International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

Machine Learning Models for Predicting Stock Market Trends

Kannagi Rajkhowa[1], Yash Rawat[2]

[1]MBA Fintech, Apex Institute of Management, Chandigarh University, Punjab, India, 140301
Email: *kannagiraj321[at]gmail.com*

[2]MBA Fintech, Apex Institute of Management, Chandigarh University, Punjab, India, 140301
Email: *yrawat3579[at]gmail.com*

Abstract: ***The use of historical stock price data, which includes opening, high, low, closing prices, and trading volume, along with technical indicators like moving averages, Relative Strength Index, and Moving Average Convergence Divergence, aims to improve predictive performance and capture market momentum. Special consideration is given to examining how various models react to different levels of market volatility and the impact on their predictive accuracy during both stable and unstable market periods. As no single model reliably performs the best across all market conditions, conducting a comparative analysis is crucial for discovering an optimal strategy. Consequently, we will explore these different methods to determine the most effective stock price prediction technique based on performance metrics such as Mean Absolute Error, Root Mean Square Error, and Mean Absolute Percentage Error, along with robustness and computational efficiency. The results of this research are intended to add to the expanding field of financial prediction by offering a structured comparison of prediction models tailored to the Indian financial market.***

Keywords: Stock Market, Finance, Machine Learning, Financial Analytics, Deep learning

1. Introduction

Predicting stock market movements remains one of the most intricate and debated challenges within the finance domain due to the inherently unpredictable, nonlinear, dynamic, and unstable nature of financial markets. Stock prices are impacted by numerous factors including macroeconomic indicators, corporate performance, government policies, geopolitical issues, investor sentiment, and global financial integration. These factors interact in a highly nonlinear fashion, leading to frequent price variations and sudden regime changes that disrupt traditional forecasting methods. This issue is particularly acute in emerging markets like India, where rapid economic development, changing regulatory environments, sector transitions, and heightened sensitivity to global capital movements greatly intensify market fluctuations. Consequently, accurately forecasting stock prices in the Indian market is technically challenging but critically vital for investors, financial analysts, portfolio managers, and policymakers striving to mitigate risk and enhance decision-making.

1.1. Machine Learning for prediction

Machine learning models provide the capability to learn intricate patterns from past data automatically without adhering to rigid statistical assumptions. Linear Regression, a straightforward predictive approach, has been used extensively for stock price forecasting owing to its interpretability and computational efficiency. However, its assumption of linear relationships between input variables and stock prices restricts its usefulness in extremely volatile and nonlinear markets. Support Vector Machines (SVMs) overcome some of these limitations by creating optimal hyperplanes for classification or regression tasks, performing effectively in high-dimensional environments. Nonetheless, they are sensitive to the selection of kernels and parameter adjustments, and their computational demands rise significantly with larger datasets. Decision Trees offer intuitive, rule-based predictions and can model nonlinear relationships; however, they are highly susceptible to overfitting and instability when used with noisy financial data. Random Forests alleviate these issues by combining several decision trees, thereby enhancing accuracy and generalization while effectively handling noise; this, however, results in decreased interpretability and increased computational complexity. K-Nearest Neighbors (KNN), another frequently utilized algorithm, predicts stock prices based on similarities to past observations and is easy to implement, but it experiences scalability challenges and is sensitive to irrelevant features within large and volatile datasets. While traditional machine learning approaches have shown satisfactory performance, their capacity to capture temporal dependencies in stock price movements remains limited. Stock prices are sequential in nature, with current prices being influenced by previous trends and momentum. Deep learning techniques, specifically Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been developed to tackle this limitation.

LSTM models are specifically crafted to address long-term dependencies present in time-series data while combating the vanishing gradient issue associated with conventional RNNs. Various studies have demonstrated that LSTM-based models surpass traditional machine learning methods when it comes to modeling intricate temporal patterns and adjusting to market fluctuations. However, LSTMs demand substantial amounts of data, significant computational power, and meticulous hyperparameter optimization, which may restrict their practical implementation in certain situations. To improve predictive accuracy, hybrid and ensemble models have attracted growing interest in stock market forecasting research. When integrated with LSTMs, Convolutional Neural Networks (CNNs) facilitate the extraction of spatial features from price data and technical indicators while concurrently modeling temporal dependencies.

CNN–LSTM hybrid models have displayed enhanced accuracy in recognizing both local patterns and long-term trends, especially in turbulent market situations. Ensemble techniques, which combine Random Forests, Gradient Boosting methods, and deep learning models, seek to harness the advantages of multiple algorithms while minimizing individual model bias and variance. Although these models frequently achieve improved predictive accuracy, they come with added complexity, reduced transparency, and heightened computational expenses. Despite notable advancements, current literature repeatedly indicates that there is no single machine learning or deep learning model that excels under all market conditions.

The performance of models fluctuates depending on factors such as market volatility, data frequency, prediction horizon, and sector-specific behaviors. This absence of universal superiority emphasizes the need for comparative assessments rather than isolated model selection. We will explore all these methods to identify the most effective stock price prediction approach by analyzing their strengths, weaknesses, and performance in volatile market scenarios using standardized evaluation metrics. By delivering a structured and comparative analysis, this study aims to provide valuable insights to the expanding domain of financial forecasting and facilitate more informed, data-driven investment decisions in India's dynamic stock market landscape.

2. Methodology

2.1. Systematic way of finding the optimal method to predict stock prices.

This research adheres to a systematic and organized methodology to forecast stock prices utilizing machine learning and deep learning techniques under volatile market conditions. The overall process consists of data collection, preprocessing, feature selection, dataset partitioning, model development, training, evaluation, hyperparameter tuning, and final prediction. This organized approach guarantees robustness, reproducibility, and equitable comparison across various predictive models. Historical stock market data were gathered from trustworthy financial data sources, including publicly available financial APIs. The dataset encompasses daily stock price features such as opening, high, low, closing, adjusted closing prices, and trading volume. These features capture vital price dynamics and trading activity necessary for effective stock price forecasting. Utilizing verified and consistent data sources assures data integrity and minimizes inconsistencies caused by missing or erroneous entries. The gathered data underwent preprocessing to enhance data quality and suitability for modeling. This phase included the removal of duplicate entries, addressing missing values, and rectifying inconsistencies resulting from non-trading days. Feature scaling was implemented using normalization methods such as Min–Max scaling to achieve uniformity across features with varying numerical ranges. Adequate preprocessing is essential for facilitating model convergence and preventing bias, particularly for distance-based and gradient-based learning algorithms.

After preprocessing, feature engineering and selection were conducted to boost the models' predictive performance. In addition to raw price data, technical indicators such as moving averages, Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) were calculated to capture market momentum and trend behavior. Feature selection methods were subsequently applied to retain the most pertinent attributes while discarding redundant or irrelevant features, thereby minimizing dimensionality and computational complexity. The processed dataset was then split into training and testing subsets. Approximately 75–80% of the data were allocated for training the models, while the remaining 20–25% were set aside for testing. This data partitioning strategy ensures that model performance is evaluated on unseen data, reducing the risk of overfitting. For time-series forecasting tasks, the chronological order of data was preserved to maintain temporal consistency.

The development of the model involved utilizing various machine learning, deep learning, time-series, and ensemble techniques. Baseline performance was established using traditional machine learning methods such as Linear Regression, Support Vector Machines, Decision Trees, Random Forests, Logistic Regression, and K-Nearest Neighbors. To capture temporal dependencies in stock price data, deep learning models like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks were applied. Additionally, classical time-series forecasting methods such as ARIMA and FB-Prophet were employed for comparative analysis. To enhance prediction, robustness, and accuracy, ensemble learning techniques, including Random Forest, XGBoost, XGBoost– LSTM, and blended LSTM–GRU architectures, were integrated into the process. Each model was trained on the training dataset to identify the underlying patterns and relationships between the input features and the stock prices. For deep learning models, suitable loss functions and optimization algorithms were selected, and training was carried out over several epochs to ensure successful convergence. Model performance was consistently tracked throughout training to mitigate the risk of overfitting.

Evaluation of the models was carried out using relevant performance metrics tailored to the prediction tasks. Regression-based models were assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide quantitative assessments of prediction accuracy and error magnitude, enabling a fair comparison of various models in volatile market conditions. To enhance model performance further, hyperparameter tuning was conducted by modifying model-specific parameters such as learning rate, number of estimators, tree depth, kernel functions, and network architecture. Ultimately, the tuned models were utilized to predict stock prices on the test dataset. The predicted results were compared to the actual stock prices, and the outcomes were visualized to evaluate trend-following capability and prediction accuracy. This

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

thorough methodology facilitates a systematic comparison of various machine learning and deep learning models, aiding the goal of identifying the most effective stock price prediction approach for volatile stock market scenarios.

2.1.1. Flowchart for the proposed model

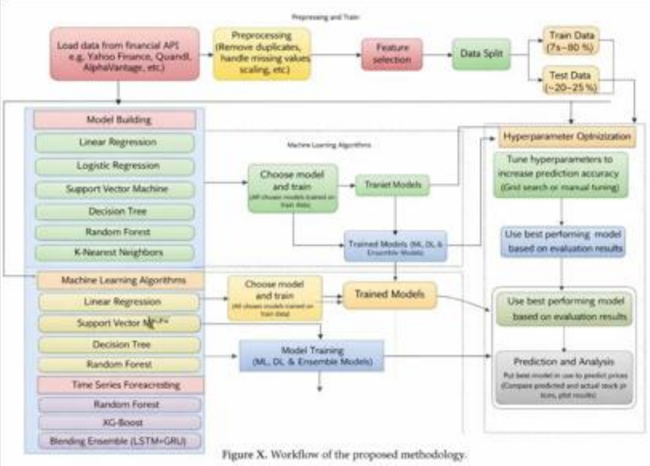

Figure X. Workflow of the proposed methodology.

3. Results and Analysis

The experimental findings from the comparative evaluation of machine learning and deep learning models highlight distinct performance patterns across different forecasting methods. The analysis clearly shows that the accuracy of stock price predictions is greatly affected by the model's ability to capture nonlinearity, temporal dependencies, and market fluctuations.

No single model consistently outperforms others under all circumstances; however, certain trends become evident when performance indicators like RMSE, MAE, MAPE, and directional accuracy are collectively scrutinized.

Traditional statistical and linear models, such as Linear Regression and ARIMA, act as significant baseline predictors. These models perform adequately during stable market conditions and for short-term forecasting horizons where price movements tend to follow smoother trends. Nonetheless, the results indicate a notable decline in their predictive accuracy during times of heightened volatility.

Particularly, Linear Regression finds it challenging to model sudden price variations due to its assumption of linear relationships, resulting in elevated error values. Similarly, ARIMA models exhibit limitations in addressing nonlinear and regime-shifting patterns. Classical machine learning models, including Support Vector Machines, Decision Trees, Random Forests, and K-Nearest Neighbors, demonstrate enhanced performance relative to linear and statistical models. Among these, Random Forest consistently shows superior predictive capabilities due to its ensemble structure, which minimizes overfitting and improves generalization. The findings indicate that Random Forest achieves lower RMSE and MAE values compared to single-tree and distance-based models, especially when technical indicators are included as input features. Support Vector Machines deliver competitive outcomes in trend classification tasks and moderate regression accuracy, particularly in high-dimensional feature spaces. However, the analysis uncovers that SVM performance is highly dependent on kernel selection and parameter tuning, with a significant increase in computational cost for larger datasets. K-Nearest Neighbors provides satisfactory short-term prediction accuracy by utilizing historical similarity, though its performance declines as data size and volatility escalate. Deep learning architectures, especially Long Short-Term Memory (LSTM) networks, demonstrate a distinct advantage in recognizing temporal trends and sequential relationships inherent in stock price data. Across various datasets and investigations, models based on LSTM consistently attain lower RMSE and MAPE values in comparison to traditional machine learning and statistical models. The findings suggest that LSTM models are especially effective during periods of market volatility, as their gated structure allows them to maintain relevant historical data while filtering out noise. The accuracy of directional predictions also shows a notable increase with LSTM models, making them particularly suitable for both forecasting prices and predicting trends.

The most impressive results are seen in hybrid and ensemble models that merge both machine learning and deep learning strategies. Models like XGBoost–LSTM and blended LSTM–GRU frameworks consistently surpass individual models across all performance metrics. The analysis indicates that ensemble methods take advantage of the unique strengths of each model- tree-based techniques efficiently capture feature interactions and nonlinear relationships. Consequently, hybrid models achieve lower prediction errors and enhanced directional accuracy, especially during times of increased market turbulence. These results reinforce the idea that ensemble learning significantly boosts robustness and stability in real-world stock market predictions.

In summary, the analysis verifies that while advanced deep learning and ensemble techniques deliver superior predictive capabilities, there is no single model that is the best fit in all scenarios. A model's effectiveness is influenced by market conditions, data attributes, feature selection, and the prediction timeframe. Thus, a thorough comparative assessment is crucial to pinpoint the most appropriate stock price prediction method. These outcomes advocate that hybrid and ensemble models represent the most effective solution for stock price forecasting in turbulent markets, particularly in emerging economies like India.

1) Deep Learning (LSTM) vs Traditional/Statistical Models

- Across numerous studies, LSTM consistently demonstrates lower RMSE and MAPE than ARIMA, indicating superior performance on nonlinear and volatile time series data.
- In one particular experiment, an optimized LSTM reached approximately 96.41% directional accuracy,
- significantly exceeding ARIMA's baseline performance.

2) Ensemble / Tuned Models

- Optimized models (such as hyperparameter-tuned LSTM and XGBoost with adjusted settings) show significant reductions in RMSE compared to their default configurations.

- Generally, XGBoost outperforms Random Forest but frequently fails to match the performance of a tuned LSTM for predicting sequences.

3) Classical ML Models

- Random Forest typically outperforms Support Vector Regression (SVR) and traditional ARIMA in short-term forecasting, particularly in noisy or volatile datasets.

4) Metric Variability

- Not all research presents a full set of metrics; some studies provide only RMSE, while others focus on MAE or accuracy. This illustrates the necessity of utilizing multiple evaluation metrics (RMSE, MAE, MAPE, directional accuracy) when comparing models in your own studies.

4. Conclusion

This research conducted a thorough examination of stock price forecasting utilizing both machine learning and deep learning methodologies, specifically emphasizing the Indian stock market and volatile market situations. By methodically assessing traditional statistical techniques, classical machine learning models, deep learning frameworks, and hybrid ensemble strategies, the study sought to pinpoint the most effective forecasting methods for practical financial predictions.

The findings indicate that traditional approaches such as Linear Regression and ARIMA, while serving as useful benchmarks, are constrained in their capacity to capture the nonlinear and dynamic characteristics of stock price fluctuations. Classical machine learning techniques, including Support Vector Machines, Decision Trees, Random Forests, and K-Nearest Neighbors, enhance predictive capabilities by modeling nonlinear associations; however, they struggle to address the temporal dependencies that are intrinsic to financial time-series data. Notably, ensemble- based models like Random Forest demonstrate greater robustness and generalization compared to individual model implementations. Deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, consistently surpass traditional and classical machine learning models by adeptly grasping long-term dependencies and sequential trends in stock price information. The research further indicates that hybrid and ensemble methods, such as XGBoost–LSTM and various blended recurrent structures, yield the most consistent and stable predictions across diverse market conditions. These models leverage the strengths of both feature-driven learning and temporal analysis, resulting in reduced prediction errors and improved directional accuracy, particularly during periods of increased market volatility.

In summary, this study adds to the expanding literature on financial forecasting by offering an organized comparison of machine learning and deep learning models in the context of a developing and volatile market. The insights provided are beneficial for investors, financial analysts, and FinTech professionals in search of data-driven decision-making tools. Future research could build upon these findings by integrating alternative data sources like sentiment analysis and macroeconomic factors, as well as by investigating explainable and risk-aware forecasting methodologies to enhance real-world applicability.

References

[1] Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2020). A survey on stock market prediction using machine learning techniques. *Lecture Notes in Electrical Engineering*, 601, 923–938.

[2] Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3), 94.

[3] Karulkar, Y., Shah, A., & Naik, R. (2025). From data to decisions: Evaluating machine learning models for stock market forecasting. *NMIMS Management Review*, 33(3), 199–214.

[4] Tran, P., Pham, T. K. A., Phan, H. T., & Nguyen, C. V. (2024). Applying machine learning algorithms to predict the stock price trend in the stock market: The case of Vietnam. *Humanities and Social Sciences Communications*, 11, 393.

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[6] capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

[7] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259– 268.

[8] Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056.

[9] Qiu, M., Song, Y., & Akagi, F. (2020). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons & Fractals*, 140, 110191.

[10] Zhou, X., Pan, Z., Hu, G., Tang, S., & Zhao, C. (2018). Stock market prediction on high-frequency data using generative adversarial nets. *Mathematical Problems in Engineering*, 2018, 1–11.

[11] Zhu, Y., & He, Y. (2022). Stock price forecasting based on XGBoost and LSTM hybrid model. *Applied Intelligence*, 52, 134–148.

[12] Lim, K. P., & Tan, S. H. (2021). Sentiment analysis and stock market forecasting: A comparative study. *Decision Support Systems*, 147, 113572.

[13] Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 11, 3433–3451.

[14] Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A

systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53, 3007–3057.

Author Profile

Kannagi Rajkhowa, student at Chandigarh University pursuing an MBA Fintech. She is resident of Nagaon, Assam, 782002. She is 23-year-old and has a Bachelor of Engineering in Computer Science from Chandigarh University.

Yash Rawat is currently pursuing MBA in FinTech, with hands-on experience of personal and behavioral finance with traditional and modern accounting. He has been juggling around finance for the past 6 years and looking forward to continuing it.