı** ile yüksek düzeyde risk ayarlı getiri sağlamıştır. Bu performans, LightGBM modelinin değerleme ve kârlılık oranlarını etkin biçimde öğrenerek yatırım açısından dengeli bir portföy sunduğunu göstermektedir.

XGBoost modeli, %20.01 getiri ile NASDAQ 100 endeksinin altında kalmıştır (**-7.41 puan fark**). Sharpe oranının **0.80** düzeyinde kalması, modelin pozitif fakat sınırlı bir risk düzeltilmiş getiri sağladığını ortaya koymaktadır. Bu durum, XGBoost'un bazı negatif

getirili hisseleri portföye dahil etmesi ve modelin önyargı düzeyinin dönemsel dalgalanmalara duyarlı olabileceğini düşündürmektedir.

Destek Vektör Regresyonu (SVR) modeli ise portföyünde yüksek zarar üreten hisseleri (örneğin WBA, BIIB, INTC) yoğun biçimde barındırması nedeniyle, **-%3.29** getirisi ve **-1.10** Sharpe oranı ile hem nominal hem de risk ayarlı getiriler bakımından başarısız olmuştur. Bu durum, SVR modelinin hisse senedi seçiminde daha dengesiz bir tahmin davranışı sergileyebileceğini, uç değerlerden ciddi şekilde etkilenebileceğini göstermektedir.

En yüksek Sharpe Oranı (1.94) ile Rastgele Orman modeli, risk başına en yüksek getiriyi sunarak en verimli strateji olarak öne çıkmaktadır.

Model performanslarının değerlendirilmesinde kullanılan RMSE (Root Mean Squared Error) ve MAPE (Mean Absolute Percentage Error) metrikleri, tahmin sonuçlarının istatistiksel doğruluğunu ortaya koymak açısından temel göstergelerdir. Bu çalışmada, özellikle **LightGBM** ve **Rastgele Orman** modelleri, göreli olarak daha düşük RMSE ve MAPE değerleri üretmiş; bu da söz konusu modellerin hisse senedi getirilerini yüksek isabetle tahmin edebildiğini göstermiştir.

Ancak değerlendirme sürecinde, modellerin pencere genişliği yaklaşımıyla eğitilip test edilmesi, hata metriklerinin mutlak düzeyde görece yüksek çıkmasına neden olmuştur. Bu yöntemde eğitim veri seti zamanla artarken, her adımda model geçmiş verilerle yeniden eğitilir ve geleceğe yönelik tahminler yapılır. Dolayısıyla her yeni tahmin dönemi, daha önce hiç görülmemiş veriler üzerinde yapılan bir dış test niteliği taşımaktadır. Bu durum, klasik çapraz doğrulamaya göre daha zorlayıcı bir yapı sunmakta ve hataların gerçek dünya koşullarına daha yakın şekilde ölçülmesini sağlamaktadır (Hyndman ve Athanasopoulos, 2018).

Sonuç olarak, RMSE ve MAPE değerlerinin görece yüksek olması model başarısızlığı olarak değil, kullanılan test yaklaşımının doğası gereği **daha gerçekçi bir hata ölçümü**

sunması olarak değerlendirilmelidir. Bu bağlamda, LightGBM ve Rastgele Orman modelleri, sadece portföy getirisi bakımından değil, aynı zamanda genişleyen zaman aralıklarında gösterdiği öngörü performansıyla da öne çıkmaktadır.

Çizelge 8.9 Modellerde en yüksek SHAP değerine sahip 10 finansal oran

Sıra	Finansal Oran	Ortalama SHAP Değeri	En Çok Etkili Olduğu Model
1	Fiyat/Satış Oranı	0.050	Rastgele Orman
2	Aktif Karlılık	0.036	LightGBM
3	Faaliyet Kâr Marjı	0.034	LightGBM
4	Net Kâr Marjı	0.032	LightGBM
5	Cari Oran	0.027	XGBoost/LightGBM
6	Özsermaye Karlılığı	0.026	LightGBM
7	Fiyat/Kazanç Oranı	0.019	LightGBM
8	FAVÖK Marjı	0.018	XGBoost
9	Likidite Oranı	0.008	Rastgele Orman
10	PD/DD	0.008	Rastgele Orman

Çizelge 8.9, farklı makine öğrenmesi modellerinde hisse senedi getirisi tahminine en çok katkı sağlayan 10 finansal oranı SHAP değerlerine göre sıralamaktadır. En yüksek SHAP değerine sahip oran **Fiyat/Satış Oranı** olup, özellikle Rasgele Orman modelinde tahminler üzerinde belirleyici rol oynamıştır (SHAP: 0.050). Onu, LightGBM modelinde öne çıkan **Aktif Karlılık (0.036)** ve **Faaliyet Kâr Marjı (0.034)** takip etmektedir.

İlginç olarak, bazı oranlar (örneğin, **Cari Oran** ve **Fiyat/Kazanç Oranı**) farklı modellerde sık kullanılmış olsalar da, tahmin çıktısına olan katkıları modelden modele değişkenlik göstermektedir. Örneğin, **Fiyat/Kazanç Oranı**, LightGBM modelinde yüksek önem düzeyine sahipken, SHAP değeri görece düşüktür (0.019). Bu durum, yapay öğrenme modellerinde sadece kullanım sıklığının değil, değişkenin karar üzerindeki marjinal etkisinin de ayrıca değerlendirilmesi gerektiğini göstermektedir.

9. SONUÇ VE TARTIŞMA

Bu çalışmada, hisse senedi getirilerinin tahmini ve yatırım portföylerinin oluşturulması amacıyla dört farklı makine öğrenmesi algoritması (Rastgele Orman, XGBoost, LightGBM ve Destek Vektör Regresyonu) kullanılmış; modellerin tahmin performansları ile finansal değişkenlerin bu performansa etkileri analiz edilmiştir. Her çeyrek yeniden eğitilen modeller, bir sonraki dönemin çeyreklik hisse senedi getirisini tahmin etmiş ve en yüksek potansiyele sahip ilk 10 hisse senedinden oluşan portföyler oluşturulmuştur.

Model sonuçları karşılaştırıldığında, Rastgele Orman ve LightGBM algoritmaları, hem tahmin doğruluğu hem de oluşturulan portföylerin performansı açısından diğer modellere göre daha başarılı sonuçlar vermiştir. Sharpe oranı gibi risk-düzeltilmiş performans ölçütlerinde de bu modellerin daha yüksek değerler sunduğu görülmüştür. Özellikle Rastgele Orman modeli, veri setindeki karmaşık ilişkileri etkili biçimde öğrenerek yüksek isabet oranlarıyla öne çıkmıştır.

Finansal değişkenler açısından, Aktif Karlılık (ROA), Cari Oran ve Fiyat/Satış Oranı modellerde öne çıkan göstergeler olmuştur. Rastgele Orman ve LightGBM modelleri açısından SHAP analizleri incelendiğinde, özellikle Fiyat/Satış Oranı'nın tahminlerde belirleyici bir rol oynadığı, ancak modelin yapısal olarak en çok başvurduğu değişkenin her zaman en etkili değişken olmadığı görülmüştür. Bu durum, geleneksel önem ölçümlerinin SHAP gibi açıklayıcı yöntemlerle desteklenmesinin gerekliliğini göstermektedir.

Modellerin seçtiği portföyler, 2024-2025 dönemi için NASDAQ 100 endeksi ile karşılaştırıldığında önemli ölçüde rekabetçi performanslar sergilemiştir. Bazı dönemlerde, seçilen portföylerin endeksin üzerinde getiriler sağladığı, ancak oynaklığın da artabileceği gözlemlenmiştir. Bu durum, makine öğrenmesi temelli aktif portföy yönetimi stratejilerinin, pasif yatırım stratejilerine alternatif olabileceğini göstermektedir.

Çalışmanın teorik katkısı, farklı makine öğrenmesi algoritmalarının finansal getiri tahminlerinde karşılaştırmalı etkinliğini ortaya koyması ve SHAP analizi gibi modern değişken yorumlama tekniklerinin finansal modellemeye nasıl entegre edilebileceğini göstermesidir. Pratik katkı açısından ise, bu yöntemlerin yatırımcılar ve portföy yöneticileri tarafından kullanılabilecek veriye dayalı hisse seçimi stratejileri geliştirilmesine olanak sunduğu belirtilmelidir.

Sonuç olarak, bu çalışma göstermiştir ki, makine öğrenmesi algoritmaları yalnızca hisse senedi getirilerini tahmin etmekte değil, aynı zamanda veri odaklı yatırım kararlarının temellendirilmesinde de etkin biçimde kullanılabilir. Özellikle SHAP gibi yöntemlerle açıklanabilirliğin artırılması, bu modellerin kurumsal yatırımcılar nezdinde kabul edilebilirliğini artırabilir. MU, AMD ve VRTX gibi seçilen bazı hisselerin negatif getiri üretmesi, modellerin geçmiş performans odaklı seçimlerinin her zaman gelecek getirileri doğru tahmin edemediğini ortaya koymuştur. 2024-2025 döneminde MU, artan yarı iletken arzı ve bellek çipi fiyat baskıları; AMD, talep dalgalanmaları ve yoğun rekabet; VRTX ise ilaç onay süreçlerinde yaşanan belirsizlikler ve AR-GE giderlerinin artması nedeniyle değer kaybı yaşamış ve modellerin pozitif sinyallerine rağmen eksi getiri üretmiştir. Bu durum, makine öğrenmesi tabanlı portföy seçim stratejilerinde makroekonomik ve sektörel risklerin dikkate alınmasının önemini vurgulamaktadır**.** Gelecek çalışmalarda, makroekonomik göstergeler, haber verisi veya duyarlılık analizleri gibi dışsal faktörlerin entegre edilmesiyle model performansının daha da iyileştirilebileceği öngörülmektedir.

Model performansı, makroekonomik göstergelerin (enflasyon, faiz, büyüme verileri) eklenmesiyle artırılabilir. Zaman serisi analizinde, pencere genişliği yöntemi dışında rolling window gibi alternatif yöntemlerin kullanımı, mevsimsellik etkilerinin daha net ortaya konmasını sağlayabilir.Derin öğrenme tabanlı modeller (LSTM, Transformer) ile mevcut modellerin performans karşılaştırmaları yapılabilir.Finansal veriler dışında, haber duyarlılığı (sentiment analysis) gibi dışsal metin verileri modele entegre edilebilir.Alternatif piyasa endeksleri ile karşılaştırmalı getiri analizleri yapılarak portföy performansının daha geniş ölçekte değerlendirilmesi sağlanabilir.

KAYNAKLAR

Bodie, Z., Kane, A., & Marcus, A. J. (2014). Investments (10th ed.). New York: McGraw-Hill Education.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Wadsworth International Group.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Candar, M. (2022). Equity Portfolio Optimization Using Reinforcement Learning: An Emerging Market Case (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, Endüstri Mühendisliği Anabilim Dalı.

Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. IEEE Transactions on Neural Networks, 14(6), 1506–1518.

Chaweewanchon, A. & Chaysiri, R. (2022). Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. International Journal of Financial Studies, 10(3), 64.

Chen, S.-H., & Yeh, C.-H. (2000). Genetic programming in the simulation of a stock market. Information Sciences, 122(2), 151–170.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference.

Damodaran, A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of Any Asset (3rd ed.). Wiley.

De Prado, M. L. (2020). Advances in financial machine learning. Wiley.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 20(177), 1–81.

Fischer, T., Krauss, C., & Treleaven, P. (2019). Deep reinforcement learning in high frequency trading: A review and potential applications. Quantitative Finance, 19(5), 833–849.

Fridson, M. S., & Alvarez, F. (2011). Financial Statement Analysis (4th ed.). Wiley.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

Gao, J., Guo, H., & Xu, X. (2022). Multifactor Stock Selection Strategy Based on Machine Learning: Evidence from China. Complexity, 2022, Article ID 7447229.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (2nd ed.). Springer.

Hu, Y., An, Y., Wang, S., & Li, J. (2022). Interpreting the black box of a stock price crash risk prediction model based on machine learning. Financial Innovation, 8(1), 106.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. Applied Statistics, 29(2), 119–127.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

Kim, Y., & Lee, H. (2021). Explainable deep learning models for stock price prediction using SHAP. *Journal of Finance and Data Science*, 7(1), 10–22

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689–702.

Lee, J., & Lee, S. (2020). Stock return prediction using a multi-factor model: Evidence from the Korean stock market. Sustainability, 12(5), 2037.

Li, X., & Ma, X. (2019). Forecasting stock returns with machine learning algorithms: Evidence from China. *International Journal of Forecasting*, 35(3), 1109–1124.

Louppe, G. (2014). Understanding Random Forests: From Theory to Practice (Doctoral dissertation, University of Liège).

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS), 30.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair,S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67.

Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91.

Metsomäki, J. (2020). Relation between Financial Ratios and Stock Returns: Machine Learning Approach (Yüksek Lisans Tezi). Lappeenranta-Lahti University of Technology LUT, School of Business and Management.

Molnar,C. (2022). Interpretable Machine Learning. *https://christophm.github*.io/ interpretable-ml-book, Erişim Tarihi: 10.06.2025

Ouyang, Z. (2022). A Study of Stock Portfolio Strategy Based on Machine Learning. 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022). Atlantis Press.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.

Penman, S. H. (2013). Financial Statement Analysis and Security Valuation (5th ed.). McGraw-Hill Education.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341–360.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.

Shapley, L. S. (1953). A Value for n-Person Games. In Contributions to the Theory of Games II (pp. 307–317). Princeton University Press.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425–442.

Sharpe, W. F. (1966). Mutual Fund Performance. The Journal of Business, 39(1), 119–138.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199–222.

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics, 8(1), 25.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer.

White, G. I., Sondhi, A. C., & Fried, D. (2003). The Analysis and Use of Financial Statements (3rd ed.). Wiley.

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, 36(2), 2473–2480.

Zhang, D., Lin, X., & Wang, F. (2022). A novel stock selection model based on explainable machine learning. Financial Innovation, 8(1), 108.

EK 1 Modellerin Seçtiği NASDAQ 100 Şirketlerinin Listesi ve Sektör Bilgileri

Sembol	Şirket Açılımı	Sektör
AAPL	Apple Inc.	Teknoloji (Tüketici Elektroniği)
AMD	Advanced Micro Devices, Inc.	Teknoloji (Yarı İletken)
BIIB	Biogen Inc.	Sağlık (Biyoteknoloji)
BKNG	Booking Holdings Inc.	Tüketici Hizmetleri (Seyahat)
CEG	Constellation Energy Corporation	Enerji
EBAY	eBay Inc.	Tüketici Hizmetleri (E-ticaret)
EA	Electronic Arts Inc.	İletişim (Video Oyunları)
EXC	Exelon Corporation	Enerji (Utilities)
FAST	Fastenal Company	Endüstriyel (Tedarik Zinciri)
FOX	Fox Corporation (Class B)	İletişim (Medya)
GILD	Gilead Sciences, Inc.	Sağlık (Biyoteknoloji)
INTC	Intel Corporation	Teknoloji (Yarı İletken)
INTU	Intuit Inc.	Teknoloji (Yazılım)
MU	Micron Technology, Inc.	Teknoloji (Yarı İletken)
NVDA	NVIDIA Corporation	Teknoloji (Yarı İletken)
ODFL	Old Dominion Freight Line	Endüstriyel (Nakliye)
PAYX	Paychex, Inc.	Finans (İnsan Kaynakları)
PCAR	PACCAR Inc.	Endüstriyel (Ağır Ekipman)
QCOM	Qualcomm Inc.	Teknoloji (Yarı İletken)
ROST	Ross Stores, Inc.	Tüketici (Perakende)
TXN	Texas Instruments Incorporated	Teknoloji (Yarı İletken)
VRTX	Vertex Pharmaceuticals Inc.	Sağlık (Biyoteknoloji)
WBA	Walgreens Boots Alliance, Inc.	Tüketici (Perakende Eczacılık)