International Journal of Engineering in Computer Science

**E-ISSN:** 2663-3590
**P-ISSN:** 2663-3582
IJECS 2024; 6(1): 61-66
[www.computersciencejournals.com/ijecs](www.computersciencejournals.com/ijecs)
Received: 17-01-2024
Accepted: 23-02-2024

**Manju Papreja**
GVM Institute of Technology and Management, DCRUST, Sonipat, Haryana, India

**Rashmi Chhabra**
GVM Institute of Technology and Management, DCRUST, Sonipat, Haryana, India

**Dr. Renu Miglani**
GVM Institute of Technology and Management, DCRUST, Sonipat, Haryana, India

**Corresponding Author:**
**Manju Papreja**
GVM Institute of Technology and Management, DCRUST, Sonipat, Haryana, India

# Analyzing the performance of machine learning in predicting stock market trends

**Manju Papreja, Rashmi Chhabra and Dr. Renu Miglani**

**DOI:** [https://doi.org/10.33545/26633582.2024.v6.i1a.113](https://doi.org/10.33545/26633582.2024.v6.i1a.113)

**Abstract**
Stock prediction has been a significant area of interest for investors, traders, and researchers. Exact stock price foretelling can lead to informed investment decisions, risk management, and potential financial gains. Machine learning techniques have gained eminence in current years as powerful tools for stock prediction due to their ability to process massive quantities of data and recognize composite patterns. This research paper presents a comprehensive study on prediction of stock price by Long Short-Term Memory (LSTM) neural networks applied to a decade's worth of historical data (2013-2023) of various companies Tesla, Netflix, Meta, Amazon, and Apple. The purpose of study is to analyze the performance of machine learning LSTM model in contrast to CNN (Convolutional Neural Networks), RNN (Recurrent Neural Network) and Random Forest models in capturing complex temporal patterns within stock price data, facilitating more accurate predictions for investment decisions and risk management. The findings of this study hold practical significance for investors, traders, and researchers, offering a basis for making well-informed investment decisions and improving risk management strategies within the dynamic landscape of stock markets.

**Keywords:** Stock market trends, machine learning, LSTM, RNN, CNN, random forest

## 1. Introduction

Forecasting stock prices has long been a difficult and complex challenge, fascinating the attention of researchers and business analysts. It is a domain that not only tempers intellectual curiosity but also requires severe investigation due to its inherent complexity. The complex task of predicting stock prices is compounded by the multifaceted influence of external factors, including psychological sentiments, social dynamics, political events, and economic situations. These external entities exert a substantial impact on the stock market's behaviour, rendering the prediction of stock prices a formidable undertaking.

A main characteristic of stock market data is its propensity to exhibit time-varying and nonlinear patterns. These intricate and dynamic data patterns challenge traditional forecasting methodologies. Accurate prediction of stock market trends and prices is pivotal for participants in the stock business, especially investors. An investment decision devoid of comprehensive information and sound predictive insights can lead to substantial financial losses, underscoring the importance of robust forecasting techniques.

In the pursuit of high profits, investors find themselves compelled to foresee the future value of stocks issued by companies. Over time, a plethora of prediction techniques has emerged to facilitate accurate forecasting in the stock market. In the earlier epochs, before the advent of computational methods for risk analysis, conventional approaches held sway. Two such methods that gained prominence are Fundamental and Technical Analysis.

**1.1 Fundamental Analysis:** To ascertain the accurate value of a product, it is imperative to possess reliable and precise information about a company's financial health, competitive position, and the economic landscape it operates within. This information serves as the bedrock for making investment decisions. Fundamentally, the notion rests on a simple principle: "Invest when the intrinsic value of a product exceeds its market value; otherwise, abstain from it as a risky investment." In addition to these parameters, metrics like book value, earnings, return on investment (ROI), price-to-earnings (P/E) ratio, among others, warrant meticulous scrutiny to glean insights into future business conditions. Fundamental analysis excels in long-term predictions due to its systematic approach and predictive capabilities.

**1.2 Technical Analysis:** Technical analysis is rooted in the premise that stock prices exhibit trends and movements in response to constantly changing investor sentiments and external factors. This methodology harnesses various quantitative parameters, including trend indicators, daily highs and lows, market indices, daily fluctuations, and trading volumes, to decipher these trends. Technical analysis seeks to uncover discernible rules and patterns from historical data, upon which investors base their future decisions. Notably, different analysts may derive distinct rules from the same data charts. Technical analysis is versatile and applicable to both short-term and long-term stock market analyses. It is often preferred over fundamental analysis as the input data for prediction systems due to its effectiveness in capturing market dynamics.

## 2. Literature Review

Stock market prediction is a perpetual challenge that has captivated the interest of business analysts and researchers. Traditional methods, such as Fundamental Analysis and Technical Analysis, have historically been used for stock forecasting. Fundamental Analysis involves evaluating a company's financial health and market position, while Technical Analysis relies on historical price patterns and trends. However, these methods have limitations in capturing the complex, non-linear relationships inherent in stock data [1]. Machine learning techniques have emerged as powerful tools for stock prediction, offering the ability to process extensive datasets and uncover intricate patterns. Approaches like Support Vector Machines, Random Forests, and artificial neural networks, particularly Long Short-Term Memory (LSTM) networks, have gained traction for their capacity to enhance prediction accuracy [2, 3, 4].

Deep learning techniques, especially LSTM neural networks, have demonstrated exceptional promise in modelling temporal dependencies in stock price data. LSTM networks excel at identifying long-term trends, short-term fluctuations, and recurrent patterns, making them suitable for various investment horizons [5]. Feature engineering has become pivotal in stock prediction, involving the creation of relevant input variables such as technical indicators, market sentiment derived from social media and news sentiment analysis, and macroeconomic indicators [6].

To evaluate the effectiveness of different machine learning models Comparative analyses have been conducted in stock prediction, often using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Outcomes from these studies vary across datasets, time periods, and model architectures, focusing on the importance of context-specific model selection [7].

The quality and availability of data are central to stock prediction. Researchers have utilized diverse data sources, including historical price data, news sentiment, economic indicators, and alternative data sources like satellite imagery [8]. Data pre-processing techniques such as normalization and feature scaling are usually applied to prepare data for model training [9].

Learning techniques such as Gradient Boosting Machines and neural networks have been applied to combine the analyse the strengths of multiple models, therefore improving prediction accuracy and robustness [10]. Moreover, hybrid models that integrates traditional and machine learning approaches have been explored [11].

As the use of machine learning models in stock prediction grows, interpretability have become critical concern. Researchers are actively developing methods to make these models more interpretable that supports the investors to gain insights into model predictions and decision-making processes [12].

In conclusion, researches on stock prediction using machine learning reflects a dynamic and progressing field. Traditional methods continue to provide valuable insights, but machine learning techniques, especially deep learning models like LSTM networks, offer enhanced capabilities for capturing the intricate patterns in stock price data. Feature engineering, data pre-processing, and model selection remain critical factors influencing prediction accuracy. The growing emphasis on interpretability and explain ability underscores the need for transparent and trustworthy models in the domain of stock market prediction. This research paper contributes to the existing body of knowledge by conducting a comprehensive analysis of LSTM-based stock prediction across various datasets and timeframes, shedding light on the practical implications of these models for investors and traders.

## 3. Methodology

The entire code is structured as a Python Notebook, leveraging essential Python libraries such as Pandas and Numpy to load the dataset and execute mathematical operations. Sklearn is utilized for implementing various machine learning algorithms. For interactive data visualization, Matplotlib and Streamlit have been incorporated into the code.

### 3.1 Data Collection

The dataset used for this analysis encompasses historical stock market data spanning the past decade, sourced from the NASDAQ finance website. The focus of this analysis is on the stock performance of five prominent companies: Amazon, Netflix, Tesla, and Apple.The data contains information about the stock such as High, Low, Open, Close, and Volume. Only the day-wise closing price of the stock has been extracted.

**Table 1:** Statistics of the Dataset for Training and Testing.

| | Date | Close/Last | Volume | Open | High | Low |
|---|---|---|---|---|---|---|
| 0 | 09/18/2023 | $265.28 | 101543300 | $271.16 | $271.44 | $263.7601 |
| 1 | 09/15/2023 | $274.39 | 133692300 | $277.55 | $278.98 | $271.00 |
| 2 | 09/14/2023 | $276.04 | 107709800 | $271.32 | $276.7094 | $270.42 |
| 3 | 09/13/2023 | $271.30 | 111673700 | $270.07 | $274.98 | $268.10 |
| 4 | 09-12-2023 | $267.48 | 135999900 | $270.76 | $278.39 | $266.60 |

### 3.2 Data Pre-processing

In our data preprocessing workflow, we commence a series of vital steps to ensure data quality and consistency. First, dataset is precisely examined for any missing or null values. When any gaps are encountered then data is promptly replaced with their mean values, ensuring that data remains complete and representative. Next, data is scrutinized for categorical values to identify any unnecessary or redundant information, which then dropped from the dataset to streamline it for analysis. Moreover, date formats in dataset is standardized to ensure uniformity throughout the dataset. Finally, for ease of numerical analysis, special symbols like dollar sign are removed from 'close', 'high', and 'low' values, subsequently converting them into integers. This comprehensive data preprocessing approach ensures that model will work on clean and consistent data, setting the stage for meaningful analysis and insights.

**Table 2:** Preprocessed Dataset

| | Date | Close | Volume | Open | High | Low |
|---|---|---|---|---|---|---|
| 0 | 18-09-2023 | 265.28 | 101543300 | 271.16 | 271.4400 | 263.7601 |
| 1 | 15-09-2023 | 274.39 | 133692300 | 277.55 | 278.9800 | 271.0000 |
| 2 | 14-09-2023 | 276.04 | 107709800 | 271.32 | 276.7094 | 270.4200 |
| 3 | 13-09-2023 | 271.30 | 111673700 | 270.07 | 274.9800 | 268.1000 |
| 4 | 09-12-2023 | 267.48 | 135999900 | 270.76 | 278.3900 | 266.6000 |

### 3.3 Data Scaling

In our research paper, we have implemented standardization and normalization techniques, with a specific focus on MinMaxScaler, as a crucial step in preparing our dataset. These techniques play a vital role in ensuring that our dataset exhibits uniform and comparable variable ranges, laying the foundation for robust machine learning methodologies. MinMaxScaler is a well-established data preprocessing approach widely embraced in the realms of both machine learning and data analysis. Its primary function is to scale and normalize numerical features, effectively transforming the original data into a new range, typically confined within the interval [0, 1]. This meticulous process harmonizes our data features, ultimately enabling us to conduct meaningful and equitable comparisons within the framework of our machine learning endeavors.

**Table 3:** Normalized Data Set

| |
|---|
| array([[0.64001794], [0.6626832 ], [0.66678833], [0.01005308], [0.01043124], [0.00952389]]) |

### 3.4 Data Splitting

The processed dataset has been partitioned into training and testing subsets using the train test split method. Specifically, 80% of the data, amounting to 2012 data points, has been allocated for training purposes, while the remaining 20%, comprising 504 data points, has been reserved for testing.

The temporal segmentation of the data is as follows: The training dataset encompasses data recorded from September 19, 2013, to September 18, 2023. Conversely, the testing dataset spans the same time frame, encompassing observations recorded from September 19, 2013, to September 18, 2023.

### 3.5 Model Building

The core of our research paper lies in the meticulous construction of predictive models utilizing the Long Short-Term Memory (LSTM) approach, specifically tailored for stock price prediction. LSTM, a recurrent neural network (RNN) variant, has been selected due to its exceptional ability to capture complex temporal dependencies inherent in time series data, making it an ideal choice for forecasting stock prices.

The LSTM model architecture is designed, encompassing the determination of LSTM layers, hidden unit configurations, activation functions, and any supplementary layers, such as dropout or dense layers. The selection of appropriate loss functions, frequently employing Mean Squared Error (MSE), and optimizers, such as Adam, is paramount to the model's efficacy.

### 3.6 Model Training

The training phase of our LSTM model is a critical step where the model is immersed in sequences of historical stock price data. Throughout this process, the model is equipped to discern and capture temporal patterns and dependencies inherent in the data. Key parameters are determined during training, including the number of training epochs (Iterations) and the batch size, which defines the volume of data sequences processed simultaneously. It is noteworthy that the model's internal weights undergo continuous adjustments during training, aimed at minimizing the specified loss function. This phase represents the core of our research, where the neural network acquires the ability to recognize and interpret complex market dynamics within the historical data, thus laying the groundwork for accurate stock price predictions.

**Table 4:** Training of designed LSTM model

| **Model: "Sequential"** | | |
|---|---|---|
| Layer (Type) | Output Shape | Param# |
| Lstm (LSTM) | (None, 50) | 10400 |
| Dense (Dense) | (None, 1) | 51 |
| Total params: 10451 (40.82 KB) | | |
| Trainable params: 10451 (40.82 KB) | | |
| Non-trainable params: 0 (0.00 Byte) | | |

### 3.7 Making Prediction

Once our LSTM model has undergone rigorous training, it is primed for action. The subsequent task is to harness its acquired knowledge to predict stock prices. To achieve this, we employ the model Prediction on the reserved testing dataset. This operation unleashes the model's ability to extrapolate future stock prices based on the intricate patterns and dependencies it has learned during its training phase.

Also the predictions generated by the model often reside in a scaled or normalized format, owing to the pre-processing steps employed earlier. To make these predictions more interpretable and align them with the actual stock price scale, a critical transformation process is invoked. This is equally applicable to the actual target values from the testing data. By employing the scaler inverse transform method, we adeptly reverse the scaling transformation, thereby restoring both predicted and actual values to their original, real-world units, such as actual stock prices. This step is pivotal in facilitating a meaningful and accurate evaluation of the model's predictive prowess against the backdrop of authentic stock price data.

### 3.8 Visualization

Visualization plays a crucial role in gaining insights into the model's performance by juxtaposing its predictions with the actual stock prices. These visual representations offer valuable perspectives on how well the model captures underlying trends, patterns, and deviations within the dataset. Common visualization techniques encompass time series plots, scatter plots, or candlestick charts, tailored to the specific nature of your data and the objectives of your analysis.

In our research paper focused on dataset from Tesla, we apply a range of visualizations including line charts, scatter plots, residuals, and distribution plots. Figure 1 shows the Tesla's stock performance analysis which gives a complete understanding and interpretation of our findings.

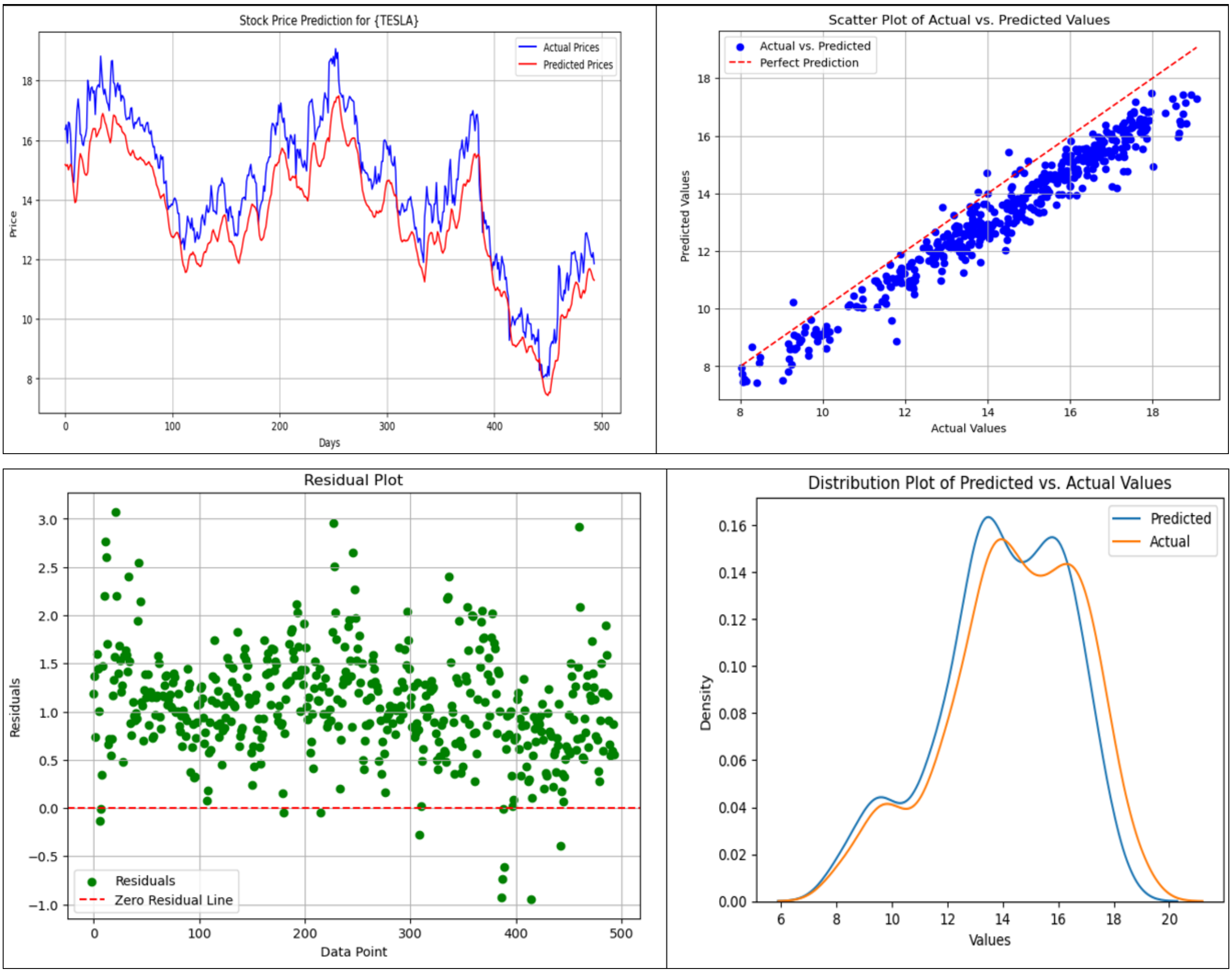


**Fig 1:** Analysis of Tesla's stock performance by training model

### 3.9 Error Calculation

In this paper, four distinct error calculation methods used for evaluation of our designed model. The formulae to compute the MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and MSE (Mean Squared Error) value are:

$$MAPE = \frac{1}{n} \sum \left( \frac{|Ai-Pi|}{|Ai|} \right) \times 100$$

$$MAE = \frac{1}{n} \sum ( |Ai-Pi| )$$

$$MSE = \frac{1}{n} \sum (A_i - P_i)^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum (Ai - Pi)^2}$$

Where:

- *n* represents the sample size.
- *Ai* denotes the actual (observed) value.
- *Pi* represents the predicted value.

## 4. Results and Discussion

The results of our stock price prediction models and their implications is presented in this section. Various models such as LSTM, RNN, Random Forest, and CNN has evaluated by error calculation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

LSTM model is emerged as the best model for stock prediction on our dataset, achieving the lowest MAE (1.09) and RMSE (1.20), MAPE (4.70%). The achievement of LSTM can be attributed to its ability to capture temporal dependencies in stock price data efficiently. Random Forest model displayed a slightly higher MAE (1.17) than LSTM that also maintained a reasonable level of accuracy with a low MAPE of 1.21%. But, it had a higher MSE (5.87) and RMSE (2.42) compared to LSTM that indicates slightly less precision of Random Forest. RNN presented intermediary performance with MAE (2.56), MAPE (5.63%) and RMSE (4.64). CNN generated the least accurate predictions model with the highest error values of MAE (5.55), MSE (67.00), RMSE (8.19), and MAPE (22.72%).

The result matrix in table 4 highlights the error calculation that helps in choosing an appropriate model for stock price prediction. LSTM and Random Forest models demonstrated higher performance, with LSTM being the most precise and accurate model for stock prediction. These models are compatible for handling the time-dependent nature of stock data, capturing trends and providing accurate forecasts.

**Table 5:** Results of stock price prediction models

| Model | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|
| LSTM | 1.09 | 1.44 | 1.20 | 4.70% |
| Random Forest | 1.17 | 5.87 | 2.42 | 1.21% |
| RNN | 2.56 | 21.55 | 4.64 | 5.63% |
| CNN | 5.55 | 67.00 | 8.19 | 22.72% |

The Comparison of these learning models on the basis of error calculation done above is shown in figure 2. These graphs indicated that LSTM model is best suited for the stock price prediction in comparison of other methods.

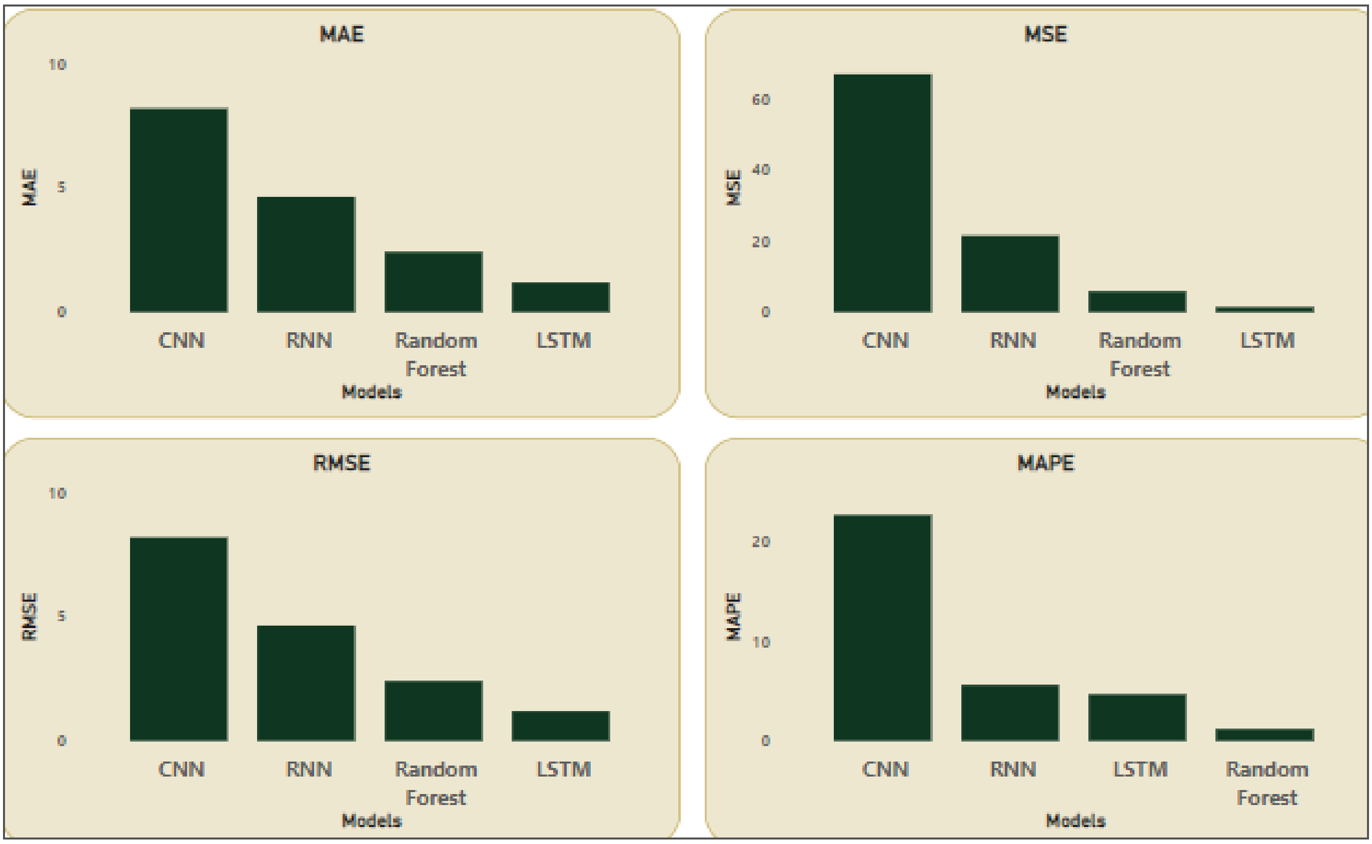


**Fig 2:** Comparison of models on the basis of error calculation

## 5. Conclusion

This research is focused to evaluate different stock price prediction models including LSTM, Random Forest, RNN, and CNN, employing a range of performance metrics. LSTM appeared as the best model, demonstrating superior accuracy with the lowest MAE and RMSE values, describing its effectiveness in capturing temporal dependencies within stock data. Random Forest with slightly less precise, maintained good accuracy. In contrast, CNN and RNN proved less effective for modeling stock price prediction. Overall, our findings emphasized the efficacy of LSTM and Random Forest in stock price prediction, with model selection tailored to specific dataset and project needs. Future researches may explore ensemble methods and feature engineering for further improvements.

## 6. References


1. Campbell JY, Lo AW, MacKinlay AC. The Econometrics of Financial Markets. Princeton University Press; c1997.
2. Daigavane VN, Patil SS. Stock Market Prediction Using Machine Learning Algorithms. Int. J Comput. Appl., 2012, 60(5).
3. Breiman L. Random Forests. Mach Learn. 2001;45(1):5-32.
4. Lipton T, Wang T, Smola A. Detecting and Interpreting Deep Neural Network Drift. arXiv preprint arXiv:2103.16148. 2021.
5. Moody J, Saffell M. Learning to Trade via Direct Reinforcement. IEEE Trans Neural Netw. 2001;12(4):875-889.
6. Deng LJ, Yeh CH. Stock Market Forecasting Using Machine Learning Algorithms. Int. J Mach Learn Comput., 2012, 2(3).
7. Zhang Y, Ma J, Zhang D. Stock Market Forecasting: A Comparison of Machine Learning and Deep Learning Models. IEEE Access. 2021;9:34498-34508.
8. Srinivasan GN. Forecasting Stock Prices using

Fundamental and Technical Analysis. IIMB Manage Rev. 2016;28(3):206-218.
9. Jin Z, Wang D, Chen J. An Empirical Study of Machine Learning Techniques for Stock Market Prediction. In: International Conference on Computational Science; c2018.
10. Wang Y, Xia Y, Song L, Yang B. Forecasting Stock Prices with Technical Indicators: A Case Study on Shanghai Stock Exchange Composite Index. Complexity, 2017, 2017.
11. Ramya PA, Geetha S. Stock Price Prediction Using Hybrid Model. Procedia Comput Sci. 2015;50:464-469.
12. Zhang Y, Dietterich TG. A Domain Adaptation Approach to Stock Price Forecasting. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; c2010.