Apple Stock Price Prediction Using Time Series and Machine Learning Models

Mannam Vijitha Reddy[1], Madhumitha R[2], Keerthikasri R[3]
[1,2,3]*Department of Artificial intelligence and Data science Dhanalakshmi Srinivasan university (DSU) Trichy*

***Abstract*—Stock market prediction has become an important research area in financial analytics due to its potential to support investment decisions and reduce financial risk. This paper presents a comparative study of statistical, machine learning, and deep learning models for predicting the stock price of Apple Inc. (AAPL). Historical stock market data containing daily trading information such as opening price, closing price, highest price, lowest price, adjusted closing price, and trading volume is analyzed. Exploratory Data Analysis (EDA) is performed to identify trends, volatility, and distribution patterns within the dataset. Classical time-series models such as AutoRegressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) are implemented as baseline forecasting models. In addition, machine learning models including Random Forest and XGBoost are applied to capture nonlinear relationships between features. A deep learning model based on Long Short-Term Memory (LSTM) networks is also implemented to learn temporal dependencies in the time series. Model performance is evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Experimental results show that the LSTM model achieves the highest prediction accuracy. Finally, the best-performing model is used to forecast Apple stock prices for the next 30 trading days. The results demonstrate that deep learning approaches provide more reliable predictions for financial time-series data.**

I. INTRODUCTION

Stock markets generate large volumes of financial data that reflect the economic activities and performance of companies. Predicting stock prices has long been an important research area in financial analytics because accurate predictions help investors and organizations make better financial decisions. Traditional statistical models such as ARIMA and SARIMA have been widely used for time-series forecasting. However, these models often struggle to capture complex nonlinear patterns and long-term dependencies that exist in financial datasets. With the advancement of machine learning and deep learning techniques, researchers have begun exploring more sophisticated models for stock price prediction.

Machine learning algorithms such as Random Forest and XGBoost can capture nonlinear relationships between variables and improve prediction accuracy. Meanwhile, deep learning architectures such as Long Short-Term Memory (LSTM) networks are specifically designed for sequential data and can effectively learn temporal dependencies in time-series datasets. This research aims to analyze and compare multiple forecasting models to determine the most effective approach for predicting Apple stock prices. The study includes data preprocessing, exploratory data analysis, time-series modeling, machine learning methods, deep learning techniques, and evaluation of prediction performance.

By analyzing historical stock price records and finding the patterns and trends of the stock data, machine learning tech- niques have demonstrated great promise for stock prediction. Several machine learning models, such as Stacked LSTM, ARIMA, and linear regression have been proposed for stock prediction. In this paper, there will be a comparison of the performance of those mentioned models in which the most effective model will be found and discussed.

II. DATASET

The dataset used in this research consists of historical stock market data of Apple Inc. (AAPL). The data was obtained from publicly available financial sources such as Yahoo Finance and represents real-world financial time-series data. The dataset contains daily trading records of Apple stock over several years and captures the price movement and trading activity of the company in the stock market. The dataset includes key financial attributes such as the opening price, highest price, lowest price, closing price, adjusted closing price, and trading volume for each trading day. These attributes provide valuable information about stock market behavior and are commonly used in financial forecasting studies. Since stock price prediction is a time-series forecasting problem, the chronological order of the data is extremely important. Therefore, the Date column was converted into datetime format and used as the index of the dataset to ensure that the temporal sequence of stock prices is preserved throughout the analysis.

The dataset provides sufficient historical depth for identifying patterns, trends, and volatility in stock price movements. It enables the application of statistical models, machine learning algorithms, and deep learning techniques for forecasting future stock prices.

A. Dataset Features

Feature	Description
Date	Trading date of the stock
Open	Opening price of Apple stock on a given trading day
High	Highest price reached during the trading day
Low	Lowest price reached during the trading day
Close	Closing price of the stock
Adj Close	Adjusted closing price accounting for dividends and stock splits
Volume	Total number of shares traded during the day

These features provide essential information about daily market activity and are widely used in stock market prediction models.

B. Dataset Characteristics

The dataset used in this study consists of historical daily stock price data of Apple Inc. (AAPL), which represents financial time-series data collected over multiple years. The data includes important attributes such as opening price, highest price, lowest price, closing price, adjusted closing price, and trading volume for each trading day. Since the dataset is time-dependent, the observations are arranged in chronological order based on the trading date to preserve the temporal relationship between consecutive records. the dataset captures daily fluctuations in stock prices and trading activity, allowing researchers to analyze trends, volatility, and market behavior over time. Among all the attributes, the closing price is considered the primary variable for prediction, as it reflects the final market value of the stock at the end of each trading day. The dataset is continuous, numerical, and highly dynamic in nature, making it suitable for applying statistical models, machine learning algorithms, and deep learning techniques for forecasting future stock prices. Additionally, the presence of trading volume data helps in understanding market participation and investor activity, which can influence stock price movements.

C. Data Preparation

Before applying prediction models, the dataset underwent several preprocessing steps to ensure data quality and suitability for time-series analysis. Initially, the dataset was cleaned by checking for missing, duplicate, or inconsistent values and correcting them where necessary. The date column was converted into a standard datetime format and used as the index to maintain the chronological order of the observations. Since stock price prediction depends heavily on temporal relationships, the data was carefully sorted according to trading dates to preserve sequential patterns. After cleaning, normalization and scaling techniques were applied to numerical features such as open, high, low, close, and volume to ensure that all variables were on a similar scale and to improve model performance. Additional features such as moving averages and daily returns were also generated to capture trends and fluctuations in stock prices. Finally, the dataset was divided into training and testing subsets using a time-based split, typically allocating around 80% of the data for training and the remaining 20% for testing. This preparation process ensures that the dataset is well-structured and optimized for implementing statistical, machine learning, and deep learning models for stock price prediction.

D Importance of the Dataset

The dataset plays a crucial role in the development and evaluation of stock price prediction models, as it provides the historical information necessary to understand market behaviour and identify price trends. In this study, the dataset contains historical trading data of Apple Inc. (AAPL), which is one of the most valuable and widely traded companies in the global stock market. The availability of detailed attributes such as opening price, closing price, highest price, lowest price, and trading volume enables researchers to analyse various aspects of market activity and price fluctuations. Historical datasets are essential for training forecasting models because they allow algorithms to learn patterns, relationships, and temporal dependencies present in financial data. By studying these patterns, predictive models can estimate future stock prices with improved accuracy. Additionally, the dataset provides valuable insights into market volatility, investor behaviour, and trading trends, which are important factors in financial analysis. The reliability and richness of the dataset make it suitable for implementing statistical methods, machine learning techniques, and deep learning models for stock prediction. Therefore, the dataset serves as the foundation of the entire research process, enabling effective analysis, model training, and evaluation of forecasting performance. Overall, the dataset provides the essential historical information needed to analyse trends and accurately predict future stock price.

III. EXPLORATORY DATA ANALYSIS (EDA)

EDA helps understand price behavior, trends, and volatility before model training.

Key Insights

- Apple stock shows a strong long-term upward trend
- Periods of high volatility occur during major market events
- Trading volume correlates with sudden price fluctuations

Important Graphs for the Paper:

1. Stock Price Trend

Graph showing Closing Price vs Time

Fig. 1. Stock Price Trend of Apple Inc. (AAPL).

Purpose

Shows overall stock growth trend.

2. Moving Average Graph

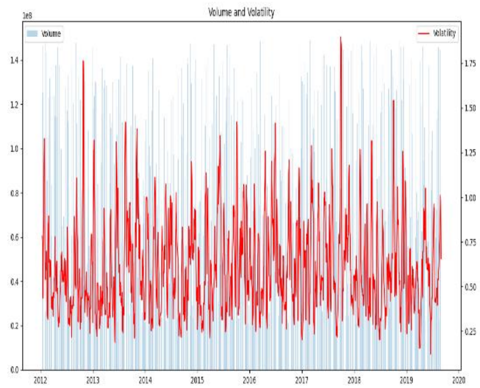
Two curves:

 7-day Moving Average

 30-day Moving Average

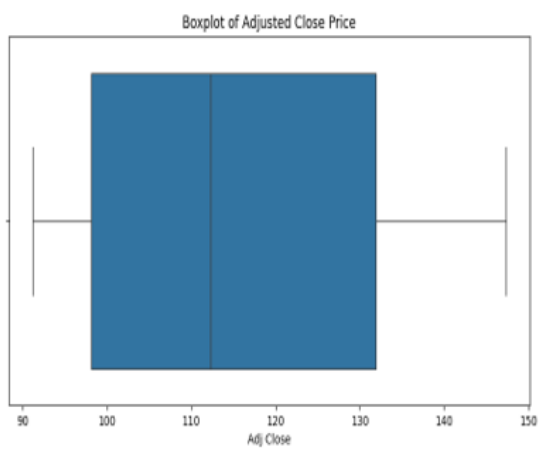
3. Volume and volatility

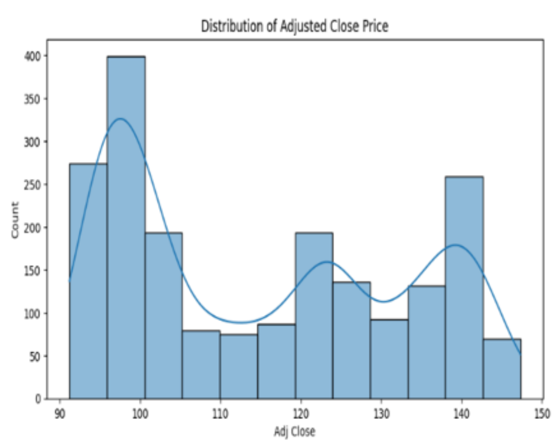
Trading volume and price volatility were analyzed together to understand market activity. High volume combined with high volatility often indicates strong market reactions, whereas low volume periods suggest stable or inactive trading phases. This analysis helps in understanding risk levels in stock movements.



4.Distribution Analysis

Histogram and boxplot analysis were used to examine the distribution of stock prices. These plots revealed skewness, outliers, and price spread over time. Understanding distribution helps assess whether assumptions of statistical models are valid.

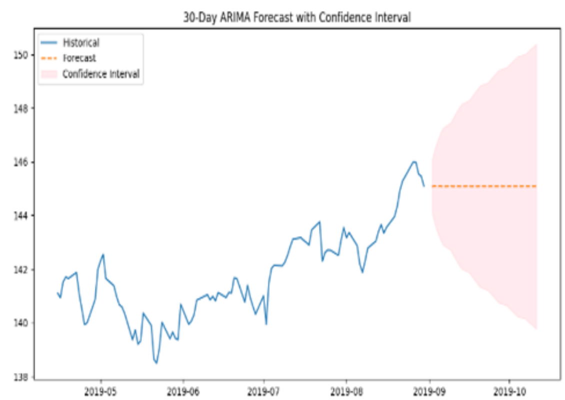




IV. FORECASTING MODELS

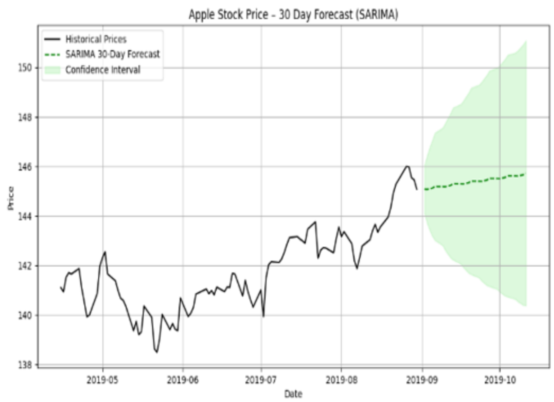
A. ARIMA Model

The AutoRegressive Integrated Moving Average (ARIMA) model is a statistical method commonly used for time-series forecasting. It combines three components: autoregression (AR), differencing (I), and moving average (MA). The autoregressive part uses past values of the variable to predict future values, while differencing helps make the data stationary by removing trends. The moving average component models the error terms in the time series. ARIMA works well for datasets with linear patterns and stable trends. However, financial datasets such as stock prices often contain nonlinear and highly volatile behavior, which can limit the predictive performance of ARIMA.



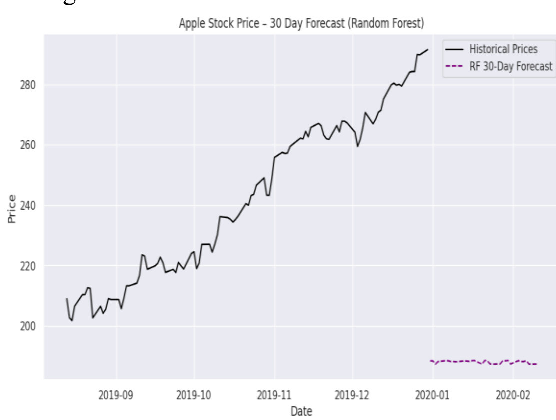
B. SARIMA Model

The Seasonal Auto Regressive Integrated Moving Average (SARIMA) model is an extension of the ARIMA model that includes seasonal components. SARIMA accounts for repeating seasonal patterns that may occur at regular intervals in time-series data. This makes it useful when the data exhibits periodic trends or cycles. In stock market forecasting, SARIMA can sometimes capture periodic movements in price trends. In this study, SARIMA improved prediction accuracy compared to ARIMA, but it still struggled to handle sudden market changes and unpredictable price fluctuations.



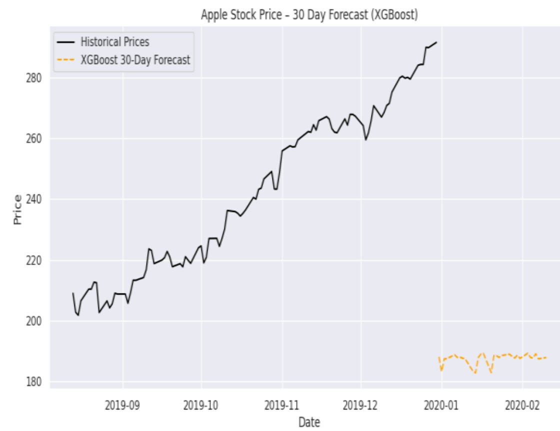
C. Random Forest Model

Random Forest is a machine learning algorithm based on ensemble learning. It constructs multiple decision trees during the training process and combines their predictions to produce a final output. Each tree is trained on a random subset of the dataset, which helps improve the model's generalization ability. Random Forest is particularly effective in capturing nonlinear relationships between input features and the target variable.



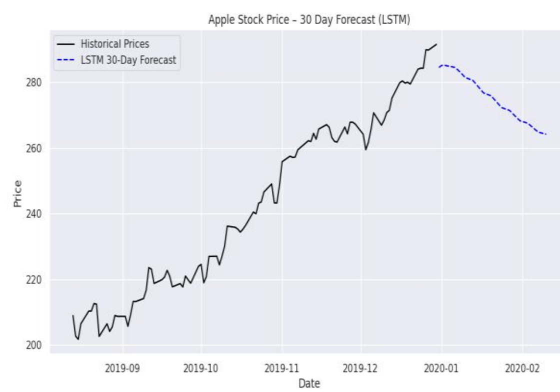
D. XGBoost Model

Extreme Gradient Boosting (XGBoost) is an advanced machine learning algorithm based on gradient boosting techniques. It builds models sequentially, where each new model focuses on correcting the errors made by previous models. XGBoost is widely used in predictive analytics due to its efficiency, scalability, and strong predictive power. In stock price prediction tasks, XGBoost can effectively model complex relationships among features such as open price, close price, volume, and historical trends. In this research, XGBoost achieved better performance than Random Forest because it can capture complex feature interactions more efficiently



E. LSTM Model

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed specifically for sequential and time-series data. Unlike traditional neural networks, LSTM networks contain memory cells that allow them to remember important information for long periods of time. This makes them highly effective for modeling temporal dependencies in time-series datasets such as stock prices.



V. SYSTEM ARCHITECTURE

A. Architecture overview

The proposed architecture for predicting the stock price of Apple Inc. consists of multiple stages that transform historical financial data into accurate future forecasts. Initially, historical stock data such as open, close, high, low prices, and trading volume are collected through data acquisition. The collected data then undergoes preprocessing, where missing values, noise, and inconsistencies are handled to ensure data quality. After preprocessing, feature engineering is applied to extract meaningful patterns and indicators such as moving averages and lag values that help improve prediction performance. The processed data is further scaled and used for training different prediction models, while hyperparameter optimization is performed to identify the best model configuration. In the model training stage, machine learning and deep learning algorithms learn patterns from historical stock trends. The trained models are then evaluated using performance metrics such as RMSE and MAPE to determine their accuracy. Finally, the best-performing model is selected in the prediction module to forecast future stock prices, and the results are visualized through graphs and charts to help analyse and interpret the predicted trends.

B. Data Collection & Preprocessing

The first stage involves gathering and preparing historical stock market data for analysis. In this stage, historical stock market data of Apple Inc. is collected from financial data sources. The dataset typically includes information such as opening price, closing price, highest price, lowest price, and trading volume over a period of time. After collecting the data, preprocessing techniques are applied to improve data quality. This includes handling missing values, removing duplicate or inconsistent records, and organizing the data into a structured time-series format. Proper preprocessing ensures that the dataset is clean and suitable for building reliable prediction models.

C. Modeling & Prediction

Once the data is prepared, different prediction models are applied to learn patterns from historical stock prices. Statistical models such as ARIMA and SARIMA are used to capture linear trends and seasonal patterns in time-series data. In addition, machine learning algorithms like Random Forest and XGBoost help model complex nonlinear relationships within the dataset. Deep learning models such as LSTM are also used because they are capable of capturing long-term dependencies in sequential financial data. These models are trained using the processed dataset to generate predictions for future stock prices.

D. Performance Evaluation

After training the prediction models, their performance is evaluated using statistical error metrics. Metrics such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are commonly used to measure the difference between predicted and actual stock prices. These evaluation measures help determine the accuracy and reliability of each model. By comparing the performance results, the system identifies which model provides the most accurate predictions for the stock price.

E. Forecasting & Visualization

In the final stage, the trained model is used to generate future stock price predictions for Apple Inc. based on historical patterns learned during the training process. The selected model analyzes previous price movements, trends, and relationships within the time-series data to estimate upcoming stock values for a specific time horizon such as days, weeks, or months. This forecasting process helps identify potential market trends and possible fluctuations in stock prices. The predicted values are then compared with historical data to assess how the stock may behave in the future under similar conditions. To make the results easier to interpret, the forecasted outputs are presented using graphical visualization techniques such as line graphs, trend plots, and comparison charts. These visualizations typically display both the actual historical prices and the predicted future prices on the same graph, allowing analysts to clearly observe the predicted trend direction and possible turning points. Additionally, visualization helps in identifying patterns such as upward trends, downward trends, or stable market behavior. By presenting the results in an intuitive visual format, investors, researchers, and financial analysts can better understand the predicted stock movements and use the insights to support decision-making, risk analysis, and investment strategy planning.

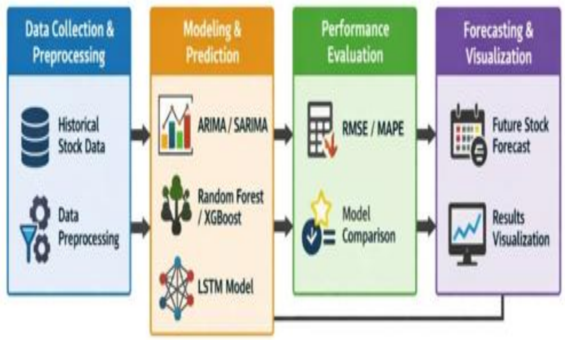


Fig. 1. System Architecture for Apple Stock Price Prediction

VI. RESULTS AND DISCUSSION

The performance of different forecasting models was evaluated using historical stock price data of Apple Inc. Several models including ARIMA, SARIMA, Random Forest, XGBoost, and LSTM were implemented to predict future stock prices. The models were trained using the training dataset and their performance was tested using unseen testing data. Evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were used to measure the prediction accuracy of each model. The experimental results indicate that traditional statistical models such as ARIMA and SARIMA were able to capture basic trends in the time-series data but showed limitations in modeling complex nonlinear relationships in stock price movements. Machine learning models such as Random Forest and XGBoost performed better than statistical models because they can capture nonlinear interactions between input features. Among all the models, the LSTM model achieved the best prediction performance due to its ability to learn long-term dependencies in sequential data.

VII. CONCLUSION

This study presents a comprehensive framework for forecasting stock price movements using historical time-series data of Apple Inc. The research involved collecting and preprocessing historical stock market data, followed by performing exploratory data analysis to identify trends, patterns, and relationships within the dataset. Several forecasting models, including ARIMA, SARIMA, Random Forest, XGBoost, and LSTM, were implemented and compared to determine their effectiveness in predicting future stock prices. The experimental results revealed that traditional statistical models such as ARIMA and SARIMA are capable of capturing basic linear trends in the time-series data but are limited when dealing with the highly nonlinear and volatile nature of financial markets. In contrast, machine learning models like Random Forest and XGBoost demonstrated improved performance by effectively modeling nonlinear relationships between different stock price features.

Among all the models evaluated, the LSTM model achieved the best prediction accuracy due to its ability to learn long-term dependencies and sequential patterns in time-series data. This makes LSTM particularly suitable for financial forecasting tasks where historical price behavior significantly influences future price movements. The comparison of evaluation metrics such as RMSE, MAE, and MAPE further confirmed that deep learning approaches provide more reliable forecasting results than traditional statistical techniques. In addition, graphical comparisons between actual and predicted stock prices indicated that the LSTM model closely follows the real market trend.

Overall, the proposed forecasting framework demonstrates the effectiveness of combining data preprocessing, exploratory analysis, and advanced machine learning techniques for stock price prediction. The results of this study highlight the importance of using deep learning models for analyzing complex financial datasets. The developed model can support investors, traders, and financial analysts in understanding market behavior and making more informed investment decisions. For future work, the forecasting system can be enhanced by incorporating additional financial indicators, macroeconomic variables, and sentiment analysis from news articles and social media platforms. Integrating hybrid deep learning architectures and real-time data processing techniques may further improve prediction accuracy and enable more robust financial forecasting systems.

REFERENCES

[1] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015.

[2] Sepp Hochreiter and Jürgen Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] Leo Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[4] Tianqi Chen and Carlos Guestrin, “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[5] Rob J. Hyndman and George Athanasopoulos, Forecasting: Principles and Practice, Melbourne, Australia: Monash University, 2018.

[6] James D. Hamilton, Time Series Analysis. Princeton, NJ, USA: Princeton University Press, 1994.

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[8] Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. Sebastopol, CA, USA: O’Reilly Media, 2019.

[9] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li, “Empirical evaluation of rectified activations in convolutional networks,” in Proceedings of the International Conference on Learning Representations, 2015.

[10] Ashish Vaswani et al., “Attention is all you need,” in Advances in Neural Information Processing Systems, 2017.

[11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[12] Jason Brownlee, Deep Learning for Time Series Forecasting. Melbourne, Australia: Machine Learning Mastery, 2018.

[13] Yahoo Finance, “Historical stock market data,”

[14] Fischer Black and Myron Scholes, “The pricing of options and corporate liabilities,” Journal of Political Economy, vol. 81, no. 3, pp. 637–654, 1973.

[15] NASDAQ, “Stock market data and financial information,”