Stock Price Prediction Using Machine Learning Algorithms

Sumukha K Nayak, Krishnamurthy Nayak and Supreetha B S

[2,3] *Department of Electronics & Communication Engineering, Manipal Institute of technology, Manipal Academy of Higher Education, Manipal,*

supreetha.bs@manipal.edu, km.nayak@manipal.edu

[1] *Student, BITS Pilani, Pilani Campus, Rajasthan, India,*

Email: f20230932@pilani.bits-pilani.ac.in

Abstract.

Stock price prediction is essential in today's dynamic financial markets. This research uses machine learning algorithms to forecast stock prices, helping investors make informed decisions. The process includes data acquisition from Tata Global Beverages for learning model for stock price prediction. The study utilizes Python and libraries like pandas, NumPy, scikit-learn, and TensorFlow. Various models are evaluated using metrics like RMSE, MAE, and R^2. The results highlight each model's effectiveness in predicting stock prices, providing valuable insights for financial decision-making. The random forest consistently outperformed other techniques, delivering the lowest RMSE and highest R-Squared values.

.

Keywords. Stock price prediction, machine learning, tata Global Beverages, RMSE, hyperparameters.

1. INTRODUCTION

Stock price prediction using machine learning has become increasingly important in finance, offering advanced solutions to the complexities and volatility of modern financial markets. The work aims to develop predictive models using historical data and machine learning algorithms to assist investors in making informed decisions. The objectives include developing and evaluating machine learning models for accurate stock price prediction, enhancing prediction accuracy, and visually representing predicted versus actual prices. The target specifications aim to deliver reliable predictive models for investors and financial analysts, enhancing investment decisions, portfolio management, and risk mitigation. The motivation for this research work arises from the limitations of traditional stock price prediction methods and the need for more accurate forecasting techniques. Previous studies often lacked comprehensive comparisons of machine learning algorithms and used limited datasets, leading to less accurate predictions. This research aims to address these gaps by employing advanced machine learning techniques, providing more reliable insights into market trends.

In today's volatile financial markets, accurate stock value estimation is essential for investors, analysts, and researchers. Recent research highlights the effectiveness of data driven approaches outperforming traditional methods. Also, Random Forest. Long Short-Term Memory (LSTM) networks show superior performance compared to simpler techniques such as linear regression, particularly in capturing temporal dependencies in stock prices. K Nearest Neighbours (KNN) has demonstrated potential with a high accuracy rate of 70%. The literature provides various insights into the effectiveness of different methodologies. Studies by Vijha et al. [1] and Pramod et al. [2]. emphasize the strengths of ANN and LSTM, respectively, in forecasting stock values. Parmar et al. [3] focus on regression and LSTM models, while Nandakumar et al. [4] highlight LSTM's advantages in handling temporal dependencies. Sonkavde et al. [5] review multiple algorithms, recommending ensemble techniques and careful hyperparameter tuning. Vanukuru et al. [6] discuss the potential of Support Vector Machines (SVM), and Latha et al. [7] propose KNN for its simplicity and effectiveness. Zhang et al. [8] find linear regression performs better with limited data compared to LSTM. Sharma et al. [9] identify XGBOOST as a highly effective model, and Al-Ali et al. [10] highlight CNN and LSTM as top performers across various sectors. Raj et al. [11] and Kumar et al. [12] explore the application of LSTM, and Wong et al.

[13] suggest integrating sentiment analysis with prediction models. Xue et al. [14] propose a novel approach using multi-branch LSTM models, while Purnama et al. [15] show that Linear Regression outperforms SVM. Liu et al. [16] develop an optimized LSTM model with regularisation techniques, improving accuracy. Gunturu et al. [17] find that hybrid models combining Linear Regression and Random Forest are particularly effective. Bathla et al. [18] emphasize LSTM's superiority over Support Vector Regression (SVR), and Patnaik et al. [19] provide a comprehensive review of various AI approaches. Rodrigues et al. [20] propose integrating sentiment analysis with ARIMA-LSTM hybrids to enhance prediction accuracy. Machine learning algorithms, including ANN, LSTM, SVM, KNN, and hybrid models, have demonstrated significant capabilities in predicting stock market trends.

2. METHODOLOGY

The process begins with the collection of historical stock market data from reputable sources such as Yahoo Finance. The dataset of Tata Global Beverages Limited (TATAGLOBAL) is considered for the study. The data collected includes daily opening and closing prices, trading volumes, and other financial measures, which provide a comprehensive view of market dynamics

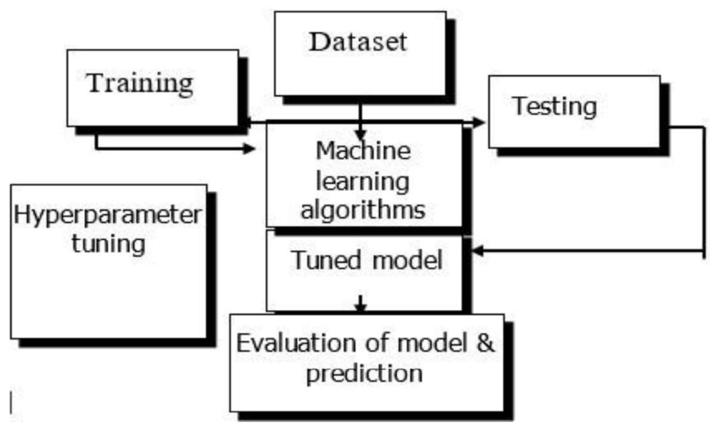


Figure 4. Methodology Block diagram

Using Python libraries like Pandas and NumPy, the data is cleaned with techniques such as Min-Max scaling. This ensures the data is consistent and ready for machine learning models. Feature engineering further enhances the data by incorporating lagged variables, technical indicators like moving averages, and volume-based features.

Figure 2: Close prize history

In this study, moving averages are utilized as key features to enhance stock price prediction models. By including these moving averages, the model is better equipped to capture underlying trends and potential turning points in the stock price. The plots of these moving averages illustrate how they track the stock's price movements, providing visual cues that help in understanding market dynamics and making more informed predictions.

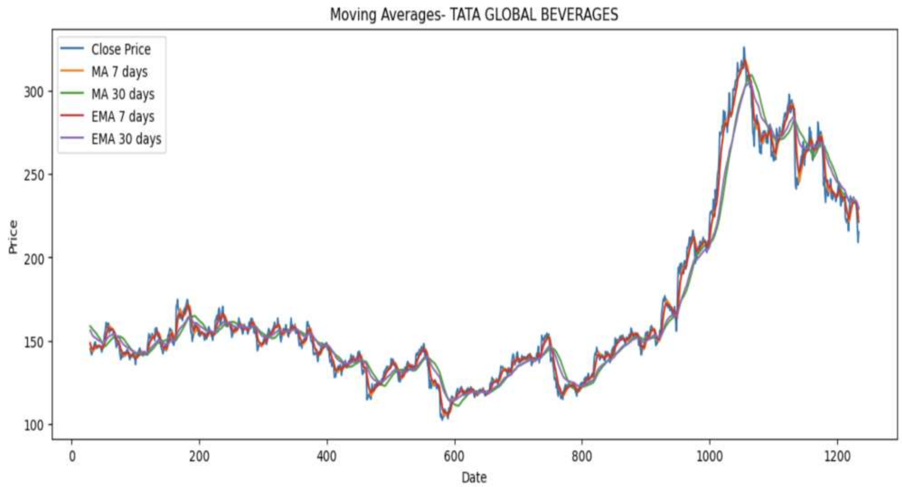


Figure 3: Moving averages plot

The methodology includes the selection and implementation of different machine learning models. Each model is carefully implemented and tuned using Python libraries such as Scikit-learn, Keras, and Statsmodels. For example, LSTM networks are designed with specific layers and regularization techniques to improve prediction accuracy, while Random Forest models leverage ensemble learning to enhance robustness.

Evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared are used to compare predictions against actual stock prices. Visualization tools are also employed to plot predicted versus actual prices and analyze error distributions, providing valuable insights into model performance.

3. RESULTS & DISCUSSION

The evaluation of stock price prediction models involves a thorough examination of numerous metrics. Root Mean Squared Error (RMSE) is used to calculate the average difference between predicted and actual values, whereas R-squared values are used to determine how well a model's independent variables explain the variation in the dependent variable. Mean Absolute Error (MAE) adds to this study by calculating the average absolute errors between predicted and actual values. These metrics are applied to a variety of models, including ARIMA, K-Nearest Neighbours, Linear Regression, Long Short-Term Memory (LSTM), Random Forest, and Support Vector Machine (SVM). By comparing these metrics and examining error distributions, effective model can be established, characterised by reduced RMSE and MAE, as well as higher R-squared values, indicating higher accuracy and predictive capabilities. Visualization in stock price prediction involves plotting train actual, validation actual and predicted data using `plt.plot()`.This comparative analysis assesses models' performance. Error distribution plots created using plt.hist(), offer insights into prediction accuracy by showcasing the frequency and magnitude of errors between actual and predicted values. Errors that are centred at 0 suggest consistent and unbiased predictions. A skewed distribution to the left implies that the model usually underestimates actual values, whereas a skew to the right indicates overestimation. Stakeholders leverage these visuals for informed decision-making, refining predictive strategies and understanding model efficacy in forecasting stock prices.

A. ***LINEAR REGRESSION:***

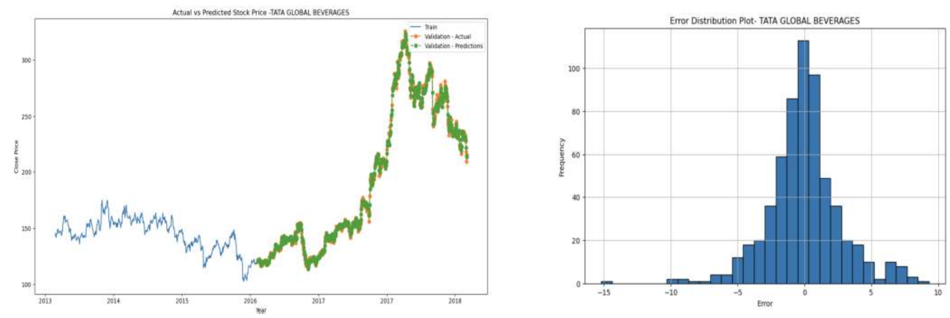


Fig.4. Actual vs predicted stock price and Error Distribution plot -Linear Regression

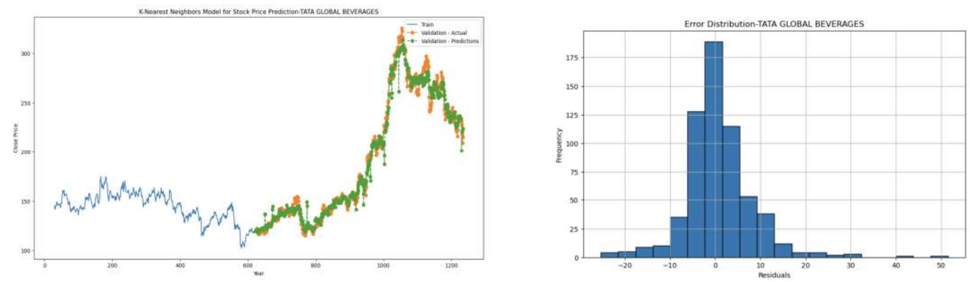


Fig.5 Actual vs predicted stock price and Error Distribution plot for K-Nearest Neighbours algorithm

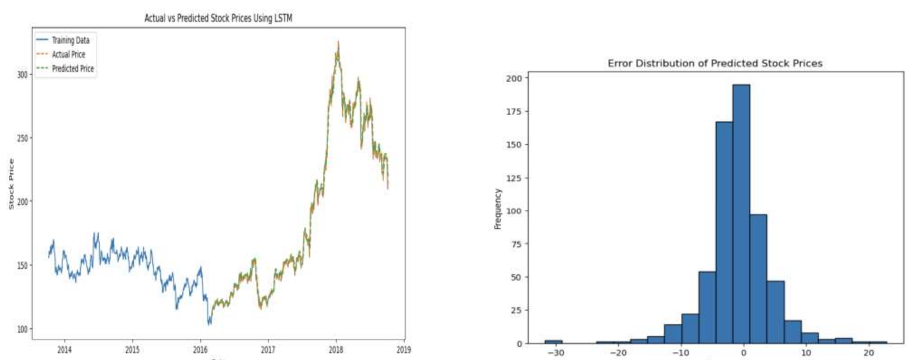


Fig.6. Actual vs predicted stock price and Error Distribution plot for LSTM

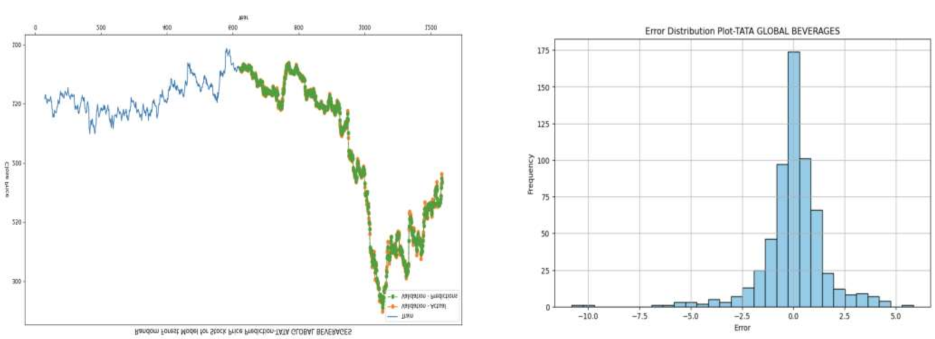


Fig.7. Actual vs predicted stock price and error distribution plot -Random Forest

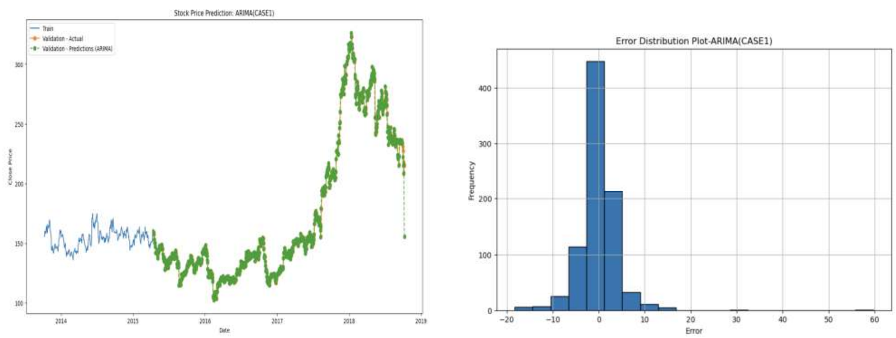
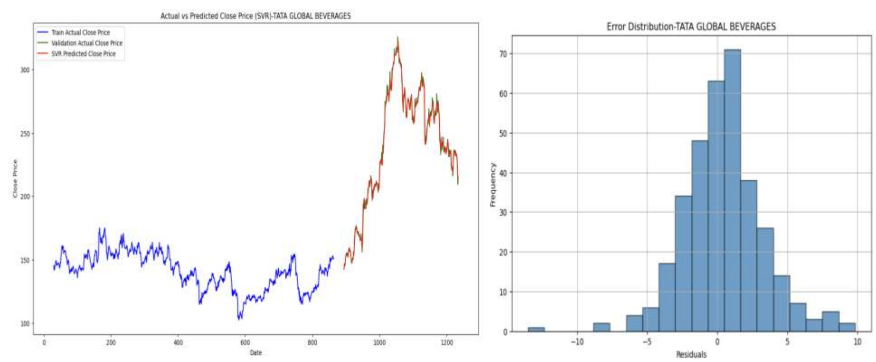


Fig.8. Actual vs predicted stock price and error distribution plot-ARIMA



CASE-1-Best Parameters for SVR: {'C': 100, 'epsilon': 0.0001, 'kernel': 'linear'}

Fig.9. Actual vs predicted stock price and error distribution plot SVR

The performance analysis through statistical indices like RMSE and MAE R2 is as presented in table#.

Table 7.1. Evaluation Metrics Values

Model	RMSE	R-squared	MAE
Linear Regression	2.73	0.998	1.93
K-N nearest neighbours	4.28	-2.580	5.09
LSTM	4.00	0.992	0.96
Random Forest	1.57	0.999	0.99
ARIMA	4.45	NaN	2.72
SVM	2.83	0.996	2.09

In the analysis of stock price prediction models, Random Forest consistently outperformed other techniques, delivering the lowest RMSE and highest R-squared values across both cases, underscoring its robustness and reliability. Linear Regression also performed well, particularly showing strong predictive power with minimal error. Support Vector Machine (SVM) achieving the lowest RMSE and highest R-squared, demonstrating its effectiveness with appropriate parameter tuning. In contrast, K-Nearest Neighbours (KNN) and ARIMA struggled, with high RMSE and negative R-squared values, likely due to their sensitivity to noise and model assumptions. Long Short-Term Memory (LSTM) models provided reasonable results, The results highlight that model efficacy varies based on dataset characteristics. Random Forest and SVM demonstrated superior performance, while Linear Regression and LSTM require careful adjustment to avoid overfitting. KNN and ARIMA performed poorly, potentially due to their sensitivity to data noise and assumptions about linearity. The poor performance of KNN suggests issues with data noise and dimensionality, while ARIMA's linearity assumptions may not suit complex, non-linear datasets. The visualizations of predicted versus actual values and error distributions further support these findings, revealing each model's strengths and weaknesses. Models that closely matched the ideal line were more accurate, while significant deviations indicated areas needing improvement. Error distribution graphs showed that narrower distributions corresponded to consistent performance, whereas broader distributions reflected variable accuracy.

4. References

[1] Vijha, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2019). Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science, 167, 599606. doi: 10.1016/j.procs.2020.03.326.

[2] B S, Pramod, and Mallikarjuna Shastry P. M. "Stock Price Prediction Using LSTM." TEST ENGINEERING AND MANAGEMENT, vol. 83, May-June 2020, pp. 5246-5251.

[3] Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 574-576, doi: 10.1109/ICSCCC.2018.8703332.

[4] Nandakumar, R., K R, U., R, V., & Lokeswari, Y. V. (2018). Stock Price Prediction Using Long Short-Term Memory. International Research Journal of Engineering and Technology (IRJET), 05(03), 3342-3348.

[5] Sonkavde, G. *et al.* (2023) Forecasting stock market prices using machine learning and Deep Learning Models: A systematic review, performance analysis and discussion of implications, *MDPI*.

[6] Vanukuru, Kranthi. (2018). Stock Market Prediction Using Machine Learning. . International Research Journal of Engineering and Technology (IRJET),05(10).

[7] R. S. Latha *et al.*, "Stock Movement Prediction using KNN Machine Learning Algorithm," *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2022.

[8] H. Zhang, "Stock Price Prediction using Linear Regression and LSTM Neural Network," *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, Guangzhou, China, 2022.

[9] P. Sharma and M. K. Jain, "Stock Market Trends Analysis using Extreme Gradient Boosting (XGBoost)," *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2023, pp. 317-322.

[10] Al-Ali and A. I. Al-Alawi, "Stock Market Forecasting Using Machine Learning Techniques: A Literature Review," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Manama, Bahrain, 2024, pp. 466-471.

[11] N. Raj, "Prediction of Stock Market Using LSTM-RNN Model," *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India, 2023, pp. 623-628.

[12] Kumar, S. Hooda, R. Gill, D. Ahlawat, D. Srivastva and R. Kumar, "Stock Price Prediction Using Machine Learning," *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, Greater Noida, India, 2023, pp. 926-932.

[13] Wong, J. Figini, A. Raheem, G. Hains, Y. Khmelevsky and P. C. Chu, "Forecasting of Stock Prices Using Machine Learning Models," *2023 IEEE International Systems Conference (SysCon)*, Vancouver, BC, Canada, 2023, pp. 1-7

[14] T. Xue, S. Qin and Q. Fu, "Stock Market's Price Movement Prediction with Multi-branch LSTM and Technical Analysis," *2023 International Conference on Neuromorphic Computing (ICNC)*, Wuhan, China, 2023, pp. 338-344.

[15] Purnama, I & Ginantra, Ni & Darma, I Wayan Agus Surya & Udayana, I. (2023). Comparison of Support Vector Machine (SVM) and Linear Regression (LR) for Stock Price Prediction. 1-6.

[16] M. Liu and Y. Zhao, "Stock Prediction Based on LSTM Model," *2023 China Automation Congress (CAC)*, Chongqing, China, 2023, pp. 1794-1798.

[17] P. A. Gunturu, R. Joseph, E. S. Revant and S. Khapre, "Survey of Stock Market Price Prediction Trends using Machine Learning Techniques," *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*, Bangalore, India, 2023, pp. 1-5.

[18] G. Bathla, "Stock Price prediction using LSTM and SVR," *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 2020, pp. 211-214.

[19] D. Patnaik, N. V. Jagarmadha Rao, B. Padhiari and S. Patnaik, "Stock Price Forecasting Using Machine Learning Techniques: A Systematic Review," *2023 International Conference on Data Science & Informatics (ICDSI)*, Bhubaneswar, India, 2023, pp. 243-251.

[20] A. P, J. Rodrigues, R. R. Naik, K. M. C. Kumar and P. Shetgaonkar, "Stock Market Prediction Using Machine Learning - A Survey," *2024 3rd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, 2024, pp. 1-7.

Biographies

Sumukha K Nayak is persuing his Engineering degree from BITS Pilani, Pilani Campus, Rajasthan. He is currently working on artificial intelligence, web development, computational intelligence and optimization, MERN stack and embedded system design in different student clubs. He can be contacted at email: f20230932@pilani.bits-pilani.ac.in.

Krishnamurthy Nayak received the B.Tech. degree in Electronics and Communication Engineering from Manipal Institute of Technology, Manipal, and the M.Tech. and PhD degree from Dr.M.G. R Educational and Research Institute, Chennai, India He is currently working as additional professor in the Department of Electronics and Communication Engineering at MIT Manipal. His current research interests include soft computing applications for environmental sustainbility, design of bio and micro sensors for environmental applications and MEMS Technology. He can be contacted at email: km.nayak@manipal.edu

Supreetha B S received B.Tech. degree in Electronics and Communication Engineering from Mangalore University, M.Tech. and Ph.D. degree from Manipal Institute of Technology, Manipal, India. She is currently working as Assistant Professor (selection grade) in the Department of Electronics and Communication Engineering, MIT Manipal. Her current research interests include soft computing applications for environmental sustainability, design of bio and micro sensors for environmental applications, analog system design and MEMS technology. She can be contacted at email: supreetha.bs@manipal.edu.