Stock Price Prediction Using Machine Learning

Bhavesh Y. Pardhi[1], Unnati L. Sonwane[2], Khushbu B. Choudhari[3] , Ashwini S. Jambhulkar[4] Akshaykumar K. Rahangdale[5] , Deepak S. Bhiogade[6]

[1]Bhavesh Y. Pardhi Computer Engineering & MPCOE Bhandara
[2]Unnati L. Sonwane Computer Engineering & MPCOE Bhandara
[3]Khushbu B. Choudhari Computer Engineering & MPCOE Bhandara
[4]Ashwini S. Jambhulkar Computer Engineering & MPCOE Bhandara
[5]Akshaykumar K. Rahangdale Computer Engineering & MPCOE Bhandara
[6]Deepak S. Bhiogade Computer Engineering & MPCOE Bhandara

---***---

Abstract - One of the most difficult and well-researched issues in data science and financial engineering is stock market prediction. Economic conditions, political developments, investor emotion, and technical indications are just a few of the many variables that affect financial markets, making them extremely unpredictable and challenging to forecast.

This research offers a thorough application of machine learning techniques for stock price prediction. Data gathering, pre-processing, and feature engineering using a variety of technical indicators are all part of the study. A variety of models are created and assessed, including deep learning architectures like Long Short-Term Memory (LSTM) networks and traditional machine learning techniques.

Additionally, the model's performance is evaluated under genuine trading conditions through the application of hyper parameter tuning and rigorous back testing strategies. Machine learning models are capable of identifying meaningful patterns in historical data and can generate profitable, risk-adjusted trading signals when applied meticulously, despite the fact that perfect prediction is not feasible. The results support this assertion.

***Key Words*:** Stock Market Prediction, Machine Learning, LSTM, Financial Time Series, Technical Indicators, Back testing.

1. INTRODUCTION

Because it facilitates wealth generation, investment, and capital production, the stock market is essential to the world economy. It enables businesses to raise money while providing chances for investors to make money. However, because a variety of factors, including the state of the economy, business performance, investor attitude, and world events, can affect stock values, it is very difficult to forecast them.

Statistical models, technical analysis, and fundamental analysis have historically been used in stock market forecasting. Although these approaches offer valuable insights, they frequently fail to convey the intricate and nonlinear character of financial data.

Machine learning has become a successful method for predicting stock prices thanks to developments in artificial intelligence. Large amounts of historical data may be analyzed by machine learning algorithms, which can also find hidden patterns and produce predictive insights. The goal of this project is to create a machine learning-based stock price prediction system, complete with appropriate validation and back testing to assess actual performance.

2. MOTIVATION AND BACKGROUND

Numerous dynamic factors, such as interest rates, inflation, corporate profitability, global demand-supply imbalances, geopolitical events, and investor attitude, have an impact on financial markets. It is very difficult to estimate stock prices accurately since these elements interact in complicated and nonlinear ways.

The rapid expansion of historical financial data availability and the development of machine learning methods that can handle high-dimensional, nonlinear, and noisy datasets are the driving forces behind this effort. Machine learning models, in contrast to conventional methods, are able to efficiently extract relationships and hidden patterns from massive amounts of data.

Algorithmic decision-making and automated trading have grown in significance in contemporary financial systems. These systems need reliable risk management techniques in addition to precise prediction models.

Algorithmic decision-making and automated trading have grown in significance in contemporary financial systems. These systems need to be able to function in real time, have strong risk management plans, and have accurate prediction models.

By combining prediction accuracy with thorough financial evaluation, including risk assessment and back testing, this initiative seeks to close the gap between theoretical machine learning models and real-world trading applications.

3. REVIEW OF LITERATURE

Because financial markets are complicated and dynamic, stock price prediction has been extensively explored using a variety of methodologies. These strategies include deep learning techniques, machine learning approaches, and statistical models.

3.1 Econometric and Statistical Models

Early studies mostly employed statistical models like ARIMA, which is helpful for short-term forecasting and captures linear correlations in time-series data. Its efficacy in actual markets is constrained by the assumption of linearity and stationary.

GARCH and its variations (EGARCH, TGARCH) were introduced to model volatility. Although these models are helpful for predicting market volatility, they are unable to completely capture nonlinear patterns and rely on significant assumptions.

Other techniques, including as VAR and exponential smoothing, have also been employed, although under extremely complicated and turbulent market conditions, their effectiveness declines. Due to these constraints, sophisticated machine learning techniques were adopted.

3.2 Initial Methods of Machine Learning

Because machine learning allows models to discover patterns directly from data without making rigid assumptions, it has opened up new possibilities for stock price prediction. Among the first models to be utilized were Artificial Neural Networks (ANNs), which frequently outperformed conventional statistical techniques and were able to capture intricate nonlinear interactions.

Because of their excellent generalization capabilities and efficiency in managing noisy financial data, Support Vector Machines (SVMs) and their version Support Vector Regression (SVR) were also extensively used.

For prediction problems, other methods including decision trees, Naïve Bayes, and k-Nearest Neighbors (k-NN) were investigated. However, these models' capacity to capture temporal dependencies in financial data was limited by their heavy reliance on handcrafted features and sensitivity to feature selection and parameter adjustment.

3.3 Models of Ensemble Learning

Because ensemble learning approaches can increase accuracy by merging numerous models, they have become popular for stock prediction. Because of its resilience, capacity to handle high-dimensional data, and resistance to over fitting, Random Forest (RF), which is based on bagging, is frequently utilized.

By systematically fixing faults, gradient boosting techniques like XGBoost, LightGBM, and CatBoost further improve performance. These models are very efficient, scalable, and frequently perform better in prediction tasks than conventional and simple machine learning models.

Ensemble approaches also offer feature importance analysis, which enhances their interpretability and qualifies them for financial applications.

3.4 Recurrent Neural Networks and Deep Learning

By automatically identifying intricate patterns in data, deep learning has greatly enhanced stock prediction. Because RNNs can capture long-term dependencies, they are frequently employed for time-series forecasting, particularly LSTM and GRU. While GRU provides comparable performance with a simplified design, LSTM models frequently outperform conventional techniques. By fusing temporal learning and feature extraction, hybrid models such as CNN–LSTM improve prediction even further.

3.5 Attention Mechanisms and Transformer Models

Transformer models capture long-range dependencies in data by using self-attention techniques. They frequently beat LSTM models in erratic markets and have the ability to dynamically focus on significant time steps. Large datasets and powerful computers are necessary, though.

3.6 Sentiment analysis and alternative data

Alternative data sources like news, social media, and financial reporting are used in recent studies. Sentiment is extracted and integrated with pricing data using NLP algorithms. Although these hybrid models increase prediction accuracy, they still have issues with data reliability and noise.

3.7 Market Efficiency and Model Validation

When predicting stocks, proper validation is crucial. To prevent data leaking, strategies including walk-forward validation and out-of-sample testing are employed. The Efficient Market Hypothesis (EMH) states that although short-term inefficiencies might occasionally be exploited, markets are difficult to forecast.

3.8 Motivation and Research Gaps

Despite improvements, problems including over fitting, non-stationary and a lack of real-world assessment still exist. Many research overlook financial risk criteria in favor of accuracy. By creating an end-to-end system that integrates prediction, validation, back testing, and risk management for real-world application, this project fills in these gaps.

4. PROPOSED WORK

This project offers a thorough machine learning-based platform for developing trading strategies and predicting stock prices. To capture a variety of market circumstances, the system makes use of historical OHLCV (Open, High, Low, Close, and Volume) data gathered from financial APIs over several years.

The suggested method starts with data pretreatment, which addresses inconsistencies, outliers, and missing values. To enhance data quality and model reliability, adjustments are made for corporate events like stock splits and data standardization.

To improve predictive performance, feature engineering is then used. To identify market trends and risk patterns, key technical indicators like Bollinger Bands, RSI, MACD, and Moving Averages are produced in addition to lagged returns and volatility metrics.

Both a regression task (predicting future prices or returns) and a classification task (predicting price direction) are used to formulate the problem. A variety of machine learning models are used, such as Support Vector Machines, Random Forest, XGBoost, Logistic Regression, and deep learning models like LSTM to capture temporal correlations.

To maintain sequence information, time-aware data splitting approaches are used during model training. Walk-forward validation and rolling window techniques are used to guarantee realistic evaluation. Grid search and random search techniques are used for hyper parameter tuning in order to maximize model performance.

To determine the system's practicality, both machine learning measures (accuracy, precision, recall, F1-score) and financial indicators (Sharpe ratio, drawdown and cumulative return) are used.

Lastly, a trading strategy that incorporates position sizing and risk management strategies like stop-loss is created based on model predictions. According to the experimental findings, ensemble and tree-based models perform better in terms of risk-adjusted returns and prediction accuracy.

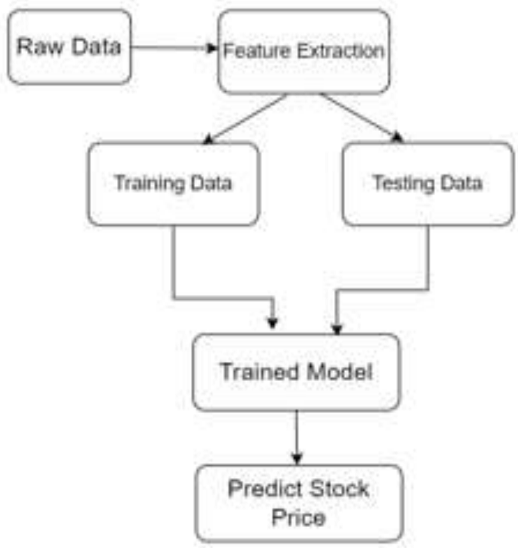
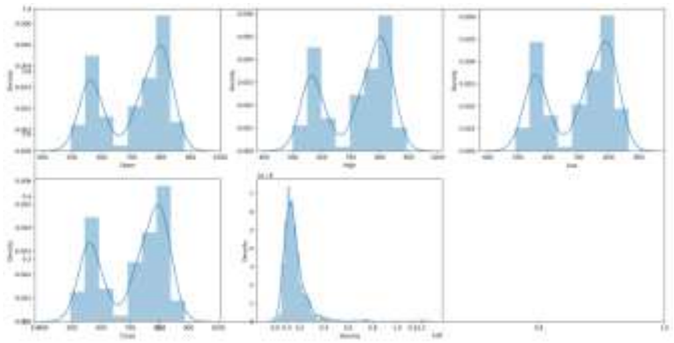


Fig -1: Data Flow Diagram.



International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

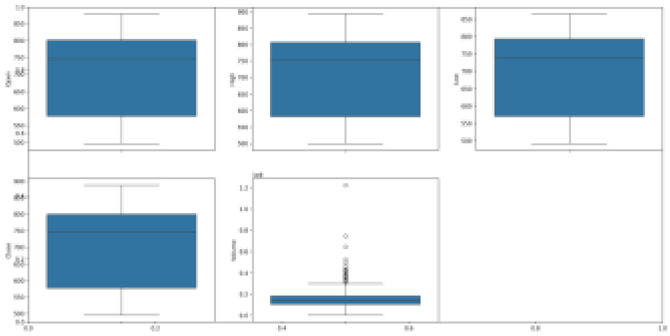


Fig -2: Data Preprocessing Workflow

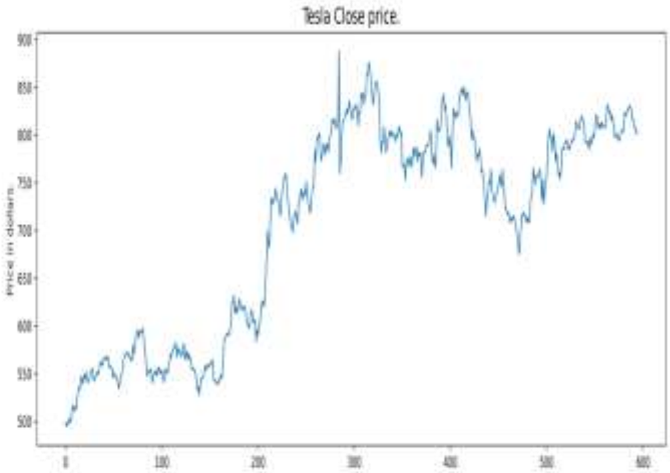


Fig -3: Time Series Validation Techniques.

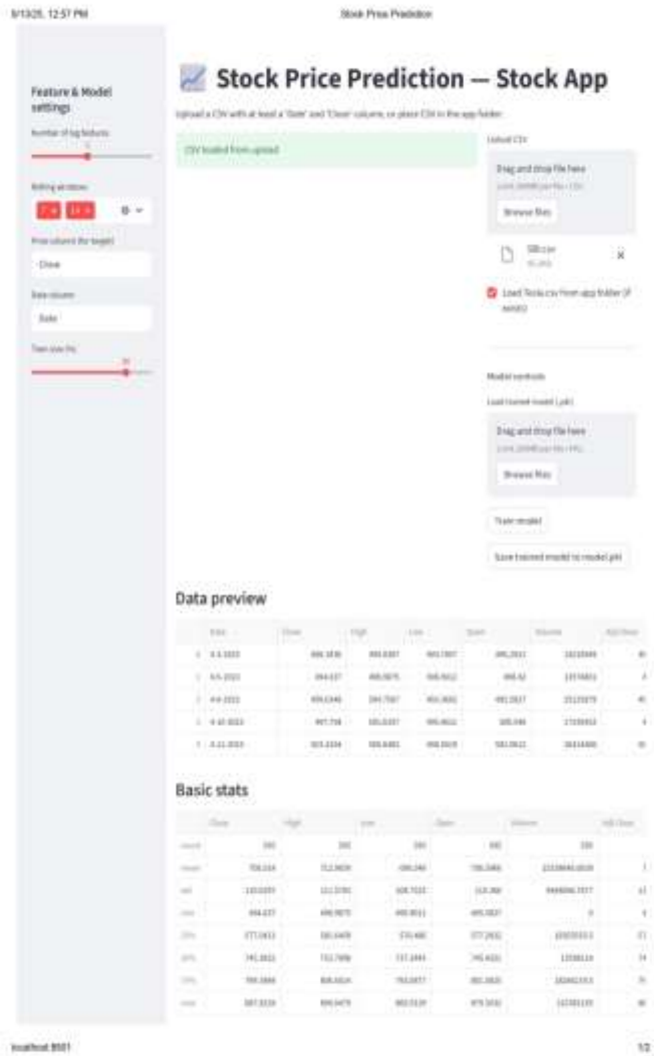


Fig -4: Stock Price prediction Page-1.

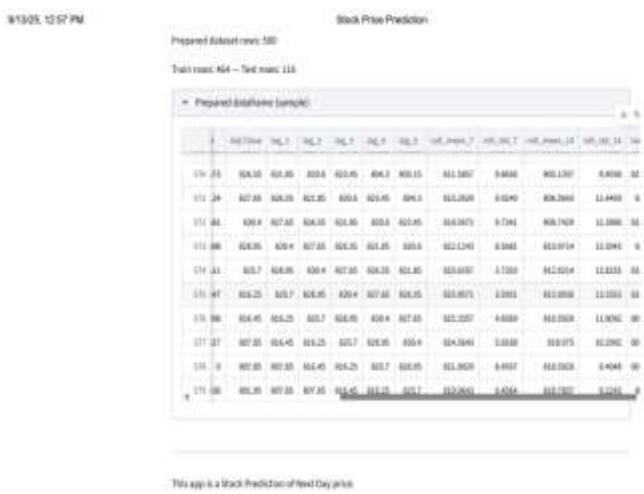


Fig -5: Stock Price prediction Page-2.

5. CONCLUSIONS

This project demonstrates that machine learning techniques can effectively be applied to stock market prediction, offering valuable insights into price trends and market behavior. Although achieving perfect accuracy in stock prediction is inherently impossible due to market volatility and the influence of unpredictable external factors, the results indicate that well-designed ML models can identify meaningful patterns and generate reasonably accurate forecasts.

By integrating data preprocessing, feature engineering (such as technical indicators), and advanced algorithms including ensemble methods and deep learning models, the system enhances predictive performance compared to traditional approaches. Furthermore, the use of proper validation strategies like walk-forward validation ensures realistic evaluation aligned with real-world trading scenarios.

Importantly, this project highlights that profitability in the stock market does not rely solely on prediction accuracy, but also on disciplined risk management, portfolio diversification, and strategic decision-making. Even moderately accurate models, when combined with robust risk control techniques, can support the development of sustainable and profitable trading strategies.

In conclusion, machine learning-based stock prediction systems are not tools for guaranteed success, but they serve as powerful decision-support systems. Future improvements may include incorporating real-time data, sentiment analysis from news and social media, and hybrid models to further enhance prediction capability and robustness.

ACKNOWLEDGEMENT

I want to sincerely thank everyone who helped and advised me while I worked on this machine learning-based stock market prediction project.

First and foremost, I am deeply thankful to my project guide for their valuable guidance, continuous encouragement, and insightful suggestions at every stage of the project. Their knowledge and assistance were essential to the successful development of this work.

Additionally, I would like to express my appreciation to the Computer Engineering Department faculty members for providing the academic expertise and resources needed for this project.

Lastly, I want to express my gratitude to my family for their unwavering support, tolerance, and encouragement, which inspired me to finish this work successfully.

REFERENCES

1. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance.
2. Tsantekidis, A., et al. (2017). Forecasting Stock Prices from the Limit Order Book using Deep Learning. IEEE.
3. Dixon, M. F., Klabjan, D., & Bang, J. H. (2017). Classification-based Financial Markets Prediction using Deep Neural Networks. WCCI.
4. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research.
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD.
6. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning in finance. arXiv.
7. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications.
8. Zhang, Y., Aggarwal, C., & Qi, G. (2018). Stock price prediction via discovering multi-frequency trading patterns. CIKM.
9. Kraaijeveld, O., & De Smedt, J. (2015). A neural network approach to stock price prediction. Expert Systems with Applications.
10. Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. Quantitative Finance.
11. Borovkova, S., & Mahakena, L. (2011). Modeling dependence structure of financial returns using copulas. Journal of Banking & Finance.
12. Lopez de Prado, M. (2018). Advances in Financial Machine Learning. Wiley.
13. Tsay, R. S. (2010). Analysis of Financial Time Series. Wiley.
14. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

15. Khandelwal, A., He, X., & Nielsen, F. (2020). Stock movement prediction using news and tweets. IEEE.
16. Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market.
17. Serrano, J. P., et al. (2020). Machine learning for financial time series forecasting: Practical guide. Journal of Financial Data Science.
18. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.
19. Kearns, M., & Nevmyvaka, Y. (2013). Machine learning for market microstructure and high frequency trading. arXiv.