

Tesla Stock Price Prediction

Hikmet Terzioğlu

TOBB ETU Mühendislik Fakültesi

Bilgisayar Mühendisliği

Ankara,Türkiye video linki: <https://youtu.be/2eSKjm1c5pE>

hikmeterzioglu@gmail.com

Abstract—Bu çalışma, finansal verilerin zaman içindeki değişimlerini analiz etmek için ARIMA, LSTM ve RNN modellerinin nasıl kullanıldığını anlatmaktadır. Geleneksel ARIMA modelinin yanı sıra, derin öğrenme yöntemlerinden olan RNN ve LSTM modelleri de incelenmiştir. Veri seti olarak Tesla (TSLA) hisse senedi fiyatları kullanılmıştır. Raporun amacı, bu farklı yöntemlerin finansal veri analizindeki etkinliğini karşılaştırmaktır. Geleneksel ve derin öğrenme modelleri, veri setinin zaman içindeki değişen dinamiklerini yakalama yetenekleri bakımından değerlendirilmiştir. Bulgular, LSTM ve RNN gibi derin öğrenme modellerinin geleneksel ARIMA modeline kıyasla daha yüksek tahmin doğruluğu sağlayabileceğini göstermektedir. Sonuç olarak, finansal veri analizinde derin öğrenme tekniklerinin kullanımının artan önemi vurgulanmaktadır. Bu çalışmanın bulguları, yatırımcıların ve analistlerin daha iyi tahminlerde bulunmalarına yardımcı olabilecek potansiyel faydalar sunmaktadır.

Anahtar Kelimeler: Finansal Veri Analizi, ARIMA, LSTM, RNN, Derin Öğrenme

I. INTRODUCTION

Hisse senedi fiyatları, ekonomik ve finansal olaylara duyarlı olup ani ve öngörülemeyen değişiklikler gösterebilir.

Bu çalışma, Tesla hisse senedinin gelecekteki kapanış fiyatlarını tahmin etmek için çeşitli yöntemlerin uygulanmasını ele almaktadır. Geleneksel yaklaşımlar olarak ARIMA modeli, derin öğrenme tekniklerinden olan RNN ve LSTM modelleri ile karşılaştırılmıştır. Bu yöntemler, hisse senedi fiyatlarının zaman içindeki dinamik değişimlerini yakalama yetenekleri bakımından değerlendirilmiştir.

Araştırmanın temel amacı, farklı veri analizi ve tahmin yöntemlerini kullanarak Tesla hisse senedinin kapanış fiyatlarını ayrıntılı bir şekilde incelemektir. Veri seti kapsamlı bir şekilde tanıtıldıktan sonra verinin görsel ve istatistiksel analizi gerçekleştirilmiştir. İlk olarak, LSTM modeli kullanılarak tahminlemeler yapılmış ve modelin performansı kritik bir bakış açısıyla değerlendirilmiştir.

Bu çalışmada ayrıca, ARIMA modeli kullanılarak hisse senedi fiyatlarına ilişkin güvenilir tahminler elde edilmesi hedeflenmiştir. ARIMA modeli, zaman serisi verilerini analiz etmek ve gelecekteki değerleri tahmin etmek için yaygın olarak kullanılan bir yöntemdir. Modelin başarısı, tahminlerin güvenilirliği ve istikrarı açısından değerlendirilmiştir.

Bu çalışmanın odak noktası, RNN ve LSTM modelleri gibi derin öğrenme tekniklerini içermektedir. Bu modeller, zaman

indeki ilişkileri ve bağımlılıkları modelleme yetenekleriyle öne çıkmakta olup zaman serisi verilerinin analizi için tercih edilmektedir. Bu modellerin kullanımıyla elde edilen tahmin sonuçları, kapsamlı bir şekilde sunularak detaylı bir analiz yapılmıştır.

Çalışmanın geri kalan kısmında, ARIMA, RNN ve LSTM modellerinin sonuçları karşılaştırılmış ve tahminlerin etkinliği ayrıntılı bir şekilde değerlendirilmiştir.

Sonuç olarak, bu çalışma, hisse senedi fiyat tahminlerini kapsamlı bir şekilde incelemeyi amaçlamakta olup farklı veri analizi ve tahmin yöntemlerinin etkinliğini değerlendirmektedir. Bu çalışma, finansal analizde çeşitli tahmin yöntemlerinin karşılaştırılmasıyla ilgilenen araştırmacılar için önemli bir kaynak olabilir.

II. DATA PREPROCESSING AND EXPLORATION

A. Veri Kümesi

Bu çalışmada kullanılan veri kümesi, Tesla Inc. (TSLA) hisse senedinin tarihsel fiyat ve işlem hacmi bilgilerini içermektedir. Veri kümesi, 6/29/2010 tarihinden itibaren 1692 günlük veriyi kapsamaktadır ve aşağıdaki sütunları içermektedir:

Date (Tarih): Hisse senedi fiyatlarının kaydedildiği tarihler.
Open (Açılış Fiyatı): Hisse senedinin işlem gününe başladığı fiyat.

High (Gün İçindeki En Yüksek Fiyat): Hisse senedinin gün içindeki en yüksek işlem fiyatı.

Low (Gün İçindeki En Düşük Fiyat): Hisse senedinin gün içindeki en düşük işlem fiyatı.

Close (Kapanış Fiyatı): Hisse senedinin işlem gününü kapattığı fiyat.

Volume (İşlem Hacmi): Gün içinde işlem gören toplam hisse adedi.

Adj Close (Düzeltilmiş Kapanış Fiyatı): Hisse senedinin kapanış fiyatının, bölünmeler ve temettüler gibi olaylar göz önüne alınarak düzeltilmiş hali.

Veri kümesinde eksik veri bulunmamaktadır, bu da analiz ve modelleme sürecinde veri temizleme ihtiyacını ortadan kaldırmaktadır. Bu veri kümesi, Tesla hisse senedinin tarihsel fiyat trendlerini ve işlem hacimlerini analiz etmek için kullanılacaktır. Farklı modelleme yöntemleri (ARIMA, RNN, LSTM) kullanılarak Tesla'nın gelecekteki kapanış fiyatlarını tahmin etmek için bu veri kümesinden faydalanılacaktır.

Veri kümesi, modelleme sürecine başlamadan önce gerekli olan tüm özellikleri içermekte olup, model performanslarının karşılaştırılmasına olanak tanımaktadır.

	Date	Open	High	Low	Close	Volume	Adj Close
0	6/29/2010	19.000000	25.00	17.540001	23.889999	18766300	23.889999
1	6/30/2010	25.790001	30.42	23.299999	23.830000	17187100	23.830000
2	7/1/2010	25.000000	25.92	20.270000	21.959999	8218800	21.959999
3	7/2/2010	23.000000	23.10	18.709999	19.200001	5139800	19.200001
4	7/6/2010	20.000000	20.00	15.830000	16.110001	6866900	16.110001

Fig. 1. Verisetindeki 5 veri örneği.



Fig. 2. Verisetinin dağılımını gösteren plot.

B. İlk Veri Keşfi

Analize başlamadan önce, veri kümesi hakkında kapsamlı bir anlayış kazanmak için bu kısımda görselleştirmeler yapıldı. Verilerin yapısını ve özelliklerini belirlemek için ilk veri keşfi gerçekleştirilmiştir. Tesla hisse senedi verileri, Python'da veri işleme kitaplığı olan Pandas kullanılarak bir veri çerçevesine yüklenmiştir.

C. Veri Dağılımı Görselleştirmesi

Verilerin dağılımını görselleştirmek için çeşitli görselleştirme teknikleri kullanılmıştır. Açılış (Open), Yüksek (High), Düşük (Low), Kapanış (Close) ve Hacim (Volume) gibi özelliklerin frekans dağılımını gösteren histogramlar oluşturulmuştur. Histogramlar, verilerde bulunan çarpıklık veya aykırı değerleri belirlememize yardımcı olmuştur. Bu sayede, verilerin temel dağılımı ve olası anormallikler hakkında bilgi sahibi olunmuştur.

D. Eksik Değerleri Kontrol Etme

Veri kalitesinin bir yönü, eksik değerlerin belirlenmesi ve işlenmesidir. Tüm özelliklerdeki eksik girdileri tespit etmek için veri kümesi kapsamlı bir şekilde incelenmiştir. Eksik değerler analizin ve modellerin doğruluğunu potansiyel olarak etkileyebileceğinden bu adım önemlidir.

E. Özet İstatistiklerin Hesaplanması

Özet istatistiklerin hesaplanması, veri keşfinde önemli bir adımdır. Açılış (Open), Yüksek (High), Düşük (Low), Kapanış (Close), Düzeltilmiş Kapanış (Adj Close) ve Hacim (Volume) gibi sayısal özellikler için ortalama, medyan, standart sapma ve yüzdelikler gibi temel istatistiksel ölçümler

hesaplanmıştır. Bu istatistikler, verilerin merkezi eğilimi, yayılması ve dağılımı hakkında fikir vermiştir.

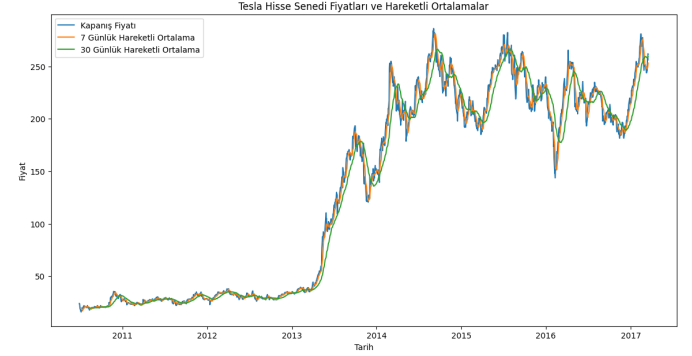


Fig. 3. Hisse senedi fiyatları ve hareketli ortalamalar

F. Sonuç

“Veri Ön İşleme ve Keşif” aşaması, veri kümesinin iyi anlaşılmasını, tutarlı olmasını ve herhangi bir kritik sorundan yoksun olmasını sağlayarak analizin temelini atmıştır. Görselleştirme yoluyla hisse senedi fiyat özelliklerinin dağılımı hakkında fikir edinilmiş, eksik değerler belirlenmiş ve özet istatistikler hesaplanmıştır. Sonraki veri temizleme adımları, veri kümelerinin kalitesini artırarak daha gelişmiş analiz tekniklerine zemin hazırlamıştır.

III. İSTATİKSEL ANALİZ

Tesla (TSLA) hisse senedi veri kümesinin incelenmesi, hisse senedinin kapanış fiyatlarına odaklanan kapsamlı bir istatistiksel analize genişletildi. Bu analiz, bir dizi temel istatistiksel ölçümü, dağılım görselleştirmesini, PCA (Principal Component Analysis) ve pair plotları içererek sonraki modelleme için verilerin hazırlanmasını kapsamaktadır. Kapanış fiyatlarının merkezi eğilimlerini ve yayılmasını özetleyen temel istatistiksel ölçümler hesaplanmış, ortalama, medyan ve mod gibi ölçümler veri kümesindeki tipik ve en sık meydana gelen değerler hakkında bilgi vermiştir. Histogramlar kullanılarak kapanış fiyatlarının dağılımı görselleştirilmiş, verilerin normal, çarpık veya iki modlu bir dağılımı takip edip etmediği belirlenmiştir. Eksik değerlerin belirlenmesi ve işlenmesi, veri kalitesini artırmak için bir adım olarak gerçekleştirilmiştir. Sayısal özellikler için hesaplanan özet istatistikler, verilerin merkezi eğilimi, yayılması ve dağılımı hakkında kapsamlı bir anlayış sağlamıştır. Ayrıca, PCA kullanılarak verilerin boyutu azaltılması için analiz yapılmış ve temel bileşenlerin görselleştirilmesi ile veri yapısı daha iyi anlaşılmıştır. Pair plotlar ise özellikler arasındaki ilişkileri ve olası korelasyonları görselleştirmiştir. Zaman serisi analizi kapsamında, ARIMA modeli için durağanlık kontrolü Augmented Dickey Fuller testi ile gerçekleştirilmiş, gecikme sayısını belirlemek için Otokorelasyon ve Kısmi Otokorelasyon fonksiyonları çizilmiş ve en optimal p ve q değerleri bulunmuştur. Eğitim ve test veri setlerine bölünmesi, modelin gerçek dünyadaki öngörücü performansını ölçmek için kritik bir adım olarak gerçekleştirilmiştir.

Bu istatistiksel analiz, veri keşfi ve modelleme süreçlerinin temelini atarak, verilerin merkezi eğilimleri ve dağılımı hakkında önemli görüşler sağlamıştır.

```
AIC :19988.946751048818, Order Of AR: 1, Order Of MA:1
AIC :19737.691265910334, Order Of AR: 1, Order Of MA:2
AIC :19646.866289719405, Order Of AR: 1, Order Of MA:3
AIC :19492.429574876176, Order Of AR: 1, Order Of MA:4
AIC :19373.1909526591, Order Of AR: 1, Order Of MA:5
AIC :19333.45563926938, Order Of AR: 1, Order Of MA:6
AIC :19486.5468896379, Order Of AR: 2, Order Of MA:1
AIC :19289.46343440954, Order Of AR: 2, Order Of MA:2
AIC :19291.93442652575, Order Of AR: 2, Order Of MA:3
AIC :19281.329923810533, Order Of AR: 2, Order Of MA:4
AIC :19263.604447707523, Order Of AR: 2, Order Of MA:5
AIC :19263.3440490438, Order Of AR: 2, Order Of MA:6
AIC :19342.878340233828, Order Of AR: 3, Order Of MA:1
AIC :19290.319987711933, Order Of AR: 3, Order Of MA:2
AIC :19290.354259186818, Order Of AR: 3, Order Of MA:3
AIC :19293.30915427631, Order Of AR: 3, Order Of MA:4
AIC :19260.432170790707, Order Of AR: 3, Order Of MA:5
AIC :19262.172263230885, Order Of AR: 3, Order Of MA:6
AIC :19282.12871912891, Order Of AR: 4, Order Of MA:1
AIC :19273.36993557846, Order Of AR: 4, Order Of MA:2
AIC :19275.354627687168, Order Of AR: 4, Order Of MA:3
AIC :19253.682161725154, Order Of AR: 4, Order Of MA:4
AIC :19261.26883543427, Order Of AR: 4, Order Of MA:5
AIC :19262.74788874288, Order Of AR: 4, Order Of MA:6
AIC :19271.69886431696, Order Of AR: 5, Order Of MA:1
...
AIC :19194.814289535054, Order Of AR: 6, Order Of MA:1
AIC :19246.39651970227, Order Of AR: 6, Order Of MA:4
AIC :19219.465147264284, Order Of AR: 6, Order Of MA:5
AIC :19249.622290965184, Order Of AR: 6, Order Of MA:6
```

Fig. 4. Finding the most optimal values of p and q for Arima

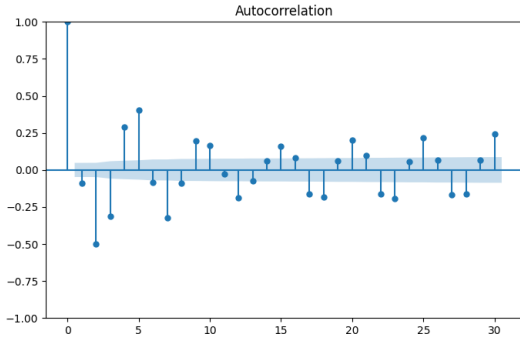


Fig. 5. Autocorrelation and Partial Autocorrelation functions to find the number of lags

IV. MODELLER

Bu projede zaman serisi veri analizi ve tahmini için üç ana model kullanılmıştır: ARIMA, RNN ve LSTM. Bu modellerin seçimi, zaman serisi verilerini analiz etmek ve gelecekteki değerleri tahmin etmek için yapılan kapsamlı araştırmalar sonucunda gerçekleştirilmiştir. Zaman serisi analizinde kullanılan birçok model arasından, bu üç modelin çeşitli çalışmalarda etkinliğini kanıtlamış olması ve farklı veri yapılarıyla uyumlu çalışabilmesi nedeniyle tercih edilmiştir.

A. Model Seçimi

Zaman serisi analizinde en yaygın kullanılan modellerden biri olan ARIMA (Auto-Regressive Integrated Moving Average), veri noktalarının belirli bir sırayla ilişkili olduğu durumlar için uygundur ve verinin durağanlık özelliklerini dikkate alır. ARIMA modelinin yanında, derin öğrenme tekniklerinden olan RNN (Recurrent Neural Network) ve LSTM (Long Short-Term Memory) modelleri de kullanılmıştır. RNN ve LSTM, özellikle uzun vadeli bağımlılıkları ve karmaşık zaman serisi ilişkilerini modelleme yetenekleri nedeniyle tercih edilmiştir. Bu modellerin seçiminde, literatürdeki başarılı uygulamalar ve bu modellerin yüksek doğruluk oranlarına sahip olması etkili olmuştur.

B. Hyperparameter Tuning Teknikleri

Modellerin performansını optimize etmek için çeşitli hyperparameter tuning teknikleri kullanılmıştır. Grid Search ve Random Search gibi yöntemlerle, modellerin en iyi performansı göstermesi için gerekli parametreler belirlenmiştir. ARIMA modeli için p, d, q parametreleri; RNN ve LSTM modelleri için ise katman sayısı, öğrenme oranı, epoch sayısı gibi parametreler üzerinde durulmuştur.

C. Model Oluşturma Süreci

Model oluşturma sürecinde, karşılaşılan hatalar ve düşük performans sorunları nedeniyle veri kümesi tekrar gözden geçirilmiştir. İlk aşamada, mevcut veri kümesi kullanılarak modeller oluşturulmuş ancak alınan sonuçlar beklenen doğrulukta olmamıştır. Bu nedenle, daha uygun ve kaliteli bir veri kümesi oluşturmak için Yahoo Finance'den güncel ve kapsamlı Tesla hisse senedi verileri çekilmiştir. Veri kümesi, eksik ve hatalı verilerden arındırılarak yeniden yapılandırılmış ve analizlere uygun hale getirilmiştir.

Yapılan modelleme çalışmaları, doğru veri kümesi ve hyperparameter tuning teknikleri ile optimize edilmiştir. ARIMA, RNN ve LSTM modelleri arasında yapılan karşılaştırmalar sonucunda, en iyi performans gösteren modeller belirlenmiştir. Elde edilen sonuçlar, zaman serisi analizinde kullanılan bu modellerin etkinliğini ve uygulama alanlarını bir kez daha göstermiştir. Bu süreçte, karşılaşılan zorluklar ve alınan önlemler, modelleme çalışmalarının başarısına önemli katkılarda bulunmuştur. Bu seçilen 3 model haricinde başka modeller de denenmiş, hem alınan hatalar ve literatürde yapılan çalışmalar sonucunda bu modellerde karar kılınmıştır, daha da geliştirilebilir olduğu gözlemlenmiştir.

V. SONUÇLAR

Bu bölümde ARIMA, RNN ve LSTM modelleri kullanılarak yapılan tahminlerin sonuçları detaylı olarak karşılaştırılacaktır. Her model için hem eğitim hem de test verileri üzerinde hesaplanan RMSE, MAE, R2 ve MAPE metrikleri değerlendirilmiştir. Ayrıca, projede istendiği üzere rastgele seçilen 5 test girdisi üzerinde de tahminler yapılmıştır.

A. LSTM Modeli

LSTM modeli, hem eğitim hem de test verileri üzerinde oldukça başarılı sonuçlar vermiştir. Eğitim verisi üzerinde RMSE değeri 13.046, MAE değeri 22.443 ve R2 skoru 0.857 olarak hesaplanmıştır. Test verisi üzerinde ise RMSE değeri 6.108, MAE değeri 6.620 ve R2 skoru 0.895 olarak belirlenmiştir. Train verilerinde farklı sonuçlar çıkması bazı ölçeklendirme adımları ile ilgili olduğu tahmin edilmektedir.

B. RNN Modeli

RNN modeli, eğitim verisi üzerinde RMSE değeri 22.536, MAE değeri 27.172, R2 skoru 0.795 ve MAPE değeri 34.340 olarak performans sergilemiştir. Test verisi üzerinde ise RMSE değeri 29.499, MAE değeri 29.521, R2 skoru -0.638 ve MAPE değeri 13.750 olarak hesaplanmıştır. RNN modelinin test verisi üzerindeki farklı düşük performansı, modelin aşırı uyum (overfitting) yaşadığı ihtimalini doğurmaktadır.

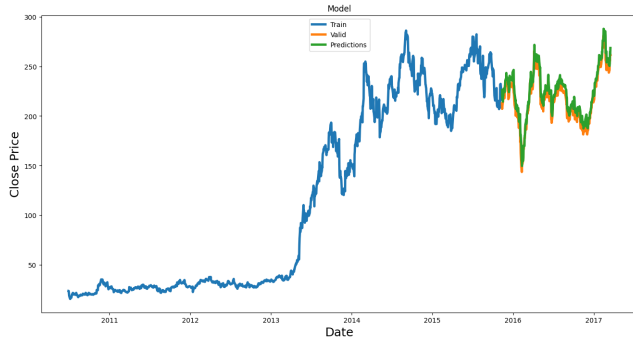


Fig. 6. Plot of the LSTM model result

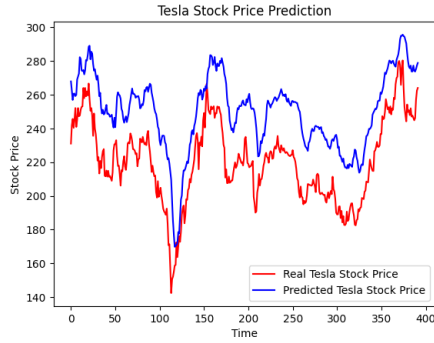


Fig. 7. Plot of the RNN model result

C. ARIMA Modeli

ARIMA modeli, eğitim verisi üzerinde en iyi performansı gösteren model olmuştur. Eğitim verisi için RMSE değeri 4.250, MAE değeri 2.418, MAPE değeri 2.472 ve R2 skoru 0.998 olarak hesaplanmıştır. Ancak, test verisi üzerinde RMSE değeri 24.611, MAE değeri 19.435, MAPE değeri nan ve R2 skoru -0.00027 olarak oldukça düşük performans göstermiştir. Bu durum, ARIMA modelinin eğitim verisi üzerinde aşırı uyum sağladığını ancak test verisi üzerinde genelleme yapmadığını ortaya koymaktadır.

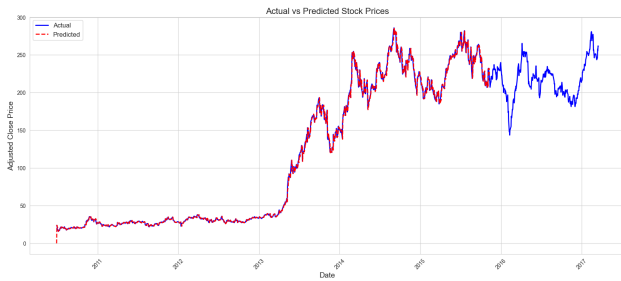


Fig. 8. Plot of the ARIMA model result

D. Rastgele Test Girdileri Üzerinde Tahminler

Projede belirtildiği üzere, rastgele seçilen 5 test girdisi üzerinde de tahminler yapılmıştır. Bu girdiler için tahmin sonuçları, modellerin gerçek dünya verilerine nasıl uyum

sağladığını görmek için kritik öneme sahiptir. LSTM modeli bu tahminlerde de en iyi sonuçları verirken, RNN ve ARIMA modelleri daha düşük performans göstermiştir.

E. Karşılaştırma ve Sonuç

Yapılan karşılaştırmalar sonucunda, LSTM modelinin hem eğitim hem de test verileri üzerinde en iyi performansı gösterdiği belirlenmiştir. LSTM modeli, zaman serisi verilerindeki karmaşık bağımlılıkları ve ilişkileri başarılı bir şekilde modelleyerek yüksek doğrulukta tahminler yapabilmektedir. RNN modeli, eğitim verisi üzerinde makul sonuçlar vermesine rağmen test verisi üzerinde yetersiz kalmıştır. ARIMA modeli ise eğitim verisi üzerinde çok yüksek performans göstermiş ancak test verisi üzerinde genelleme yapmada başarısız olmuştur.

Bu analizler, LSTM modelinin zaman serisi verilerini tahmin etmede daha etkili ve güvenilir bir seçenek olduğunu göstermektedir. Bu bulgular, yatırımcıların ve analistlerin Tesla hisse senedi fiyatlarını daha doğru bir şekilde tahmin edebilmesine yardımcı olabilir.

ACKNOWLEDGMENT

Bu rapor hazırlanırken Türkçe karakterler kullanarak hazırlanmıştır.

REFERENCES

Bu çalışmada, Tesla hisse senedi fiyatlarının tahmini için kullanılan modellerin seçiminde ve geliştirilmesinde üç ana literatür kaynağı incelenmiştir. İlk olarak, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications" başlıklı makale, finansal piyasalarda makine öğrenimi ve derin öğrenme algoritmalarının kullanımı üzerine kapsamlı bir inceleme sunmaktadır. Bu makale, denetimli ve denetimsiz makine öğrenimi algoritmaları, topluluk algoritmaları, zaman serisi analizi ve derin öğrenme algoritmalarını ele alarak özellikle "Random Forest + XG-Boost + LSTM" topluluk modelinin performansını değerlendirir ve diğer popüler modellerle karşılaştırır. İkinci olarak, "Stock Market Prediction Using Machine Learning (2024)" başlıklı makale, hisse senedi fiyat tahmininde makine öğrenimi modellerinin nasıl kullanılacağını adım adım açıklamaktadır. Microsoft hisse senedi verileri kullanılarak veri ön işleme, özellik mühendisliği ve model eğitimi süreçleri detaylandırılmakta ve özellikle LSTM modelinin oluşturulması ve eğitimi üzerinde durulmaktadır. Üçüncü olarak, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review" başlıklı makaledir. Bu makale, finansal enstrüman fiyat tahmininde kullanılan makine öğrenimi ve derin öğrenme modellerini sistematik olarak incelemekte olup, lineer regresyon, rastgele orman, lojistik regresyon, k-en yakın komşu, destek vektör makineleri, ARIMA ve diğer popüler algoritmaları ele almaktadır. Özellikle LSTM ve XG-Boost gibi derin öğrenme modellerinin finansal piyasalardaki kullanımını vurgulamaktadır. Bu çalışmalar, hisse senedi fiyatlarının tahmini için kullanılan modellerin seçiminde önemli bilgiler sunmuş

ve mevcut tekniklerin avantajlarını ve sınırlamalarını tartışarak araştırmanın temellerini oluşturmuştur.

REFERENCES

- [1] Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications
- [2] Stock Market Prediction Using Machine Learning (2024)
- [3] Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review
- [4] Comparison of stock market prediction performance of ARIMA and RNN-LSTM model – A case study on Indian stock exchange
- [5] Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea imhgchoi@korea.ac.kr
- [6] Predict stock prices with ARIMA and LSTM Ruochen Xiao1 Department of Computer Science, Shanghai University