



Giriş

Finansal piyasalardaki belirsizlik ve volatilité, yatırımcıların doğru tahminler yapmasını ve stratejik kararlar almasını oldukça zorlaştırmaktadır. Özellikle COVID-19 pandemi sonrası piyasalarda yaşanan dalgalanmalar, risk yönetimi ve portföy optimizasyonu gibi konuları daha da kritik hale getirmiştir. Geleneksel finansal analiz yöntemleri, piyasa dinamiklerini yeterince doğru şekilde yansıtamayarak, mevcut veri setlerinin yüzeysel analiz edilmesi ise piyasa trendlerini ve hisse senedi performanslarını anlamayı güçleştirmektedir. Ayrıca, pandemi sürecinin piyasalara etkileri tam olarak incelenmemiş ve firmaların bu dönemdeki performans değişimleri belirsiz kalmıştır.

Makine öğrenimi tekniklerinden yararlanılarak, borsa tahmini ve portföy optimizasyonu süreçlerinde daha etkili ve yenilikçi modellerin geliştirilmesi amaçlanmaktadır.

Veri odaklı yaklaşımlar ve gelişmiş makine öğrenimi algoritmaları sayesinde finansal piyasalardaki belirsizlikler daha doğru bir şekilde öngörülebilir.

Geliştirilen tahmin modelleri, yatırımcıların risk-getiri dengesini daha doğru bir şekilde analiz etmelerine ve daha sağlam temellere dayalı yatırım kararları almalarına yardımcı olacaktır.

Bu tahminler, yatırımcıların risk-getiri dengesini en iyi şekilde sağlayacak optimal portföy stratejilerinin geliştirilmesine katkı sağlayacaktır.

Literatür Taraması

- Bu çalışma için toplamda 22 adet makale araştırılmıştır. Bu makaleler, finansal tahminleme alanındaki yeniliklerin yatırımcılar için stratejik kararlar almada kritik bir rol oynadığını ortaya koymaktadır.
- Sağ tarafta bulunan QR kod ile bu çalışma için oluşturulan literatür tablosuna ve referanslara ulaşabilirsiniz.

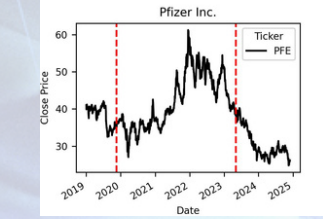


Yöntem ve Tasarım

VERİ SETİ

Sağlık sektöründe hisse senedi fiyatları için en uygun modeli belirlemek amacıyla, 20 büyük şirketin son 5 yıla ait günlük piyasa verileri Yahoo Finance platformundan elde edilmiştir.

Veri setimiz aşağıdaki niteliklere sahiptir:



- 01 Açılış Fiyatı
- 02 En Yüksek Fiyat
- 03 En Düşük Fiyat
- 04 Kapanış Fiyatı
- 05 Düzeltilmiş Kapanış
- 06 Hacim

YÖNTEMLERİN UYGULANMASI

| LSTM | XGBoost | ARIMA | RASTGELE ORMAN | SVM |
|--|--|---|--|--|
| LSTM hücreci, uzun vadeli bağımlılıkları öğrenmek için kullanılan bir tür tekrarlayan sinir ağı modelidir. Girdi, unutmaya ve çıktı kapısından oluşur. 1. Unutma Kapısı (Forget Gate): Hangi bilginin unutulacağına karar verir. 2. Giriş Kapısı (Input Gate): Hangi bilginin hücreye alınacağına belirir. 3. Aday Bellek (Candidate Memory): Yeni hücre durumunu oluşturur. 4. Hücre Durumu (Cell State): Önceki hücre durumu ve yeni bilgi birleşimi. 5. Çıkış Kapısı (Output Gate): Yeni gizli durumu oluşturur. Bu mekanizma sayesinde uzun dizilerle etkili bir şekilde çalışabilir. | XGBoost, makine öğreniminde yüksek doğruluk ve verimlilik için kullanılan bir algoritmadır. 1. Temel Yapı: Birden çok karar ağacı ardışık olarak oluşturulur, her ağaç bir önceki hatayı düzeltmeye çalışır. 2. Amaç Fonksiyonu: Tahmin hatalarını ve model karmaşıklığını minimize eder. 3. Düzenleme (Regularization): Aşırı öğrenmeyi önlemek için ceza terimleri içerir. 4. Paralel Hesaplama: Süreci, büyük veri setleriyle hızlı çalıştırabilir. | Kapsam: Zaman serisi tahmininde kullanılan bir modeldir. Bileşenler: 1. AR (AutoRegressive): Geriye dönük gecikmeler. 2. I (Integrated): Fark alarak durağanlık sağlama. 3. MA (Moving Average): Hata terimleri üzerinden modelleme. Kullanım Alanı: Durağan olmayan seriler, fark alarak modele uygun hale getirilir. | Rastgele Orman, birden fazla karar ağacından oluşan ve tahminlerin birleştirilmesiyle sonuç üreten bir yöntemdir. Hem sınıflandırmada hem de regresyon problemlerinde kullanılır. 1. Veri Girişi: Eğitim verisi hazırlanır. 2. Ağaçların Oluşturulması: Rastgele örneklerle karar ağaçları oluşturulur. 3. Tahmin: Her ağaç için tahmin yapılır. 4. Sonuçların Birleştirilmesi: Regresyonda tahminlerin ortalaması, sınıflandırmada çoğunluk oylaması alınır. 5. Çıktı: Nihai tahmin veya sınıflandırma sonucu elde edilir. | SVM, verileri optimal bir hiper düzlem ile ayırarak sınıflandırmaya ve regresyon problemlerini çözme gücü bir algoritmadır. Hem doğrusal hem doğrusal olmayan problemlerde etkilidir. 1. Veri Girişi: Eğitim verisi hazırlanır. 2. Hiper Düzlem Belirleme: Verileri sınıflara ayıran optimal hiper düzlem hesaplanır. 3. Çekirdek Fonksiyonu: Kullanım: Doğrusal olmayan problemlerde çekirdek fonksiyonu uygulanır. 4. Model Eğitimi: Performans iyileştirme için marjın maksimuma çıkarılır. 5. Tahmin/Çıktı: Yeni veriler sınıflandırılır veya tahmin yapılır. |

HİBRİT MODEL (LSTM+XGBoost)

Bu yöntemde, LSTM ve XGBoost modellerinin tahmin sonuçları birleştirilerek, hata oranlarına göre ağırlıklandırma yapılmıştır. Daha düşük hata oranına (MAPE) sahip model, birleştirilmiş tahmine daha fazla katkı sağlamaktadır. Bu sayede, modellerin güçlü yönlerini birleştirerek daha doğru ve genel hatası düşük bir tahmin modeli elde edilmiştir.

METRİKLERİN DEĞERLENDİRİLMESİ

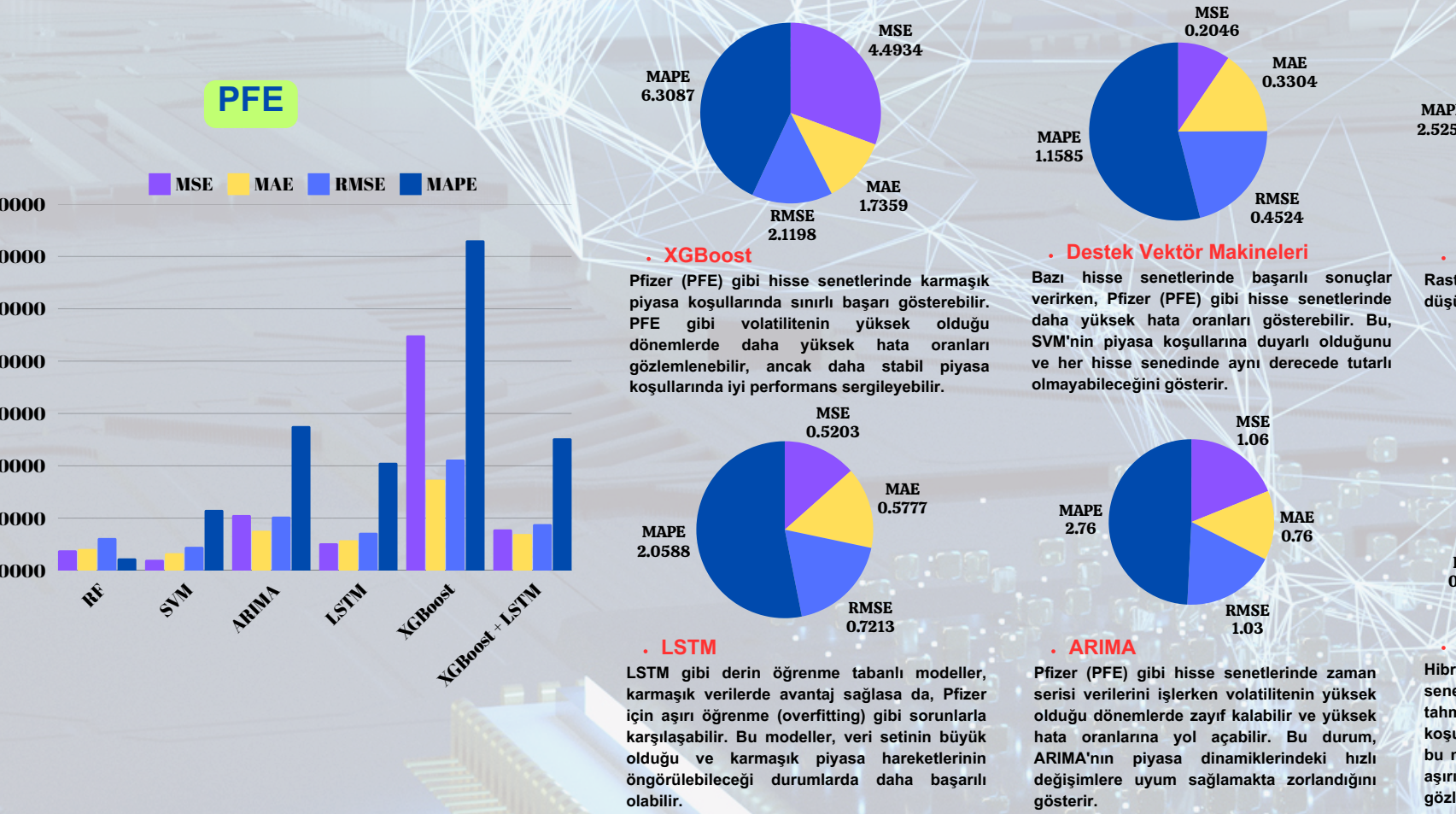
| MSE | MAE | RMSE | MAPE |
|---|--|---|---|
| MSE, tahmin edilen değer ile gerçek değer arasındaki farkın ortalamasıdır. Hata karelerinin büyüklüğünü vurgular. $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ | MAE, tahmin edilen değer ile gerçek değer arasındaki farkların mutlak değerlerinin ortalamasıdır. Hata büyüklüğünü ölçmek için kullanılır. $MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $ | RMSE, MSE'nin kareköküdür. Büyük hatalara MSE'ye göre daha hassastır. $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ | MAPE, mutlak hatanın gerçek değerlere oranının yüzdesel ortalamasıdır. Yüzdellik olarak hata oranını verir. $MAPE = \frac{1}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{ y_i } \times 100\%$ |

MODEL SEÇİMİ

Python ve yapay zekâ kütüphaneleriyle geliştirilen modeller, her algoritmanın farklı avantajlara sahip olduğunu ve hibrit yaklaşımların borsa tahminlerinde umut verici olduğunu göstermiştir. İstatistiksel değerlendirmeler, modellerin güçlü ve zayıf yönlerini belirleyerek daha etkin tahmin modellerinin geliştirilmesine katkı sağlamıştır.

DeneySEL Sonuçlar

- Modelimizin performans sonuçlarını anlayabilmek için PFE firmasını esas alalım.

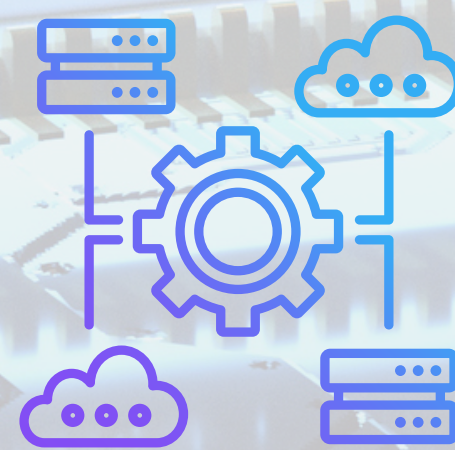


| Endeks/Hisse Senedi | Kriter | Modellerin Tahmin Performansı | | | | | |
|---------------------|--------|-------------------------------|--------|-------|--------|---------|----------------|
| | | RF | SVM | ARIMA | LSTM | XGBoost | XGBoost - LSTM |
| PFE | MSE | 0.3855 | 0.2046 | 1.06 | 0.5203 | 4.4934 | 0.7852 |
| | MAE | 0.4120 | 0.3304 | 0.76 | 0.5777 | 1.7359 | 0.6970 |
| | RMSE | 0.6209 | 0.4524 | 1.03 | 0.7213 | 2.1198 | 0.8861 |
| | MAPE | 0.2315 | 1.1585 | 2.76 | 2.0588 | 6.3087 | 2.5258 |
| JNJ | MSE | 2.4138 | 2.1428 | 2.07 | 4.5418 | 2.4862 | 2.5942 |
| | MAE | 1.1511 | 1.0857 | 1.04 | 1.6987 | 1.2071 | 1.2658 |
| | RMSE | 1.5536 | 1.4658 | 1.44 | 2.1311 | 1.5768 | 1.6107 |
| | MAPE | 0.1020 | 0.6982 | 0.67 | 1.0944 | 0.7772 | 0.8152 |
| BMY | MSE | 0.4857 | 0.7855 | 0.19 | 1.6459 | 2.9027 | 1.6612 |
| | MAE | 0.5172 | 0.5865 | 0.3 | 0.9456 | 1.1215 | 0.9279 |
| | RMSE | 0.6969 | 0.8863 | 0.44 | 1.2829 | 1.7037 | 1.2889 |
| | MAPE | 0.1781 | 1.1813 | 0.58 | 1.9444 | 2.4353 | 1.9641 |
| GSK | MSE | 0.2767 | 0.2867 | 0.31 | 0.4254 | 0.3284 | 0.3299 |
| | MAE | 0.3652 | 0.3734 | 0.38 | 0.4879 | 0.4104 | 0.4127 |
| | RMSE | 0.5260 | 0.5354 | 0.56 | 0.6522 | 0.5731 | 0.5731 |
| | MAPE | 0.1125 | 0.9473 | 1.05 | 1.2338 | 1.0364 | 1.0422 |
| TAK | MSE | 0.0454 | 0.0276 | 0.15 | 0.0417 | 0.0320 | 0.0324 |
| | MAE | 0.1459 | 0.1194 | 0.34 | 0.1529 | 0.1315 | 0.1334 |
| | RMSE | 0.2131 | 0.1663 | 0.39 | 0.2043 | 0.1788 | 0.1801 |
| | MAPE | 0.1552 | 0.8541 | 2.46 | 1.0995 | 0.9464 | 0.9599 |
| FREDE | MSE | 0.3453 | 0.1665 | 0.18 | 0.3689 | 0.1874 | 0.1915 |
| | MAE | 0.4061 | 0.2912 | 0.37 | 0.4821 | 0.3367 | 0.3440 |
| | RMSE | 0.5876 | 0.4081 | 0.42 | 0.6159 | 0.4329 | 0.4376 |
| | MAPE | 0.2836 | 1.0152 | 1.3 | 1.6878 | 1.1668 | 1.1998 |



Sonuçlar

ARIMA modeli stabil piyasa koşullarında başarılı sonuçlar verirken, yüksek volatilité dönemlerinde performans düşüşü yaşamıştır. LSTM modeli karmaşık verileri işleme avantajı sunsa da aşırı öğrenme (overfitting) nedeniyle doğruluğu olumsuz etkilenmiştir. Geleneksel modellerden Rastgele Orman (RF), düşük hata oranları ve tutarlı performansı ile öne çıkarken, Destek Vektör Makineleri (SVM) bazı hisse senetlerinde düzensiz sonuçlar vermiştir. XGBoost hızlı tahmin sağlasa da karmaşık piyasa koşullarında sınırlı kalmıştır. Hibrit modeller, dalgalı piyasalarda daha dengeli ve doğru sonuçlar elde etse de yüksek hesaplama maliyetleri ve aşırı öğrenme riski taşımaktadır. Çalışma, piyasa koşullarına uygun model seçiminde hibrit yaklaşımların önemli bir avantaj sağladığını göstermektedir.



Gelecek Çalışmalarımız

TÜBİTAK 2209-A Projesi kapsamında, modellerin performansını iyileştirmek için hiperparametre optimizasyonu yapılacak ve sonuçlar yeniden değerlendirilecektir. Ayrıca, endüstri mühendisliği perspektifiyle geliştirilen karar destek aracı, finansal tahminlerde doğruluğu artırarak stratejik kararların alınmasına katkı sağlamaktadır.