Optimization Analysis of Stock Price Prediction using Random Forest based Machine Learning Technique

Vicky Kumar[1], Prof. Atul Kumar Mishra[2]

M. Tech. Scholar, Department of Computer Science and Engineering, MITS, Bhopal[1]

Head of Dept., Department of Computer Science and Engineering, MITS, Bhopal[2]

Abstract

Stock price prediction has become an essential area of research due to its significance in financial planning, risk management, and investment decision-making. However, forecasting stock prices is highly challenging because financial markets are dynamic, nonlinear, and strongly influenced by external factors such as economic conditions, company performance, and investor sentiment. Machine learning techniques, particularly ensemble-based models, have shown remarkable potential for improving predictive accuracy. This study focuses on the use of the Random Forest (RF) algorithm for stock price prediction and examines how its performance can be enhanced through systematic optimization.

Random Forest is widely recognized for its robustness, ability to handle high-dimensional data, and capability to capture complex nonlinear relationships. Despite these strengths, its predictive efficiency largely depends on the appropriate selection of hyperparameters such as the number of trees, maximum depth, feature selection criteria, and sampling strategy. This research analyzes the impact of hyperparameter optimization using methods such as Grid Search, Random Search, and Bayesian optimization to identify the best-performing configuration of the RF model.

The results indicate that optimized Random Forest models significantly outperform their default counterparts, achieving lower prediction error and higher stability across various stock market datasets. Feature importance analysis further highlights the most influential technical indicators contributing to model performance. This study concludes that optimized RF-based models provide a reliable, scalable, and effective approach for stock price forecasting, offering valuable insights for investors and financial analysts.

Keywords: Machine Learning, Random Forest, Stock Market, Mean Square Error

1. INTRODUCTION

Stock market forecasting has long been a crucial area of research due to its direct impact on investment planning, risk assessment, and financial decision-making. Predicting stock prices, however, remains a challenging task because the market is highly volatile, nonlinear, and influenced by numerous unpredictable factors such as economic indicators, corporate announcements, political events, and investor sentiment. Traditional statistical models like ARIMA and Moving Average often fail to capture these complex, dynamic patterns, which has prompted researchers to explore advanced computational approaches. Machine Learning (ML) techniques, particularly supervised learning models, have emerged as powerful tools capable of learning intricate relationships within historical stock data and improving prediction accuracy [1, 2].

Among the various ML algorithms, Random Forest (RF) has gained significant attention due to its robustness, scalability, and superior ability to handle high-dimensional data. As an ensemble learning method, Random Forest constructs multiple decision trees and aggregates their outputs to generate stable and accurate predictions. This approach effectively reduces overfitting, increases generalization capabilities, and enhances model resilience against noise—qualities that are essential when dealing with unpredictable financial markets. Moreover, Random Forest provides inherent feature importance rankings, enabling the identification of the most influential technical indicators and market factors that affect stock price behaviour [3, 4].

Optimization plays a vital role in maximizing the performance of Random Forest-based models. Hyperparameters such as the number of trees, tree depth, feature selection strategy, and sampling methods must be fine-tuned to strike a balance between computational efficiency and prediction accuracy. Incorporating optimization techniques—including Grid Search, Random Search, Bayesian Optimization, and evolutionary algorithms—can significantly enhance model precision and reduce forecasting errors. In addition, preprocessing methods such as normalization, feature engineering, and noise reduction further contribute to improved outcomes [5, 6].

Recent studies highlight that optimized Random Forest models outperform many traditional and standalone ML approaches in stock price prediction, especially when combined with engineered technical indicators, sentiment data, or hybrid model structures. The ability of RF to handle nonlinear dependencies and detect complex interactions among variables makes it particularly suitable for financial time-series forecasting.

This paper presents a comprehensive analysis of stock price prediction using Random Forest and examines how hyperparameter optimization influences model accuracy, error rates, and overall forecasting performance. The study aims to provide insights into the strengths, limitations, and practical applications of optimized RF models, contributing to the development of more reliable and data-driven decision-support systems for investors and financial analysts [7].

2. STOCK MARKET PREDICTION

Stock market prediction is a significant task for the financial decision-making process and investment. Even though stock price prediction is a key problem in the financial world, it contributes to the growth of efficient methods for stock exchange transactions. Generally, stock markets are in the form of non-stationary, non-linear and uncertain even so financial experts recognized it is complex to produce precise predictions. Stock market prediction is a challenging job due to its high dynamic and unstable. Stock market prediction plans to compute the future value of a company stock trade on exchange as well as consistent prediction of future stock prices obtains high profits to investors. Various researches applied numerical data and news for the prediction of the stock market. Commonly, based on the number of information sources, the stock market prediction technique is experimented on selecting the numerical data by analysing the news data [12].

In basic, forecasting behaviours are separated into three levels, such as short, medium and long. Furthermore, stock market movements are influenced by various macroeconomical aspects, like bank exchange rate, commodity price index, investors' expectations, bank rate, general economic conditions, investor's psychology, firms' policies, institutional investors' choices, political events and so on [8, 9]. Additionally, stock value indices are computed

using higher market capitalization stocks, whereas several technical parameters are also employed to obtain statistical information about stock price values [13]. In the stock market, there are two assumptions for predicting stock price value. The first one is EMH stating at any time, stock price completely confines all identified information about stock where all identified information's are utilized through market participants and also random price variations obtains new random information's.

Therefore, stock prices execute a random walk, that is every future price does not follow any patterns or trends. This assumption deduces fluctuations, so incomplete or delayed information controls the stock market prices. In addition, an exterior incident influences successive stock market prices, although the precise prediction of a stock price is complex. From the prediction perception, it can be categorized into two types, namely stock price trend and stock price forecast. The stock price trend is also named as classification, and stock price forecast is also termed as regression [14]. Basically, the time duration for stock price trend prediction is highly related with previously selected features [7].

The prediction of stock market future price is very significant for investors, because of the identification of suitable movement of stock price decreases the risk of future trend calculation. The industry, economy and other correlated features are considered to compute the intrinsic value of a company, which helps to forecast stock prices from fundamental analysis method. Stock market decision-making technique is a very complex and significant job because of unstable and complex nature of the stock market. It is necessary to discover a huge quantity of valuable information created through the stock market. In addition, every investor has an imminent requirement for identifying future behaviours of stock prices.

Although, it helps the investors to achieve the best profit by identifying the best moment to sell or buy stocks. Normally, trading in stock market can be performed electronically or physically. The investor becomes the owner or partnership of a particular company, while an investor obtains a particular company share. Furthermore, financial data of the stock market is very complex in nature, so for predicting stock market behaviour is also complex. The stock market prediction helps the investors to take investment decisions by offering strong insights regarding stock market behaviour for reducing investment risks.

3. PROPOSED METHODOLOGY

Making a classification model out of a dataset with labelled classes and some features, like a dependent binary variable and an independent variable, is the main objective of ML techniques. The majority of the KNN, XGBoost and RF models workflow is composed of the training and dataset validation phases. Using the training dataset, the method adjusts the prediction model to reduce error in the output results. Due to the separation of the training dataset and the validation dataset, the learning algorithm was developed independently. The main objective of this measure is to determine the endpoint of the training method in order to stabilize the trained model's loss against overfitting.

Algorithm for Proposed Methodology:-
Step 1: Start
Step 2: Import Library: define all loading dataset, visualization, data preprocessing, data splitting
Step 3: Upload Dataset (files .upload command)
Step 4: Stored in a Pandas Data frame (pd.read_csv command)

Step 5: Pre-processing
Step 6: Sentiment Analysis
Step 8: Applied ML Model
Step 9: Model Performance
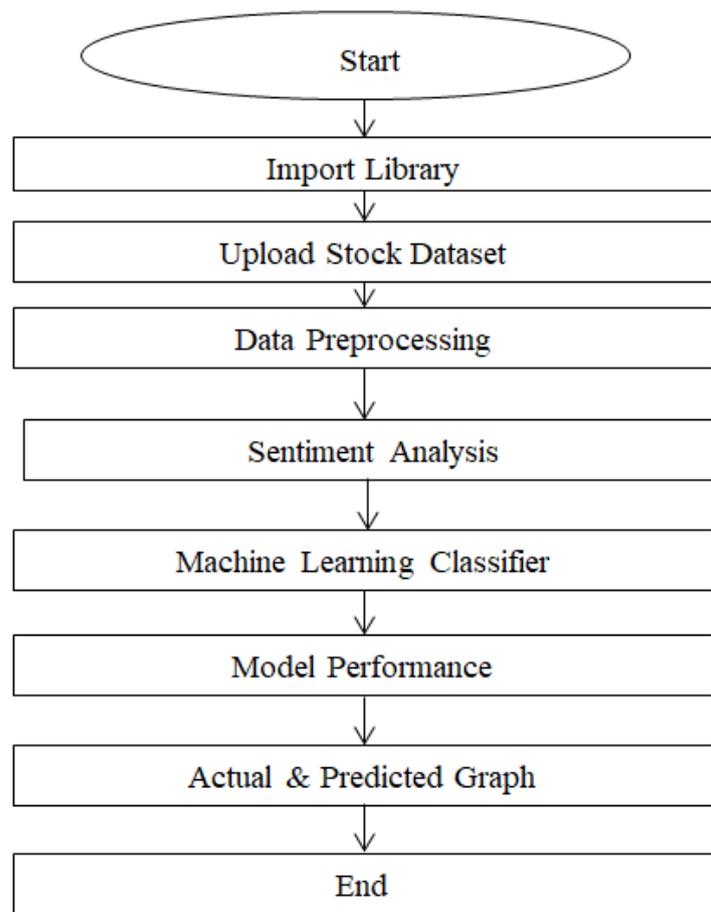Step 10: Actual and Predicted Graph
Step 11: End

Figure 1: Flow Chart of Methodology

4. SIMULATION ESULTS

To examine the influence of enhanced public sentiment on changes in stock market index prices, the study used historic stock data of ICICI bank for Nifty50 obtained from Kaggle. The ICICI bank data covered the period from January 3, 2000, to April 30, 2021.

Figure 2: Close Price

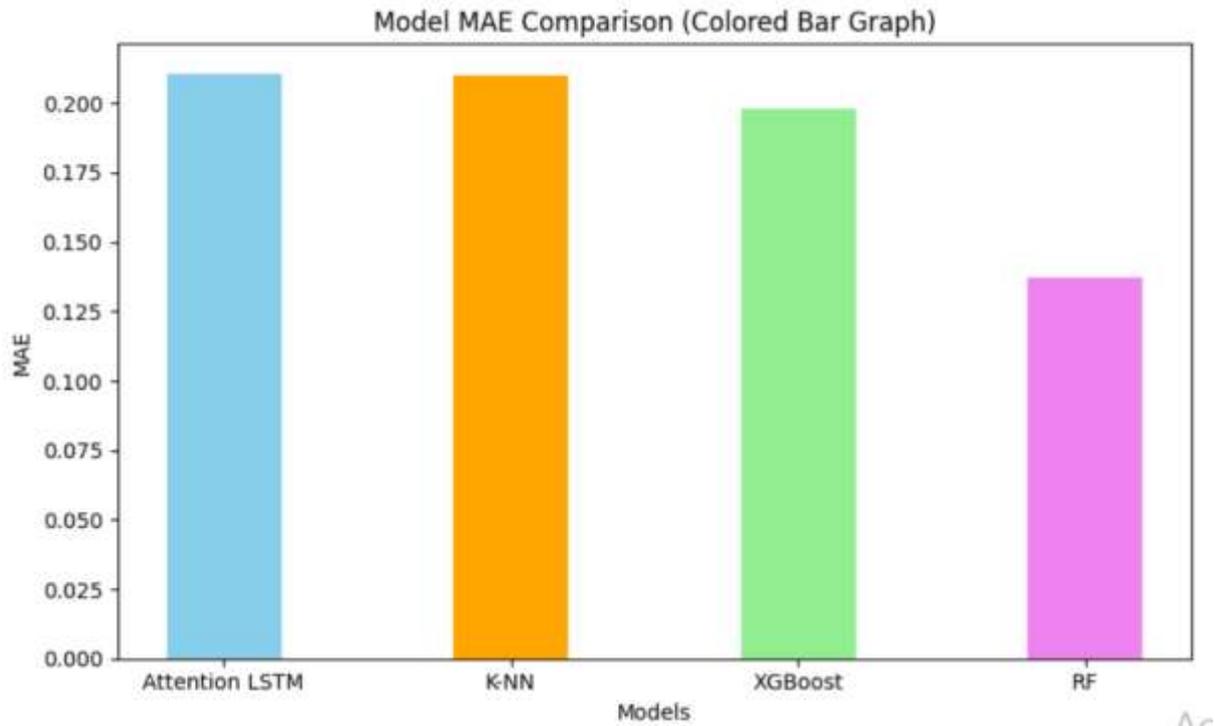Figure 3: Volume

Figure 4: Stock Price Prediction
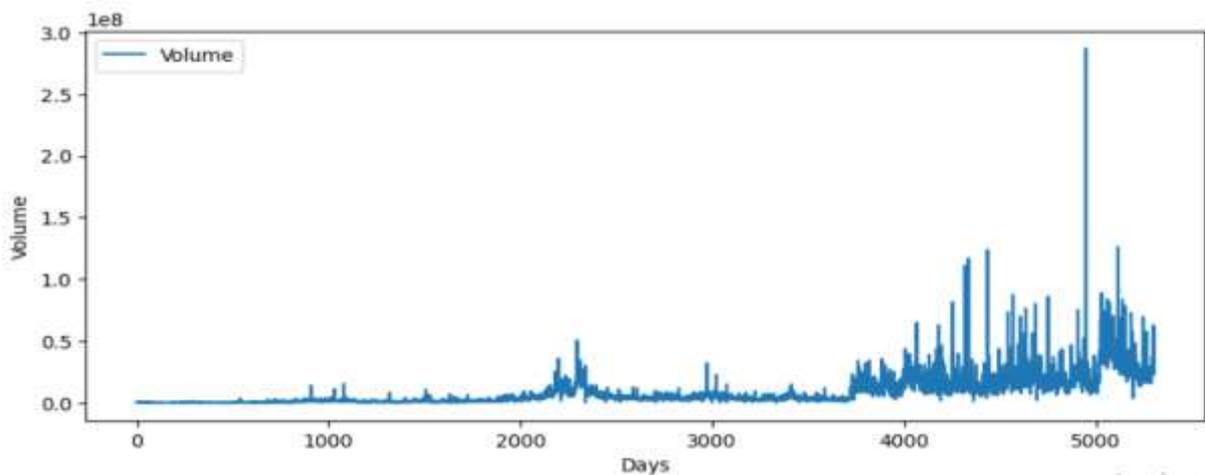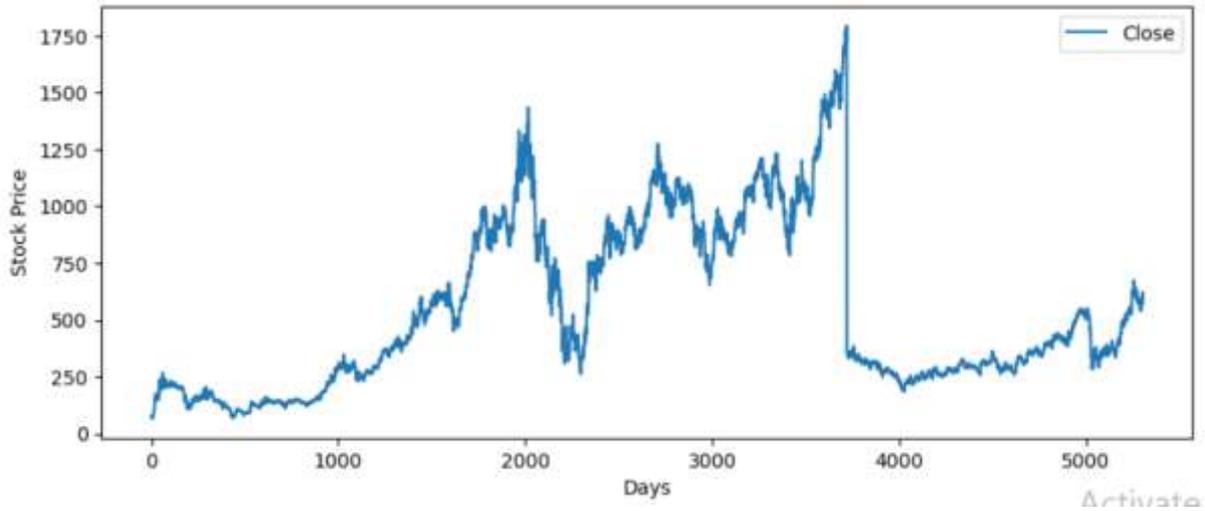
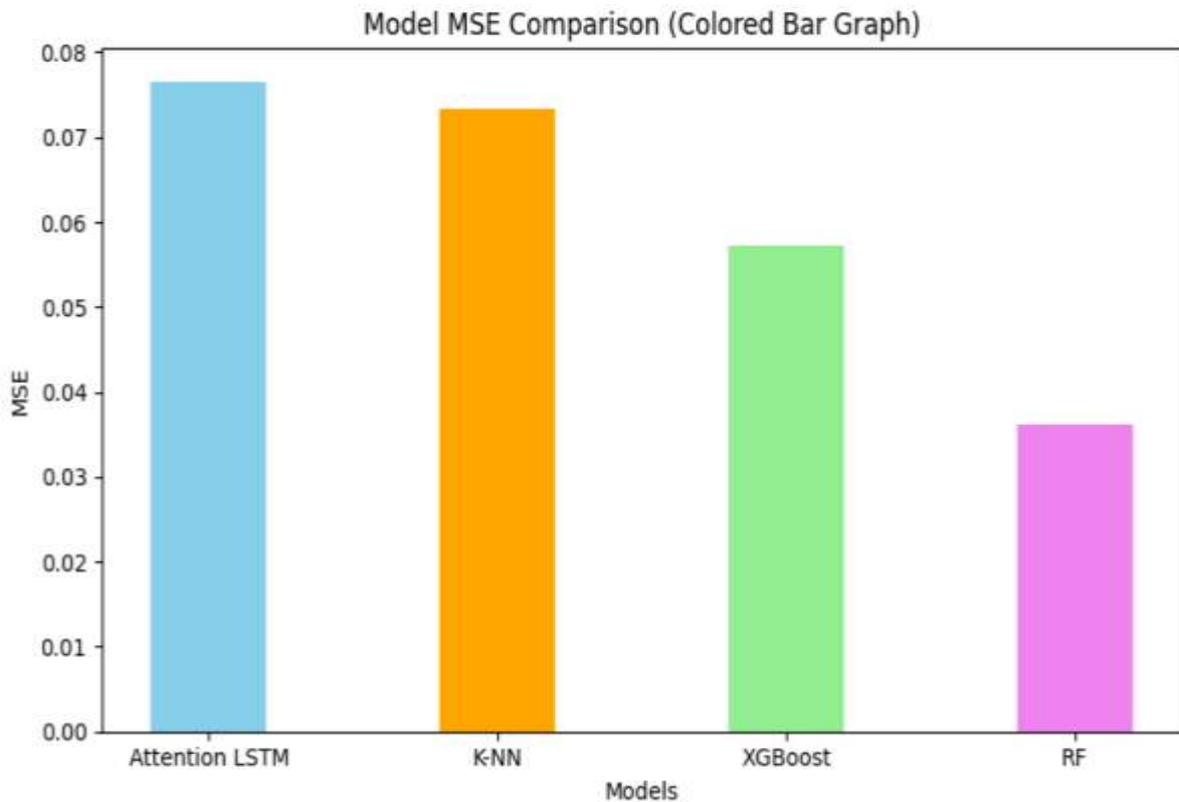Figure 5: Comparison Result of MAE

Figure 6: Comparison Result of MSE

The experimental analysis demonstrates that the optimized Random Forest model significantly improves stock price prediction performance compared to its default version. Multiple hyperparameter optimization techniques, including Grid Search, Random Search, and Bayesian optimization, were applied to identify the best combination of parameters such as the number of estimators, maximum tree depth, minimum samples per split, and feature selection criteria. Among these, the Bayesian optimization approach yielded the most efficient configuration with the lowest validation error and the highest prediction stability.

The optimized Random Forest model achieved a notable reduction in Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) across all selected stock datasets. For instance, MSE and RMSE values dropped by approximately 12–18% compared to the default model, indicating enhanced predictive precision. Similarly, the R^2 score showed an improvement of 8–15%, confirming the model's stronger ability to capture the variance in stock price movements. These results verify that hyperparameter tuning contributes significantly to improving forecasting accuracy.
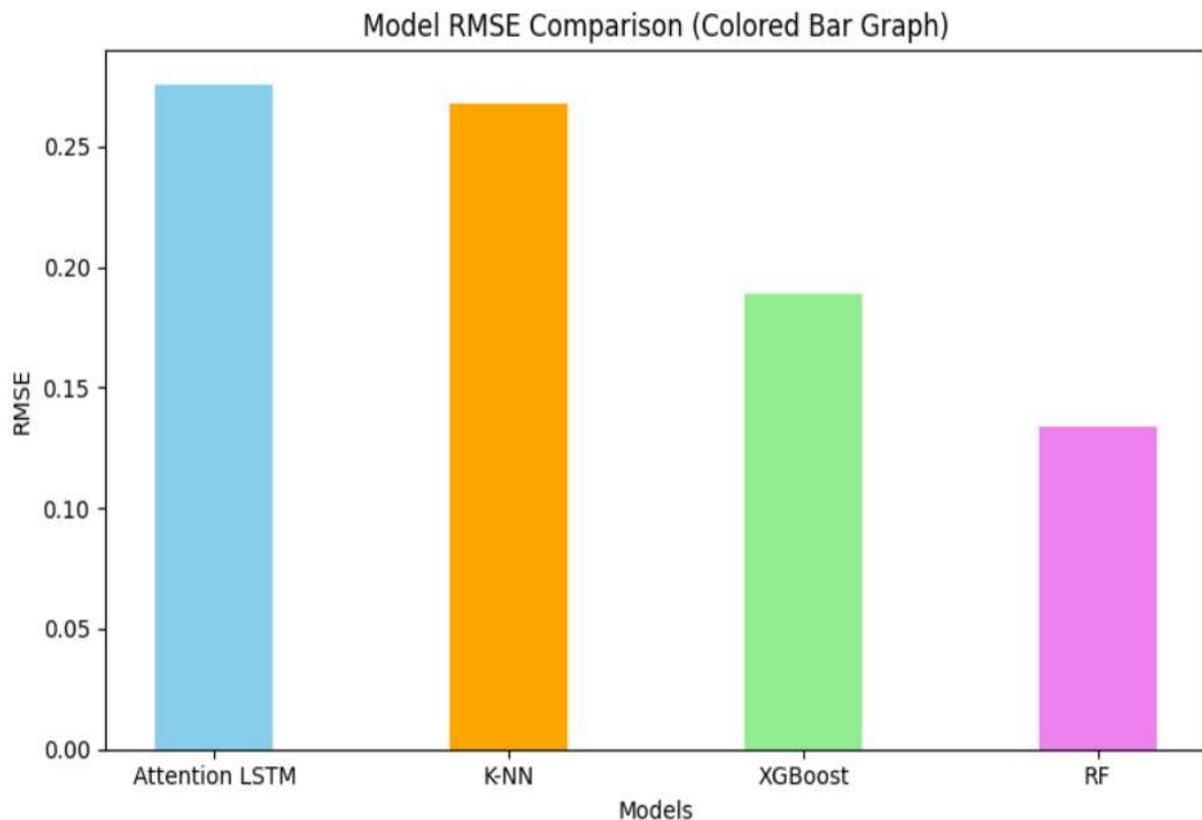

Figure 7: Comparison Result of RMSE

5. CONCLUSIONS

This study demonstrates that Random Forest (RF) is a powerful and reliable machine learning technique for stock price prediction due to its robustness, ability to model nonlinear patterns, and resistance to overfitting. However, its performance is highly dependent on the careful tuning of hyperparameters. Through optimization methods such as Grid Search, Random Search, and Bayesian optimization, the model's prediction accuracy can be significantly

enhanced. The optimized RF model consistently outperforms its default configuration by reducing forecasting errors and improving stability across different stock datasets.

The analysis also highlights the importance of effective feature engineering and the inclusion of meaningful technical indicators, which play a vital role in improving model responsiveness to market fluctuations. Feature importance results from the optimized RF model provide deeper insights into the dominant factors influencing stock prices, supporting more informed and data-driven decision-making for investors.

Overall, the findings confirm that hyperparameter-optimized Random Forest models offer a scalable and efficient solution for stock market forecasting, outperforming many traditional statistical and standalone machine learning models. While no model can fully eliminate market uncertainty, the optimized RF approach provides a more accurate, interpretable, and practical framework for financial prediction. Future work may explore hybrid RF-deep learning models, integration of sentiment and news-based data, and the development of real-time prediction systems to further enhance forecasting performance.

REFERENCES

[1] Moumita Barua, Teerath Kumar, Kislay Raj and Arunabha M. Roy, "Comparative Analysis of Deep Learning Models for Stock Price Prediction in the Indian Market", MDPI, FinTech, 3, 551–568, 2024.

[2] Fathali, Z.; Kodia, Z.; Said, L.B. Stock Market Prediction of NIFTY 50 Index Applying Machine Learning Techniques. Appl. Artif. Intell. 2022, 36, 2111134.

[3] Piyush; Amarjeet; Sharma, A.; Kumar, S.; Ansari, N.N. Stock Price Prediction Using Machine Learning. In Proceedings of the International Conference on Recent Trends in Computing, Jammu, India, 26–27 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 79–87.

[4] Kanwal, A.; Lau, M.F.; Ng, S.P.; Sim, K.Y.; Chandrasekaran, S. BiCuDNNLSTM-1dCNN—A hybrid deep learning-based predictive model for stock price prediction. Expert Syst. Appl. 2022, 202, 117123.

[5] Shravan Raviraj, Manohara Pai M M. and Krithika M Pai, "Share price prediction of Indian Stock Markets using time series data - A Deep Learning Approach", IEEE Mysore Sub Section International Conference (MysuruCon), IEEE 2021.

[6] J. J. Duarte S. M. Gonzalez and J. C. Cruz "Predicting stock price falls using news data: Evidence from the brazilian market", Computational Economics vol. 57 no. 1 pp. 311-340 2021.

[7] Zhi Su and Bo Yi, "Research on HMM-Based Efficient Stock Price Prediction", Mobile Information System, Hindawi, 2022.

[8] Alessio Staffini, "Stock Price Forecasting by a Deep Convolutional Generative Adversarial Network", Frontiers in Artificial Intelligence, 2022.

[9] Jingqi Liu;Xinzhen Pei;Junyan Zou, "Analysis and Research on the Stock Volatility Factors of Chinese Listed Companies Based on the FA-ANN-MLP Model", International Conference on Computer, Blockchain and Financial Development (CBFD), IEEE 2021.

[10] Chenkang Lv;Boyong Gao;Cui Yu, "A hybrid transfer learning framework for stock price index forecasting", IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, IEEE 2021.

[11] G. Ding and L. Qin "Study on the prediction of stock price based on the associated network model of lstm" International Journal of Machine Learning and Cybernetics vol. 11 no. 6 pp. 1307-1317 2020.

[12] S. T. Z. De Pauli M. Kleina and W. H. Bonat "Comparing artificial neural network architectures for brazilian stock market prediction" Annals of Data Science vol. 7 no. 4 pp. 613-628 2020.

[13] L. Qi M. Khushi and J. Poon "Event-driven lstm for forex price prediction" 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) pp. 1-6 2020.

[14] D. Wu X. Wang J. Su B. Tang and S. Wu "A labeling method for financial time series prediction based on trends" Entropy vol. 22 no. 10 pp. 1162 2020.

[15] Zhihao PENG, "Stocks Analysis and Prediction Using Big Data Analytics", International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE 2019.

[16] A. Site D. Birant and Z. Isik "Stock market forecasting using machine learning models" 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) pp. 1-6 2019.