

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ALGORİTMİK TİCARETTE KANTİTATİF YATIRIM STRATEJİLERİ VE
MAKİNE ÖĞRENİMİ KULLANARAK FİNANSAL KARAR ALMA
YETENEĞİNİN GÜÇLENDİRİLMESİ

Hammam Jameel Mahmoud ABURAI

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2025

Her hakkı saklıdır

ÖZET

Yüksek Lisans Tezi

ALGORİTMİK TİCARETTE KANTİTATİF YATIRIM STRATEJİLERİ VE MAKİNE ÖĞRENİMİ KULLANARAK FİNANSAL KARAR ALMA YETENEĞİNİN GÜÇLENDİRİLMESİ

Hamam Jameel Mahmoud ABURAI

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. İman ASKERBEYLİ

Makine öğrenimi tekniklerinin finansal piyasalara entegrasyonu, geleneksel ticaret ve risk yönetimini dönüştürmüştür. Fiyat Eylem Stratejisi, Temel Analiz Stratejisi ve Yüksek Frekans Stratejisi gibi çeşitli algoritmik ticaret stratejilerinin mevcut olmasına rağmen, bilinçli alım veya satım kararları için hisse senedi hareketlerini doğru bir şekilde tahmin etmek zor olmaya devam etmektedir. Bu tez, S&P500 hisse senedi hareketlerini tahmin etmek için Teknik Göstergeler Stratejisine odaklanarak bu tahmin sorununu ele almaktadır. Teknik Göstergeler Stratejisi, bir hisse senedinin fiyat ve hacim verilerine dayalı matematiksel hesaplamalar kullanarak ticarete alım ve satım kararları vermek için kullanılır. Yatırımcıların kalıpları ve gelecekteki potansiyel fiyat hareketlerini belirlemelerine yardımcı olur. Yahoo Finance API'den hisse senedi fiyat verilerini topladık ve Bollinger Bantları, MACD ve RSI gibi çeşitli teknik göstergelerle zenginleştirdik. Daha sonra hisse senetlerinin aylık getirilerini farklı zaman ufukları için özellik olarak hesapladık. Verileri Göreceli Güç Endeksi (RSI) ve diğer özelliklere göre etiketlemek için bir kendi kendini etiketleme yöntemi kullandık. Ardından test ve değerlendirme için spektral kümeleme modelimizi oluşturduk. Yeni kendi kendini etiketleme yaklaşımımız, %89'un üzerinde test doğruluğu, %87'nin üzerinde hassasiyet ve %86'nın üzerinde geri çağırma ile umut verici sonuçlar gösterdi. Bu araştırma, nicel stratejileri makine öğrenimi ile birleştirmenin veri odaklı finansal karar verme sürecini önemli ölçüde geliştirebileceğini göstermektedir.

Ocak 2025, 58 sayfa

Anahtar Kelimeler: Algoritmik ticareti, Kantitatif Yatırım Stratejileri, Makine Öğrenimi, Finansal Araç Seti Geliştirme, Göstergeler, S&P 500 Endeksi, Hisse Senedi Piyasası Tahmini.

ABSTRACT

Master Thesis

EMPOWERING FINANCIAL DECISION-MAKING THROUGH LEVERAGING QUANTITATIVE INVESTMENT STRATEGIES AND MACHINE LEARNING IN ALGORITHMIC TRADING

Hamam Jameel Mahmoud ABURAI

Ankara University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Prof. Dr. İman ASKERBEYLİ

The integration of machine learning techniques in financial markets has transformed traditional trading and risk management. Despite the availability of various algorithmic trading strategies, such as Price Action Strategy, Fundamental Analysis Strategy, and High-Frequency Strategy, accurately predicting stock movements for informed buy or sell decisions remains challenging. This thesis addresses this prediction issue by focusing on the technical indicators strategy to predict S&P500 stock movements. Technical indicators strategy is used to make buy and sell decisions in trading by using mathematical calculations based on a stock's price and volume data. It helps traders identify patterns and potential future price movements. We collected stock price data from the Yahoo Finance API and enriched it with several technical indicators such as Bollinger Bands, MACD, and RSI. Then we calculated the monthly returns of the stocks for different time horizon as features. We used a self-labeling method to label the data based on the Relative Strength Index (RSI) and other features. Then we build our spectral clustering model for testing and evaluation. Our novel self-labeling approach showed promising results, with test accuracy over 89%, precision over 87%, and recall over 86%. This research demonstrates that combining quantitative strategies with machine learning can significantly enhance data-driven financial decision-making.

January 2025, 58 pages

Key Words: Algorithmic Trading, Quantitative Investment Strategies, Machine Learning, Financial Toolkit Enhancement, S&P 500 Index Fund, Stock Market Prediction.

TEŞEKKÜR

Her şeyden önce, bu tezi tamamlamam için bana güç, bilgi ve sabır veren Yüce Allah'a şükürler olsun. Onun ilahi rehberliği ve lütfu olmadan bu başarı mümkün olmazdı.

Saygıdeğer hocam, Prof. Dr. İman ASKERBEYLİ'e en derin minnettarlığımı ifade etmek isterim. Bu süreç boyunca verdiğiniz paha biçilemez rehberlik, değerli geri bildirimleriniz ve sarsılmaz desteğiniz bu tezi şekillendirmede çok önemli rol oynadı. Uzmanlığınızdans ders almaktan inanılmaz derecede mutluyum ve sabrınız ve teşvikleriniz için minnettarım.

Ankara Üniversitesi'ndeki tüm hocalarıma, özellikle Bilgisayar Mühendisliği Bölümü'ndeki hocalarıma en içten teşekkürlerimi sunuyorum. Çalışmalarım sırasında edindiğim bilgi ve temel, bu araştırmada çok önemli bir rol oynadı.

Akademik çalışmalarım boyunca bana verdikleri sevgi ve destek için aileme en derin şükranlarımı sunuyorum. Sürekli teşvikleri ve bana olan inançları motivasyonumu artırdı ve zorlukların üstesinden gelmeme yardımcı oldu.

Son olarak, anlayışları ve destekleri için özel bir teşekkür de arkadaşlarıma ve meslektaşlarıma borçluyum. Varlıkları bu yolculuğu daha keyifli hale getirdi.

Hamam Jameel Mahmoud ABURAI

Ankara, Ocak 2025

İÇİNDEKİLER

TEZ ONAY SAYFASI

ETİK.....	i
ÖZET.....	ii
ABSTRACT	iii
TEŞEKKÜR	iv
KISALTMALAR DİZİNİ.....	vii
ŞEKİLLER DİZİNİ	viii
1. GİRİŞ	1
2. TEORİK TEMELLER & İLGİLİ ÇALIŞMALAR	5
2.1 Teorik Temeller.....	5
2.1.1 Ticaret algoritmalarına giriş.....	5
2.1.2 Finansal piyasalar terminolojileri	13
2.1.3 Gösterge çeşitleri	15
2.1.4 Makine öğrenimine giriş.....	19
2.1.5 Finans ve algoritmik ticaret uygulamaları	20
2.1.6 Makine öğrenimi modelleri	21
2.2 İlgili Çalışmalar	23
3. MATERYAL ve METOD.....	26
3.1 Veri Seti	26
3.1.1 Veri kaynağı ve veri tanımı	26
3.1.2 Verilerin ön işlenmesi.....	29
3.2 Özellik Mühendisliği	32
3.2.1 Teknik göstergeler	32
3.2.2 Özellik normalleştirilmesi.....	35
3.2.3 Aylık getiriler	35
3.3 Makine Öğrenimi Modelleri.....	36
3.3.1 K-Means ile kendi kendini etiketleme	36
3.3.2 Spektral kümeleme	39
3.4 Donanım ve Yazılım	41
3.4.1 Yazılım	41
3.4.2 Donanım	42
4. BULGULAR ve TARTIŞMA	43

4.1 Bulgulara Giriş	43
4.2 Sonuçların Sunumu.....	43
4.2.1 K-Means modeli	43
4.2.2 Spektral modeli	44
4.3 Tartışma	49
4.5 Kısıtlamalar	51
5. SONUÇ	53
KAYNAKLAR.....	55
ÖZGEÇMİŞ.....	58



KISALTMALAR DİZİNİ

MACD	Areketli Ortalama Yakınsama Uzaklaşma (Moving Average Convergence Divergence)
RSI	Göreceli Güç Endeksi (Relative Strength Index)
BB	Bollinger Bantları (Bollinger Bands)
ATR	Ortalama Gerçek Aralık (Average True Range)
GKV	Garman Klass Oynaklığı (Garman Klass Volatility)
LR	Lojistik Regresyon (Logistic Regression)
GDA	Gauss Diskriminant Analizi (Gaussian Discriminant Analysis)
NB	Naive Bayes sınıflandırıcısı (Naive Bayes)
SVM	Destek Vektör Makinesi (Support Vector Machine)
RBF	Radyal Temel Fonksiyonu (Radial Basis Function)
LSTM	Uzun Kısa Süreli Bellek (Long Short-Term Memory)
XGB	Extreme Gradyan Arttırma (Extreme Gradient Boosting)
MSE	Ortalama Kare Hatası (Mean Squared Error)
RMSE	Kök Ortalama Kare Hatası (Root Mean Squared Error)
N-BEATS	Sinir Temeli Genişleme Analizi Zaman Serisi (Neural Basis Expansion Analysis Time Series)

ŞEKİLLER DİZİNİ

Şekil 2.1 Sınıflandırma ve kümeleme arasındaki fark	20
Şekil 3.1 APPL hisse senedi verileri 1	30
Şekil 3.2 APPL hisse senedi verileri	30
Şekil 3.3 APPL hisse senedi verileri	31
Şekil 3.4 APPL hisse senedi verileri	31
Şekil 3.5 APPL hisse senedi verileri	31
Şekil 3.6 APPL hisse senedi için RSI.....	33
Şekil 3.7 APPL hisse senedi için BB.....	33
Şekil 3.8 APPL hisse senedi için MACD.....	34
Şekil 3.9 APPL hisse senedi için ATR.....	34
Şekil 3.10 Veri seti	35
Şekil 3.11 Veri seti	36
Şekil 3.12 Etiketli veri seti.....	38
Şekil 4.1 Model değerlendirme metrikleri	44
Şekil 4.2 Model görselleştirme	45
Şekil 4.3 PCA'ya dayalı özelliklerin önemi	48
Şekil 4.4 PCA'ya dayalı özelliklerin önemi	49
Şekil 4.5 Mevcut çalışmalarla ile karşılaştırma	51

1. GİRİŞ

Finansal piyasalar, yatırımcıların finansal refahları ve daha geniş ekonomi üzerinde önemli etkileri olan kararlar aldıkları dinamik ekosistemlerdir. Güçlü finansal bilgi ve karar verme becerilerine sahip olmak, bireylerin seçenekleri tartmalarına ve nasıl ve ne zaman tasarruf edip harcayacaklarına karar vermelerine yardımcı olur. Büyük bir satın almanın maliyetlerini önceden karşılaştırmak ve emeklilik veya diğer uzun vadeli tasarruflar için plan yapmak gibi finansal durumlar için bilinçli seçimler yapmalarını sağlar.

Son yıllarda, finansal karar verme ortamı, teknolojiye ilerlemeler ve verilerin artan kullanılabilirliği ile beslenen derin dönüşümlerden geçmiştir. Bu değişim, yatırımcıların karmaşık piyasa koşullarında daha iyi kararlar alabilmeleri için erişilebilir ve etkili araçlara olan talebi vurgulamaktadır. Geleneksel yatırım stratejileri, bireysel yatırımcılar için genellikle önemli zorluklar teşkil eder. Büyük miktarda finansal veriyi analiz etmek, piyasa trendlerini takip etmek ve kapsamlı araştırmalar yapmak, özellikle yoğun programları olanlar için son derece zaman alıcı olabilir. Ayrıca, duygusal önyargılar muhakemeyi gölgeleyebilir ve optimal olmayan kararlara yol açabilir.

Bu gibi nedenlerle, geleneksel yatırım ve alım satım yöntemleri, büyük miktarlardaki finansal verilerden eyleme geçirilebilir içgörüler çıkarmayı vaat eden sofistike algoritmalar ve makine öğrenimi modelleri ile değiştirilmese de güçlendirilmektedir. Algoritmik ticaret, otomatik ticaret olarak da bilinir, önceden tanımlanmış kural ve talimat setlerine dayalı olarak ticaret yapmak için bilgisayar programlarının kullanılması anlamına gelir. Algoritmik ticaret stratejileri arasında Fiyat Eylem Stratejisi, Temel Analiz Stratejisi, Yüksek Frekans Stratejisi ve Teknik Göstergeler Stratejisi gibi çeşitli yaklaşımlar bulunur. Fiyat Eylem Stratejisi, fiyat hareketlerine dayalı olarak ticaret yapmayı içerir. Bu strateji, geçmiş fiyat hareketlerine bakarak gelecekteki fiyat hareketlerini tahmin etmeye çalışır. Yatırımcılar, belirli kalıpları ve trendleri tanımlayarak alım ve satım kararları verirler. Fiyat eylemi stratejisi, göstergeler veya diğer analiz araçları kullanmadan sadece fiyat ve hacim verilerine dayanarak ticaret yapmayı amaçlar. Temel Analiz Stratejisi, bir şirketin finansal durumunu, yönetimini ve piyasa konumunu

değerlendirerek hisse senetlerinin gerçek değerini belirlemeye çalışır. Bu strateji, şirketlerin finansal raporlarını, ekonomik göstergeleri ve sektörel trendleri analiz ederek hisse senetlerinin alım veya satım kararlarını verir. Temel analiz, uzun vadeli yatırımcılar için özellikle önemlidir çünkü şirketlerin uzun vadeli performansını tahmin etmeye çalışır. Yüksek Frekans Stratejisi, saniyeler veya daha kısa süreler içinde yüksek sayıda işlem yapmayı içerir. Bu strateji, küçük fiyat farklarından yararlanarak kısa vadeli kazançlar elde etmeyi amaçlar. Yüksek frekans ticareti, genellikle karmaşık algoritmalar ve yüksek hızlı bilgisayarlar kullanarak gerçekleştirilir ve büyük miktarda veri işleme kapasitesi gerektirir. Bu strateji, piyasadaki en küçük fiyat hareketlerini bile yakalayarak kar elde etmeye çalışır. Teknik Göstergeler Stratejisi, bir hisse senedinin fiyat ve hacim verilerine dayalı matematiksel hesaplamalar kullanarak ticarete alım ve satım kararları vermek için kullanılır. Teknik göstergeler, yatırımcıların kalıpları ve gelecekteki potansiyel fiyat hareketlerini belirlemelerine yardımcı olur. Bu göstergeler arasında Göreceli Güç Endeksi (RSI), Bollinger Bantları, MACD (Hareketli Ortalama Yakınsama İraksama) ve ATR (Ortalama Gerçek Aralık) gibi çeşitli göstergeler bulunur. Bu strateji, yatırımcılara piyasa eğilimlerini ve potansiyel geri dönüş noktalarını tanımlamada yardımcı olur.

Bu stratejiler, yatırımcıların piyasadaki fırsatları belirlemelerine ve riskleri yönetmelerine yardımcı olmayı amaçlar. Ancak, bu stratejiler içinde bile, bilinçli alım veya satım kararları için hisse senedi hareketlerini doğru bir şekilde tahmin etmek zor olmaya devam etmektedir.

Bu tez, S&P500 hisse senedi hareketlerini tahmin etmek için teknik göstergeler stratejisini ve makine öğrenimi tekniklerini birleştirerek, yatırımcıların daha bilinçli ve veriye dayalı kararlar almasına yardımcı olmayı amaçlamaktadır. Yahoo Finance API'den hisse senedi fiyat verilerini toplayarak, Bollinger Bantları, MACD ve RSI gibi çeşitli teknik göstergelerle zenginleştirdik. Hisse senetlerinin aylık getirilerini farklı zaman ufukları için özellik olarak hesapladık. Tezimizde iki ana model geliştirdik: İlk olarak, verileri etiketlemek için bir K-ortalama kümeleme modeli kullandık. Bu model, teknik göstergeleri dikkate alarak verileri kategorize etti. İkinci olarak, test ve değerlendirme için bir Spektral Kümeleme modeli oluşturduk. Bu model, kendi kendine etiketleme

yöntemiyle oluşturulan verileri kullanarak yüksek doğruluk, hassasiyet ve geri çağırma oranları elde etti.

Yahoo Finance API: Bu API, kullanıcıların hisse senedi fiyatları ve finansal veriler gibi piyasa verilerine erişimini sağlayan bir araçtır. Yahoo Finance API kullanılarak, çeşitli hisse senetlerinin tarihsel fiyat bilgileri, işlem hacmi ve diğer finansal metrikler toplanabilir. Tezimizde bu API, S&P500 hisse senetleri için fiyat verilerini toplamak amacıyla kullanılmıştır.

Bollinger Bantları: Bu teknik gösterge, bir hisse senedinin fiyat hareketlerinin volatilitasını değerlendirmek için kullanılır. Bollinger Bantları, bir hareketli ortalama etrafında iki standart sapma bantı oluşturur. Fiyat, bu bantların dışında hareket ettiğinde aşırı alım veya aşırı satım koşulları hakkında bilgi verir. Tezimizde, Bollinger Bantları, hisse senedi fiyatlarının aşırı değerlenip değerlenmediğini belirlemek için kullanılmıştır.

MACD (Moving Average Convergence Divergence): MACD, bir hisse senedinin kısa ve uzun vadeli hareketli ortalamalarının arasındaki farkı ölçen bir trend takip göstergesidir. Bu gösterge, potansiyel alım ve satım sinyalleri üretmek için kullanılır. MACD çizgisi ve sinyal çizgisi arasındaki kesişmeler, piyasa momentumunu ve trend yönünü gösterebilir. Tezimizde, MACD, hisse senedi fiyatlarının yönü ve momentumunu değerlendirmek için kullanılmıştır.

RSI (Relative Strength Index): RSI, bir hisse senedinin aşırı alım veya aşırı satım koşullarını belirlemek için kullanılan bir momentum göstergesidir. 0 ile 100 arasında değişen bu gösterge, genellikle 70'in üzerindeki değerler aşırı alım, 30'un altındaki değerler ise aşırı satım olarak değerlendirilir. Tezimizde, RSI, hisse senedi fiyatlarının mevcut aşamasını ve olası geri dönüş sinyallerini belirlemek amacıyla kullanılmıştır.

Sonuç olarak, bu tez, makine öğrenimi tekniklerinin ve teknik göstergeler stratejisinin birleştirilmesinin, yatırımcıların finansal kararlarını güçlendirebileceğini ve piyasa trendlerini daha doğru bir şekilde tahmin edebileceğini göstermektedir. Bu araştırma,

algoritmik ticaret ve veri odaklı finansal karar verme süreçlerine önemli katkılarda bulunmayı amaçlamaktadır.

Bu tez şu şekilde yapılandırılmıştır: Bölüm 1'de borsada sağlıklı finansal karar almanın önemi ve bireysel yatırımcıların karşılaştığı zorluklar vurgulanmıştır. Bölüm 2'de bu araştırmanın teorik temelleri ve makine öğrenimi, algoritmik ticaret kavramları ve stratejileri ele alınmıştır. Bölüm 3, S&P500 hisse senetleri için veri toplama, veri temizleme, özellik mühendisliği ve kullanılan makine öğrenimi modellerini detaylandırmaktadır. Bölüm 4, uygulamamızdan elde edilen bulguları ve modellerin performansını sunmaktadır. Sonuçlar bölümünde ise araştırmadan elde edilen temel çıkarımlar özetlenmekte ve gelecekteki araştırma yönleri tartışılmaktadır.

2. TEORİK TEMELLER & İLGİLİ ÇALIŞMALAR

2.1 Teorik Temeller

2.1.1 Ticaret Algoritmalarına giriş

Genellikle algo ticareti veya otomatik ticaret olarak adlandırılan algoritmik ticaret, finansal piyasaların manzarasını değiştirerek menkul kıymetlerin alınıp satılma biçiminde devrim yaratmıştır. Bu bölümde, algoritmik ticaretin temellerini, gelişimini ve tarihini, avantajlarını, zorluklarını ve çeşitli ticaret stratejilerini inceleyeceğiz.

Algoritma ve Algoritmik Ticaretin Tanımı: Bir algoritma, değerleri girdi olarak alan ve değerleri çıktı olarak üreten iyi tanımlanmış herhangi bir hesaplama prosedürüdür. Bir başka tanıma göre algoritma, belirli bir görevi yerine getirmek için birisi tarafından yazılan bir dizi talimattır.

Algoritmik ticaret (otomatik ticaret, kara kutu ticareti veya algo ticareti olarak da adlandırılır), bir işlem yapmak için tanımlanmış bir dizi talimatı (bir algoritma) izleyen bir bilgisayar programı kullanır. Ticaret, teorik olarak, bir insan tüccar için imkânsız olan bir hız ve sıklıkta kar üretebilir. Tanımlanan talimat setleri zamanlama, fiyat, miktar veya herhangi bir matematiksel modele dayanır. Tüccar için kâr fırsatlarının yanı sıra, algo ticareti, insan duygularının ticaret faaliyetleri üzerindeki etkisini ortadan kaldırarak piyasaları daha likit ve ticareti daha sistematik hale getirir. Kısacası Algoritmik Ticaret, yatırım kararları almak için bilgisayarların kullanılması anlamına gelmektedir.

Algoritmik Ticaretin Tarihi: Finansal piyasalarda algoritmaların kullanımı, borsalarda işlem yapmak için basit kural tabanlı sistemlerin kullanıldığı 1970'lere kadar uzanmaktadır. Bu ilk algoritmalar nispeten basitti ve öncelikle mevcut en iyi fiyatlardan işlem yapmak için kullanılıyordu. 1980'lerde daha sofistike algoritmalar geliştirilmeye başlandı ve piyasa verilerini analiz etmek ve alım satım fırsatlarını belirlemek için bilgisayarların kullanımı daha yaygın hale geldi. Bu dönemde ayrıca, yatırımcıların insan

aracilar yerine elektronik ortamda emir girmelerine ve işlem yapmalarına olanak tanıyan elektronik alım satım sistemleri de kullanılmaya başlanmıştır. 1990'larda elektronik ticaret platformlarının yaygın olarak benimsenmesi ve 2000'lerde teknoloji ve veri analizinde daha fazla ilerleme kaydedilmesi, daha karmaşık ve sofistike algoritmaların geliştirilmesine yol açmıştır. Algoritmik ticarete makine öğrenimi ve yapay zekâ kullanımı da bu dönemde ortaya çıkmaya başlamıştır. Son yıllarda, finansal piyasalarda adaleti ve şeffaflığı sağlamaya çalışan düzenleyicilerle birlikte algoritmik ticarete yönelik düzenleyici incelemeler artmıştır. Bu artan incelemeye rağmen, algoritmik ticaret finansal piyasalarda popüler ve etkili bir güç olmaya devam etmektedir.

Algoritmik Ticaretin Avantaj ve Dezavantajları:

Algo-Ticaretinin Avantajları: İşlemler bir bilgisayar programı tarafından yürütüldüğünden, algo ticaretinin geleneksel çevrimiçi ticaret stratejilerine göre bir dizi avantajı vardır. Öyleyse, bazı önemli faydalarını okuyalım.

Anlık Hızlar: Algoritmalar, milisaniyeler ve mikro saniyeler içinde anlık hızlarda işlem yapabilmektedir. Motor organlarıyla ticaret yapan bir insan asla algo yazılımının hızında ticaret yapamaz. Bu nedenle algoritmik yatırımcılar menkul kıymet fiyatlarındaki en küçük dalgalanmalardan bile faydalanabilmektedir. Aynı zamanda, algoritmalar bir grafiği bir saniye içinde bile analiz edebilir.

Yüksek Doğruluk: Algo ticareti, alım veya satım kararları vermek için insan müdahalesi gerektirmediğinden, algo ticareti çok daha yüksek bir doğruluğa sahiptir. İnsan kaynaklı tüm hatalardan arındırılmışlardır. Örneğin, algoritma işlem yapılması gereken birim miktarını yanlış girmeyecektir. Her zaman doğru sayıda sıfır girecek ve dikkatsizce 1000 yerine 100 birim işlem yapmayacaktır.

Duygusal Müdahalelerden Arındırılmış: Geleneksel çevrimiçi ticaret stratejilerini kullanan bir tüccar olarak, hangi stratejiyi kullanırsanız kullanın, duygularınız işin içine girerse her şey alt üst olabilir. Duygularımız stratejiyi rayından çıkarabilir ve disiplini

bozarak olumsuz sonuçlara yol açabilir. Ancak algoritmik ticaret, bir bilgisayar programı duygulardan yoksun olduğu için bu büyük sorunu çözer. Önceden tanımlanmış koşullar yerine getirilirse, bilgisayar programı işlemi otomatik olarak gerçekleştirecektir. Bu durumda, ikinci düşünceler tüccarın daha sonra pişman olacağı eylemleri gerçekleştirmesini veya gerçekleştirmekten kaçınmasını engelleyemez. Dolayısıyla, aşırı alım satım ve düşük alım satımın kontrol altında tutulmasına da yardımcı olur.

Uzun Vadede Azalan Maliyet: Algo-ticaret altyapısını kurmanın ilk maliyeti pahalı görünebilir. Ancak, sistem bir kez kurulup çalışmaya başladığında, uzun vadede maliyet tasarrufu sağladığı kanıtlanır. Algo ticareti kısa bir süre içinde büyük hacimli işlemlerin gerçekleştirilmesine olanak tanıdığı için çoklu işlemler gerçekleştirilir ve işlem maliyetleri azalır.

Çeşitlendirme: Algoritma ve bilgisayar programı birkaç dakika içinde birden fazla grafiği tarayabildiğinden, aynı anda birden fazla işlem gerçekleştirecek şekilde de programlanabilirler. Ve bu tek bir borsa veya coğrafya ile sınırlı değildir, çünkü bilgisayar dünyanın her yerindeki borsalarda grafikleri tarayabilir ve alım satım yapabilir.

Algo-Ticaretinin Dezavantajları: Yüksek doğruluk ve hıza sahip olmasına ve duygulardan yoksun olmasına rağmen, algoritmik ticaretin bazı kayda değer dezavantajları vardır.

Teknolojiye Güvenmek: Teknoloji, algo ticaretinin bugünkü haline gelmesine yardımcı olmuştur, ancak bunun eksikliği en büyük dezavantajıdır. Teknolojik altyapınız yoksa veya teknolojiye erişiminizi kaybederseniz, algo ticaretinden yararlanamazsınız. Bazı durumlarda, İnternet bağlantınızdaki bir kesinti, tarih yerel olarak saklanıyorsa emrinizin yerine getirilmemesine neden olur.

Yüksek Başlangıç Giderleri:Günlük olarak birden fazla ticari emir vermeyi planlıyorsanız, uzun vadede algoritmik ticaretin uygun maliyetli olduğu doğrudur. Ancak

algo-ticaret altyapısını kurmanın ilk maliyeti maliyetlidir. Algo-tüccarlar, anlık hızlarda işlem yapma kapasitesine sahip en hızlı bilgisayarlara sahip olmayı arzuluyorlar.

Programlama Becerilerine İhtiyaç Var: Yüksek başlangıç maliyetinin yanı sıra, algoritmayı yazmak için programlama bilgisine de ihtiyacınız olacaktır. Bunun için C++ veya Python gibi bir bilgisayar dili bilmeniz gerekir. Eğer bilmiyorsanız, algoritmayı sizin için yazması için bir profesyonel tutmanız gerekecektir.

Algoritmik Ticaretin Riskleri ve Zorlukları: Diğer tüm ticaret stratejileri gibi algoritmik ticaret de kendi risk ve zorluklarıyla birlikte gelir. Bu risk ve zorluklardan bazılarını ve bunların algoritmik ticareti nasıl etkileyebileceğini tartışacağız.

Piyasa Oynaklığı :Algoritmik ticaretteki en büyük risklerden biri piyasa oynaklığıdır. Borsa çok öngörülemez olabilir ve ani piyasa değişimleri önemli kayıplara yol açabilir. Bunun nedeni, algoritmik alım satım sistemlerinin önceden belirlenmiş bir dizi kural dahilinde çalışmak üzere tasarlanmış olmasıdır. Piyasa bu kurallardan saptığında, algoritma yeterince hızlı ayarlama yapamayabilir ve bu da büyük kayıplara neden olabilir. Bu riski azaltmak için yatırımcılar, belirli bir zarar eşiğine ulaştıklarında işlemlerden otomatik olarak çıkmak için zarar durdur emirlerini kullanabilir.

Teknik Arızalar:Algoritmik ticaretin bir diğer riski de teknik arızalardır. Bu arızalar donanım veya yazılım arızaları, elektrik kesintileri veya internet bağlantısı sorunlarından kaynaklanabilir. Teknik arızalar, kaçırılan işlemlere, yanlış işlem uygulamalarına ve hatta sistem çökmelerine neden olabilir. Teknik arızaları önlemek için yatırımcılar yedek sunucular ve internet bağlantıları gibi yedekleme önlemleri uygulayabilir.

Aşırı Optimizasyon:Aşırı optimizasyon, algoritmik ticarete sık karşılaşılan bir sorundur. Tüccarlar algoritmalarını geçmiş veriler üzerinde iyi performans gösterecek şekilde ince ayarladıklarında, ancak algoritma gerçek zamanlı ticarete aynı performansı gösteremediğinde ortaya çıkar. Algoritma değişen piyasa koşullarına uyum sağlayamayabileceğinden bu durum kayıplara neden olabilir. Aşırı optimizasyondan

kaçınmak için yatırımcılar, algoritmalarını görünmeyen veriler üzerinde test etmek üzere örneklem dışı test ve ileriye dönük analiz kullanabilir.

Veri Kalitesi: Algoritmik ticarete kullanılan verilerin kalitesi, bu ticaretin başarısı için çok önemlidir. Düşük kaliteli veriler yanlış alım satım kararlarına yol açabilir ve önemli kayıplara neden olabilir. Veri kalitesini sağlamak için yatırımcılar, verilerdeki hataları ve tutarsızlıkları gidermek amacıyla veri temizleme ve doğrulama tekniklerini kullanabilir.

Özet olarak, algoritmik ticaretin birçok faydası vardır, ancak aynı zamanda kendi risk ve zorluklarını da beraberinde getirir. Yatırımcılar bu riskleri ve zorlukları anlayarak bunları azaltmaya ve algoritmalarının performansını artırmaya yönelik stratejiler uygulayabilir. Algoritmik ticaretin herkese uyan tek bir çözüm olmadığını ve yatırımcıların başarılarını garantilemek için algoritmalarını sürekli olarak izlemeleri ve ayarlamaları gerektiğini unutmamak önemlidir.

Algoritmik Ticaret Çeşitleri:

İstatistiksel Arbitraj : İstatistiksel arbitraj, binlerce finansal araç arasındaki göreceli fiyatlandırma verimsizliklerinden yararlanma konseptiyle gelişen ilgi çekici bir algo ticaret stratejisidir. Bu stratejiyi kullanan yatırımcılar, tarihsel olarak güçlü bir korelasyon sergileyen menkul kıymet çiftlerini veya gruplarını belirler. Aynı anda düşük değerli menkul kıymeti satın alarak ve aşırı değerli olanı satarak geçici fiyat farklılıklarından yararlanırlar. Beklenti, zaman içinde bu enstrümanların fiyatlarının birbirine yaklaşarak kârlı işlemlerle sonuçlanmasıdır. İstatistiksel arbitraj, bu fiyat farklılıklarını belirlemek ve hassas zamanlamayla alım satım gerçekleştirmek için sağlam istatistiksel modellere ve nicel analize dayanır. Çiftler ticareti stratejisi, kısa vadeli ortalama geri dönüş ilkelerine ve riskten korunma stratejilerine dayanan böyle bir istatistiksel arbitraj stratejisidir.

Trend Takibi: Finansal piyasaların sürekli değişen ortamında, trend takibi, diğer algoritmik ticaret stratejilerine kıyasla işlevselliği ve kullanım kolaylığı nedeniyle popüler bir algo ticaret stratejisi olarak ortaya çıkmaktadır. Bu strateji, yerleşik piyasa

trendlerini belirleyip bu trendler doğrultusunda alım satım yaparak kâr elde etmeyi amaçlar. Trend takip stratejilerini kullanan algo ticaret sistemleri, hâkim piyasa trendini belirlemek için teknik göstergelere ve fiyat modellerine güvenir. Bir trend onaylandığında, bu sistemler pozisyonlara girmek ve trendi sürdürmek için alım sinyalleri üretir. Tersine, trend tersine dönme belirtileri gösterdiğinde, pozisyonlardan çıkmak için satış sinyalleri üretilir. Trend takip algoritmaları değişen piyasa koşullarına uyum sağlayarak yatırımcıların sürekli piyasa trendlerinden kâr etmesine ve piyasa türbülansı dönemlerinde kayıpları en aza indirmesine olanak tanır.

Momentum: Yatırımcılar tarafından takip edilen en temel ve yaygın algoritmik yatırım sistemlerinden biri momentum yatırım stratejisidir. Bu yatırım türü, piyasa trendinin yüksek hacimde tek bir yönde önemli ölçüde hareket etmesini bekler. Bu yatırım sistemi çok basit ya da önemli ölçüde zor olabilir. Basit bir momentum yatırım stratejisi, 12 aylık performansa dayalı bir endekste en iyi performans gösteren beş hisseye yatırım yapabilir. Daha zor bir strateji ise hem göreceli hem de mutlak momentumdan faydalanarak zaman içinde momentumu harmanlayabilir. Ayrıca, bu sistemi kullanmak yatırımcıların momentum sistemlerini haftalık, aylık, üç aylık ve hatta yıllık olarak yeniden dengelemelerini sağlar.

Momentum ve Trend Takibi arasındaki fark: Trend Takibi, piyasaların uzun süreler boyunca belirli bir yönde hareket etme eğiliminde olduğu inancıyla çalışan bir ticaret stratejisidir. Bu trendlerin niceliksel olarak belirlenmesini ve bunlardan yararlanılmasını içerir. Zaman Serisi momentumu olarak da bilinen bu strateji, öncelikle zaman serisi verileri kategorisine giren ve birbirini izleyen bir dizi zaman birimini gösteren fiyat verilerine odaklanır. Trend Takibini Momentum stratejilerinden ayıran fark, trendleri belirlemek için kullanılan referans noktasında yatmaktadır. Momentum stratejileri, hisse senetlerinin performansını belirli bir evrendeki diğer hisse senetlerine göre karşılaştırır. Buna karşılık, Trend Takibi veya Zaman Serisi momentumu, bir trendin mevcut olup olmadığını belirlemek için bir hisse senedinin veya endeksin mevcut fiyatını geçmiş fiyatıyla karşılaştırır.

Duygu Analizi: Duyarlılık analizi ticaret stratejisi, kısa vadeli fiyat değişikliklerini tahmin etmek ve piyasa hareketlerinden faydalanmak için kalabalık tepkilerinden ve haber makaleleri ve sosyal paylaşımlar gibi yapılandırılmamış verilerin analizinden yararlanmayı içerir. Bu strateji, alım satım kararları için yön sinyali olarak kullanılabilir. Duyarlılık puanları atamak için bilgisayar doğal dil işleme ve anlamadaki ilerlemeleri kullanır. Kantitatif hedge fonları ve diğer tüccarlar, piyasada avantaj elde etmek için duyuruluk analizini stratejilerine dahil etmişlerdir.

Sosyal duyuruluk analizinin popülerliđi, insan yorumunun zaman alıcı doğasından ve etkin piyasa hipotezinin önerdiği gibi finansal piyasaların her zaman tam verimli olmayabileceğinin kabul edilmesinden kaynaklanmaktadır. Ayrıca araştırmalar, çevrimiçi sosyal medyadan alınan haberlerin ekonomik ve ticari göstergelerdeki değişikliklerin erken göstergelerini sağlayabileceğini göstermiştir. Duygu analizinin kısa vadeli bir faktör olarak dahil edilmesi, getirileri tahmin etmek için geleneksel olarak fiyat ve hacme güvenen yatırımcılar için yeni bir bakış açısı sağlar.

Bir duyuruluk analizi modeli geliştirmek, hedef makalelerin aranması, metin verilerinin temizlenmesi ve hazırlanması, metin belgelerinden özellikler oluşturulması, duyuruluk sınıflandırmak için modelin eğitilmesi ve performansının büyük ölçekli veri kümeleriyle test edilmesi gibi çeşitli adımları içerir. Yatırımcılar, duyuruluk analizini kullanarak piyasa duyuruluk hakkında bilgi edinebilir ve potansiyel olarak alım satım kararlarını iyileştirebilir. Duyuruluk analizi yapmak için yapay zekanın kullanılması, duyuruluğa dayalı içgörülere dayalı gerçek zamanlı karar vermeye olanak tanıyacaktır.

Ortalama Dönüş: Ortalama dönüş, fiyatların zaman içinde ortalamalarına veya ortalamalarına dönme eğiliminde olduğu ilkesine dayanan zorlayıcı bir algo ticaret stratejisidir. Bu strateji, bir varlığın fiyatının aşırı alım veya aşırı satım nedeniyle tarihsel ortalamasından önemli ölçüde saptığı durumları belirler ve fiyatın sonunda uzun vadeli ortalama fiyatına geri döneceği beklentisiyle pozisyonlara girme fırsatı sunar.

Ortalama dönüş ticaret stratejileri genellikle Hareketli Ortalamalar ve Bollinger Bantları gibi teknik göstergeleri kullanır. Hareketli ortalamalar bir varlığın ortalama geçmiş

fiyatını saęlarken, Bollinger Bantları standart sapma gibi volatilit e  l  mlerine dayanarak bir piyasanın ortalamasından ne zaman  nemli  l  de saptıęını belirlemeye yardımcı olur. Bu t r bir ok teknik g sterge g n m zde alım satım platformlarında mevcuttur ve manuel yatırımcılar tarafından yaygın olarak kullanılmaktadır. Ortalama d n ş prensibine dayalı karmaşıık stratejiler kodlama ve programlama gerektirir.

Ortalama d n ş stratejileri, kısa vadeli g n i i alım satımlardan uzun vadeli ufuklara kadar  eřitli zaman dilimlerinde uygulanabilir. Strateji, istatistiksel analize ve aşıırı geniřlemiş fiyat hareketlerinin belirlenmesine dayanır. Bir fiyat sapması tespit edildięinde, algo alım satım sistemi, beklenen fiyat d zeltmesinden kar elde etmeyi ama layarak sapmanın tersi y nde alım satımları bařlatmak i in sinyaller  retir.

Bařarılı ortalama d n ş stratejileri, risk y netimi ve piyasa kořullarının dikkatli bir şekilde deęerlendirilmesini gerektirir. Ortalama d n ş stratejilerini kullanan yatırımcılar, fiyatın tersine d nme olasılıęını deęerlendirmeli ve ortalama d n ř n ger ekleřmemesi durumunda potansiyel kayıpları azaltmak i in uygun stop-loss seviyeleri belirlemelidir. Yatırımcılar istatistiksel analiz, risk y netimi teknikleri ve piyasa dinamikleri anlayışını akıllıca bir araya getirerek algo ticareti  abalarında ortalama d n ş olgusundan potansiyel olarak faydalanabilir.

Ticaret Stratejileri:

- 1- Fiyat Hareketi Stratejisi (Price Action Strategy): Bu strateji, teknik g stergelere dayanmak yerine, fiyat hareketleri ve grafiklerdeki desenler temel alınarak ticaret kararlarının alınmasını i erir. T ccarlar, destek ve diren  seviyeleri, mum  ubuk desenleri ve dięer fiyat davranışlarını g zlemleyerek alım veya satım kararı verirler.
- 2- Teknik G stergeler Stratejisi (Technical Indicators Strategy): Bu strateji, fiyat, hacim veya a ık pozisyon verilerine dayalı matematiksel hesaplamalar kullanarak ticaret kararlarının alınmasını i erir. Yaygın teknik g stergeler arasında hareketli ortalamalar, G receli G   Endeksi (RSI), Hareketli Ortalama

Yakınsama/Uzaklaşma (MACD) ve Bollinger Bantları bulunur. Bu göstergeler, tüccarlara trendleri ve potansiyel dönüş noktalarını belirlemede yardımcı olur.

- 3- Temel Analiz (Fundamental Analysis): Bu strateji, bir şirketin finansal sağlığı, kazançları, ekonomik koşullar ve diğer niteliksel faktörleri analiz ederek bir hissenin içsel değerini belirlemeyi içerir. Bu stratejiyi kullanan tüccarlar, şirketin değerini mevcut piyasa fiyatına göre değerlendirerek alım veya satım kararı verirler.
- 4- Yüksek Frekanslı Ticaret (High-Frequency Trading - HFT): Bu strateji, güçlü bilgisayarlar ve algoritmalar kullanarak çok yüksek hızlarda, genellikle milisaniyeler içinde, çok sayıda emir gerçekleştirmeyi içerir. HFT, küçük fiyat farklılıklarından ve piyasa verimsizliklerinden yararlanarak her ticarete küçük kârlar elde etmeyi amaçlar ve bu kârlar zamanla birikir.
- 5- Mevsimsel Desenler (Seasonal Patterns): Bu strateji, belirli zaman dilimlerinde tekrarlayan tarihsel desenleri veya eğilimleri analiz etmeyi içerir. Bu stratejiyi kullanan tüccarlar, tatil sezonunda perakende satışlarının artması veya hasat dönemlerinde tarımsal emtia fiyat hareketleri gibi mevsimsel eğilimlere dayanarak ticaret kararları alırlar.

2.1.2 Finansal piyasalar terminolojileri

Finansal piyasalar, yatırımcılar, şirketler ve hükümetler arasında sermaye ve risk alışverişini kolaylaştırarak küresel ekonomilerin bel kemiği olarak hizmet eder. Bu bölümde, Borsalar, Hisse Senetleri, Endeks ve Göstergeler dahil olmak üzere finansal piyasaların temel bileşenlerini ve bunların algoritmik ticaretteki önemini inceleyeceğiz.

Hisse Senedi Piyasaları:Borsa, en hafif tabiriyle, hisse senetlerinin alınıp satıldığı bir yerdir. Aynı zamanda ekonomik döngünün bir göstergesi olarak da işlev görür. Ekonomi iyi gittiğinde, hisse senetlerinin fiyatları genellikle piyasada artma eğilimindedir. Ancak, ekonomi kötüye gittiğinde, hisse senedi fiyatlarının düşmesini de bekleyebilirsiniz, bu çok iyi bir hisse senedi için bile geçerli olabilir. Hisse senedi fiyatlarının öncelikle işletmenin performansına bağlı olduğunu da belirtmek gerekir. Bir işletme iyi iş yaptığında, hisse senetlerinin fiyatı da artma eğiliminde olacaktır; işletme iyi

kazanmıyorsa bunun tersi olur. Elbette, spekülâtorlerin hiç de iyi gitmeyen bir hisse senedini satın aldığı ve bunun da berbat bir şirket için bile yüksek bir hisse senedi fiyatıyla sonuçlanacağı garip durumlar vardır, ancak bu başka bir zaman için başka bir hikâye. Şimdi, soru devam ediyor: İşletmeler neden hisse senedi sunar? Bunun nedeni, işletmelerin projelerini finanse etmek için ister işletmenin genişlemesi için ister günlük faaliyetleri için ya da sadece paydaşlarını mutlu etmek için olsun, fon toplamalarına yardımcı olmasıdır. Sonuçta para, her işletmenin can damarıdır.

Borsa Endeksi: İnsanlar hisse senetleri hakkında konuşurken, piyasanın bir gün içinde nasıl yükseldiğinden veya düştüğünden bahsettiklerini duyabilirsiniz. Ancak borsaya baktığınızda, tüm hisse senetlerinin o gün yükselmediğini veya düşmediğini görebilirsiniz. Bunu anlamanın yolu borsa endeksini ya da endekslerini (çoğul) bilmektir.

Borsada listelenen birçok hisse senedi vardır. Benzer hisse senetleri bir araya getirilerek bir endeks oluşturulur. Bu hisse senetlerinin bir araya getirilme şekli kapitalizasyon, endüstri, şirket büyüklüğü ve diğer faktörlere dayalı olabilir. Birlikte gruplandırıldıkları için, belirli bir endekse ait hisse senetlerinin değerindeki herhangi bir değişiklik endeksin değerini etkileyecektir. Dünyadaki ünlü endeksler, bunlarla sınırlı olmamak üzere, şunlardır: Dow Jones, Nasdaq ve S&P 500.

S&P 500 Endeksi: S&P 500, Amerika Birleşik Devletleri'ndeki borsalarda listelenen 500 büyük şirketin performansını ölçen bir borsa endeksidir. ABD hisse senedi piyasasının geniş bir kesitini temsil eder ve genel hisse senedi piyasasının performansı için yaygın olarak bir ölçüt olarak kabul edilir.

Teknik Göstergeler ve Teknik Analiz: Öncelikle teknik analiz konusunu tanımlayalım. Teknik analiz, gelecekteki fiyat eğilimlerini tahmin etmek amacıyla, öncelikle grafiklerin kullanılması yoluyla piyasa hareketinin incelenmesidir. “Piyasa hareketi” terimi teknisyenin kullanabileceği üç temel bilgi kaynağını içerir- fiyat, hacim ve açık faiz. (Açık faiz sadece vadeli işlemler ve opsiyonlarda kullanılır.) Sıklıkla kullanılan “fiyat hareketi” terimi çok dar görünmektedir çünkü çoğu teknisyen hacim ve açık faizi piyasa analizlerinin ayrılmaz bir parçası olarak dahil etmektedir. Bu ayrımın yapılmasıyla

birlikte, bu tartışmanın geri kalanında “fiyat hareketi” ve “piyasa hareketi” terimleri birbirinin yerine kullanılacaktır.

Şimdi teknik analiz terimini anladıktan sonra, teknik Göstergenin ne olduğunu anlayalım.

Teknik göstergeler, teknik analizi takip eden yatırımcılar tarafından kullanılan bir menkul kıymetin veya sözleşmenin fiyatı, hacmi ve/veya açık faizi tarafından üretilen sezgisel veya model tabanlı sinyallerdir. Teknik analistler geçmiş verileri analiz ederek gelecekteki fiyat hareketlerini tahmin etmek için göstergeleri kullanır. Yaygın teknik göstergelere örnek olarak Göreceli Güç Endeksi (RSI), Ortalama Gerçek Aralık (ATR), Hareketli Ortalama Yakınsama Sapması (MACD) ve Bollinger Bantları (BB) verilebilir. Algoritmik ticaret bağlamında göstergeler, piyasa verilerinin matematiksel hesaplamaları veya görsel temsilleridir. Hisse senedi fiyatlarındaki eğilimleri, kalıpları ve gelecekteki potansiyel hareketleri analiz etmek için kullanılırlar.

2.1.3 Gösterge çeşitleri

İki temel teknik gösterge türü vardır:

Overlays: Fiyatlarla aynı ölçeği kullanan teknik göstergeler bir hisse senedi grafiğindeki fiyatların üzerine çizilir. Örnekler arasında hareketli ortalamalar ve Bollinger Bantları bulunur.

Oscillators: Yerel bir minimum ve maksimum arasında salınan teknik göstergeler bir fiyat grafiğinin üstüne veya altına çizilir. Örnekler arasında MACD veya RSI bulunur.

Şimdi çalışmamızda kullandığımız her bir hisse senedine kısaca bakalım:

Göreceli Güç Endeksi (RSI): Bu, son fiyat hareketlerine bakarak bir hisse senedinin aşırı alım (çok yüksek fiyatlı) veya aşırı satım (çok düşük fiyatlı) olup olmadığını

belirlemeye yardımcı olur. Bunu hisse senedinin fiyatı için bir sıcaklık kontrolü olarak düşünün- sıcak mı (aşırı alım) yoksa soğuk mu (aşırı satım)?

RSI indikatörünün hesaplaması için aşağıdaki Formül 1 ve Formül 2 kullanılır.

$$RSI = 100 - (100 / (1 + RS)) \text{ (İlk 14 periyotta)} \quad (1)$$

$$RS = (\text{Pozitif kapanışların ortalaması} / \text{Negatif kapanışların ortalaması}) \quad (2)$$

Bollinger Bantları:Bunlar bir hisse senedinin fiyatı etrafında hareket eden kılavuz raylar gibidir. Fiyat daha değişken (inişli çıkışlı) hale geldiğinde bantlar genişler ve daha sakin olduğunda daralır. Bollinger Bantlarını bir kanal olarak düşünün- daha geniş bir kanal fiyatın daha fazla dalgalanabileceğini gösterirken, dar bir kanal daha istikrarlı olabileceğini gösterir. BB indikatörünün hesaplaması için aşağıdaki Formül 3 kullanılır.

$$BBi = Xi \pm (\sigma_i * d) \quad (3)$$

Xi: Belirli bir dönem için fiyatın basit hareketli ortalaması (SMA).

σ_i : Belirli bir dönem için SMA'nın standart sapması.

d: Ortalama sapmadan uzaklaşma miktarı (genellikle 2).

Ortalama Gerçek Aralık (ATR): Bu, bir hisse senedinin fiyatının belirli bir süre boyunca ortalama olarak ne kadar dalgalandığını ölçer. Bunu, hisse senedinin tipik günlük fiyat dalgalanmalarını gösteren bir cetvel olarak düşünün.

Ortalama Gerçek Aralığı (ATR) hesaplamak için öncelikle belirli bir dönem için Gerçek Aralığı (TR) hesaplamamız gerekir. TR, aşağıdaki üç değerden en büyüğüdür:

1- Mevcut en yüksek ile bir önceki kapanış arasındaki fark.

2- Mevcut düşük ile önceki kapanış arasındaki fark.

3- Mevcut en yüksek ile mevcut en düşük arasındaki fark.

Belirli bir dönem için Gerçek Aralığı (TR) hesaplama adımları aşağıda verilmiştir:

- 1- Mevcut en yüksek ile mevcut en düşük arasındaki farkı bulun.
- 2- Mevcut en yüksek ile bir önceki kapanış arasındaki farkın mutlak değerini bulun.
- 3- Mevcut düşük ile önceki kapanış arasındaki farkın mutlak değerini bulun.
- 4- Gerçek Aralık (TR) yukarıda hesaplanan üç değer en büyüğüdür.

Örneğin, mevcut en yüksek değerin 50\$, mevcut en düşük değerin 40\$ ve bir önceki kapanışın 45\$ olduğunu varsayalım. TR'nin hesaplanması aşağıdaki gibi olacaktır:

- 1- Mevcut en yüksek ve mevcut en düşük arasındaki fark = 50 \$ - 40 \$ = 10 \$.
- 2- Mevcut en yüksek ve önceki kapanış arasındaki farkın mutlak değeri = |\$50- \$45| = \$5.
- 3- Mevcut düşük ile önceki kapanış arasındaki farkın mutlak değeri = |\$40 - \$45| = \$5.
- 4- Üç değer en büyüğü 10 '\$dır ve bu Gerçek Aralıktır (TR).

TR belirli bir dönem için hesaplandıktan sonra, Ortalama Gerçek Aralık (ATR) aynı dönemdeki TR değerlerinin ortalaması alınarak hesaplanabilir. En yaygın kullanılan dönem 14'tür, ancak yatırımcılar bu değeri kendi alım satım tarzlarına ve tercihlerine uyacak şekilde ayarlayabilir.

Ortalama Gerçek Aralık (ATR) Formülü

Belirli bir dönem için Gerçek Aralığı (TR) hesapladıktan sonra, bir sonraki adım aşağıdaki formülü kullanarak Ortalama Gerçek Aralığı (ATR) hesaplamaktır:

$$ATR = [(\text{Önceki ATR} * (n - 1)) + \text{Mevcut TR}] / n \quad (4)$$

Burada:

Önceki ATR: önceki dönem için ATR değeri.

n: hesaplamada kullanılan dönem sayısı. En yaygın olanı 14 döneme dayanır.

Geçerli TR: geçerli dönem için Gerçek Aralık (TR) değeri.

İlk dönem için ATR'yi hesaplamak için ATR değeri olarak Gerçek Aralık (TR) değeri kullanılır.

Örneğin, 14 günlük bir dönem için ATR hesapladığımızı ve ilk 14 gün için Gerçek Aralık (TR) değerlerini zaten hesapladığımızı varsayalım. İşte 15. gün için ATR hesaplama adımları:

- 1- Daha önce açıklanan formülü kullanarak 15. gün için TR değerini hesaplayın.
- 2- Önceki ATR, bir önceki günün (14. gün) ATR değeridir.
- 3- n 14'tür (hesaplamada kullanılan dönem sayısı).
- 4- Değerleri ATR formülünde yerine koyarak 15. günün ATR değerini hesaplayın.

$$\begin{aligned} \text{ATR} &= [(\text{Önceki ATR } (n - 1)) + \text{Mevcut TR}] / n \\ &= [(14. \text{ gün için ATR } 13) + 15. \text{ gün için TR}] / 14 \end{aligned}$$

Bu hesaplamanın sonucu 15. gün için ATR değeridir. Bu işlem, tüm dönem için ATR'yi hesaplamak üzere sonraki her gün için tekrarlanabilir.

Hareketli Ortalama Yakınsama İraksaması (MACD): Bu, iki ortalama karşılaştırarak hisse senedi fiyatındaki eğilimleri aradığını söylemenin süslü bir yoludur. MACD'yi iki hareketli ortalama arasındaki bir çekişme olarak düşünün- bir ortalama diğerinin önemli ölçüde önüne geçtiğinde, potansiyel bir eğilime işaret edebilir. MACD, uzun vadeli üstel hareketli ortalamanın (EMA) (26 dönem) kısa vadeli EMA'dan (12 dönem) çıkarılmasıyla hesaplanır.

MACD'yi hesaplamak için aşağıdaki Formül 5 kullanılır.

$$\text{MACD} = 12\text{-Period EMA} - 26\text{-Period EMA} \quad (5)$$

Garman-Klass Volatilitesi (GKV): Bu, bir hisse senedinin fiyatının gelecekte ne kadar yukarı veya aşağı hareket edebileceğini tahmin etmeye çalışır. Bunu, hisse senedinin fiyat yolculuğunun ne kadar inişli çıkışlı olabileceğine dair bir gösterge olarak düşünün.

GKV'yi hesaplamak için aşağıdaki Formül 6 kullanılır.

$$\sigma_{GK}^2 = \frac{1}{n} = \sum_{i=1}^n \left[\frac{1}{2} \left(\ln \frac{H_i}{L_i} \right)^2 - (2 \ln 2 - 1) \left(\ln \frac{C_i}{O_i} \right) \right] \quad (6)$$

Dolar Hacmi : Bu basitçe, belirli bir günde bir hisse senedini almak ve satmak için harcanan toplam para miktarını ölçer. Bunu, hisse senedine ne kadar para girip çıktığını takip eden bir sayaç olarak düşünün.

2.1.4 Makine öğrenimine giriş

Makine öğrenimi, verilerden öğrenme ve açıkça programlanmadan tahminler veya kararlar verme yeteneğine sahip algoritmaların ve modellerin geliştirilmesine odaklanan yapay zekanın (AI) bir alt alanıdır. Bu bölümde, makine öğreniminin temellerini, çeşitli alanlardaki uygulamalarını ve algoritmik ticaretle ilgisini keşfedeceğiz.

Tanım ve Amaçlar: Makine öğrenimi, faydalı olabilmek için bazı fenomenlerin örneklerinin bir koleksiyonuna dayanan algoritmalar oluşturmakla ilgilenen bilgisayar biliminin bir alt alanıdır. Bu örnekler doğadan gelebilir, insanlar tarafından elle işlenebilir veya başka bir algoritma tarafından oluşturulabilir. (Andriy B.)

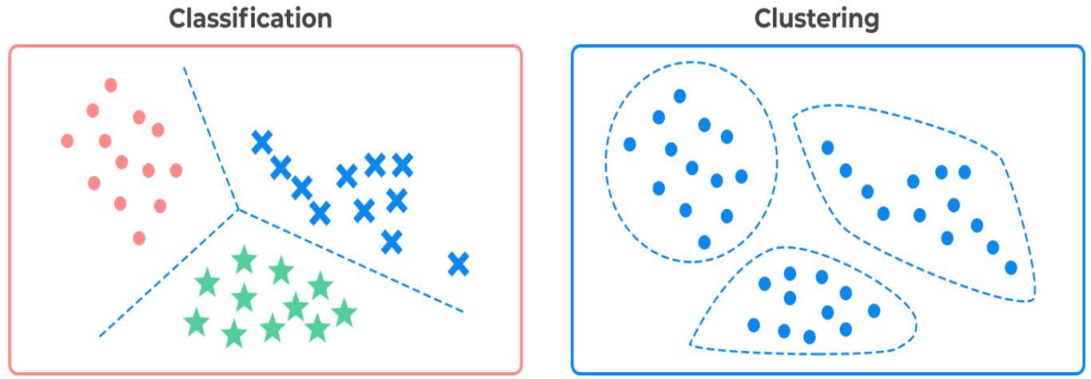
Makine öğrenimi, verilerden bilgi çıkarmakla ilgilidir. İstatistik, yapay zekâ ve bilgisayar bilimlerinin kesiştiği bir araştırma alanıdır ve tahmine dayalı analitik veya istatistiksel öğrenme olarak da bilinir. (Kevin P.). Makine öğreniminin birincil hedefleri şunlardır:

Tahmin (Prediction): Geçmiş verilere dayanarak gelecekteki sonuçları veya eğilimleri tahmin etmek.

Sınıflandırma (Classification): Veri noktalarının önceden tanımlanmış kategorilere veya sınıflara atanması.

Kümeleme (Clustering): Benzer veri noktalarının özelliklerine göre gruplandırılması.

Regresyon (Regression): Girdi değişkenleri ile çıktı değerleri arasındaki ilişkinin tahmin edilmesi.



Şekil 2.1 Sınıflandırma ve kümeleme arasındaki fark

Makine Öğrenimi Çeşitleri: Makine öğrenimi algoritmaları, öğrenme süreçlerinin doğasına bağlı olarak üç ana türe ayrılabilir:

Denetimli Öğrenme: Denetimli öğrenmede algoritmalar, her bir girdi örneğinin karşılık gelen bir çıktı veya hedef değişkenle ilişkilendirildiği etiketli verilerden öğrenir. Denetimli öğrenme görevleri sınıflandırma ve regresyonu içerir.

Denetimsiz Öğrenme: Denetimsiz öğrenme algoritmaları, veri içindeki gizli kalıpları veya yapıları keşfetmeye çalışarak etiketsiz verilerden öğrenir. Kümeleme ve boyut azaltma yaygın denetimsiz öğrenme görevleridir.

Pekiştirmeli öğrenme: Pekiştirmeli öğrenme, bir ajanın deneme yanılma yoluyla sıralı kararlar vermeyi öğrenmesini ve kümülatif ödülleri en üst düzeye çıkarmayı amaçlamasını içerir. Pekiştirmeli öğrenme dinamik ve etkileşimli ortamlar için çok uygundur.

2.1.5 Finans ve algoritmik ticaret uygulamaları

Makine öğrenimi teknikleri, büyük miktarda finansal veriyi analiz etme, kalıpları belirleme ve veriye dayalı kararlar alma becerileri nedeniyle finans sektöründe ve algoritmik ticarete ilgi görmüştür. Finans ve algoritmik ticarete makine öğreniminin bazı yaygın uygulamaları şunlardır:

Tahmine Dayalı Modelleme: Gelecekteki fiyat hareketlerini veya trendleri tahmin etmek için geçmiş piyasa verilerinin kullanılması.

Risk Yönetimi: Kantitatif modeller ve algoritmalar aracılığıyla portföy riskinin değerlendirilmesi ve yönetilmesi.

Portföy Optimizasyonu: İstenen risk-getiri profillerine ulaşmak için varlık dağılımlarının ve alım satım stratejilerinin optimize edilmesi.

Piyasa Duyarlılığı Analizi: Piyasa duyarlılığını ölçmek ve bilinçli alım satım kararları almak için haber makalelerini, sosyal medya akışlarını ve diğer metinsel verileri analiz etme.

Yüksek Frekanslı Ticaret: Kısa vadeli piyasa verimsizliklerinden ve fiyat anomalilerinden yararlanan algoritmik stratejilerin uygulanması.

2.1.6 Makine öğrenimi modelleri

Bu alt bölümde, verileri etiketlemek ve değerlendirmek için uygulamada kullanılan ve benzerlerini makine öğrenimi modellerine kısa ve açık bir genel bakış sunuyoruz.

GMM (Gaussian Mixture Model): GMM, gözlemlenen verilerin birden fazla normal dağılımdan oluştuğunu varsayar. Her dağılım, bir kümeyi temsil eder ve her kümenin ortalama vektörü ve kovaryans matrisi model parametreleri olarak belirlenir. GMM, EM algoritması gibi özdeşleşmeli öğrenme yöntemleri kullanılarak modellenebilir. GMM, gürültülü verilerle bile sağlam kümeleme performansı sağlayabilir ve kümelerin karmaşık şekillerde dağıldığı durumlarda faydalıdır.

Avantajlar:

- Gürültülü verilere karşı sağlamdır.

- Karmaşık şekilli kümeleri ele alabilir.
- Olasılık modelleri sunar.

Dezavantajlar:

- Parametrelerin yorumlanması zor olabilir.
- Hesaplama maliyetli olabilir.
- Önceden küme sayısını belirlemek gerekir.

K-Means Clustering: K-Means, verileri önceden belirlenmiş sayıda k kümeye ayırmayı amaçlayan basit bir kümeleme algoritmasıdır. Algoritma, her küme için bir merkez noktası seçerek ve ardından her veri noktasını en yakın merkeze atayarak çalışır. Daha sonra merkezler, kümelere atanan veri noktalarının ortalaması olarak güncellenir ve bu işlem sabit bir duruma gelene kadar tekrarlanır. K-Means, basit ve hızlı bir algoritmadır ve büyük veri kümeleri üzerinde bile kullanılabilir.

Avantajlar:

- Basit ve hızlıdır.
- Büyük veri kümeleri üzerinde kullanılabilir.
- Küme merkezleri somuttur ve yorumlanabilir.

Dezavantajlar:

- Önceden küme sayısını belirlemek gerekir.
- Kümelerin küresel şekilli olması varsayılır.
- Gürültülü verilere karşı hassastır.

Spectral Clustering: Spectral Clustering, verileri bir grafikte temsil ederek ve grafiğin Laplacian matrisinin özdeğerlerini ve özvektörlerini kullanarak kümeleri belirleme yöntemidir. Bağlantılar, veri noktaları arasındaki benzerlikleri temsil eder ve Laplacian matrisi grafiğin bağlantı yapısını kodlar. Spectral Clustering, verilerde doğal olarak bulunan kümeleri bulmada oldukça etkili olabilir ve önceden küme sayısını belirleme ihtiyacı duymaz.

Avantajlar:

- Doğal kümeleri bulmada etkilidir.
- Önceden küme sayısını belirleme ihtiyacı yoktur.
- Kümelerin karmaşık şekillerde dağıldığı durumlarda faydalıdır.

Dezavantajlar:

- Hesaplama maliyetli olabilir.
- Gürültülü verilere karşı hassas olabilir.
- Büyük veri kümeleri üzerinde kullanımı zor olabilir.

Sonuç: GMM, K-Means ve Spectral Clustering, veri kümeleme problemini çözmek için kullanılan üç popüler yöntemdir. Her yöntemin kendine özgü avantajları ve dezavantajları vardır ve en iyi yöntem, veri kümesinin özelliklerine ve istenen sonuçlara bağlıdır.

2.2 İlgili Çalışmalar

Algoritmik ticarete makine öğreniminin (ML) uygulanması, veriye dayalı stratejiler aracılığıyla alfa (riske göre ayarlanmış üstün getiriler) üretme potansiyelini araştıran çok sayıda çalışma ile önemli bir araştırma alanı haline gelmiştir. Bu bölümde, finansal piyasalarda alım/satım sinyalleri üretmek için makine öğreniminin benzer uygulamalarını araştıran ilgili araştırmalar incelenmektedir. Mevcut çalışmaları inceleyerek, teknolojinin mevcut durumunu belirleyebilir, mevcut yaklaşımların güçlü yönlerini ve sınırlamalarını anlayabilir ve kendi araştırmamızı makine öğrenimi ile algoritmik ticaretin daha geniş bağlamı içinde konumlandırabiliriz.

Bu bölümde algoritmik ticaret bağlamında borsa tahmini için makine öğrenimi modellerinin uygulanmasına ilişkin mevcut araştırmalar incelenmektedir.

(Liu vd., 2016) tarafından yazılan çalışma, Lojistik Regresyon (LR), Gauss Ayırıştırma Analizi (GDA), Naive Bayes (NB) ve Destek Vektör Makinesi (DVM) dahil olmak üzere S&P 500 hisse senedi endeksini tahmin etmek için denetimli öğrenme modellerinin

uygulanmasını araştırmaktadır. Yazarlar, özellikleri ve modelleri optimize etmeyi amaçlayan bir dizi deney aracılığıyla, SVM çekirdek seçimi ve farklı modeller arasında özellik seçimi de dahil olmak üzere çeşitli yönleri araştırmaktadır. Makalede, Radyal Taban Fonksiyonu (RBF) çekirdeğine sahip bir DVM modelinin S&P 500 endeksinin gelecekteki piyasa eğilimini tahmin etmede %62,51'lik kayda değer bir doğruluk oranına ulaştığı sonucuna varılmıştır.

(Ghosh vd., 2021) tarafından hazırlanan çalışma, gün içi hisse senedi fiyat hareketlerini tahmin etmek için makine öğrenimi modellerinin etkinliğini araştırmaktadır. Yazarlar, S&P500 hisse senetlerinin geçmiş kapanış fiyatları, açılış fiyatları ve gün içi getirilerinin bir kombinasyonunu kullanarak yön hareketlerini tahmin etmek için LSTM ağları ve Rastgele Ormanlar kullanmaktadır. Bulguları, her iki modelin de 0,74'ü aşarak umut verici bir doğruluk elde ettiğini ve bu yaklaşımın gün içi ticaret uygulamaları için potansiyelini vurguladığını göstermektedir. Bu çalışma, borsa tahmininde model performansını iyileştirmek için kapanış fiyatlarının ötesinde daha geniş bir özellik yelpazesinin kullanılması anlayışına katkıda bulunmaktadır.

(Liu vd., 2023) tarafından hazırlanan çalışma, S&P 500 endeksini tahmin etme konusundaki zorlu görevi ele almaktadır. Önceki algoritmalarda gözlemlenen sınırlamaların üstesinden gelerek, hassas zaman serisi tahmini için geçmiş bilgilerden yararlanmak üzere tasarlanmış N-BEATS algoritmasını tanıtmaktadırlar. Deneysel veri seti, Nisan-Ağustos 2017 tarihleri arasında 500 hisse senedi ve S&P 500 endekslerine ilişkin ayrıntılı verilerden oluşmaktadır. Giriş verileri, 500 hisse senedini ve bunlara karşılık gelen S&P 500 endekslerini önceki 15 dakika için içermekte ve veri kümesini (41,251, 15, 501) boyutlarına şekillendirmektedir. Sonuçlar, N-BEATS algoritmasının S&P 500 endeksini tahmin etmede diğer tahmincilerden daha iyi performans gösterdiğini ortaya koymaktadır. COR, R2, MAPE ve RMSE gibi değerlendirme ölçütleri, sırasıyla 0,9998, 0,9996, %0,02 ve 0,8242 puanlarıyla N-BEATS modelinin üstünlüğünü göstermektedir. Bu bulgular, N-BEATS'in geniş kapsamlı uygulama potansiyeli ile borsa tahmini için umut verici bir yöntem olarak etkinliğinin altını çizmektedir.

(Shi vd., 2024) tarafından yapılan çalışma, S&P 500 borsasının tahmin edilmesinde makine öğrenimi modellerinin uygulanmasını araştırmaktadır. Çalışma, iki makine öğrenimi modeli kullanarak temel bir analiz sağlamayı amaçlamaktadır: doğrusal regresyon ve XGBoost. Son yıllarda S&P 500 borsasından elde edilen geçmiş verileri kullanan araştırma, veri setini sırasıyla %70, %15 ve %15 oranlarında dağıtılan eğitim, test ve doğrulama setlerine ayırmaktadır. Doğrusal regresyon modeli, Ortalama Karesel Hata (MSE) ve R-kare (R^2) metriklerini kullanarak tahmin doğruluğunu değerlendirir ve test grubunda $3,051 \times 10^{17}$ MSE ve 0,316 R^2 verir. Buna karşılık, XGBoost modeli doğruluğu MSE ve Ortalama Karesel Hatanın Kökü (RMSE) kullanarak değerlendirir ve test grubunda 14816.886 MSE ve 121.725 RMSE verir. Bulgular, XGBoost modelinin doğruluk açısından doğrusal regresyon modelinden daha iyi performans gösterdiğini ve borsa tahmini için uygun olduğunu göstermektedir. Bu sonuç, XGBoost modelinin doğrusal regresyona kıyasla daha düşük MSE ve RMSE değerleri ile kanıtlanan üstün performansından çıkarılmaktadır.

(Htun vd., 2024) tarafından yapılan çalışma makine öğrenimi (ML) tekniklerini kullanarak S&P 500 hisse senetleri için hisse senedi fiyatlarındaki değişiklikleri tahmin etme gibi zorlu bir görevi araştırmaktadır. Yazarlar, birden fazla önceki göreceli getiriden oluşan özellik vektörleri oluşturmakta ve sınıflandırıcı olarak üç ML yöntemini (rastgele orman (RF), destek vektör makinesi (SVM) ve uzun kısa vadeli bellek (LSTM)) kullanmaktadır. Amaç, bir hisse senedinin takip eden 10 gün içinde endeksinden %2 daha fazla getiri sağlayıp sağlayamayacağını tahmin etmektir. Analiz, 2017-2022 dönemi boyunca tüm S&P 500 şirketlerini kapsamaktadır. Performans değerlendirmesi doğruluk, kesinlik ve geri çağırma metrikleri kullanılarak yapılmış ve sonuçlar rastgele seçim stratejisiyle karşılaştırılmıştır. Bulgular, LSTM sınıflandırıcısının RF ve SVM'den daha iyi performans gösterdiğini ve ML yöntemlerinin rastgele seçim sınıflandırıcısından önemli ölçüde daha iyi performans gösterdiğini ortaya koymaktadır (LSTM'nin doğruluğu için $p = 8.46e-17$). Özellikle, LSTM yüksek bir doğruluk seviyesine ulaşarak ML tekniklerinin borsa tahminindeki potansiyelini vurgulamaktadır.

3. MATERİYAL ve METOD

3.1 Veri Seti

Bu bölümde, veri setinin toplanması ve analiz için hazırlanması süreci detaylandırılmaktadır. Veri toplama süreci, S&P 500 hisse senetleri için geçmiş piyasa verilerinin alınmasını, gerekli veri temizliğinin yapılmasını ve verilerin daha fazla analiz için uygun bir formatta toplanmasını içermektedir.

3.1.1 Veri kaynağı ve veri tanımı

Bu bölümde, veri setinin toplanması ve analiz için hazırlanması süreci detaylandırılmaktadır. Veri toplama süreci, S&P 500 hisse senetleri için geçmiş piyasa verilerinin alınmasını, gerekli veri temizliğinin yapılmasını ve verilerin daha fazla analiz için uygun bir formatta toplanmasını içermektedir.

Veri Kaynağı: Bu çalışma için birincil veri kaynağı, yfinance Python kütüphanesi kullanılarak erişilen Yahoo Finance API'dir. Bu kütüphane, Yahoo Finance API'sine bir arayüz sağlayarak belirli bir zaman aralığı için hisse senedi fiyat verilerini indirmemize olanak tanır. Yahoo Finance API, hisse senetleri, ETF'ler ve endeksler dahil olmak üzere çeşitli finansal araçlar için geçmiş piyasa verileri sağlar. Son üç yılda S&P 500 hisse senetleri hakkında ayrıntılı bilgi almak için API'yi kullandık ve kapsamlı analiz ve modelleme yapmamızı sağladık.

Veri Tanımı: Bu S&P 500 hisse senetleri tahmin modeli için kullanılan veri seti, her biri float64 veri türünde olan 13 sütun ve 17904 girişten oluşmaktadır. Veriler, belirli bir tarih ve hisse senedini gösteren bir zaman damgası ve hisse senedi sembolü ile indekslenen, pandas'ta çoklu indeksli bir DataFrame olarak organize edilmiştir. Bu çoklu indeksleme, çoklu hisse senetlerinin zaman serisi analizi ve kesitsel karşılaştırması için sofistike bir yapı sağlar.

Veri Sütunları:

1- RSI (Göreceli Güç Endeksi)

RSI, fiyat hareketlerinin hızını ve değişimini ölçen bir momentum osilatörüdür. 0 ile 100 arasında dalgalanır ve geleneksel olarak bir hisse senedinin aşırı alım veya aşırı satım koşullarını belirlemek için kullanılır. RSI değeri 30'un altında olduğunda aşırı satım, 70'in üzerinde olduğunda ise aşırı alım durumu olduğu kabul edilir.

2- bb_low (Bollinger Bantları Alt)

Bollinger Bantları üç hattan oluşur: orta bant basit bir hareketli ortalama (SMA) iken, üst ve alt bantlar bu SMA'nın standart sapmalarıdır. Alt bant (bb_low), potansiyel destek seviyelerini gösterir.

3- bb_mid (Bollinger Bantları Orta)

Orta bant (bb_mid) genellikle fiyatın 20 günlük SMA'sıdır. Üst ve alt bantlar için bir temel olarak hizmet eder ve belirli bir dönem boyunca ortalama fiyatı yansıtır.

4- bb_high (Bollinger Bantları Üst)

Üst bant (bb_high), orta banttan belirli bir standart sapma yukarıda olan banttır ve potansiyel direnç seviyelerini gösterir.

5- macd (Hareketli Ortalama Yakınsama İraksama)

MACD, bir hisse senedinin fiyatının iki hareketli ortalaması arasındaki ilişkiyi gösteren bir trend takip eden momentum göstergesidir. 26 dönemlik EMA'dan 12 dönemlik EMA çıkarılarak hesaplanır. Genellikle 9 dönemlik EMA olan bir sinyal hattı, MACD'nin üzerine çizilerek yükseliş veya düşüş sinyallerini belirlemek için kullanılır.

6- atr (Ortalama Gerçek Aralık)

ATR, piyasa volatilitesini ölçen bir teknik analiz göstergesidir. Mevcut yüksek ile mevcut düşük arasındaki fark, mevcut yüksek ile önceki kapanış arasındaki mutlak değer ve

mevcut düşük ile önceki kapanış arasındaki mutlak değerden en büyüğü alınarak hesaplanır.

7- garman_klass_vol (Garman-Klass Volatilitesi)

Bu, yüksek, düşük, açılış ve kapanış fiyatlarını kullanarak bir hisse senedinin volatilitesi tahmin etmek için bir yöntemdir. Daha fazla fiyat bilgisi içerdiğinden diğer volatiliteler tahmincilerinden daha doğru olduğu kabul edilir.

8- return_1m (1 Aylık Getiri)

1 aylık dönem boyunca getiri, hisse senedinin fiyatındaki son bir ay içindeki yüzde değişim olarak hesaplar. Kısa vadeli fiyat hareketleri hakkında bilgi sağlar.

9- return_2m (2 Aylık Getiri)

1 aylık getiriye benzer şekilde, bu sütun son iki ay içindeki hisse senedinin fiyatındaki yüzde değişimi temsil eder ve biraz daha uzun vadeli bir fiyat performansı görünümü sunar.

10- return_3m (3 Aylık Getiri)

Bu sütun, 3 aylık dönem boyunca hisse senedinin performansını izler ve genellikle üç aylık eğilimleri ve yatırımcı duyarlılığını değerlendirmek için kullanılır.

11- return_6m (6 Aylık Getiri)

Altı ay boyunca getiri hesaplar ve orta vadeli bir perspektif sağlar, eğilimleri ve döngüleri tanımlamak için kullanışlıdır.

12- return_9m (9 Aylık Getiri)

Bu sütun, dokuz aylık dönem boyunca hisse senedinin fiyat değişimini yansıtır ve orta vadeli ile uzun vadeli performans analizini köprüler.

13- return_12m (12 Aylık Getiri)

Yıllık getiri, son bir yıl içindeki hisse senedinin fiyatındaki yüzde değişim olarak hesaplanır. Bu uzun vadeli ölçüm, hisse senedinin genel performansını ve tam bir piyasa döngüsü boyunca eğilimleri değerlendirmeye yardımcı olur.

Tüm sütunlar, eksik değer içermeyen girişlere sahiptir, bu da eğitim sürecinin bütünlüğü için kritik önem taşır, çünkü eksik veriler, önyargıları tanıtabilir ve modelin tahmin gücünü azaltabilir. Veri setinin bellek kullanımı 211.0+ KB'dir ve bu, tipik bir Jupyter Notebook ortamında işlemek için verimlidir.

Teknik Göstergelerin Önemi: Seçilen teknik göstergeler ve getiriler, sağlam bir tahmin modeli oluşturmak için kritik öneme sahiptir. Bunlar, momentum (RSI), volatilité (ATR ve Garman-Klass Volatilitesi) ve trend takip eden özellikler (Bollinger Bantları ve MACD) dahil olmak üzere çeşitli piyasa koşullarını ve yatırımcı davranışlarını kapsar. Birden fazla getiri döneminin dahil edilmesi, modelin zaman bağımlılıklarını ve kalıplarını yakalamasına olanak tanır ve bu da gelecekteki hisse senedi hareketlerini doğru bir şekilde tahmin etme yeteneğini artırır.

Bu çeşitli özellikleri entegre ederek model, hem tarihsel fiyat hareketlerinden hem de türetilmiş istatistiksel ölçülerden yararlanarak bilgilendirilmiş tahminler yapabilir ve bu da kantitatif finans ve algoritmik ticaret alanına katkıda bulunur.

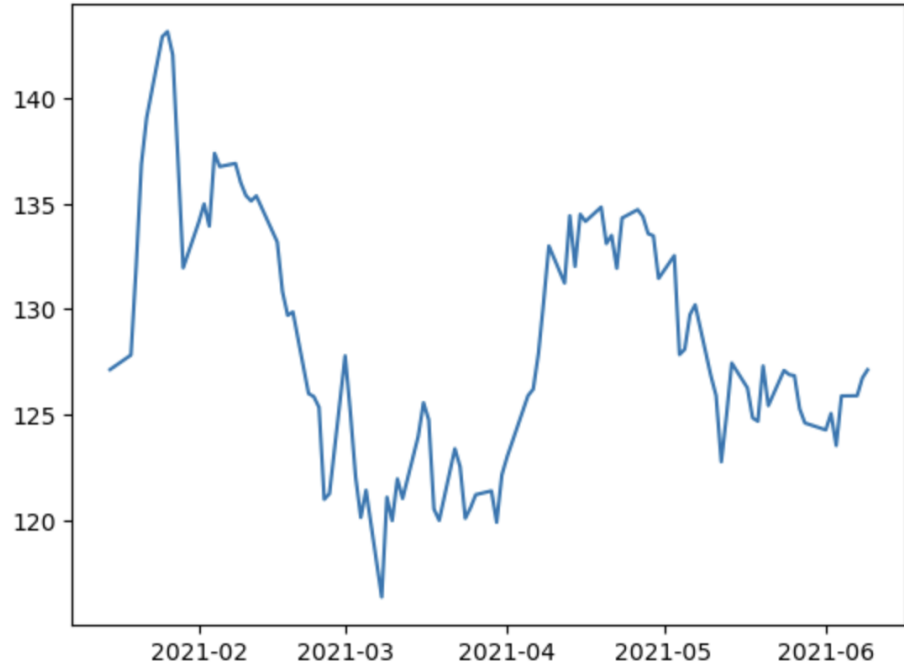
3.1.2 Verilerin Ön İşlenmesi

Veri Temizleme: Verileri aldıktan sonra, veri kalitesini ve tutarlılığını sağlamak için gerekli temizleme ve ön işleme adımlarını gerçekleştirdik. Buna sütun adlarının değiştirilmesi, eksik değerlerin ele alınması dahildi.

Veri Görselleştirme: Aşağıdaki şekiller bir hisse senedi (APPL) için verilerin nasıl görüldüğünü göstermektedir.

Date	Open	High	Low	Close	Adj Close	Volume
2021-01-15	128.779999	130.220001	127.000000	127.139999	124.605324	111598500
2021-01-19	127.779999	128.710007	126.940002	127.830002	125.281586	90757300
2021-01-20	128.660004	132.490005	128.550003	132.029999	129.397842	104319500
2021-01-21	133.800003	139.669998	133.589996	136.869995	134.141357	120150900
2021-01-22	136.279999	139.850006	135.020004	139.070007	136.297485	114459400
...	...	...	...	...	...	...
2024-01-08	182.089996	185.600006	181.500000	185.559998	185.072495	59144500
2024-01-09	183.919998	185.149994	182.729996	185.139999	184.653595	42841800
2024-01-10	184.350006	186.399994	183.919998	186.190002	185.700836	46792900
2024-01-11	186.539993	187.050003	183.619995	185.589996	185.102417	49128400
2024-01-12	186.059998	186.740005	185.190002	185.919998	185.431549	40444700

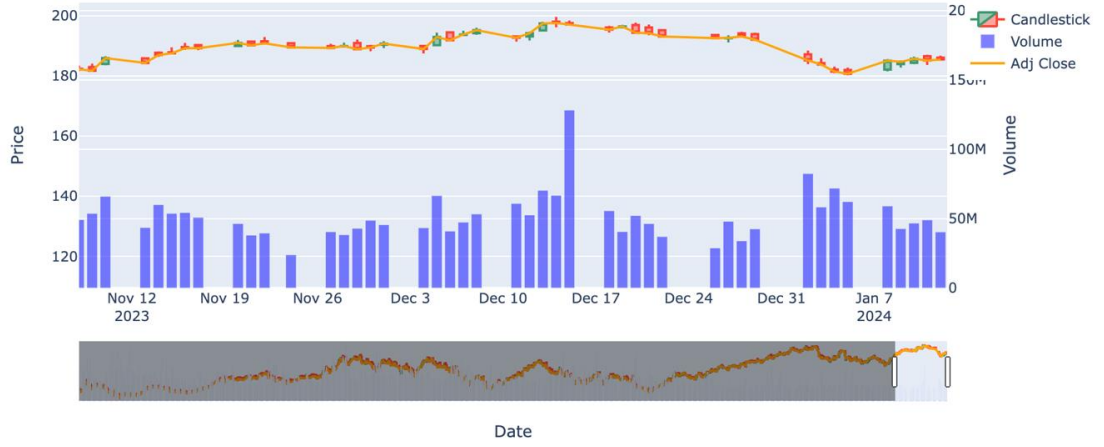
Şekil 3.1 APPL hisse senedi verileri 1



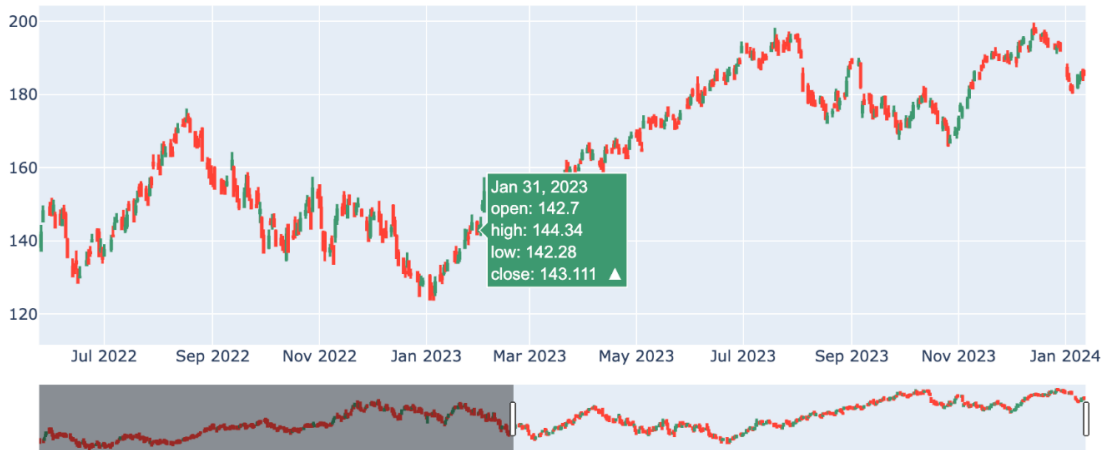
Şekil 3.2 APPL hisse senedi verileri



Şekil 3.3 APPL hisse senedi verileri



Şekil 3.4 APPL hisse senedi verileri



Şekil 3.5 APPL hisse senedi verileri

Veri Toplama: Analizi kolaylaştırmak için verileri aylık düzeyde topladık.

Veri Filtreleme: Her ay için en likit 150 hisse senedini dolar hacmine göre filtreledik.

3.2 Özellik Mühendisliği

Özellik mühendisliği, ham verileri ilgili kalıpları ve ilişkileri yakalayan bilgilendirici özelliklere dönüştürerek makine öğrenimi modellerinin başarısında çok önemli bir rol oynar. Bu bölümde, tahmine dayalı modelleme için anlamlı özellikler çıkarmak üzere veri kümemize uygulanan özellik mühendisliği sürecini özetliyoruz.

3.2.1 Teknik göstergeler

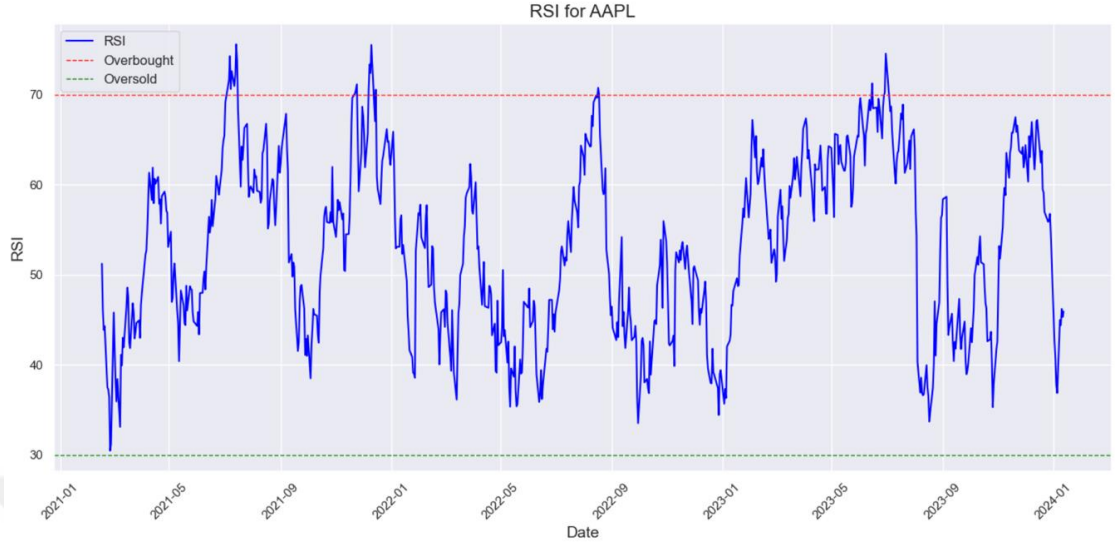
Her bir hisse senedi için piyasa eğilimleri ve fiyat hareketleri hakkında değerli bilgiler sağlayan çeşitli teknik göstergeler hesapladık. Bu göstergeler şunları içerir:

- Göreceli Güç Endeksi (RSI)
- Bollinger Bantları (BB)- Alt, Orta ve Üst Bantlar
- Hareketli Ortalama Yakınsama Sapması (MACD)
- Ortalama Gerçek Aralık (ATR)
- Garman-Klass Volatilitesi
- Dolar Hacmi

Bu teknik göstergeler, hisse senedi fiyatlarını tahmin etmede önemli faktörler olan piyasa duyarlılığı, volatilité ve momentum için vekil olarak hizmet eder.

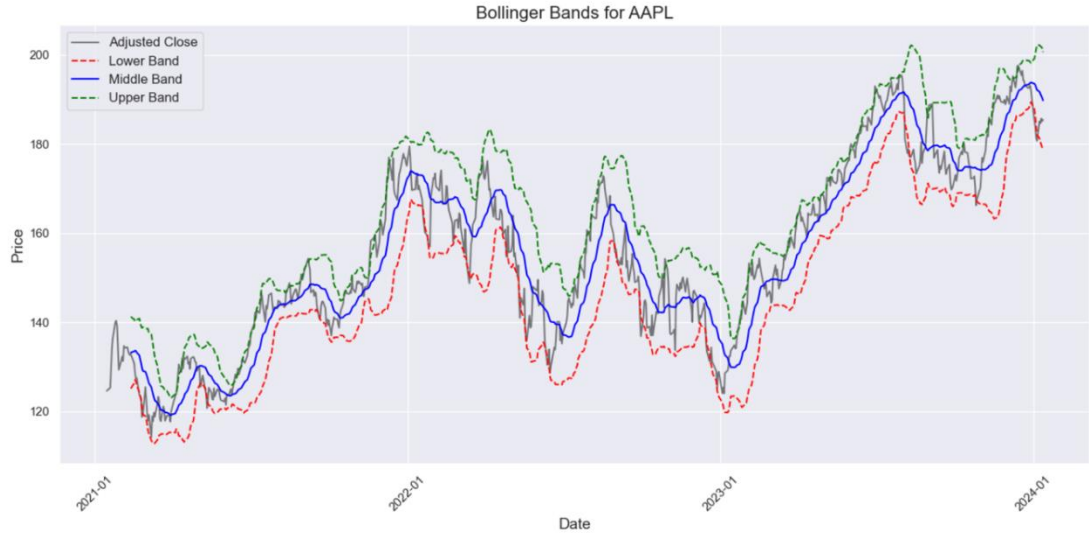
Teknik Göstergeler Görselleştirme: Aşağıdaki şekiller yukarıda bahsettiğimiz teknik göstergelerin görselleştirilmesini göstermektedir.

A- RSI'yi görselleştirme örneği



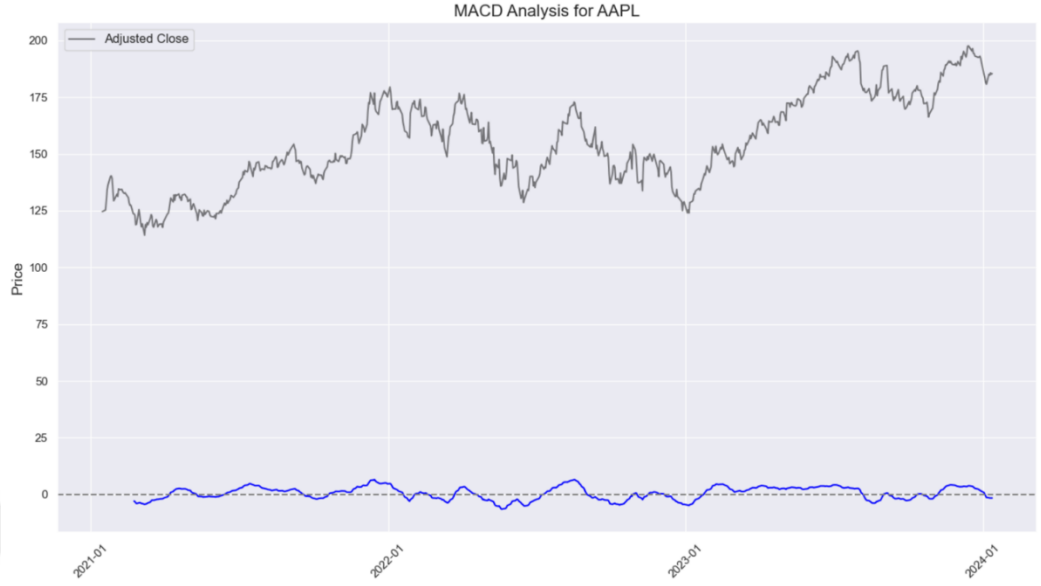
Şekil 3.6 APPL hisse senedi için RSI

B- BB'yi görselleştirme örneği



Şekil 3.7 APPL hisse senedi için BB

C- MACD'yi görselleştirme örneği



Şekil 3.8 APPL hisse senedi için MACD

D- ATR'yi görselleştirme örneği



Şekil 3.9 APPL hisse senedi için ATR

3.2.2 Özellik normalleştirilmesi

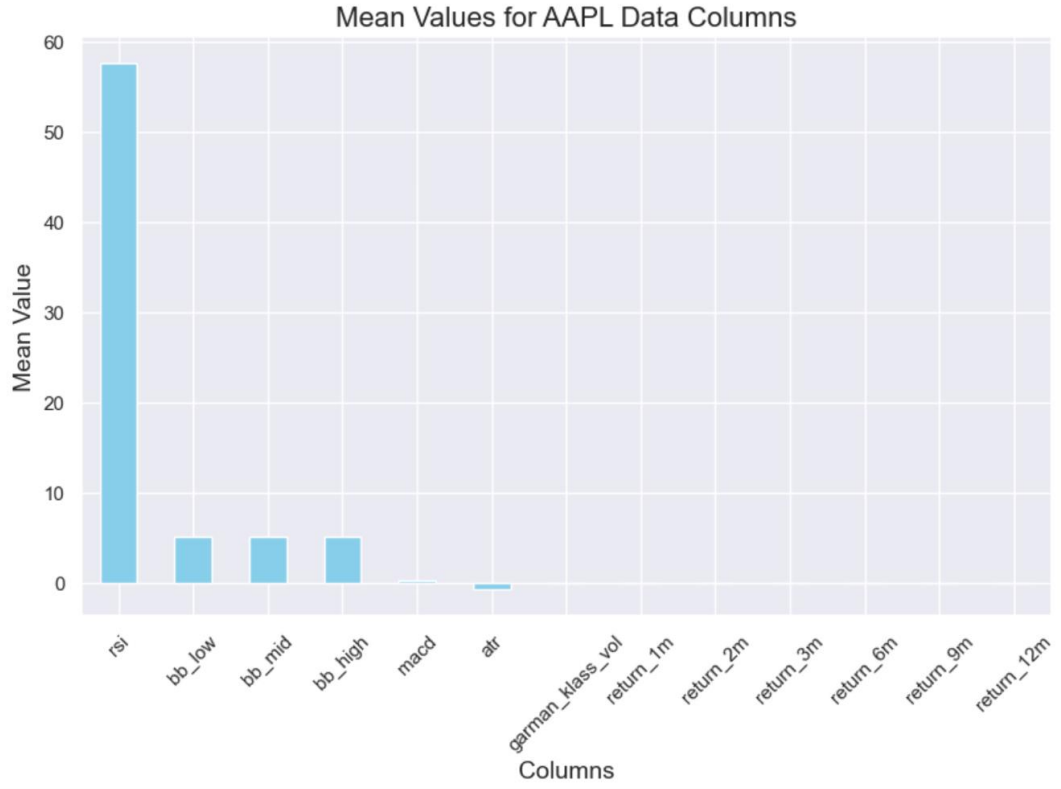
Özellikler arasında tutarlılık ve karşılaştırılabilirlik sağlamak için uygun dönüşümler ve normalleştirme teknikleri uyguladık. Bu, özellikleri ortak bir aralığa ölçeklendirmeyi ve sıfır ortalama (Zero Mean) ve birim varyansa (Unit variance) sahip olacak şekilde standartlaştırmayı içeriyordu.

3.2.3 Aylık getiriler

Hisse senedi fiyat hareketlerindeki zamansal kalıpları ve eğilimleri yakalamak için çeşitli zaman ufukları (1, 2, 3, 6, 9 ve 12 ay) için aylık getirileri hesapladık. Bu getiriler, hisse senedi performansı hakkında değerli tarihsel bilgiler sağlayarak gelecekteki fiyat değişikliklerinin tahmin edilmesine yardımcı olur.

Veri Seti

date	ticker	rsi	bb_low	bb_mid	bb_high	macd	atr	gamma_klass_yol	return_1m	return_2m	return_3m	return_6m	return_9m	return_12m
2023-01-01	AAJL	59.7760237228974	2.8609942725481677	2.819035700648904	2.9628771287496403	1.209262720966977	-0.483581841572936	0.000185466985153417	0.262535329327426	0.26759004675797883	0.044100795543534885	0.0275960373629344062	-0.01963336253657173	-0.0016803348504602038
2023-01-01	AAPL	57.7047857216723	4.815472786389511	4.910443411977488	5.025414673565485	0.6373967864476424	-0.025051862251491	8.63985725938856-05	0.11952127541202994	-0.01271335781770857	-0.0196317973748487105	-0.019126182461981833	-0.00309414402485183	-0.01497245097415112
2023-01-01	ABEV	38.838149648419254	4.910211359797884	4.98794619696488	5.065681038132076	1.812545791084371	0.09147255056887374	0.000182528531146543	-0.07077116558264016	-0.0382750642269148	0.006223806032629015	0.008150374894158754	0.0038742867976131468	0.009520234479635177
2023-01-01	ABNB	62.68115888386562	4.627402727933239	4.602043654339232	4.77880002745407	1.326122807164632	-0.8513151864038063	0.0001555433895317014	0.2933583120888219	0.012927782024846305	0.00019513102813011313	0.00309807729606823	-0.026802450752088202	-0.006802450752088202
2023-01-01	ABT	52.70797682942654	4.675988308797	4.699317766134511	4.723027181281052	0.4329102274026479	-0.5338118973897259	5.6088516640881234-06	0.011481315900545621	0.01600202139341114	0.03824069595054881	0.004144217884758629	-0.00140735662043182	-0.01033761813243589
2023-01-01	ACN	51.89430466601752	5.581254148020043	5.602756306383582	5.644258465469682	-0.0902910832448934	0.17842254004138347	0.0001655017351729283	0.05003538290306878	-0.0300725238802563	-0.0043734151887711	-0.018888363071232	-0.006812335924873847	-0.0179862598462486
2023-01-01	ADBE	81.95714701306554	5.78702453248303	5.858914018508927	5.930803513069551	0.5914471375198297	-1.0482793448583074	7.516035398227834-05	0.10046651912528914	0.02617910278402636	0.05155112840706414	-0.01680951724578244	-0.00742027743306513	-0.0300828889502581
2023-01-01	ADI	56.718668863822	5.06883912300938	5.102604708121382	5.138380303741827	0.43006789548020559	-0.7770852948199952	8.962716225440584-05	0.04353756253875172	0.000369849529159177	0.06406850342951227	0.010065158930996667	0.01320644712819655	0.005324788759750065
2023-01-01	ADSK	60.24299257446214	5.208483413588483	5.30057273818132	5.36266113403777	1.013829948372281	-0.4681129621720853	0.00029336825648861413	0.1513887155997482	0.03218808573829873	0.001335904040489795	-0.0008957439628826114	0.014342158738530582	0.0106109293137147
2023-01-01	AMAT	56.05013864296402	4.58805557250127	4.682488450450575	4.778973329651023	0.6934217821488511	-0.7447835678979913	0.00030314615135170265	0.14489626581275148	0.008585415994231790	0.08174881548381019	0.00932157715143167	-0.014200255617645007	-0.0166109293137147
2023-01-01	AMD	57.70138725870094	4.1447255770082	4.24635243457485	4.384544826989495	0.43915121494375897	-0.897195522271597	0.000333368226780326	0.16025943830524585	-0.01610282646054988	0.0775763844515887	-0.01474832972045666	-0.014200255617645007	-0.03430623889098617
2023-01-01	AMGN	55.43678201732665	5.48196297055559	5.53898746902304	5.592011920961249	-1.3702981457433587	-0.23845895381677943	-0.0007320400212232975	-0.03898780835206265	-0.06073219122370399	-0.0204011791430391	0.005730173874055833	0.011346430210106193	0.01144378695962011
2023-01-01	AMT	54.2362727173205	5.31209866401301	5.362887407795458	5.413678120108575	2.9857388911086236	-0.03651807906382829	0.000167025805065197	0.05422272060646858	0.00853409791756855	0.0279326786948363	-0.0292207542792947	-0.009134892188041876	-0.007688310212448979
2023-01-01	AMZN	81.8056759075506	4.43034476412799	4.554327997898733	4.67813518980967	0.899589930803349	-0.6708276521887532	8.992200588348864-05	0.2277380254008018	0.03356752163093969	0.00240177634126095	-0.04302948937703913	-0.02051539944028237	-0.03052807655649025
2023-01-01	AON	58.49437655847895	5.890760832780075	5.743167378441286	5.7955738238045845	0.8380196899102625	-0.06488455788015353	8.63447386456838-05	0.063646888872067	0.01762558955571582	0.042631380869537086	0.015862854706412133	0.0119146315014851105	0.0125166680969406
2023-01-01	AVGO	58.29664286971892	6.293125608307315	6.3330119352614729	6.372898256922127	0.2849934082026993	-0.5626621560177991	0.0012041295973177705	0.042696295862245015	0.034661674120055164	0.07585057020358837	0.017675203983318566	0.00878011597826302	0.00258323841457006
2023-01-01	AXP	70.7193909681555	4.93148967882523	5.0236359490929	5.13380230935551	1.38742704239031	0.8706239684834836	0.0001793810849782503	0.18591880454762166	0.055405137199429744	0.05745596028822764	0.02268234736978308	-0.001353597474248953	-0.00114328214276263
2023-01-01	BA	64.73966658333	5.05888895061381	5.347990158270089	5.3900914214967965	1.0191040438147934	-0.5158522740778416	0.0001971058212049833	0.181868090771578	0.09121140568553789	0.1434370968784314	0.0495974383149654	0.040628103281958206	0.005161231087154939
2023-01-01	BAC	58.2312840993107	3.4948234346484316	3.531176405970738	3.567525958820044	0.470136614488856	-0.73955838889284	-0.000229865648167188	0.07125605589191396	-0.028897737303258	-0.003271607855662245	0.010152780581873216	-0.01977631484627783	-0.01977631484627783
2023-01-01	BIB	54.948282870030915	5.60315743332949	5.653154443813919	5.703148454094889	0.3120385778803866	-0.438360748218898	3.527306414878254-05	0.050483821287254906	-0.02366637492225289	0.008687295380688752	0.051832532362073216	0.03828553271880917	0.021258883955746295
2023-01-01	BKNG	71.20799446607715	7.630873273981409	7.736777700488745	7.756777700488745	1.840482128998081	1.8723603943570386	-0.9020448276856116	0.1997581919435305	0.08191957890537643	0.0819578286800783	0.0819578286800783	0.01077385453144832	-0.000116191846532194
2023-01-01	BOWL	59.6818056955095	6.538135888458889	6.57942633797191	6.622716786484831	0.6602704721371071	0.172758896160187	-0.3206236294002615	0.0371729950514016	0.0332759443044947	0.057788916900515586	0.023896641332837196	0.02458592760705328	-0.004367117656918218
2023-01-01	BMV	46.9886042917588	4.230244317963314	4.246745705219702	4.26724705044609	-0.31123363618146	0.3548785490025874	0.000901922719337412	0.0177174115340961	-0.0449518705669037	-0.01826137603266446	5.3826210485164576-06	-0.001447752969602963	0.0143908373428313
2023-01-01	BRK-B	62.40961890553768	5.72292641722057	5.74487105753475	5.74505696078638	-0.2519557388238125	0.1221348587969039	0.0001355993136337412	0.00848168951523736	0.01173561981829646	0.01825132683764666	0.05984868763613051	-0.00356452883848335	-0.004503215293197588
2023-01-01	BXN	54.40961890553768	4.2360267701134	4.40069651853842	4.458846812375706	1.90362844382485	-0.29487227197380054	0.00017832839154810965	0.28482056647467821	0.023913785943517373	0.01732659732148913	-0.09843748137188086	-0.0023844738830271447	-0.0235240111484963
2023-01-01	C	67.3048024383708	3.780245284582947	3.8628763672035028	3.94350714882408	1.439620156682964	-0.96761363148138	0.001183580374859383	0.15454351211828087	0.038681632796675	0.04850410305047104	0.0029464838155941715	0.0113088784684883	-0.0148146193515768235
2023-01-01	CAT	56.28699574068246	5.4495476788238615	5.507608724516132	5.566168772108302	0.9975561136263943	0.6100505024151242	0.0002602450300071607	0.05818139338875814	0.0350302411000305	0.054054031873802216	0.0249773927864432	0.023388171712544824	0.02085140265745317
2023-01-01	CCL	60.53531162104124	2.249308817850232	2.4028158618702057	2.568950968018193	0.70918306259172	-1.18135801002893	0.00026742051741633783	0.34343165405442677	0.04385215788536425	0.000691627842111676	0.000209591993759815	-0.0580395275409687	-0.0191502454403595
2023-01-01	CHTR	60.53531162104124	5.859630830635885	5.9472560540732164	6.034876024181044	1.2004059534452576	-0.739103235793379	0.00017894651060140064	0.13332347579278324	-0.00895936488015825	0.01490997080430811	-0.0193450270455915	-0.01201807302617715	-0.035540634801287374
2023-01-01	CI	51.0943981827904	3.8628763672035028	3.71760731871228	3.747346611002128	-0.7849122343706188	0.134307316871821	0.00011894651620245488	-0.044274778253330676	-0.017089733579635815	0.024821453772029267	0.00238889410337498	0.02788022938350964	-0.017297411966067
2023-01-01	CMAA	61.8853348622886	3.5797029673411	3.6433653867299548	3.7069868411884894	1.632748145777922	-0.769960460413033	-0.0007347414248508306	0.1340565559571955	0.0404347202651903	0.0756418806383552	0.01071955323107732	0.0014414121809191963	-0.017297411966067
2023-01-01	CMG	64.53402907200298	7.206402320860557	7.321639174710714	7.43687604873637	0.9138426331884941	-0.20450903300840264	0.000159383676691997	0.1868874404182613	0.005960495510030989	0.030190774837381994	0.0058684467903954	0.0137778372938112	0.0086012554865316
2023-01-01	COF	68.2730418446784	4.47823557398647	4.827050116945845	4.77785678493243	1.803547251568199	0.3631071884832625	0.000493377021250231	0.2678486479150586	0.07381726207023064	0.04130577852732742	0.01540808695807886	0.00230250774467517426	-0.0156122583567601
2023-01-01	COPI	53.817839507891984	4.702135884000381	4.74724808613834	4.79427629027287	0.24875285826948074	0.6074802386948074	0.032796580794789934	-0.035674136309079854	-0.00707934363409969	0.430020119654375994	0.032130013552505	0.0034869812982323	-0.0034869812982323
2023-01-01	COST	61.8918403684354	6.082901479016275	6.148770273991749	6.21463086875584	0.4842276992367454	-0.0561524286222094	0.4053453477785954-05	0.1199591473223637	-0.02541271885912366	0.00368817894707919	-0.009217147831376467	-0.00368817894707919	-0.00368817894707919
2023-01-01	CRM	68.816984651357	4.894204040319											



Şekil 3.11 Veri seti

3.3 Makine Öğrenimi Modelleri

Bu bölümde, iki makine öğrenimi modelinin metodolojisini ve uygulamasını özetliyoruz. Biri S&P 500 verilerini etiketlemek için, diğeri ise çeşitli özelliklere ve teknik göstergelere dayalı olarak test etmek için.

3.3.1 K-Means ile kendi kendini etiketleme

Bu hisse senedi tahmin modelinin bu bölümünde, önceden belirlenmiş hedef Göreceli Güç Endeksi (RSI) değerlerine dayalı olarak verileri kendi kendine etiketlemek için bir kümeleme tekniği kullanıyoruz. Bu adımın amacı, verileri belirli RSI eşiklerini temsil eden farklı kümelere ayırmaktır. Bu kendi kendine etiketleme süreci, verileri daha uygun bir şekilde yapılandırarak sonraki tahmin modelleri için hazırlamaktadır. Burada kullanılan kümeleme tekniği, popüler bir denetimsiz makine öğrenme algoritması olan K-Means'tir.

- K-Means Kodu

Başlangıç Merkezlerinin Tanımlanması (Initial Centroids Definition)

```
# Define initial centroids based on target RSI values
target_rsi_values = [30, 70]
initial_centroids = np.zeros((len(target_rsi_values), 13))
initial_centroids[:, 0] = target_rsi_values
initial_centroids
```

Kod, K-Means algoritması için başlangıç merkezlerini tanımlayarak başlar. Merkezler, kümeleme süreci için başlangıç referans noktalarıdır. Bu durumda, 30 ve 70 hedef RSI değerleri, aşırı satım ($RSI < 30$) ve aşırı alım ($RSI > 70$) koşullarını belirlemek için kullanılan tipik RSI eşiklerine karşılık gelir. Merkezler, kümelerin (2) ve özelliklerin (13) sayısına karşılık gelen bir sıfır matrisi olarak başlatılır ve ardından ilk sütun hedef RSI değerlerine ayarlanır.

Kümeleme Fonksiyonu (Clustering Function)

```
def get_clusters(df):
    df['cluster'] = KMeans(n_clusters=2,
                           random_state=0,
                           init=initial_centroids).fit(df).labels_
    return df
```

Bir 'get_clusters' fonksiyonu, K-Means kümeleme algoritmasını veri setine uygulamak için tanımlanır. Fonksiyon, bir DataFrame 'df' yi girdi olarak alır ve K-Means algoritması tarafından atanan küme etiketlerini içeren yeni bir 'cluster' sütunu ekler. 'KMeans' yapıcı aşağıdaki parametrelerle çağrılır:

'n_clusters=2': Algoritmanın 2 küme oluşturması gerektiğini belirtir.

'random_state=0': Rastgele tohumun sabitlenmesiyle tekrar üretilebilirliği sağlar.

'init=initial_centroids': Önceden tanımlanmış merkezleri hedef RSI değerlerine göre kullanır.

'fit' yöntemi, K-Means modelini veri üzerinde eğitmek için kullanılır ve 'labels_' öznetiliği, her veri noktası için küme etiketlerini sağlar.

Kümeleme Fonksiyonunun Uygulanması (Applying the Clustering Function)

```
dataONE = dataONE.dropna().groupby('date', group_keys=False).apply(get_clusters)
```

Kümeleme fonksiyonu, eksik değer içeren satırları düşürdükten sonra 'dataONE' veri setine uygulanır. Veriler, her tarih grubunda kümeleme yapılmasını sağlamak için 'date' sütununa göre gruplandırılır, bu da verilerin zamansal yapısını korur. 'groupby' yöntemi verileri gruplandırır ve 'apply' yöntemi 'get_clusters' fonksiyonunu her gruba uygular. 'group_keys=False' parametresi, orijinal indeksin korunmasını sağlar.

'Cluster' sütununun eklenmesi, verileri RSI değerlerine göre segmentlere ayırarak, kalıpları tanımlamaya ve modelin sonraki adımlarında daha doğru tahminler yapmaya yardımcı olur.

Etiketli Veri Seti

Data, labeled															
date	ticker	ret	lb_low	lb_mid	lb_high	meand	atr	german_bkxw_idx	return_1m	return_3m	return_6m	return_9m	return_12m	cluster	
2023-01-01	AAL	59.770327228974	2.50094672541077	2.81093750044804	2.9005771287496403	1.326257220666577	-0.48305184157006	0.000184500581152417	0.202533339527425	0.50750204675797883	0.04410079543234885	0.0276003735234402	-0.016653023255857173	-0.001882348504400038	1
2023-01-01	AAPL	57.79478548718733	4.81021126380511	4.910448411977448	5.005414825505464	0.037089784446371	-0.0505186225251491	9.03051750059860e-05	0.11022127541202994	-0.01271335781170867	-0.0185317072754847105	-0.01912001484198183	-0.00265414482482183	-0.0138724309715112	1
2023-01-01	ABV	38.83814904818254	1.91021126380511	4.9878413889448	5.005081038130278	-1.812543791084371	0.0914732559587374	-0.00018232823114543	-0.37077115308264818	-0.0382750642289148	0.00523026603620215	0.008130014894198754	0.0038234478835177	0	
2023-01-01	ABNB	62.6811508886502	4.427427783309	4.80048454329353	4.77800052747407	1.326123871648332	-0.851315184403003	0.000135433953517014	0.29258312588219	0.342984398841405	0.0129778304443055	0.0001951102913811315	-0.0080243573089302	-0.0080243573089302	1
2023-01-01	ADT	52.107160942654	4.87618620396787	4.89811778154511	4.732027181281052	0.403912377406479	-0.530118873697298	5.808811650589123e-05	0.21148115160824821	0.2160202130411114	0.006409583095481	0.00414421788475828	-0.00140735602547818	-0.00103278143328389	0
2023-01-01	ADN	51.894366601752	5.16125414820243	5.65278308658832	5.84425844546862	-0.060291885344004	0.178425454133847	0.000146501751719263	0.2053353082035878	-0.035725023820565	-0.0404734151887711	-0.0001233064873847	-0.017862584842284	-0.017862584842284	0
2023-01-01	ADRE	61.9571477300554	5.787452394033	5.85891410868827	5.83082031309651	0.59144713718897	-1.048279344848074	7.51389398227834e-05	0.1004651812528914	0.0061781027843838	0.05151128407986414	-0.01688095124578244	-0.007402073743005313	-0.0030628888850281	1
2023-01-01	ADI	56.7186668322	5.6682611250308	5.10264708121382	5.13883033741827	0.4030879596465059	-0.777082948190952	8.962718254540583e-05	0.0483756253875172	0.000389944926159177	0.0640858234961227	0.00165515883006667	0.01329644713819555	0.00532478878750065	1
2023-01-01	ADSK	66.242827446214	5.20848241358463	5.306511343277	5.3905611343277	1.013820948372281	-0.4681126821720883	0.0002938264489413	0.151387155987462	0.00218805973826873	0.001335950444038795	-0.00089543962882814	-0.01434125816353052	-0.01212576259251337	1
2023-01-01	AMAT	56.031364499432	4.5880557129237	4.68248945450375	4.77872329651023	0.854371782148811	-0.744785815697913	0.0000314615135170625	0.1448626561275148	0.0068541594421793	0.08174897454931919	0.0003215771514317	0.01198692878750268	-0.01861090933127147	1
2023-01-01	AOD	57.701887289784	4.14472556173302	4.25463243246785	4.384544829992495	0.489132149473887	-0.88719552271597	0.000333896099174028	0.1020294032242495	0.0771759454151987	0.0301414822249586	-0.014482025261954507	-0.0343623699596617	-0.0343623699596617	1
2023-01-01	AMZN	35.4382071732958	5.18126267805659	5.35897468623454	5.50071832691248	-1.370284182432047	-0.2345895381677245	-0.0001945021223975	-0.388488730862825	-0.16073216722372099	-0.0504011176143091	0.00513071301781063	0.01138643210718015	0.0114737869682911	0
2023-01-01	AMT	54.33287227613205	5.31208865420191	5.34268740754938	5.413676129108575	0.2857785811808356	-0.0361380719005281	0.0001470251608036197	0.0544227542978947	0.006342097917562654	0.02783367869446352	-0.0026207545279347	-0.00613482186641878	-0.00788810212144879	1
2023-01-01	AMZN	61.8085768735506	4.5402417841296	4.55427367895673	4.6787331888067	0.9886889630863349	-0.870627821887352	8.96220088604884e-05	0.2277380254089018	0.0036752136308869	0.00224017765412005	-0.00281948507703913	-0.00261538694028257	-0.0005801656806825	1
2023-01-01	ADN	58.49433955847895	5.800768287307073	5.74314723841286	5.79027382304554	0.808611808910285	-0.36489402813553	8.434778346188330e-05	0.030648888872067	0.01762255005571592	0.042633185858037053	0.01184261501481105	0.01251669500688450	-0.01184261501481105	1
2023-01-01	AVGO	58.2995428787882	0.293155683073731	6.330311832814729	6.37288256822127	0.284894382820883	-0.562821580177871	-0.0001294120391317705	0.04828623086245015	0.00406167410005194	0.078802700308837	0.0176752083318586	0.00878011381830322	0.00283223444187006	1
2023-01-01	ADP	70.71983080152	4.814047186523	5.03203942940019	5.12642203003521	1.32472475229531	0.87652964824584	0.0017881904879503	0.1881188424762106	0.3044821571994428	0.021400928227764	0.0236352173878608	0.0118003971214485	0.00114332821472653	1
2023-01-01	BA	64.78905865333	3.30088806306121	5.247899158278089	5.300301414907865	1.0181404328147324	-0.515822719778149	0.0001913860217164023	0.11885860025775178	0.3812114098025715	0.1434740887842314	0.04859740831488544	0.00408233281898208	0.0001412327154339	1
2023-01-01	BAC	58.231264899317	3.494624343494871	3.531175420970738	3.56726382820044	4.070138514688958	-0.77390588898994	-0.000239890094416788	0.0172360509191396	-0.02898730737330258	-0.00327187895582240	0.01013276058181828	0.00139130555743912	-0.01877631494622783	1
2023-01-01	BID	54.84828287300315	5.80315743333348	5.023134343813819	5.7031484343888	0.310385177860308	-0.430087444021888	3.52372641487825e-05	0.05084881287254828	-0.02386032178002589	0.00887268308887752	0.05163253232027216	0.00288052718809175	0.01215883855748295	1
2023-01-01	BKNG	71.2078846507115	7.83827273881409	7.755871720486745	7.84848212899801	1.872382834527386	-0.820348278656119	0.8422384125124578	0.1987581918435055	0.081857860637643	0.081857860637643	0.038282862647608	0.011077395433144832	-0.00011618185032194	1
2023-01-01	BLK	59.6851804853089	5.3381388894889	6.57842833787181	6.82271678648831	0.8607084721317071	0.172783889160187	-0.00028339824864e-05	0.0713272930514016	0.0332758433044847	0.05778891898081886	0.02389941332837186	0.02434890762192173	-0.00438711700891818	1
2023-01-01	BBY	46.88984217817989	4.22543437982314	4.2461487352781932	4.2524793248408	-0.311233838108148	0.2849785400538274	-0.00089122719823732	0.017717711151438981	-0.0448918789589307	-0.0186103703289448	0.29623108818437408	0.01118989367242813	0.01118989367242813	0
2023-01-01	BWB	52.282166535196	5.722816741327287	5.7148271138153178	5.77456543097808	-0.2118817882851238	0.123149819887839	0.001194890132874713	0.00848146388162378	-0.0117546718126646	0.0182013283176888	0.00948481938188301	-0.0006215267017198	-0.0006215267017198	1
2023-01-01	BK	64.6889136949935	4.2603687570134	4.40069851853842	4.548494381375708	1.1594836381024895	-0.2942727150073064	0.0001193328813413968	0.2846386846787821	0.023913786343617973	0.073369976231877	0.00363718137188008	-0.003847478636174417	0.0025324911158463	1
2023-01-01	CAT	67.3040328437878	7.8202485885847	3.862878367303038	3.84350743882406	1.408629188808864	-0.96781382148158	-0.00118388031489383	0.154433121026087	0.000881627007335	0.0483402198204734	0.0028486488156841715	0.0113088784884882	-0.01146119515878235	1
2023-01-01	CAT	56.2898957408846	4.44407418620615	5.02760872451632	5.0616971720002	0.9676891432102943	0.0150006241512142	0.000024900200077362	0.058167388878184	0.005002241100055	0.0540452728864452	0.0429772827886452	0.0223887171254084	0.0088143268514317	1
2023-01-01	CCL	66.53387183104124	2.24828817803232	2.4038158818102057	2.588628088907193	1.11813880100803	-0.000281420114132763	0.34243143402442877	0.04385171885384257	0.006916138242111676	0.006916138242111676	0.03020915993768815	-0.008082954043967	-0.0491502416433895	1
2023-01-01	CHTR	54.1933136489124	8.89036830630885	5.947280548732184	6.05476024181904	1.2054036304462378	-0.73973232579379	0.000117894788918084	0.1320245787827833	-0.0899358488015625	0.0148909788420311	0.01281807320617175	0.03954054801287374	0.03954054801287374	1
2023-01-01	CMA	68.61884651377	4.884284433820615	5.01757575258915	5.143853114257893	1.320071504182678	-0.447747718710271	0.18420184162205458	-0.3447417882330876	-0.7889232587802816	-0.3008191566717143	0.04482145377200587	0.02438289110327488	0.027890328585994	0
2023-01-01	CDCO	55.954869283449	3.82410218119101	3.94440411424332	3.88601541982367	0.178418826547403	-0.66601541982367	0.0001193328813413968	0.023913786343617973	0.00363718137188008	0.0148909788420311	0.01281807320617175	0.03954054801287374	0.03954054801287374	1
2023-01-01	CSC	47.47700680383174	3.4644440398187054	3.4644440398187054	3.518724215032471	-0.0047167810343	0.9183235175123	0.000443868813754949	-0.001808647512025217	-0.02744861403885378	0.01871788628221358	0.00037483308728321	-0.01058729143143953	-0.0073808383068622	1
2023-01-01	CVS	40.622300122330	4.131481818888858	4.425323227387	4.49192472820883	-1.37478828720883	-0.8447388494288	0.0001930328888937126	0.346788488782243	-0.0682433424488786	-0.0214785026651812	0.00742584304842737	-0.0011777888073396	-0.0011777888073396	1
2023-01-01	COF	47.0805652324229	5.08712809313878	4.62700311894545	4.77780678430243	1.803472151669138	0.393137188432058	0.00064837701252021	0.28784864749158956	0.0736172820782384	0.0413207783273742	0.01400800829057886	-0.0032021714467317428	-0.001022838367601	1
2023-01-01	CAL	62.6730418447034	4.74024537389647	4.62700311894545	4.77780678430243	1.803472151669138	0.393137188432058	0.00064837701252021	0.28784864749158956	0.0736172820782384	0.0413207783273742	0.01400800829057886	-0.0032021714467317428	-0.001022838367601	1
2023-01-01	CVS	40.622300122330	4.131481818888858	4.425323227387	4.49192472820883	-1.37478828720883	-0.8447388494288	0.0001930328888937126	0.346788488782243	-0.0682433424488786	-0.0214785026651812	0.00742584304842737	-0.0011777888073396	-0.0011777888073396	1
2023-01-01	COF	47.0805652324229	5.08712809313878	4.62700311894545	4.77780678430243	1.803472151669138	0.393139								

K-Means kümelemesinin kullanımı, özellikle finansal zaman serileriyle çalışırken verileri ön işlemek ve kendi kendine etiketlemek için etkili bir yöntemdir. Verileri RSI eşiklerine göre segmentlere ayırarak, model piyasa koşullarının ve yatırımcı davranışının inceliklerini daha iyi yakalayabilir. Bu kümeleme adımı, yalnızca veri yapısını basitleştirmekle kalmaz, aynı zamanda modelin anlamlı kalıpları ve trendleri tespit etme yeteneğini artırarak nihai tahmin doğruluğunu artırır.

3.3.2 Spektral kümeleme

Hisse senedi tahmin modelinin ikinci aşamasında, verilerin ilk modelde kendi kendine etiketlendiği yöntemin etkinliğini test etmek ve değerlendirmek için Spektral Kümeleme kullanıyoruz. Spektral Kümeleme, benzerlik matrisinin özdeğerlerini kullanarak boyut azaltma işlemi yapar ve ardından K-Means gibi geleneksel kümeleme algoritmalarını uygular. Bu model, tüm S&P 500 hisseleri üzerinde eğitim ve değerlendirme yapılacak şekilde tasarlanmıştır ve geniş bir finansal veri yelpazesinde kümeleme performansını kapsamlı bir şekilde analiz etmektedir.

Spektral Kümeleme Kodu:

Veri Hazırlığı (Data Preparation)

```
data0 = dataONE.iloc[:, 2:]
X = data0.iloc[:, :-1].values
y = data0.iloc[:, -1].values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Veri seti 'dataONE', ilk modelden eklenen 'cluster' sütunu ile birlikte, Spektral Kümeleme modeli için hazırlanır. Kod, özellik sütunlarını ('X') ve hedef etiketleri ('y') izole eder. Özellikler teknik göstergeleri ve getirileri içerirken, hedef K-Means algoritması tarafından atanan küme etiketidir.

Veri seti, eğitim ve test setlerine 80-20 oranında ayrılır; bu, modelin geniş bir veri kısmında eğitim yapmasını ve bağımsız bir alt kümede test edilmesini sağlar.

'train_test_split' fonksiyonu, tekrar üretilebilirliği sağlamak için sabit bir 'random_state' ile kullanılır.

Model Eğitimi ve Tahmin (Model Training and Prediction)

```
spectral_clustering = SpectralClustering(n_clusters=2, affinity='nearest_neighbors', random_state=42)
spectral_clustering.fit(X_train)
y_pred = spectral_clustering.fit_predict(X_test)
```

Spektral Kümeleme, 'n_clusters=2' parametresiyle başlatılır; bu, modelin verileri iki kümeye ayırmayı hedeflediğini belirtir. 'affinity='nearest_neighbors' parametresi, kümeleme için kullanılan benzerlik matrisinin en yakın komşulara dayalı olduğunu belirtir, bu da kümeleri belirgin olmayan şekillerde tanımlamak için etkili olabilir. Model, 'X_train' veri seti üzerinde eğitilir ve 'X_test' veri setinde tahminler yapılır.

Değerlendirme Ölçütleri (Evaluation Metrics)

```
# Evaluate
silhouette_avg = silhouette_score(X_test, y_pred)
ari = 1-adjusted_rand_score(y_test, y_pred)
nmi = 1-normalized_mutual_info_score(y_test, y_pred)

accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
```

Modelin performansı, çeşitli metriklerle değerlendirilir:

Doğruluk (Accuracy): Doğru şekilde sınıflandırılmış örneklerin toplam örneklerin oranını ölçer. 0.895 doğruluk, modelin test verilerinin %89.5'ini doğru etiketlediğini gösterir.

Kesinlik (Precision): Pozitif tahminler arasında doğru pozitif tahminlerin oranını gösterir. 0.872 kesinlik, model pozitif bir etiket tahmin ettiğinde, bu tahminin %87.2 doğru olduğunu belirtir.

Geri Çağırma (Recall): Gerçek pozitifler arasında doğru pozitif tahminlerin oranını ölçer. 0.860 geri çağırma, modelin tüm gerçek pozitif örneklerin %86.0'ını başarıyla tanımladığını gösterir.

Silhouette Skoru: Kümelerin bağlılık ve ayrılmasını değerlendirir. 0.550'lik bir silhouette skoru, kümelerin orta düzeyde ayrıldığını ve bağlılık gösterdiğini, ancak daha da iyileştirilebileceğini gösterir.

Düzeltilmiş Rand Endeksi (ARI): Gerçek etiketlerle tahmin edilen kümeler arasındaki benzerliği ölçer, şansa göre düzeltilmiştir. 0.375 ARI, gerçek ve tahmin edilen kümeler arasında orta düzeyde bir uyum olduğunu belirtir.

Normalleştirilmiş Karşılıklı Bilgi (NMI): Gerçek etiketlerle tahmin edilen kümeler arasındaki paylaşılan bilgi miktarını ölçer. 0.477 NMI, orta düzeyde paylaşılan bilgi olduğunu, gerçek ve tahmin edilen kümeler arasında bazı uyumların olduğunu gösterir.

Spektral Kümeleme, kümelerin doğrusal olarak ayrılabilir olmadığı veri setleri için özellikle kullanışlıdır; bu da onu finansal verilerle çalışırken etkili bir yöntem yapar. Tüm S&P 500 hisseleri içeren veri setine Spektral Kümeleme uygulayarak, model, piyasa davranışlarının ve trendlerinin geniş bir yelpazesini yakalayabilir. Değerlendirme metrikleri, kümeleme modelinin ilk modelden gelen kendi kendine etiketlenmiş verilerle ne kadar uyumlu olduğunu ortaya koyar ve iyileştirme alanlarını vurgular. Bu kapsamlı değerlendirme, kümeleme kalitesini anlamaya ve modelin finansal verileri etkili bir şekilde segmentlere ayırma yeteneğini değerlendirmeye yardımcı olur.

3.4 Donanım ve Yazılım

3.4.1 Yazılım

Uygulama, Jupyter Notebook ortamında Python programlama dili kullanılarak geliştirilmiştir. Python, çok yönlülüğü, kapsamlı kütüphaneleri ve veri analizi ve makine

öğrenimi görevlerinde kullanım kolaylığı nedeniyle seçilmiştir. Geliştirme süreci boyunca aşağıdaki Python kütüphaneleri kullanılmıştır:

- **statsmodels:** Yuvarlanan regresyonlar gerçekleştirmek için kullanılır.
- **sklearn:** Kümeleme, sınıflandırma ve değerlendirme metrikleri dahil olmak üzere çeşitli makine öğrenimi görevleri için kullanılır.
- **pandas_datareader:** Yahoo Finance gibi çevrimiçi kaynaklardan finansal verileri almak için kullanılır.
- **matplotlib:** Veri görselleştirme ve çizim için kullanılır.
- **statsmodels:** İstatistiksel analiz ve modelleme için kullanılır.
- **numpy:** Sayısal hesaplama ve Sayısal hesaplama ve dizi işlemleri için kullanılır.
- **datetime:** Tarihleri ve saatleri işlemek için kullanılır.
- **yfinance:** Yahoo Finance'den geçmiş piyasa verilerini almak için kullanılır.
- **pandas_ta:** Teknik analiz göstergelerini hesaplamak için kullanılır.
- **warnings:** Uyarılar: Çalışma zamanı sırasında gereksiz uyarıları bastırmak için kullanılır.

Bu kütüphaneler veri ön işleme, özellik mühendisliği, model eğitimi, değerlendirme ve görselleştirme için kapsamlı bir araç seti sağlamıştır.

3.4.2 Donanım

Uygulama, veri analizi ve makine öğrenimi görevleri için güçlü ve verimli bir donanım platformu sağlayan bir MacBook Pro M1 üzerinde geliştirilmiş ve yürütülmüştür. MacBook Pro M1, enerji verimliliğini korurken yüksek performanslı bilgi işlem yetenekleri sunan en yeni Apple Silikon çipine sahiptir. Bu donanım kurulumu, karmaşık veri işleme görevlerinin sorunsuz bir şekilde yürütülmesini sağlamış ve makine öğrenimi modellerinin hızlı bir şekilde geliştirilmesini ve test edilmesini kolaylaştırmıştır.

4. BULGULAR ve TARTIŞMA

4.1 Bulgulara Giriş

Bu bölümde makine öğrenimi modellerini kullanarak S&P 500 borsasının hareketinin tahmin edilmesine yönelik araştırmamızın sonuçlarını sunuyoruz. Borsa verilerini analiz etmek ve yaklaşımımızın etkinliğini değerlendirmek için K-Means kümelemesiyle kendi kendini etiketleme ve Spectral kümelemeyle test etme dahil olmak üzere çeşitli teknikler kullandık.

Modelin performansı Doğruluk, Kesinlik, Geri Çağırma, Siluet Skoru, Düzeltilmiş Rand İndeksinin Karşılaştırması (ARI) ve Normalleştirilmiş Karşılıklı Bilgi (NMI) gibi çeşitli ölçümlerle değerlendirilir.

4.2 Sonuçların Sunumu

4.2.1 K-Means modeli

Önceden belirlenmiş hedef Göreceli Güç Endeksi (RSI) değerlerine dayalı olarak verileri kendi kendine etiketlemek için bir kümeleme tekniği kullanıyoruz. Bu adımın amacı, verileri belirli RSI eşiklerini temsil eden farklı kümelere ayırmaktır. Bu kendi kendine etiketleme süreci, verileri daha uygun bir şekilde yapılandırarak sonraki tahmin modelleri için hazırlamaktadır. Burada kullanılan kümeleme tekniği, popüler bir denetimsiz makine öğrenme algoritması olan K-Means'tir.

Model, K-Means algoritması için başlangıç merkezlerini tanımlayarak başlar. Merkezler, kümeleme süreci için başlangıç referans noktalarıdır. Bu durumda, 30 ve 70 hedef RSI değerleri, aşırı satım ($RSI < 30$) ve aşırı alım ($RSI > 70$) koşullarını belirlemek için kullanılan tipik RSI eşiklerine karşılık gelir. Merkezler, kümelerin (2) ve özelliklerin (13) sayısına karşılık gelen bir sıfır matrisi olarak başlatılır.

K-Means kümelemesinin kullanımı, özellikle finansal zaman serileriyle çalışırken verileri ön işlemek ve kendi kendine etiketlemek için etkili bir yöntemdir. Verileri RSI eşiklerine göre segmentlere ayırarak, model piyasa koşullarının ve yatırımcı davranışının inceliklerini daha iyi yakalayabilir. Bu kümeleme adımı, yalnızca veri yapısını basitleştirmekle kalmaz, aynı zamanda modelin anlamlı kalıpları ve trendleri tespit etme yeteneğini artırarak nihai tahmin doğruluğunu artırır.

4.2.2 Spektral modeli

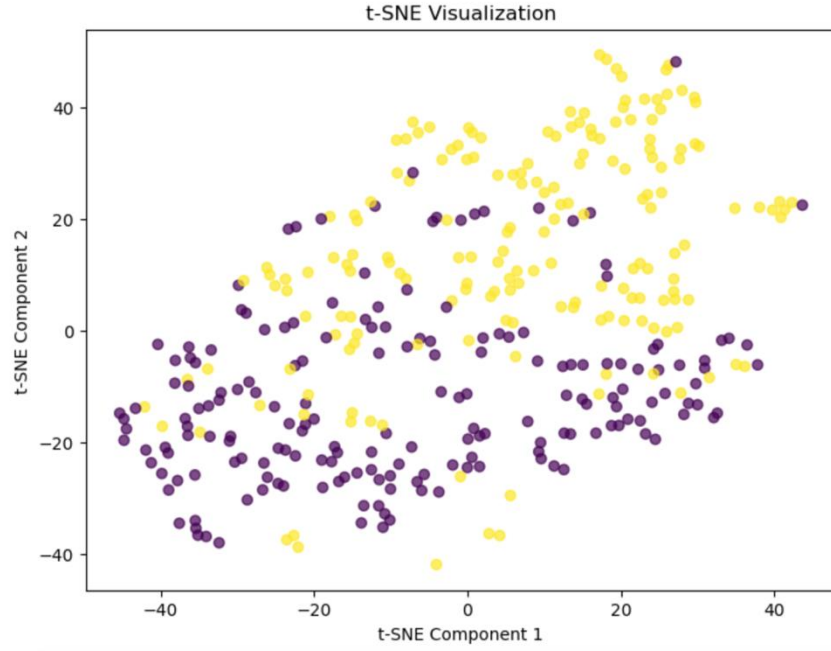
K-means kümeleme kullanılarak gerçekleştirilen ilk kendi kendine etiketleme aşamasının ardından, ikinci adımda test ve değerlendirme için Spektral Kümeleme uygulanmıştır. Bu model, daha önce etiketlenmiş verileri kullanarak tahmin çerçevesinin doğruluğunu ve sağlamlığını artırmak amacıyla tüm S&P 500 hisseleri üzerinde eğitilmiştir. Spektral Kümeleme, verilerdeki doğrusal olmayan yapılarla başa çıkma konusundaki etkinliği ile bilinen güçlü bir tekniktir ve benzerliklerine dayalı olarak veri noktalarını ayırt etmek için kullanılır. Spektral Kümeleme modeli, çeşitli metrikler boyunca güçlü performans sergilemiştir:

Model adı	Doğruluk	Hassasiyet	Hatırlama	Siluet Puanı	(ARI)	NMI
Spektral Kümeleme	89.5%	87.2%	86.0%	55%	37.5%	47.7%

Şekil 4.1 Model değerlendirme metrikleri

Bu sonuçlar, Spektral Kümeleme modelinin verideki belirgin desenleri etkili bir şekilde tanımladığını ve yüksek doğruluk, kesinlik ve geri çağırma değerleriyle desteklendiğini göstermektedir. Silhouette skoru, kümelerin oluşumunun oldukça iyi tanımlandığını öne sürerken, ARI ve NMI değerleri, gerçek ve tahmin edilen kümeler arasındaki orta düzeyde bir anlaşmayı işaret etmektedir.

- Model Görselleştirme



Şekil 4.2 Model Görselleştirme

Özellikler Öneminin Detaylı Açıklaması:

PCA (Temel Bileşen Analizi): Temel Bileşen Analizi (PCA), büyük bir değişken setini daha küçük bir set haline dönüştüren ve büyük setteki bilgilerin çoğunu içeren yaygın olarak kullanılan bir boyut indirgeme tekniğidir. PCA'nın amacı, verilerdeki varyansı en çok açıklayan temel bileşenleri belirlemektir. Bu bağlamda, hisse senedi piyasası hareketlerini tahmin ederken çeşitli özelliklerin önemini belirlemek için PCA uygulanmıştır. PCA'dan elde edilen özellik önem değerleri, hangi özelliklerin veri setindeki varyansa en önemli katkıyı sağladığını anlamamıza yardımcı olur.

Özellik Öneminin Yorumlanması:

Özelliklerin önem değerleri aşağıda listelenmiştir

Özellik: bb_high, Değer: 0.8609

Özellik: bb_mid, Değer: 0.8590

Özellik: bb_low, Değer: 0.8559

Özellik: return_9m, Değer: 0.8267
Özellik: return_12m, Değer: 0.8242
Özellik: macd, Değer: 0.7763
Özellik: rsi, Değer: 0.7138
Özellik: return_1m, Değer: 0.7035
Özellik: return_6m, Değer: 0.6913
Özellik: atr, Değer: 0.6548
Özellik: return_2m, Değer: 0.5904
Özellik: cluster, Değer: 0.5861
Özellik: return_3m, Değer: 0.4746
Özellik: garman_klass_vol, Değer: 0.2040

Sonuçların Analizi:

1- En Yüksek Katkı Sağlayan Özellikler:

Bollinger Bantları (bb_high, bb_mid, bb_low): En yüksek önem değerleri, Bollinger Bantları göstergeleriyle ilişkilidir, özellikle 'bb_high' (0.8609), 'bb_mid' (0.8590) ve 'bb_low' (0.8559). Bu göstergeler, hisse senedi fiyat hareketlerini tahmin etmede kritik öneme sahiptir çünkü geçmiş fiyat hareketlerine göre volatilitiyi ve fiyat seviyelerini temsil ederler. Yüksek önem değerleri, Bollinger Bantlarının hisse senedi fiyat verilerindeki varyansı açıklamada önemli olduğunu göstermektedir.

2- Uzun Vadeli Getiriler:

9 aylık getiri (return_9m) ve 12 aylık getiri (return_12m): Bu özellikler sırasıyla 0.8267 ve 0.8242 önem değerlerine sahiptir. Bu, uzun vadeli getirilerin hisse senedi fiyatlarındaki temel trendleri yakalamada kritik olduğunu ve varyansa önemli ölçüde katkıda bulunduğunu göstermektedir.

3- Momentum Göstergeleri:

MACD (Hareketli Ortalama Yakınsama Diverjansı): 0.7763 önem değeri ile MACD, hisse senedi fiyat varyansını açıklamada önemli bir özelliktir. MACD, bir hissenin fiyatının iki hareketli ortalaması arasındaki ilişkiyi gösteren bir momentum göstergesidir ve potansiyel alım ve satım sinyallerini tanımlamaya yardımcı olur.

4- Göreceli Güç Endeksi (RSI):

RSI: RSI, 0.7138 önem değerine sahiptir. RSI, fiyat hareketlerinin hızını ve değişimini ölçer ve bir hissenin aşırı alım veya aşırı satım koşullarını belirlemek için kullanılır.

5- Kısa Vadeli Getiriler:

1 aylık getiri (return_1m) ve 6 aylık getiri (return_6m): Bu özellikler sırasıyla 0.7035 ve 0.6913 önem değerlerine sahiptir. Kısa vadeli trendler ve fiyat tersine dönüşleri hakkında içgörüler sağlarlar ve temel bileşenler tarafından açıklanan genel varyansa katkıda bulunurlar.

6- Volatilite ve Aralık:

Ortalama Gerçek Aralık (ATR): 0.6548 önem değeri ile ATR, volatilitenin önemli bir ölçüsüdür. Fiyat volatilitesinin derecesini gösterir ve hisse senedi fiyat hareketlerini anlamada önemli bir rol oynar.

7- Orta Vadeli Getiriler:

2 aylık getiri (return_2m) ve 3 aylık getiri (return_3m): Bu özellikler sırasıyla 0.5904 ve 0.4746 gibi daha düşük önem değerlerine sahiptir. Hâlâ ilgili olsalar da, uzun vadeli ve kısa vadeli getirilerle karşılaştırıldığında genel varyansa daha az katkıda bulunurlar.

8- Küme Atamaları:

Küme: 0.5861 önem değeri ile bu özellik, kendi kendine etiketleme aşamasında atanan küme etiketlerini yansıtır. Kümeleme işlemi, benzer özelliklere sahip hisseleri gruplandırır ve temel bileşenler tarafından açıklanan varyansa katkıda bulunur.

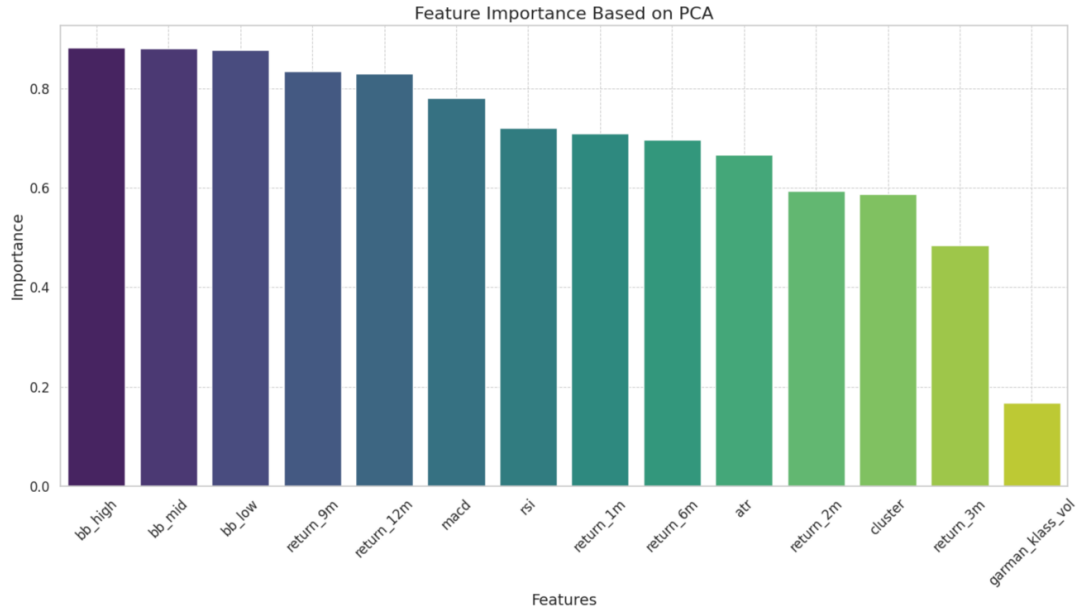
9- Garman-Klass Volatilitesi:

Garman-Klass Volatilitesi: 0.2040 ile en düşük önem değerine sahiptir. Bu, modele katkıda bulunsun da, veri setindeki varyansı açıklama açısından diğer özelliklerle karşılaştırıldığında nispeten minimal bir etkiye sahip olduğunu gösterir.

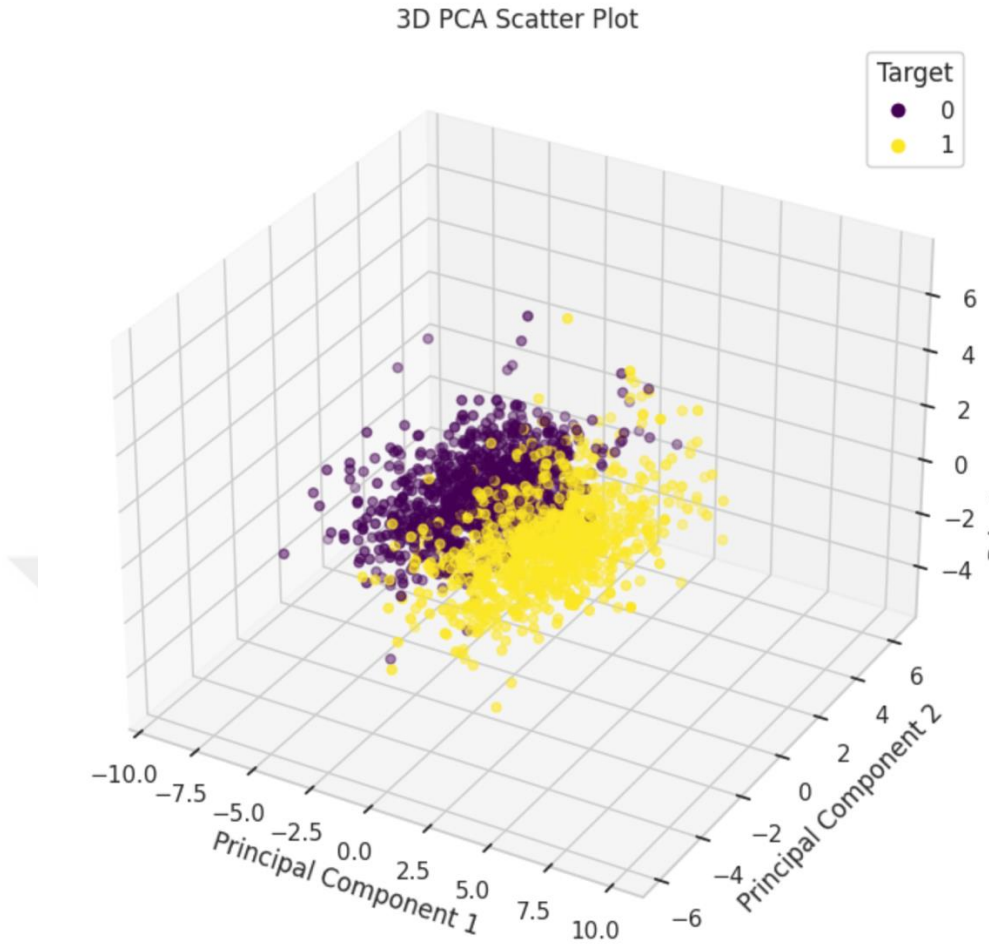
PCA'dan elde edilen özellik önem değerleri, hangi değişkenlerin hisse senedi fiyat verilerini açıklamada en etkili olduğunu anlamamıza yardımcı olur. Bollinger Bantları, fiyat volatilitesi ve trendleri yakalamada en önemli göstergeler olarak öne çıkar. Uzun vadeli getiriler de, hisse senedi hareketlerini anlamak için kritik bir rol oynar. MACD ve RSI gibi momentum göstergeleri ve ATR gibi volatilité ölçüleri, modelin piyasa dinamiklerini kapsamlı bir şekilde anlamasını sağlar. Bu önemli özelliklere odaklanarak,

model hisse senedi fiyat hareketlerini daha doğru ve sağlam bir şekilde tahmin edebilir ve algoritmik ticaret stratejilerinin etkinliğini artırabilir.

Bu detaylı özellik önem analizi, hisse senedi fiyatlarını yönlendiren temel faktörleri anlamaya yardımcı olur ve temel bileşen analizi üzerine kurulu sofistike ticaret modelleri geliştirmek için sağlam bir temel sağlar.



Şekil 4.3 PCA'ya dayalı özelliklerin önemi



Şekil 4.4 PCA'ya dayalı özelliklerin önemi

4.3 Tartışma

Bulgularımız bağlamında, araştırmamızın ilerlemelerini ve katkılarını açıklamak için yöntemlerimizi ve sonuçlarımızı önceki çalışmalarla karşılaştırmak ve karşılaştırmak önemlidir. Modelimizde Spektral Kümeleme uygulaması, literatürde araştırılan geleneksel denetimli öğrenme yöntemlerinden saparak hisse senedi piyasası tahminine yeni bir yaklaşım sunmaktadır.

(Liu vd., 2016), S&P 500 endeksini tahmin etmek için RBF çekirdeği ile SVM dahil olmak üzere çeşitli denetimli öğrenme modellerini kullandılar. Yaklaşımları, özellik seçimi ve model parametrelerini optimize etmeye odaklanarak %62.51 doğruluk elde etti.

Buna karşılık, modelimiz, önceden tanımlanmış etiketler olmadan verilerdeki yapıları belirlemek için denetimsiz bir öğrenme tekniği olan Spektral Kümelemeyi kullandı. Bu yaklaşım, yalnızca etiketleme sürecini basitleştirmekle kalmaz, aynı zamanda denetimli yöntemlerle kolayca tanımlanamayan karmaşık desenleri de yakalar. Liu ve ark. SVM'nin performansını vurgularken, modelimiz kümeleme yöntemlerinin önemli piyasa eğilimlerini ortaya çıkarmadaki etkinliğini vurgulamaktadır.

(Ghosh vd., 2021), LSTM ağları ve Random Forest kullanarak içgünlük hisse senedi fiyat hareketlerini tahmin ettiler ve %74'ün üzerinde doğruluk elde ettiler. Çalışmaları, kapanış fiyatları dışında geniş bir özellik yelpazesinin model performansını artırmadaki önemini vurguladı. Çalışmamız da Bollinger Bantları ve MACD gibi çeşitli teknik göstergelerden yararlanmaktadır ancak yöntem olarak Spektral Kümelemeyi kullanmaktadır. Bu zaman serisi tahmin modellerinden kümeleme tekniklerine geçiş, veri yapısına yeni bir bakış açısı sunarak zaman serisi analizinde ortaya çıkmayan piyasa rejimlerinin belirlenmesini sağlamaktadır.

(Liu vd., 2023), zaman serisi tahmini için N-BEATS algoritmasını tanıttı ve yüksek doğruluk ve düşük hata metrikleri ile üstün performans sergiledi. Yöntemlerinin, tarihsel veriler üzerine odaklanması, zaman serisi tahminini hassas bir şekilde yapma açısından modelimizin kümeleme temelli yaklaşımıyla tezat oluşturmaktadır. Bu karşılaştırma, hisse senedi piyasası tahmininde etkili tekniklerin çeşitliliğini vurgulamakta olup, modelimiz denetimsiz öğrenmenin piyasa kalıplarını belirlemede faydasını göstererek bu alana katkıda bulunmaktadır.

(Shi vd., 2024), doğrusal regresyon ve XGBoost modellerini karşılaştırarak, XGBoost'un daha düşük MSE ve RMSE değerleriyle doğrusal regresyondan daha üstün olduğunu belirtti. Araştırmamız, ileri makine öğrenimi modellerinin geleneksel olanlardan üstünlüğü konusundaki bulguları ile uyumludur. Ancak, modelimiz doğrudan tahmin yerine veri kümelemeye odaklanmaktadır. Bu yöntem, benzer özelliklere sahip kümelerin belirlenmesini sağlayarak, XGBoost gibi tahmin modellerini tamamlayan farklı bir analiz boyutu sunmaktadır.

(Htun vd., 2024), rastgele orman, SVM ve LSTM sınıflandırıcılarının performansını değerlendirdi ve LSTM'nin diğer yöntemlerden daha üstün olduğunu belirtti. Çalışmaları, ardışık verileri işleme konusunda LSTM'nin gücünü vurgularken, modelimiz benzer davranışlar sergileyen veri noktalarını gruplandırmayı öneren kümeleme yaklaşımı sunmaktadır. Bu kümeleme yaklaşımı, belirli kalıplar sergileyen sektörlerin veya piyasa etkinliği dönemlerinin belirlenmesinde özellikle yararlı olabilir ve stratejik yatırım kararları için kritik olabilir.

Çalışma	Yıl	Metot	Doğruluk
Liu vd.	2016	Lojistik Regresyon (LR), Gauss Ayırıştırma Analizi (GDA), Naive Bayes (NB) ve Destek Vektör Makinesi (DVM)	62%
Ghosh vd.	2021	Uzun Kısa Süreli Bellek (LSTM), Rastgele Ormanlar	74%
Liu vd.	2023	Sinir Temeli Genişleme Analizi (NBEATS)	99%
Shi vd.	2024	XGBoost	14816.86
Htun vd.	2024	Uzun Kısa Süreli Bellek (LSTM)	$P=8.46e-17$
Önerdiğimiz model	2024	Spektral Kümeleme Modeli	89%

Şekil 4.5 Mevcut Çalışmalarla ile karşılaştırma

4.5 Kısıtlamalar

Bu araştırmada elde edilen umut verici sonuçlara rağmen, çalışmamızın kapsamı ve kısıtlamalarının kapsamlı bir şekilde anlaşılmasını sağlamak için kabul edilmesi gereken çeşitli sınırlamalar vardır.

İlk olarak, birincil sınırlama modellerimizi eğitmek ve test etmek için kullanılan verilerle ilgilidir. Veri setimiz kapsamlı olmakla birlikte, yalnızca belirli bir dönemdeki S&P 500 hisse senetlerini kapsamaktadır. Bu zamansal ve endekse özgü kısıtlama, bulgularımızın diğer endekslere veya daha geniş piyasa koşullarına genellenebilirliğini sınırlandırabilir. Piyasa davranışları farklı dönemlerde ve farklı endeksler arasında önemli ölçüde değişiklik gösterebilir ve modellerimiz bu değişiklikleri etkili bir şekilde

yakalayamayabilir. Ayrıca, teknik göstergelerin kullanımı, sağlam olmasına rağmen, hisse senedi fiyatlarını önemli ölçüde etkileyebilecek makroekonomik faktörleri, haber olaylarını veya diğer temel verileri hesaba katmaz.

İkinci olarak, çalışmamızın yaklaşımı ağırlıklı olarak niceliksel ve büyük ölçüde geçmiş fiyat verilerine ve teknik göstergelere dayanmaktadır. Bu yaklaşım, piyasa duyarlılığı, jeopolitik olaylar ve hisse senedi fiyatlarını etkileyebilecek diğer dış etkiler gibi niteliksel faktörleri göz ardı edebilir. Bu tür niteliksel verilerin entegre edilmesi, tahmin modellerinin sağlamlığını ve doğruluğunu potansiyel olarak artırabilir.

Özetle, araştırmamız borsa tahmininde denetimsiz öğrenme tekniklerinin potansiyelini ortaya koysa da, sonuçları yorumlarken bu sınırlamaları göz önünde bulundurmamak önemlidir. Gelecekteki araştırmalar, daha geniş bir veri yelpazesini dahil ederek, alternatif kümeleme algoritmalarını keşfederek, değerlendirme metriklerini geliştirerek ve daha bütünsel ve geliştirilebilir tahmin modelleri geliştirmek için nitel faktörleri entegre ederek bu sınırlamaları ele alabilir.

5. SONUÇ

Bu tezde, S&P 500 endeksini tahmin etmek için makine öğrenimi tekniklerinin uygulanmasını araştırdık. Çalışmamızda, verileri kendi kendine etiketlemek için K-Means kümeleme ve test ve değerlendirme için Spektral Kümeleme olmak üzere iki farklı model kullandık. Bu yöntemlerin entegrasyonu, finansal tahminlerde denetimli öğrenme yöntemlerinden saparak ve finansal öngöründe denetimsiz öğrenme tekniklerinin potansiyelini göstererek piyasa tahminine yönelik yenilikçi bir yaklaşımı sergilemektedir.

K-Means kümeleme modeli, RSI, Bollinger Bantları, MACD, ATR ve çeşitli getiri metrikleri gibi temel teknik göstergelere dayalı olarak hisse senetlerini kategorize etmek için kullanıldı. Hedef RSI değerlerine dayalı başlangıç merkezlerini tanımlayarak, verileri anlamlı kümelerle etkin bir şekilde ayırdık. Bu kümeleme yaklaşımı, veriseti için önceden tanımlanmış etiketlere gerek kalmadan hisse senetlerini düzenlememize olanak tanıyan temel bir yapı sağladı. Modelin, verilerdeki gizli yapıları ortaya çıkarma yeteneği, denetimsiz öğrenmenin finansal piyasalardaki gizli yapıları ortaya çıkarmadaki faydasını vurgulamaktadır.

Bu temelin üzerine inşa ederek, tahminlerimizi değerlendirmek ve test etmek için Spektral Kümeleme uyguladık. Karmaşık ve doğrusal olmayan veri dağılımlarını işleme yeteneği ile bilinen Spektral Kümeleme, hisse senedi hareketi tahminlerimizi daha da geliştirmek için kümeleme verilerine uygulandı. Model, tüm S&P 500 hisse senetleri üzerinde, tahmin doğruluğunu artırmak için çeşitli teknik göstergelerden yararlanarak eğitildi. Değerlendirme metrikleri, doğruluk, kesinlik, geri çağırma, siluet skoru, düzeltilmiş rand indeksi (ARI) ve normalize karşılıklı bilgi (NMI) dahil olmak üzere yaklaşımımızın sağlamlığını ve etkinliğini gösterdi. Elde edilen yüksek doğruluk ve kesinlik oranları, modelin piyasa trendlerini tahmin etmedeki güvenilirliğini doğrulamaktadır.

Araştırmamız, finansal piyasa tahmininde denetimsiz öğrenme tekniklerinin etkinliğini göstererek mevcut bilgi birikimine katkıda bulunmaktadır. İlgili çalışmalarla yapılan karşılaştırma, yaklaşımımızın yenilikçi yönünü vurgulamaktadır. SVM, LSTM ve

XGBoost gibi geleneksel yöntemler, hisse senedi piyasası tahmininde önemli başarılar göstermiş olmasına rağmen, K-Means ve Spektral Kümeleme kullanımı tamamlayıcı bir bakış açısı sunmaktadır. Bu denetimsiz yöntemler, etiketleme sürecini basitleştirmenin yanı sıra piyasanın yapısal dinamikleri hakkında daha derinlemesine içgörüler sağlamaktadır.

Özellik önemine yönelik başlıca bileşen analizi (PCA), seçtiğimiz göstergelerin geçerliliğini daha da pekiştirdi. Bollinger Bantları (bb_high, bb_mid, bb_low), uzun vadeli getiriler (return_9m, return_12m) ve MACD'ye atfedilen yüksek önem değerleri, tahmin modelindeki önemlerini vurgulamaktadır. Bu bulgular, bu teknik göstergelerin piyasa trendlerini yakalamada kritik olduğunu doğrulayan daha geniş literatürle uyumludur.

Sonuç olarak, bu tez, makine öğrenimi uygulamalarının hisse senedi piyasası tahminindeki gücünü vurgulayarak kapsamlı bir analiz sunmaktadır. K-Means ve Spektral Kümeleme'nin başarılı bir şekilde uygulanması, finansal tahminlerin doğruluğunu ve güvenilirliğini artırma potansiyellerini vurgulamaktadır. Gelecekteki araştırmalar, tahmin performansını daha da artırmak için ek denetimsiz teknikler veya hibrit modeller entegre ederek bu çalışmayı genişletebilir. Bulgularımız, piyasa analizine yönelik daha sofistike ve ayrıntılı yaklaşımlar için yol gösterici niteliktedir ve yatırımcılar ile finansal analistlere piyasanın karmaşıklıklarını anlamada değerli araçlar sunmaktadır.

KAYNAKLAR

- Ahmad, A.Y.B., Gongada, T.N., Shrivastava, G., Gabbi, R.S., Islam, S. and Nagaraju, K., 2023. E-commerce trend analysis and management for Industry 5.0 using user data analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 11(11s), pp.135-150.
- Alderighi, S., Gurrola-Perez, P., Lin, K. and Speth, B., 2021. Circuit breakers and other market safeguards. London: World Federation of Exchanges.
- Aldridge, I., 2013. High-frequency trading: a practical guide to algorithmic strategies and trading systems (Vol. 604). John Wiley & Sons.
- Algorithmic developments in contemporary financial markets. *Economy and Society*, 45(2), pp.149-165.
- Aloud, M.E. and Alkhamees, N., 2021. Intelligent algorithmic trading strategy using reinforcement learning and directional change. *IEEE Access*, 9, pp.114659-114671.
- Andreas C. Müller, Sarah Guido, (2016), *Introduction to Machine Learning with Python: A Guide for Data Scientists*.
- Andriy Burkov, (2019), *The Hundred-Page Machine Learning Book*. Machine Learning.
- Applications in Finance and Algorithmic Trading <https://medium.com/@f4rid/the-10-most-useful-algorithms-and-their-applications-in-finance-3fb5afdb42d2>, Erişim tarihi: 13.01.2024
- Arnoldi, J., 2016. Computer algorithms, market manipulation and the institutionalization of high frequency trading. *Theory, Culture & Society*, 33(1), pp.29-52.
- Bollinger Bands: A powerful technical tool for traders www.britannica.com/money/bollinger-bands-indicator, Erişim tarihi: 06.11.2023
- Bradley, A.P. (1997) The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, 30, 1145-1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Chaboud, A.P., Chiquoine, B., Hjalmarsson, E. and Vega, C., 2014. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5), pp.2045-2084.
- Chan, E., 2013. Algorithmic trading: winning strategies and their rationale (Vol. 625). John Wiley & Sons.
- Choudhry, R. and Garg, K. (2008) A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology*, 39, 315-318. <http://waset.org/publications/8952/a-hybrid-machine-learning-system-for-stock-market-forecasting>
- classification vs clustering in machine learning <https://www.analyticsvidhya.com/blog/2021/11/quick-tutorial-clustering-data-science/> Erişim tarihi: 17.01.2024

- Cooper, R., Davis, M., Kumiega, A. and Van Vliet, B., 2020. Ethics for automated financial markets. Handbook on ethics in finance, pp.1-18.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to algorithm (3rd edition).
- Definition of Algorithmic Trading www.investopedia.com/articles/active-trading/101014/basics-algorithmic-trading-concepts-and-examples.asp, Erişim tarihi: 10.10.2023
- Evolution of Algorithmic Trading www.algomojo.com/blog/evolution-of-algorithmic-trading/, Erişim tarihi: 10.10.2023
- Ewim, D.R.E., Okwu, M.O., Onyiriuka, E.J., Abiodun, A.S., Abolarin, S.M. and Kaood, A., 2021. A quick review of the applications of artificial neural networks (ANN) in the modelling of thermal systems.
- Forecasting directional movements of stock prices for intraday trading using LSTM and random forests <https://arxiv.org/abs/2004.10178v2>
- Forecasting relative returns for S&P 500 stocks using machine learning <https://jfin-swufe.springeropen.com/articles/10.1186/s40854-024-00644-0>
- Forecasting S&P 500 Stock Index Using Statistical Learning Models <https://www.scirp.org/journal/paperinformation?paperid=72604>
- Galeshchuk, S., 2017. Technological bias at the exchange rate market. Intelligent Systems in Accounting, Finance and Management, 24(2-3), pp.80-86.
- GMM https://en.wikipedia.org/wiki/Gaussian_mixture_model, Erişim tarihi: 23.01.2024
- Henke, N. and Jacques Bughin, L., 2016. The age of analytics: Competing in a data-driven world.
- Huang, W., Nakamori, Y. and Wang, S.Y. (2005) Forecasting Stock Market Movement Direction with Support Vector Machine. Computers & Operations Research, 32, 2513-2522. <https://doi.org/10.1016/j.cor.2004.03.016>
- J. Welles Wilder, Jr. (1978), New Concepts in Technical Trading Systems.
- John J. Murphy, (1999) Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance.
- Johnson, B. (2010). Algorithmic trading & DMA: an introduction to direct access trading strategies. 4Myeloma Press, London.
- Kevin P. Murphy, (2012), Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series).
- Kirilenko, A.A. and Lo, A.W., 2013. Moore's law versus murphy's law: Algorithmic trading and its discontents. Journal of Economic Perspectives, 27(2), pp.51-72.
- K-Means https://en.wikipedia.org/wiki/K-means_clustering, Erişim tarihi: 23.01.2024
- Lange, A.C., Lenglet, M. and Seyfert, R., 2016. Cultures of high-frequency trading: Mapping the landscape of
- Nazareth, N. and Reddy, Y.Y.R., 2023. Financial applications of machine learning: A literature review. Expert Systems with Applications, p.119640.

- Pole, A., 2011. Statistical arbitrage: algorithmic trading insights and techniques. John Wiley & Sons.
- Predicting the S&P 500 stock market with machine learning models https://www.researchgate.net/publication/379076543_Predicting_the_SP_500_stock_market_with_machine_learning_models
- Pros and Cons of Algorithmic Trading www.shareindia.com/knowledge-center/algo/pros-and-cons-of-algorithmic-trading, Erişim tarihi: 17.10.2023
- Risks and Challenges of Algorithmic Trading <https://fastercapital.com/startup-topic/Risks-and-Challenges-of-Algorithmic-Trading.html>, Erişim tarihi: 02.11.2023
- Sanni, O., Adeleke, O., Ukoba, K., Ren, J. and Jen, T.C., 2024. Prediction of inhibition performance of agro-waste extract in simulated acidizing media via machine learning. Fuel, 356, p.129527.
- Spectral Clustering https://en.wikipedia.org/wiki/Spectral_clustering, Erişim tarihi: 23.01.2024
- Stefan Jansen, (2020), Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python.
- Technical Indicators www.investopedia.com/terms/t/technicalindicator.asp, Erişim tarihi: 06.11.2023
- The S&P 500 Index Prediction Based on N-BEATS <https://www.atlantispress.com/proceedings/icbis-23/125989798>
- Trevor Hastie, Robert Tibshirani, Jerome Friedman. (2001), The Elements of Statistical Learning: Data Mining, Inference, and Prediction. (Second Edition).
- Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M. and Iosifidis, A., 2017, August. Using deep learning to detect price change indications in financial markets. In 2017 25th European signal processing conference (EUSIPCO) (pp. 2511-2515). IEEE.
- Tyler Yamazaki, (2019), STOCK MARKET INVESTING FOR BEGINNERS, Blue Labyrinth Publishing.
- Types of Algorithmic Trading www.wrightresearch.in/blog/what-are-the-different-types-of-algo-trading/#Types_of_Algo_trading_strategies, Erişim tarihi: 06.11.2023
- What Is the Relative Strength Index (RSI) In Stocks www.nasdaq.com/articles/what-is-the-relative-strength-index-rsi-in-stocks, Erişim tarihi: 06.11.2023
- Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S. and Stoica, I. (2010) Spark: Cluster Computing with Working Sets. HotCloud, 10, 10-10. http://static.usenix.org/legacy/events/hotcloud10/tech/full_papers/Zaharia.pdf