OSTİM TEKNİK UNIVERSITY ENGINEERING FACULTY
GRADUATION PROJECT REPORT

STOCK PRICE PREDICTION APP

MEGUMİ SELİN TERZİ
200201033
ELİF SENA DALDAL
200201015
İREMNAZ ULU
190201024

Department: Computer Engineering (English)
Project Advisor: Asst. Prof. Ali Metin BALCI
Course Code and Name: CENG412 Graduation Project
Project Starting Date: 2023-2024 Spring Semester
Report Presentation: 2023-2024 Spring Semester

This report has been accepted by the jury whose names are written below on …./…./ 20…

Title	Name Surname	Sign
………….	………………………………………………….	………………………..
………….	………………………………………………….	………………………..
………….	………………………………………………….	………………………..

ABSTRACT

This study presents the Stock Price Prediction App, designed to forecast stock prices using both the Prophet and Long Short-Term Memory (LSTM) models. The application, built on the Streamlit platform, leverages historical data from Yahoo Finance and incorporates technical indicators such as EMA, MACD, and RSI to enhance prediction accuracy. Users can interactively select stock tickers to view and predict prices, with the app displaying both historical trends and future predictions.

The performance of the forecasting models is evaluated using metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared. This project aims to combine traditional statistical methods and advanced deep learning techniques to improve the predictive capabilities of financial time series analysis, providing valuable insights for investors and market analysts.

List of Figures

List of Tables

Tables

List of Symbols and Abbrevations

Symbol / Abbreviation	Explanation
LSTM	Long Short-Term Memory
RMSE	Root Mean Squared Error
MACD	Moving Average Convergence Divergence
EMA	Exponential Moving Average
SMA	Simple Moving Average
RSI	Relative Strength Index
MAE	Mean Absolute Error
R^2	R-squared(Coefficient of Determination)
MAPE	Mean Absolute Percentage Error

Table of Contents

1. INTRODUCTION

The financial market is an ever-evolving and dynamic field where investment decisions are crucially dependent on accurate predictions of stock prices. The ability to predict stock prices not only serves the financial analysts and investors but also enhances the strategic planning of companies and provides insight into market trends. This project aims to develop the "Stock Price Prediction App," a robust tool designed to forecast stock prices using advanced statistical methods and machine learning techniques.

The primary objective of this project is to integrate and compare two distinct forecasting models: the Prophet model and the Long Short-Term Memory (LSTM) network. Each model offers unique advantages in handling time series data, and their performance in predicting stock prices is of particular interest. This report will detail the development of an interactive application using Streamlit, which allows users to select stocks, visualize historical data, and view predictions generated by the implemented models.

The scope of this study is to provide a comprehensive analysis of the models' accuracy through detailed backtesting with historical data from Yahoo Finance. By evaluating metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared values, the project aims to identify which model most effectively predicts future stock prices under varying market conditions. Additionally, this study aims to increase user interaction by providing a user-friendly platform for exploring and engaging with the predictive outcomes, thereby simplifying the intricacies of stock market forecasting for both experienced traders and beginners.

2. LITERATURE REVIEW

2.1. Role of Machine Learning in Enhancing Predictive Accuracy

The advent of machine learning has introduced more sophisticated analytical techniques to stock price prediction. Unlike traditional statistical methods, machine learning can model complex non-linear interactions between variables without requiring explicit pre-defined model structures. For instance, neural networks, particularly Long Short-Term Memory (LSTM) networks, have gained prominence for their ability to remember long-term dependencies within the data — a critical feature given the sequential nature of stock prices (Hochreiter & Schmidhuber, 1997). LSTM networks have been shown to outperform conventional models in capturing the dynamics of volatile financial markets, as they can learn from the long-term temporal sequences without the risk of gradient vanishing problems typical in standard recurrent networks (Fischer & Krauss, 2018). Furthermore, machine learning models can integrate a wide range of input features, including technical indicators such as Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and others, which are often used to capture market sentiment and trader behavior, thus providing a more holistic view of market dynamics.

2.2. Enhancing Machine Learning Models with Technical Indicators

Technical analysis remains a cornerstone of financial forecasting, employing various statistical measures to predict future movements of stock prices based on historical trading data. When combined with machine learning, these indicators serve as powerful features that enhance the predictive capabilities of models. Research has demonstrated that models integrating price data with technical indicators can yield more accurate and robust predictions (Chong, Han, & Park, 2017). This synthesis allows models to not only capture underlying trends and patterns but also to react to market conditions in real-time, making them highly advantageous for automated trading systems.

2.3. Comparative Studies and Performance Metrics

Comparative studies between traditional and machine learning-based forecasting models generally favor machine learning approaches, particularly when combined with large datasets and enriched feature sets. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are commonly used to evaluate the accuracy of predictions, with recent studies incorporating measures like the Sharpe Ratio to assess risk-adjusted returns (Rasekhschaffe & Jones, 2019).

3. DATA COLLECTION

For the experimental study, we downloaded live datasets for stocks such as Amazon, Apple, and Eregli from the Yahoo Finance website (https://finance.yahoo.com/). The table below summarizes the minimum and maximum values for opening (Open), lowest (Low), highest (High), and closing (Close) prices of the stocks over their entire history.

Table 3.1 Amazon

Attribute Name	Min	Max
Open	181.649994	184.339996
Low	180.080002	183.279999
High	182.440002	186.669998
Close	180.750000	183.539993

Table 3.2 Apple

Attribute Name	Min	Max
Open	188.820007	192.270004
Low	186.630005	190.919998
High	190.580002	192.820007
Close	186.880005	192.350006

Table 3.3 Eregli

Attribute Name	Min	Max
Open	46.860001	48.480000
Low	46.860001	48.400002
High	48.560001	49.500000
Close	47.939999	49.419998

Sample Input

Table 3.4 Sample Input

Date	Trade Open	Trade Low	Trade High	Trade Close
20-May-2024	184.339996	183.279999	186.669998	183.539993
21-May-2024	182.300003	180.750000	183.259995	183.149994
22-May-2024	183.880005	181.970001	185.220001	183.130005
23-May-2024	183.660004	180.080002	184.759995	181.050003
24-May-2024	181.649994	180.300003	182.440002	180.750000

4. METHODOLOGIES

4.1. Integration of Technical Indicators with Machine Learning Models

Technical analysis remains a cornerstone of financial forecasting, employing various statistical measures to predict future movements of stock prices based on historical trading data. When combined with machine learning, these indicators serve as powerful features that enhance the predictive capabilities of models. Research has demonstrated that models integrating price data with technical indicators can yield more accurate and robust predictions (Chong, Han, & Park, 2017). This synthesis allows models to not only capture underlying trends and patterns but also to react to market conditions in real-time, making them highly advantageous for automated trading systems.

4.1.1. Exponential Moving Average (EMA)

The Exponential Moving Average (EMA) is a widely used technical analysis tool in stock and other financial markets. Similar to the Simple Moving Average (SMA), the EMA gives more weight to recent prices, making it more responsive to new information (Kuang, P., Li, W., & Wei, Y. ,2014). This increased sensitivity helps in better capturing short-term trends. In this project, we observe two EMA lines:

EMA Fast: Typically calculated over a shorter period (e.g., 12 days). It reacts more quickly to price changes.
EMA Slow: Typically calculated over a longer period (e.g., 26 days). It reacts more slowly to price changes.

The calculation of the EMA begins with the determination of the SMA for a specific period. Then, the EMA is calculated using the following formula:

$$EMA_{Today} = \frac{Value_{Today} - EMA_{Yesterday}}{\text{Days} + 1} \times Smoothing + EMA_{Yesterday}$$

Where:

- EMA_{Today} is the EMA value for the current day.
- $Value_{Today}$ is the closing price of the current day.
- $EMA_{Yesterday}$ is the EMA value for the previous day.
- *Smoothing* is a constant, usually set to 2.
- *Days* is the number of days over which EMA is calculated.

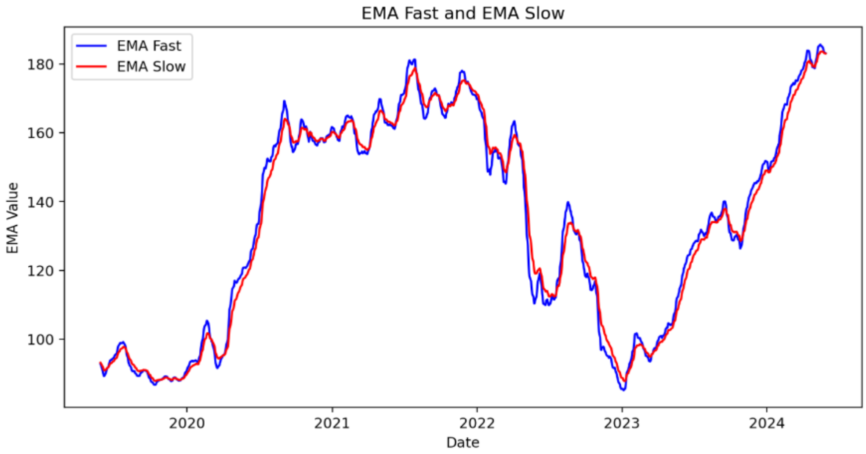

Fig 4.1.1.1

The crossover points of the fast and slow EMAs for Amazon stock can be interpreted as potential buy and sell signals. When the fast EMA crosses below the slow EMA, it is considered a sell signal, whereas when the fast EMA crosses above the slow EMA, it is considered a buy signal. The EMA is particularly useful for identifying the direction and strength of trends. By providing more recent price information, it helps traders make timely decisions.

4.1.2. Moving Average Convergence Divergence (MACD) Indicator

The Moving Average Convergence Divergence (MACD) is a popular momentum and trend-following indicator used in technical analysis. It helps traders identify the direction and strength of a trend, as well as potential buy and sell signals. The MACD is calculated by subtracting the 26-period EMA (EMA Slow) from the 12-period EMA (EMA Fast). The result is the MACD line. Additionally, a 9-period EMA of the MACD line, known as the signal line, is plotted to generate trading signals.

The MACD consists of three main components:

MACD Line: The difference between the 12-period EMA and the 26-period EMA.
Signal Line: A 9-period EMA of the MACD line.
Histogram: The difference between the MACD line and the signal line, often plotted as a bar chart.

The formula for calculating the MACD is as follows:

$$MACD = EMA_{12} - EMA_{26}$$

Where:

- EMA_{12} is the 12-period Exponential Moving Average.

- EMA_{26} is the 26-period Exponential Moving Average.

The Signal Line is calculated as:

$$Signal\ Line = EMA_9(MACD)$$

Where:

- EMA_9 is the 9-period Exponential Moving Average of the MACD line.

The Histogram is calculated as:

$$Histogram = \text{MACD} - \text{Signal Line}$$

Usage in Technical Analysis
Bullish Signal: When the MACD line crosses above the signal line, it indicates a potential buy signal.
Bearish Signal: When the MACD line crosses below the signal line, it indicates a potential sell signal.
The MACD is a versatile indicator that can be used to identify trend direction, momentum, and potential reversal points. Its ability to provide timely buy and sell signals makes it a valuable tool for traders and analysts.

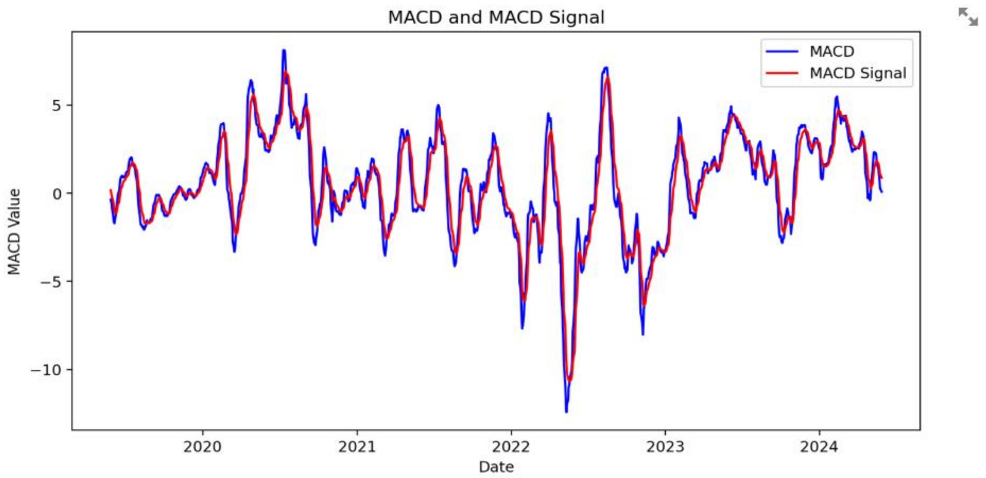

Fig 4.1.2.1

Figure 4.1.2.1 shows the MACD chart, showcasing the interaction between the MACD line (blue) and the signal line (red) over the period from 2020 to 2024. The chart aids in identifying potential buy and sell signals based on the crossover points of these lines. In mid-2020, the MACD line crosses above the signal line, indicating a potential buy signal. Early 2021 sees another crossover where the MACD line moves above the signal line, suggesting a bullish trend. Similarly, early 2023 features a crossover indicating a buy signal as the MACD line rises above the signal line. Conversely, in late 2020, the MACD line crosses below the signal line, indicating a potential sell signal. Mid-2021 witnesses another crossover where the MACD line moves below the signal line, suggesting a bearish trend. Finally, in late 2022, a bearish crossover occurs as the MACD line falls below the signal line.

4.1.3. Relative Strength Index (RSI)

The Relative Strength Index (RSI) is an indicator that predicts the direction of short-term and medium-term trends using previous periods' closing prices. Created by J. Welles Wilder in 1978, the RSI shows signals of whether a stock is overbought or oversold. It measures the anomalies of price movements. Typically, a fourteen-day period is used, and when this period falls below 30, it indicates that the investor should buy, and when it rises above 70, the investor should sell the asset.

The formula for the RSI is as follows:

$$RSI = 100 - \frac{100}{1 + \frac{\text{Average Gain}}{Average\ Loss}}$$

Where:

- *Average Gain* is the average of all gains over the selected period.
- *Average Loss* is the average of all losses over the selected period.

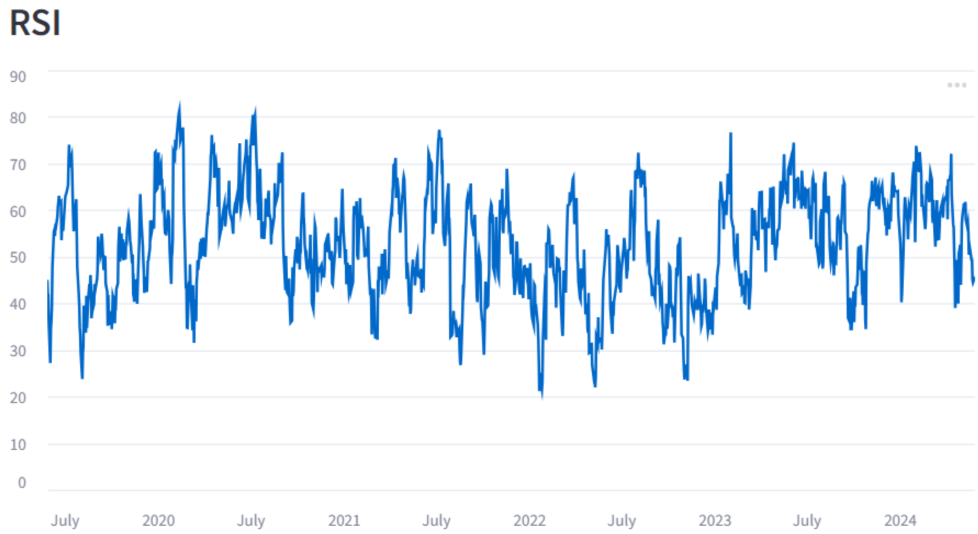

Fig. 4.1.3.1

As shown in the figure, the RSI index for Amazon stock fluctuates between overbought and oversold conditions, providing valuable insights into potential trading opportunities. In mid-2020, the RSI exceeded 70, indicating overbought conditions and suggesting that the stock might be overvalued, leading to a potential price correction. Early 2020 saw the RSI drop below 30, signaling oversold conditions and suggesting that the stock was undervalued, presenting a buying opportunity. Late 2020 featured another rise in the RSI above 70, indicating overbought conditions and implying possible overvaluation and a likely price correction. In mid-2022, the RSI fell below 30 again, indicating oversold conditions and suggesting that the stock was potentially undervalued, presenting another buying opportunity. Late 2022 saw another drop below 30 in the RSI, indicating oversold conditions and signaling another potential buying opportunity. Early 2023 had the RSI exceed 70, indicating overbought conditions and suggesting that the stock might be overvalued, with a price correction likely. In the recent period (2024), the RSI is trending downwards towards 30, suggesting that Amazon stock might be approaching oversold conditions, potentially indicating a buying opportunity if the RSI dips below 30.

4.2. System Architecture

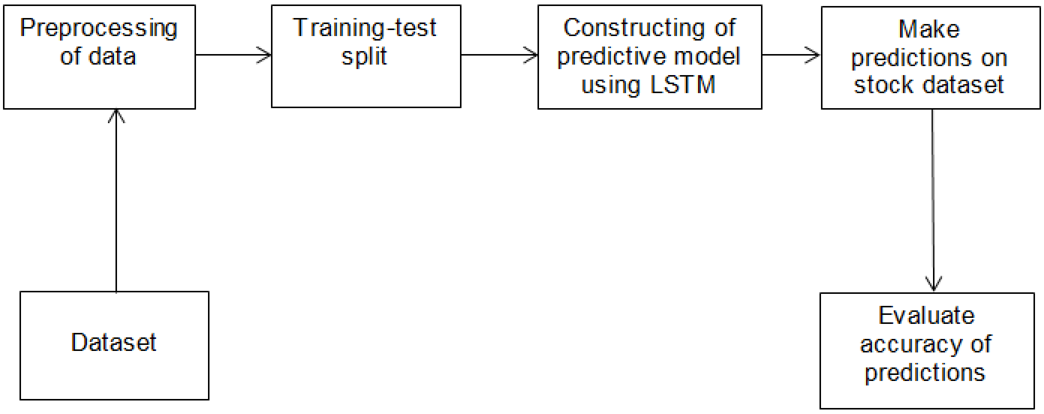

Fig 4.2.1 System Architecture

Data Selection:
The initial step involves the selection of data pertinent to the study and its subsequent division into training and testing sets. Specifically, 80% of the data is allocated for training purposes, while the remaining 20% is reserved for testing.

Pre-processing of Data:
During the pre-processing stage, relevant attributes necessary for the predictive algorithm are selected, and extraneous attributes are excluded. The chosen attributes encompass 'Close', 'Volume', 'trend_ema_fast', 'trend_ema_slow', 'trend_macd', 'trend_macd_signal', 'trend_adx', 'momentum_rsi', 'volatility_bbm', 'volatility_bbh', 'volatility_bbl', and 'volume_obv'. Normalization techniques are employed to scale the attribute values within a specific range, ensuring uniformity and improving model performance.

Prediction using LSTM:
In this phase, the Long Short-Term Memory (LSTM) algorithm is utilized to forecast stock prices. Initially, the training dataset is used to train the LSTM model. Post-training, the model is employed to predict stock prices, and these predicted values are subsequently compared against the actual values in the testing phase to evaluate the model's performance.

Evaluation:
During the evaluation phase, key performance metrics such as Accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are calculated. These metrics serve to assess and compare the efficacy of the predictive model.

5. EXPERIMENTAL RESULTS

5.1. AMAZON

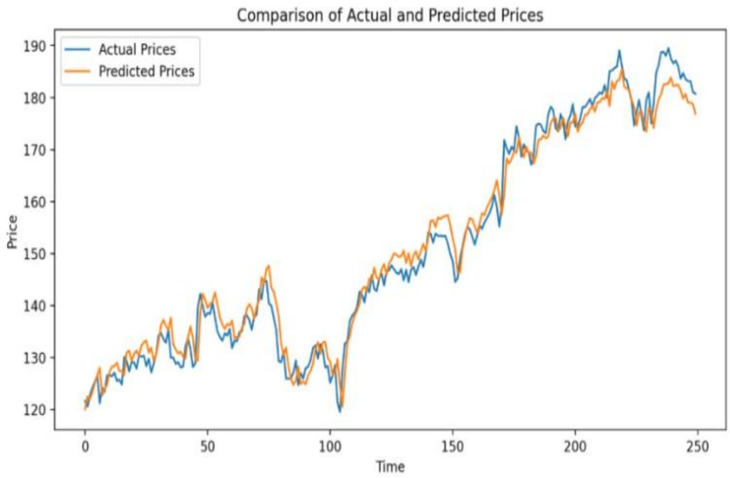

Fig 5.1.1 Amazon Graph

epochs	R^2	RMSE	MAE	MAPE
60	0.96	4.01	3.12	2.01%
80	0.98	3.26	2.60	1.76%
100	0.98	3.33	2.67	1.79%

Table 5.1.2 Amazon Epochs

In the results, as we have shown in Fig 5.1.1, the graph shows the Close Price value for the Amazon dataset. In this graph, the blue line indicates the actual prices, and the orange line represents the predicted values from the test data. Table 5.1.2 illustrates the R-squared, RMSE, MAE, and MAPE values for different numbers of epochs (iterations).

5.2. APPLE

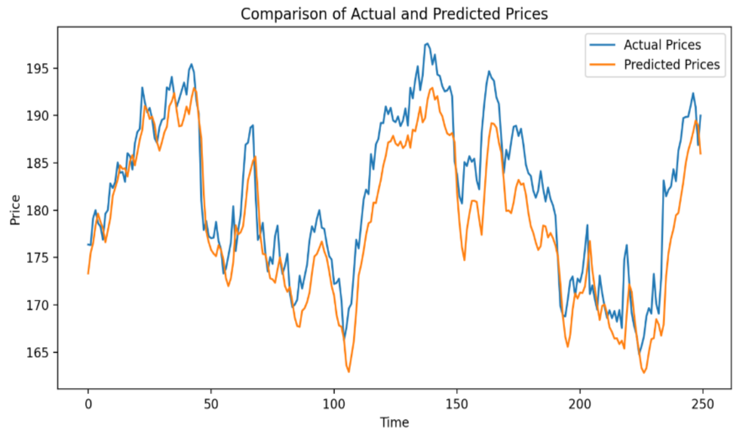

Fig 5.2.1 Apple Graph

epochs	R^2	RMSE	MAE	MAPE
60	0.19	7.51	6.73	3.68%
80	0.81	3.63	2.84	1.58%
100	0.76	4.12	3.40	1.86%

Table 5.2.2 Apple Epochs

Above graph 5.2.1 shows Close Price value for the Apple dataset and table 5.2.2 illustrates the R-squared, RMSE, MAE, and MAPE values for different numbers of epochs (iterations).

5.3. EREGLI

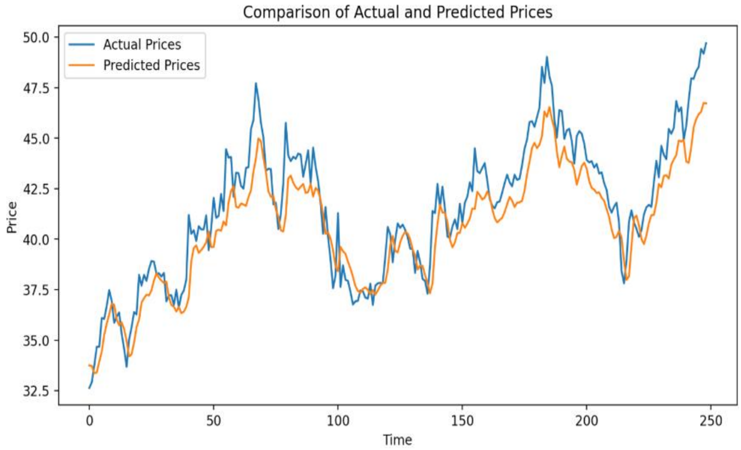

Fig 5.3.1 Eregli Graph

epochs	R^2	RMSE	MAE	MAPE
60	0.80	1.58	1.28	3.02%
80	0.80	1.57	1.28	3.01%
100	0.82	1.46	1.20	2.82%

Table 5.3.2 Eregli Epochs

Above graph 5.3.1 shows Close Price value for the Eregli dataset and table 5.3.2 illustrates the R-squared, RMSE, MAE, and MAPE values for different numbers of epochs (iterations).

The interactive application developed on the Streamlit platform underscores the practical relevance of this project. By enabling users to select stock tickers, visualize historical data, and view future predictions interactively, the app provides a valuable tool for investors and market analysts, simplifying the complexities of stock market forecasting (Box, Jenkins, & Reinsel, 2015).

6. CONCLUSION

Trading in the stock market is experiencing rapid growth, and investors and analysts are eager to discover effective methods and techniques to predict future stock market trends accurately. This study presents the "Stock Price Prediction App," which employs both Prophet and Long Short-Term Memory (LSTM) models to forecast stock prices. While the LSTM neural network model has shown exceptional performance in capturing trends and seasonality in long-term forecasts, Prophet excels in handling seasonality and holiday effects, providing clear and interpretable results.

Our comparative analysis of these models demonstrates that incorporating technical indicators such as Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), and Relative Strength Index (RSI) significantly enhances predictive accuracy. Despite the strengths of the LSTM model in capturing complex patterns in stock prices, it is not a perfect prediction method for all scenarios. Nevertheless, it has proven to be highly effective in long-term forecasting.

The interactive application developed on the Streamlit platform underscores the practical relevance of this project. By enabling users to select stock tickers, visualize historical data, and view future predictions interactively, the app provides a valuable tool for investors and market analysts, simplifying the complexities of stock market forecasting.

This study contributes to the ongoing discourse on financial forecasting and offers practical benefits for real-world applications. Future research could explore the integration of additional machine learning models, the refinement of technical indicators, and the expansion of datasets to include a wider range of stocks and market conditions. Such efforts would enhance prediction accuracy and broaden the applicability of forecasting methods, ultimately benefiting investors, analysts, and anyone interested in the stock market by providing a deeper understanding of its future trends.

7. REFERENCES

1. Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research*, 270(2), 654-669.

2. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

3. Rasekhschaffe, K. C., & Jones, R. (2019). Machine Learning for Stock Price Prediction: A Comparative Study. *Journal of Financial Data Science*, 1(4), 1-20.

4. Chong, E., Han, C., & Park, F. C. (2017). Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. *Expert Systems with Applications*, 83, 187-205.

5. Brownlee, J. (2018). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

6. Liu, G., & Wang, J. (2019). Applications of Machine Learning in Stock Price Prediction: A Survey. *IEEE Access*, 7, 123231-123247.

7. De Prado, M. L. (2018). *Advances in Financial Machine Learning*. Wiley.

8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

9. Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159-175.

10. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

11. Kuang, P., Li, W., & Wei, Y. (2014). The profitability of moving average strategies in the stock markets of emerging countries. International Review of Financial Analysis, 35, 29-39.

12. Wilder, J. W. (1978). New Concepts in Technical Trading Systems. Trend Research.