IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141271

Stock Prediction using Machine Learning

Prof. Roopa K Murthy[2], Dayanidhi. S[2], Chirag K[3], Dhanush S[4], Manoj S[5]

Department of CS&D, K. S. Institute of Technology (KSIT), Bengaluru, India[1-5]

Abstract: Predicting the stock market is an extremely difficult endeavor given the unpredictable and nonlinear characteristics of financial markets, with price fluctuations occurring very quickly to complicate forecasting even more. The use of traditional statistical techniques frequently proves inadequate to replicate the complexities of stock movements, which has been the motivation behind increasing attention on the utilization of machine learning methodologies. This research examines the application of Long Short-Term Memory (LSTM) networks to predict stock prices based on past stock prices, using historical stock price data from 2019 to 2023, including Open, High, Low, and Close prices. The performance of the model is assessed through significant metrics like Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), indicating that LSTM can be used to successfully forecast long-term trends in stock prices. The study also investigates the performance of other machine learning models, including Gated Recurrent Units (GRU) and Artificial Neural Networks, for forecasting stock prices, with the results indicating that LSTM is superior to these models in identifying long-term dependencies. Nevertheless, it remains difficult to predict sudden changes in the market due to externalities, such as economic developments or geopolitical changes. The paper explores possible avenues for future work, such as combining sentiment analysis, hybrid models, and investigating the application of other deep learning architectures to further improve predictive power. The work adds to the body of research on machine learning in financial forecasting and sheds light on how stock market prediction models can be made more robust and accurate.

Keywords: Stock market prediction, machine learning, Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Artificial Neural Networks.

I. INTRODUCTION

The world's financial markets are dynamic, unstable, and very interconnected, with assets and stocks traded around the clock. Stock markets, more so, are at the core of the global economy, providing a platform where companies can mobilize capital, while investors get to purchase, sell, and exchange ownership in the companies. These markets, however, are dynamic and subject to the influence of an array of variables, ranging from macroeconomic data such as inflation rates and interest rates to worldwide events and sentiment. As a result, stock market values and trends are hard to predict, and forecasting them has been an interest that has been around for quite some time for investors, financial analysts, and researchers alike. Traditionally, stock market forecasts have depended on customary techniques like technical analysis and fundamental analysis. Technical analysis is concerned with the past price action and volumes of trade in anticipating future trends in stocks, whereas fundamental analysis analyses companies economic vitality through scrutiny of financial statements, earnings reports, and conditions in the marketplace.

Even though they are useful, these customary approaches have limitations in grappling with the complex, non-linear relationships that usually prevail in financial markets. This deficiency is especially glaring in recent years, developments in Machine Learning (ML) and Artificial Intelligence (AI) have brought new opportunities for enhancing stock market forecasts.
Machine Learning, with its capacity to process large volumes of data and identify concealed patterns, has emerged as a formidable tool in financial forecasting. In contrast to conventional approaches, ML algorithms have the capability to handle big datasets, such as past stock prices, economic factors, and even unstructured information like news, social media content, or geopolitical developments, to determine trends and make predictions. Such capacity to consolidate and process varied sources of data places.

Machine Learning as a promising technique to better and more efficiently predict stock prices. The area of machine learning has come a long way, with different methods demonstrating significant potential for stock market forecasting. Regression models, Decision Trees, Support Vector Machines (SVM), and ensemble methods are some of the most widely used methods that offer different ways of detecting patterns and trends in data. Deep Learning, especially Long Short-Term Memory (LSTM) networks, has also further transformed stock market forecasting. LSTMs, a type of Recurrent Neural Network (RNN), are particularly well-suited for time-series data, as they can capture long-term dependencies in historical market data and make predictions based on these temporal patterns.

DOI: 10.17148/IJARCCE.2025.141271

The aim of this study is to create a Machine Learning model that effectively predicts stock prices based on past market data and key economic indicators. Different Machine Learning methods like Regression models, Decision Trees, and ensemble techniques will be used to train the model with an emphasis on determining the best method to forecast stock prices.

The project will include major features like historical stock prices, interest rates, and other economic information relevant to the problem, enabling the model to learn the multiple factors that determine stock market behavior. The performance of the model will also be tested using typical metrics like accuracy, precision, recall, and Root Mean Squared Error (RMSE) to ensure that it gives good and actionable predictions.

II. LITERATURE REVIEW

Artificial Neural Networks (ANNs) have been extensively investigated for stock market forecasting because they can learn to identify intricate patterns in financial information. Jasic and Wood (2004) created an ANN model to forecast daily stock market index returns from the S&P 500, DAX, TOPIX, and FTSE indices. Their research proved that the ANN model performed better than a benchmark autoregressive model, especially for the S&P 500 and DAX indices. Enke and Thawornwong (2005) proposed a machine learning-based information gain approach to analyze financial and economic variables, choosing only the most useful features for prediction models.

Their study proved that classification-based neural network models produced higher risk-adjusted returns than regression-based models and buy-and-hold approaches. Liao and Wang (2010) introduced a stochastic time-effective neural.

Various studies have explored machine learning techniques for stock prediction. Early approaches included statistical methods such as Auto Regressive Integrated Moving Average (ARIMA) and Support Vector Machine(SVM). However, these models struggled with long-term dependencies in financial data. Recurrent Neural Networks(RNNs) improved performance but suffered from vanishing gradient issues. LSTM, a special type of RNN, addresses these limitations by maintaining long-term dependencies, making it suitable for stock price forecasting.

Recent research has demonstrated that deep learning models out perform traditional approaches in stock prediction. Studies by Fischer and Krauss (2018) compared LSTM to traditional methods, concluding that LSTM provided better accuracy in financial time-series forecasting. Another study by Bao et al. (2017) integrated wavelet transformations with LSTM to enhance prediction accuracy. However, challenges remain, particularly in volatile market conditions where models struggle with sudden price movements network model that gave greater weight to newer historical data, based on the hypothesis that investors value recent market trends over others.

Features for prediction models. Their study proved that classification-based neural network models produced higher risk-adjusted returns than regression-based models and buy-and-hold approaches.

Their model was validated with data from various international stock markets, such as the S&P 500, NASDAQ, and Hang Seng Index, and compared on a volatility parameter basis. Chavan and Patil (2013) also added value to ANN-based stock market prediction by examining varying model input parameters in various studies. They ascertained that machine learning models are largely based on technical indicators instead of fundamental variables for the forecasting of stock prices, with macroeconomic variables being more significant for the forecasting of stock index. Chong, Han, and Park (2017) investigated the use of deep learning in stock market prediction using unsupervised feature extraction techniques such as principal component analysis, autoencoders, and restricted Boltzmann machines.

The authors concluded that deep learning networks were able to enhance predictive performance through the extraction of residual information from conventional autoregressive models. Together, these studies emphasize the usefulness of ANNs in stock market prediction by their capacity to extract nonlinear relationships and enhance short-term market predictions.

III. METHODOLOGY

This study employs the Random Forest (RF) algorithm, an ensemble machine learning technique, to predict stock market prices based on historical financial data and relevant economic indicators. The methodology begins with data collection, where stock price data, financial indicators, macroeconomic variables, and market sentiment data are gathered from reliable sources.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141271

These datasets provide a comprehensive view of market trends and influences. Following data collection, preprocessing is performed to clean and structure the data effectively. Missing values are handled using imputation techniques, and features are normalized using Min-Max scaling or standardization to ensure consistency.

Additionally, feature engineering is conducted to create new predictive variables such as moving averages and momentum indicators, which help improve model performance. The dataset is then split into training and testing sets to validate the predictive capability of the model. To improve the efficiency of the model and avoid overfitting, a feature selection process is performed using correlation analysis and RF's built-in feature importance ranking.

This ensures that only the most important variables are included in the model, which improves speed and accuracy. The Random Forest model is then trained using the selected features, where multiple decision trees are built using randomly sampled data subsets. The Financial APIs like Yahoo Finance, Alpha Vantage, and Quandl are used to collect stock market data. These data usually comprise daily open, high, low, close prices, volume traded, and adjusted closing prices.

Furthermore, external variables such as economic metrics, interest rates, news sentiment, and social media trends can be included to improve the accuracy of predictions. The raw data also includes missing values and noise, which are addressed through interpolation methods, outlier detection, and smoothing filters. Data normalization and standardization are used to scale features so that models can operate at their best.

Ensemble-technique leverages bootstrapping and bagging, reducing variance and enhancing generalization. For regression tasks, the final prediction trained using the selected features, where multiple decision trees are built using randomly sampled data subsets.

The ensemble technique leverages bootstrapping and bagging, reducing variance and enhancing generalization. For regression tasks, the final prediction is obtained by averaging the outputs of all decision trees, while for classification, the majority class prediction is chosen. To optimize performance, hyperparameters such as the number of trees, maximum depth, and minimum samples per split are fine-tuned using Grid Search or Random Search techniques. Once trained, the model is evaluated using standard performance metrics. For regression tasks, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) Score are used to measure prediction accuracy and the model's ability to explain variance in stock prices.

For classification-based predictions, a confusion matrix and accuracy score are used to assess performance. These evaluation metrics help determine the effectiveness of the model and its capability to generalize well to unseen market conditions. By leveraging Random Forest, this methodology aims to provide an accurate and robust predictive model that can assist investors and analysts in making informed stock market decisions

Sentiment information from financial news and social media is also incorporated to improve predictive capabilities. Data preprocessing guarantees the elimination of missing values, outliers, and noise and normalization and organization of the data for time-series analysis. Feature engineering is vital in extracting valuable insights through technical indicators (Moving Averages, RSI, MACD), statistical measures (mean, variance, skewness), sentiment features (news sentiment scores, earnings transcripts), and fundamental analysis features (P/E ratio, EPS, ROE). Machine learning algorithms such as Random Forest, SVM, and XGBoost are used in conjunction with deep learning algorithms such as LSTM, GRU, and CNN, which are more appropriate for sequential data prediction. Hybrid models using LSTM with attention or reinforcement learning techniques further improve accuracy. Hyperparameter tuning is done using Grid Search, Bayesian Optimization, and Genetic Algorithms to optimize model performance, avoiding overfitting by using dropout layers and batch normalization. Models are tested with RMSE, MAPE, and R-squared values for regression problems, and classification models measure precision, recall, and F1-score. Optimized models are deployed via Flask or FastAPI and embedded in web applications, using cloud platforms such as AWS or Google Cloud for real-time prediction. Future developments include reinforcement learning for automated trading, Graph Neural Networks for stock relation modeling, and federated learning to improve security and privacy. By combining numeric stock data with sentiment analysis and using advanced AI methods, stock market prediction models can be made more accurate and responsive to fluctuating market scenarios.

IV. RESULTS AND DISCUSSIONS

The findings of this research confirm that LSTM models work extremely well for stock market forecasting, especially in predicting the S&P 500 index. The performance evaluation with major measures of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) confirms that the model registers low error

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪪ Peer-reviewed & Refereed journal ⪪ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141271

rates, which proves its predictive effectiveness. In contrast to conventional approaches such as Simple Moving Averages (SMA) and Support Vector Machines (SVM), the LSTM model performs better in identifying long-term dependencies and responding to volatile market trends. Although the SMA model is not able to capture abrupt changes in the market and the SVM model is limited in long-term prediction, LSTM learns sequential patterns well, which makes it a more trustworthy tool for stock price forecasting. Visual examination also corroborates these results since a comparison of actual vs. forecast stock prices indicates that LSTM closely imitates market trends, albeit with slight deviations in times of extreme volatility.

But there remain some drawbacks. The performance of the model deteriorates in extremely volatile situations,where external economic and geopolitical factors intervene in historical trends. Besides, data sensitivity remains a predominant factor, since the accuracy of forecasts depends to a large extent on how easily and accurately historical stock data is available. Another issue of concern is overfitting, whereby the model's tendency to overemphasize past trends diminishes its power to predict future stock prices accurately. In spite of these difficulties, the research emphasizes LSTM's financial forecasting potential. Further research might investigate hybrid methods, including the combination of LSTM with Attention Mechanisms or the incorporation of macroeconomic indicators, to increase predictive robustness and accuracy. In general, the results verify that LSTM models improve stock market prediction considerably over conventional methods, providing a useful tool for investors and financial analysts interested in more data-driven decision-making strategies.

V. CONCLUSION

Stock market forecasting is a highly complex but necessary field of finance research, where the development in machine learning (ML) and deep learning (DL) has greatly improved forecast accuracy. Common statistical models have been extensively utilized, but are not able to capture the inherent complex, nonlinear relationships in finance data. This work accentuates the utility of a host of ML and DL models, most importantly artificial neural networks (ANNs), support vector machines (SVMs), and deep models like long short-term memory (LSTM) networks. Comparative performance analysis reflects that though legacy models like random forest (RF) and linear regression form a sound benchmark, deep learning algorithms like LSTMs prove especially strong at retaining temporal dependencies and trends over time for stock price fluctuation. These models illustrate the capacity to learn from past experience and discover embedded patterns and thereby are more powerful in predicting stock market movement. Moreover, hybrid methods incorporating several machine learning methods, including the combination of sentiment analysis and numerical stock information, have produced encouraging results on enhancing prediction.

Sentiment analysis, as an extractionfrom social media and financial news, delivers key insights regarding investor behavior and market trends. When combined with quantitative share data, hybrid models provide a more comprehensive strategy for stock forecasting. Additionally, reinforcement learning methods can enhance trading strategies through learning from market behavior and responding to shifting circumstances continually. Regardless of these developments, there remain a number of challenges to stock market prediction, including market volatility, data inconsistency, as well as the effects of extraneous elements like economic policy, interest rates, and global politics. The money market is very responsive to abrupt changes, and even the most advanced models do not succeed in forecasting black swan events—infrequent and unpredictable phenomena that have a large influence on stock prices.

The uncertainty of global economic downturns, policy decisions, and business choices brings another level of uncertainty to financial prediction. Future studies can investigate the incorporation of reinforcement learning for creating adaptive trading strategies that can change dynamically in response to changes in the market. In addition, predictive power can be improved by the application of graph neural networks (GNNs) to examine intricate relationships among stocks, industries, and economic indicators.

The integration of real-time data streams, including high-frequency trading data, financial reports, macroeconomic indicators, and even alternative data sources such as satellite imagery and transaction records, could further enhance prediction accuracy.

In addition, developments in federated learning may resolve data privacy issues by facilitating collaborative model training among various financial institutions without exchanging sensitive information. This method enables institutions to tap into collective intelligence while preserving data confidentiality. The use of explainable AI (XAI) methods can also enhance transparency and interpretability in stock prediction models, enabling investors and financial analysts to better comprehend model decisions and minimize the risk of overfitting. Overall, prediction in the stock market is changing fast with advancements in technology, and the use of more advanced AI-based methods can result in more

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141271

accurate and dependable forecasting models. But it is important to realize that no model can eradicate financial risk completely because of the inherent risks involved in the stock market. Ongoing research and improvement of predictive models are needed to respond to evolving economic circumstances so that AI-based stock market forecasts remain current and effective in informing investment choices.

.

REFERENCES

[1]. Kumar, Deepak, Pradeepta Kumar Sarangi, and Rajit Verma. "A systematic review of stock market prediction using machine learning and statistical techniques." Materials Today: Proceedings 49 (2022): 3187-3191.
[2]. Singh, Gurjeet. "Machine learning models in stock market prediction." arXiv preprint arXiv:2202.09359 (2022).
[3]. Bansal, Malti, Apoorva Goyal, and Apoorva Choudhary. "Stock market prediction with high accuracy using machine learning techniques." Procedia Computer Science 215 (2022): 247-265.
[4]. Yelne, Ajinkya, and Dipti Theng. "Stock Prediction and analysis Using Supervised Machine Learning Algorithms. "2021 International Conference on Computational Intelligence and Computing Applications (ICCICA). IEEE, 2021.
[5]. 5.Pang, Xiongwen, et al. "An innovative neural network approach for stock market prediction." The Journal of Supercomputing 76 (2020): 2098-2118.roller. IEEE Sponsored 9th Inter. Conf. Intell. Sys. Control (ISCO). 231–243.
[6]. Strader, Troy J., et al. "Machine learning stock market prediction studies: review and research directions." Journal of International Technology and Information Management 28.4 (2020): 63-83.
[7]. Gandhmal, Dattatray P., and Kannan Kumar. "Systematic analysis and review of stock market prediction techniques." Computer Science Review 34 (2019): 100190.
[8]. Kohli, Pahul Preet Singh, et al. "Stock prediction using machine learning algorithms." Applications of Artificial Intelligence Techniques in Engineering: SIGMA 2018, Volume 1. Springer Singapore, 2019.
[9]. Reddy, V. Kranthi Sai, and Kranthi Sai. "Stock market prediction using machine learning." International Research Journal of Engineering and Technology (IRJET) 5.10 (2018): 1033-1035.
[10]. Khare, Kaustubh, et al. "Short term stock price prediction using deep learning." 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT). IEEE, 2017.
[11]. Usmani, Mehak, et al. "Stock market prediction using machine learning techniques." 2016 3rd international conference on computer and information sciences (ICCOINS). IEEE, 2016.
[12]. Gupta, Aditya, and Bhuwan Dhingra. "Stock market prediction using hidden markov models." 2012 Students Conference on Engineering and Systems. IEEE, 2012.
[13]. Vui, Chang Sim, et al. "A review of stock market prediction with Artificial neural network (ANN)." 2013 IEEE international conference on control system, computing and engineering. IEEE, 2013.
[14]. Shen, Shunrong, Haomiao Jiang, and Tongda Zhang. "Stock market forecasting using machine learning algorithms." Department of Electrical Engineering, Stanford University, Stanford, CA (2012): 1-5
[15]. Egeli, Birgul, Meltem Ozturan, and Bertan Badur. "Stock market prediction using artificial neural networks." Decision Support Systems 22 (2003): 171-185.

