Journal of Applied Information Science
14 (1) 2026, 01-06
www.publishingindia.com/jais
https://doi.org/10.21863/jais/2026.14.1.001

Submitted: 16 May 2025
Accepted: 09 June 2025

An Empirical Study on Machine Learning Approaches for Indian Stock Market Forecasting

Nirav Shukla[1*] and Vishal Dahiya[2]

[1]Research Scholar & Assistant Professor, CPICA, SVG University, Ahmedabad, Gujarat, India.
Email: niravshukla@cpi.edu.in
[2]Professor, Department of MCA, CPICA, SVG University, Ahmedabad, Gujarat, India.
Email: director-i-mca@cpi.edu.in
[*]Corresponding Author

Abstract: With the mixing of machine learning (ML) algorithms, stock market forecasting has become an area of interest for researchers. This paper presents a review of the literature on ML techniques for price prediction in Indian stock markets. It analyses major algorithms such as linear regression, support vector machines (SVM), random forest, and neural network, as well as discussing their theoretical, advantageous and disadvantageous aspects. This research offers an exhaustive assessment of the implementation of diverse ML techniques for predicting trends in the Indian stock market. The swift changes and the nonlinear nature of stock markets make the problem of prediction extremely difficult. The author demonstrates how ML techniques are used to improve the accuracy of predictions by analysing historical data, sentiments of the market, and various technical indicators. Furthermore, it discusses issues such as crash in data, unpredictability of the market, and overfitting that create challenges for forecasting in ML. This review is aimed at providing substantial evidence in the implementation of various ML techniques adjusted to the Indian stock market for predicting stock price movements and providing information where further research is needed.

Keywords: Artificial intelligence in finance, Indian stock market, Machine learning, Stock market forecasting, Stock price prediction.

I. Introduction

One of the most studied financial activities is the stock market, which is supposed to simplify the difficult decision-making process for brokers, investors, and even financial analysts. Stock market is one of the most researched financial activities with the hope that it will make it easier for investors, brokers, or even financial analysts to ease their complex decision making. The Indian stock market's dynamic nature and price volatility present a significant challenge for forecasters. Old-style predicting methods, such as basic analysis and technical analysis, face limitations due to their dependency on linear models, which cannot accurately represent the stock price movement's complex and nonlinear dynamics. This is why there is a decrease in using traditional methods and there is a rise in using machine learning (ML) techniques, which are effective in using huge historical data to identify patterns and trends to enhance accuracy in forecasting.

ML models, such as linear regression, support vector machines (SVM), random forest, and neural networks, have been generally used in stock market forecasting. These models leverage time series analysis, sentiment analysis, and deep learning techniques to enhance predictive performance. The Indian stock market, represented by the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE), provides a rich dataset for exploring these ML-driven forecasting methods.

This paper's goal is to conceptually examine the various ML techniques used to forecast the Indian stock market. It analyses the underlying principles, benefits, drawbacks, and issues of these methods. The research also notes new developments such as the application of natural language processing (NLP) for sentiment analysis and the adoption of hybrid ML models for increased accuracy.

More precisely, this study's main goal is to examine how different ML algorithms are used to predict the Indian stock market. The study aims to relate the act of models – including linear regression, SVM, random forest, and neural networks – based on their accuracy, tendencies towards overfitting, and adaptability to market fluctuations. It will also assess each ML model's benefits and drawbacks in relation to stock price prediction. The research also explores the probability for integrating sentiment analysis and technical indicators to enhance forecasting accuracy. A further goal is to identify emerging opportunities and ongoing challenges in applying ML to India's evolving financial ecosystem. To achieve these objectives, the study employs a literature review and comparative analysis methodology.

Gathering pertinent research articles that highlight the application of ML algorithms for stock prediction in Indian markets – specifically the NSE and the BSE – is part of the process. A comparative analysis will be conducted on techniques including linear regression, SVM, random forest, convolutional neural networks (CNNs), long short-term memory (LSTM), and hybrid models. This analysis will use performance metrics such as forecast accuracy, mean absolute percentage error (MAPE), and generalisation capability. Emphasis will be placed on empirical studies that utilise real-world stock data, technical indicators, and sentiment features. This qualitative synthesis aims to consolidate current knowledge and highlight the best ML techniques for stock market forecasting.

II. Limitation and Challenges

- *Data Noise:* The data we use to make predictions can be jumbled or random, which can confuse the model and make its predictions less accurate.
- *Market Unpredictability:* The market is always changing and it is hard to predict what will happen next. This means even the best models cannot always give accurate predictions when things are unreliable.
- *Overfitting:* This happens when a model is too focused on the data it was trained on. It might do great with that data but struggle to make good predictions on new and unseen data because it is too sleek to the training data.
- *Data Quality Matters:* For more accurate predictions, it is important to use clean and consistent data. Inaccurate results may arise from any interference or chance mistakes in the data.
- *Market Fluctuations:* As we all know that the market is unpredictable, even the best models might not always deliver perfect outcomes. Stakeholders should be prepared for some level of uncertainty and adjust strategies accordingly.
- *Balancing Training and Reality:* It should be possible for models to generalise to new scenarios even though they can be quite accurate given the data they are trained on. Avoiding overfitting is crucial because models that work well on training data but poorly in real-world situations might result in inadequate decision making.

III. Stock Market Forecasting Workflow Using ML

- *Choose the Market and the Time Frame:* Specify the prediction period (daily, weekly, or monthly) and the stock market (NSE, BSE, and so on).
- *Features Choice:* Choose the relevant features that will be used for prediction, such as volume, technical indicators, open price, and close price.
- *Collecting Data:* Collect past stock market data from reputable application programming interfaces (APIs) or sources.
- *Preprocessing:* Clean the data (handle missing values, outliers), normalise it, and prepare it for training.

- *Feeding the Model:* Train an ML model (for example, LSTM, random forest, XGBoost) using the pre-processed data with algorithms.
- *Error Near to 0?* Evaluate the model performance. If the prediction error is not effectively low, return to the training step to fine-tune.
- *Use the Forecasting Results:* Once the error is acceptable, deploy the model to make predictions and inform investment decisions.

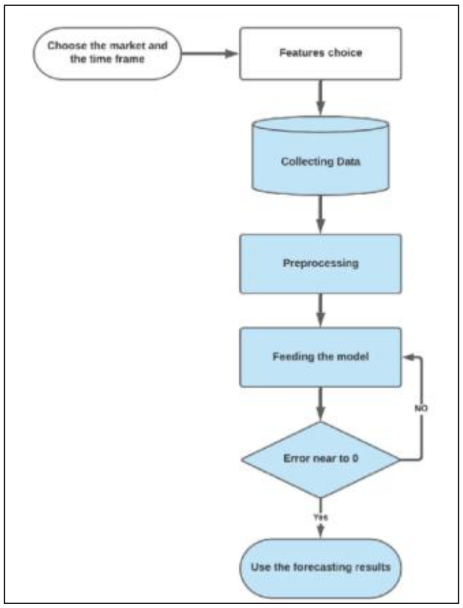

Fig. 1: Stock Market Forecasting Workflow Using ML

IV. Related Work

[1] P. Prasad *et al.* – "Evaluates the efficiency of algorithms for machine learning – XGBoost, LSTM, SVM, and random forest – for stock price prediction in the Indian share market. Using 12,600 daily data points over five years, models were trained with technical indicators and optimised through hyperparameter tuning. Among them, XGBoost achieved the highest accuracy (86.5%), followed by LSTM (82.3%), random forest (85.1%), and SVM (79.6%). Key features influencing predictions were open and close prices. The findings highlight the potential of ML, particularly XGBoost and LSTM, in improving stock market forecasting and informing investment strategies, with opportunity for upcoming research on advanced techniques," *International Journal of Engineering and Advanced Technology*, vol. 5, no. 6, 2024.

[2] V. R. Surjuse *et al.* – "Explore the challenge of predicting stock prices, focusing on the Indian stock market's NIFTY50 index. Using Python and linear regression, the authors developed a stock prediction website. They employed a genetic algorithm to extract features from the NIFTY50 lag index and used linear regression to find correlations with stock trends. Testing on three stocks showed an 82.55% accuracy, demonstrating the effectiveness of their approach compared with existing forecasting methods for daily stock price changes," *Journal of Electrical Systems*, vol. 20, no. 2s, pp. 614–624, 2024.

[3] A. M. Priyatno, W. F. R. Sudirman and R. J. Musridho – "Enhances stock price prediction by improving the recursive feature elimination (RFE) method used for feature selection. Traditional RFE may not always select the most impactful features, so the study proposes combining important features with nonparametric correlation indicators. Using technical indicators and stock price history, various weighting strategies – such as average, 25:75%, and 75:25% – were tested. The proposed method was applied to predict PT Bank Rakyat Indonesia Tbk (BBRI) stock prices and achieved low error rates, including a 1.78% mean absolute percentage error. Results show this modified RFE outperforms standard RFE in predictive accuracy for stock price forecasting," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, pp. 1906–1915, 2024.

[4] B. P. Ghosh *et al.* – "Explore this study and evaluate four deep learning models – Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), LSTM, and CNN – for stock price prediction using NSE and New York Stock Exchange (NYSE) data. Trained on NSE data, the CNN model outperforms others, effectively predicting NYSE prices. Compared with autoregressive integrated moving average (ARIMA), neural networks show superior performance, highlighting their potential and revealing shared patterns across different

stock markets," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, pp. 68–75, 2024.

[5] A. Ghosh, S. Bose, G. Maji, N. Debnath and S. Sen – "Explore this research that studies the complexity of stock market forecast, influenced by various rational and irrational factors. Built on the efficient market theory, it emphasises that historical stock prices reflect all market events and can be used to forecast future trends. The study proposes a framework using ML, specifically the LSTM model, along with a company net growth calculation algorithm. By analysing historical data, the approach aims to uncover hidden patterns and improve prediction accuracy, enabling more reliable forecasts of a company's future performance," 2024.

[6] J. Singh and G. Singh – "Explores stock price forecast via deep learning, focusing on the Indian stock market. Traditional models such as ARIMA struggle with nonlinear patterns and long-term dependencies, prompting a comparison between CNN and hybrid CNN-LSTM models. While CNNs are effective at identifying local patterns, CNN-LSTM models better capture long-term trends. Using historical data from NIFTY50 constituents, the study evaluates predictive accuracy, efficiency, and adaptability. Results show CNNs work well for short-term forecasting, but CNN-LSTM replicas are more robust for medium- to long-term predictions in dynamic market conditions," *Journal of Management World*, vol. 2024, no. 5, pp. 217–237, 2024.

[7] X. Chen *et al.* – "Explore bibliometric analysis of 223 articles on sentiment analysis in stock market research from 2010 to 2022, sourced from the Web of Science database. It identifies key affiliations, countries, publication sources, and subject areas, along with top-cited papers and collaboration networks. Results show computer science journals dominate the field, with broad global participation and stronger intra-regional collaboration. Keyword mapping reveals major research themes, including deep learning in prediction, financial news sentiment, investor sentiment effects, and microblog sentiment analysis. The findings offer valuable insights for scholars and practitioners exploring sentiment analysis in financial markets," *Natural Language Processing Journal*, vol. 10, p. 100125, 2025.

[8] J. M. Sangeetha and K. J. Alfia – "Focuses on forecasting stock prices using an Evaluated linear regression-based machine learning (ELR-ML) approach. Aimed at predicting the S&P 500 index, the model leverages factors such as open, close, low, high, and volume data. By addressing the nonlinear and discontinuous nature of stock market influences, the research highlights the importance of carefully selecting global financial data to enhance prediction accuracy," *Measurement: Sensors*, vol. 31, p. 100950, 2024.

[9] N. Ayyildiz and O. Iskenderoglu – "Explore the study, which compares the performance of many ML algorithms in predicting the directional movements of stock market indices in developed countries. Indices analysed include NYSE 100, Nikkei Stock Average (Nikkei 225), Financial Times Stock Exchange (FTSE) 100, CAC 40, DAX 30, FTSE MIB, and Toronto Stock Exchange (TSX). Algorithms tested include decision tree, random forest, k-nearest neighbour, naive Bayes, logistic regression, support vector machines, and artificial neural networks (ANN). Outcomes show that ANN achieved best for most indices, while logistic regression excelled for others. Overall, ANN, logistic regression, and SVM achieved over 70% accuracy across all indices," *Heliyon*, vol. 10, no. 2, p. e24123, 2024.

[10] H. Wang, Z. Xie, D. K. W. Chiu and K. K. W. Ho – "Explore multimodal approach to forecast stock market variations in China, highlighting the importance of stock market forecasting in the country's economic growth. The LSTM + Transformer model outperformed others in accuracy, F1-score, precision, and recall. Using Granger causality and impulse response tests, the research found causal links between investor sentiment, COVID-19 indicators, and stock trends – particularly in pharmaceutical stocks – offering valuable insights for investors and regulators in strategy formulation and market analysis," *Applied Intelligence*, vol. 55, no. 1, p. 77, 2025.

V. Discussion and Analysis

It is clear from the examined studies that the size of the dataset, the state of the market, and the features employed affect how well ML algorithms perform.

- XGBoost and LSTM have shown superior results in most cases because of their capacity to manage nonlinearity and long-term dependencies.
- SVM and linear regression, while easier to implement, often fall short in capturing complex market behaviour.
- Random forest offers good performance but can be prone to overfitting if not properly tuned.
- Hybrid models (for example, CNN-LSTM) demonstrate robustness in adapting to different market scenarios by leveraging strengths of both architectures.

Moreover, incorporating sentiment analysis and external factors such as economic indicators has improved prediction accuracy, especially in volatile market conditions. However, challenges such as data quality, overfitting, and market unpredictability persist.

A tabulated comparison of model performance from literature is shown in Table I.

Table I : Performance Comparison of Machine Learning Models for Time Series Prediction

Model	Accuracy (%)	Strengths	Weaknesses
XGBoost	86.5	High accuracy, fast training	Sensitive to outliers
LSTM	82.3	Long-term memory, sequence-aware	Computationally intensive
Random forest	85.1	Robust to noise, less overfitting	Limited interpretability
SVM	79.6	Good for classification tasks	Poor performance with large datasets
CNN-LSTM hybrid	84–88	Captures both local and sequential patterns	Complex architecture

This analysis reinforces the relevance of model selection built on the specific forecasting goal – short-term trends vs long-term movements.

VI. Conclusions

Every prediction-making model has advantages and disadvantages. In this instance, maintaining data accuracy while taking the market's uncertainty into account helps produce risk calculations that are accurate. It is also critical to avoid overfitting. Not only historical data, but also real-world application must support the model's validity. Everyone needs to take the initiative to solve these problems and build accurate models that enable well-informed choices. To make well-informed, trustworthy, and logical judgements, stakeholders should continue to be aware of these difficulties, modify their approaches appropriately, and improve their models on a regular basis.

References

[1] N. Ayyildiz, and O. Iskenderoglu, "How effective is machine learning in stock market predictions?," *Heliyon*, vol. 10, no. 2, p. e24123, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e24123.

[2] B. P. Ghosh, M. S. Bhuiyan, D. Das, T. N. Nguyen, M. Jewel, M. T. Mia, and D. M. Cao, "Deep learning in stock market forecasting: Comparative analysis of neural network architectures across NSE and NYSE," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, pp. 68–75, 2024, doi: https://doi.org/10.32996/jcsts.2024.6.1.8.

[3] X. Chen, H. Xie, Z. Li, H. Zhang, X. Tao, and F. L. Wang, "Sentiment analysis for stock market research: A bibliometric study," *Natural Language Processing Journal*, vol. 10, p. 100125, 2025, doi: https://doi.org/10.1016/j.nlp.2025.100125.

[4] A. Ghosh, S. Bose, G. Maji, N. Debnath, and S. Sen, "Stock price prediction using LSTM on Indian share market," pp. 101–190, n.d. doi: https://doi.org/10.29007/qgcz.

[5] P. Prasad, and M. Mishra, "Analysis of stock price predictions and forecasting using machine learning," *International Journal of Engineering and Advanced Technology*, vol. 5, no. 6, 2024.

[6] A. M. Priyatno, W. F. R. Sudirman, and R. J. Musridho, “Feature selection using nonparametric correlations and important features on recursive feature elimination for stock price prediction,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, p. 1906, 2024, doi: https://doi.org/10.11591/ijece.v14i2.pp1906-1915.

[7] J. M. Sangeetha, and K. J. Alfia, “Financial stock market forecast using evaluated linear regression-based machine learning technique,” *Measurement: Sensors*, vol. 31, p. 100950, 2024, doi: https://doi.org/10.1016/j.measen.2023.100950.

[8] J. Singh, and G. Singh, “Deep learning for financial forecasting: Evaluating CNN and CNN-LSTM in Indian stock market prediction,” *Journal of Management World*, vol. 2024, no. 5, pp. 217–237, 2024, doi: https://doi.org/10.53935/jomw.v2024i4.1070.

[9] V. R. Surjuse, “Stock price prediction and stock behaviour using GA and LR with NIFTY50 index based on daily charts,” *Journal of Electrical Systems*, vol. 20, no. 2, pp. 614–624, 2024, doi: https://doi.org/10.52783/jes.1528.

[10] H. Wang, Z. Xie, D. K. W. Chiu, and K. K. W. Ho, “Multimodal market information fusion for stock price trend prediction in the pharmaceutical sector,” *Applied Intelligence*, vol. 55, no. 1, p. 77, 2025, doi: https://doi.org/10.1007/s10489-024-05894-0.