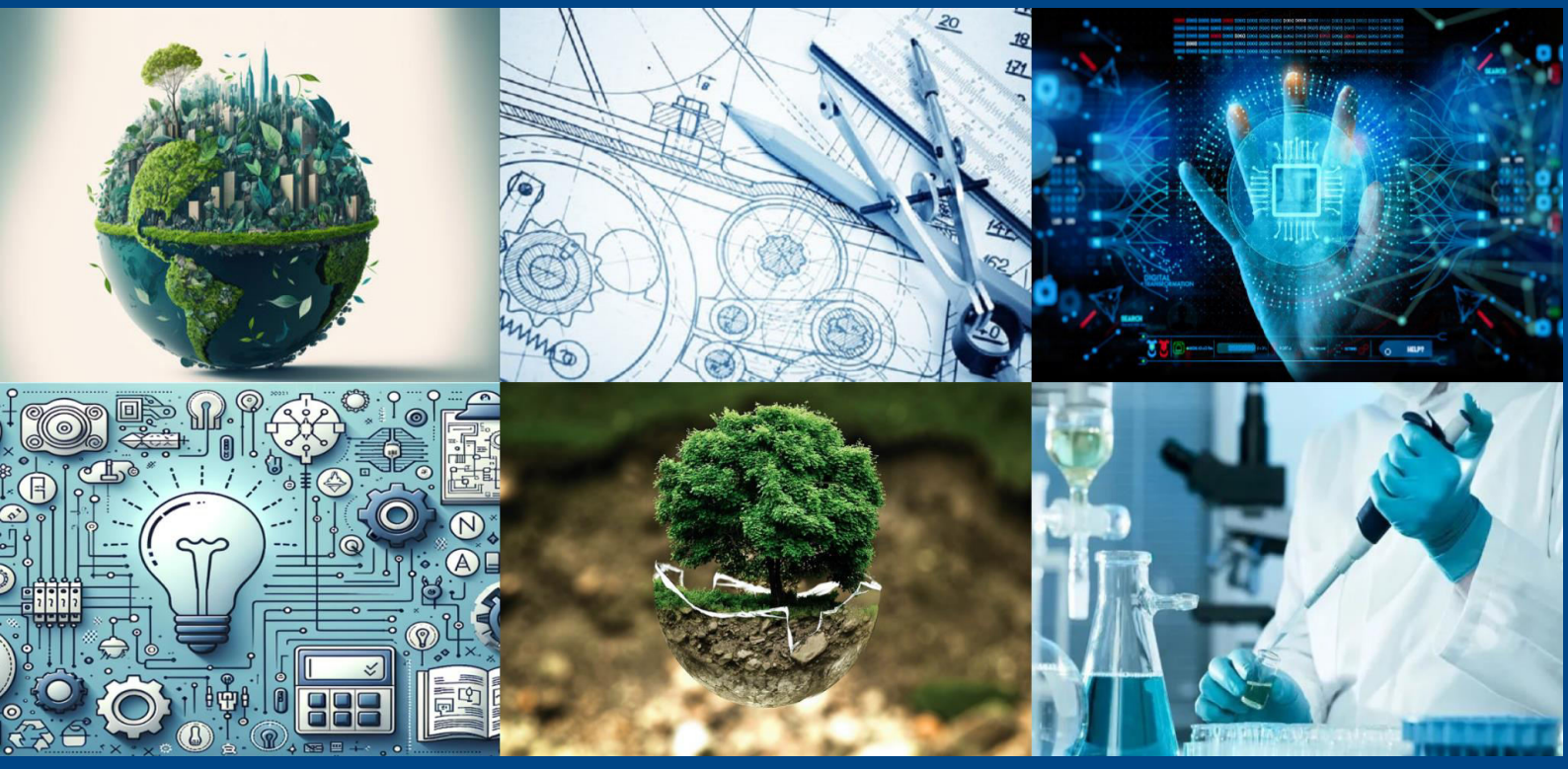
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 3, March 2025**

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803196

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Title: Stock Price Prediction by Machine Learning using Python

Ms. Reshma R, R. Karupusamy

Asst. Prof., Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India

UG Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India

ABSTRACT: Researchers have been exploring various methods to effectively forecast stock market prices, aiming to provide traders with insights into future trends and equip investors with valuable information about market conditions. One such approach involves utilizing machine learning algorithms for prediction. This study aims to enhance the accuracy of stock market forecasts by analyzing stock values. While traditional methods like Artificial Neural Networks (ANNs) have been employed to uncover hidden patterns and classify data for stock market prediction, this research proposes an alternative technique. Instead of fitting data to a specific model, our approach focuses on identifying latent dynamics within the data using machine learning architectures.

In this work, we apply machine learning models—Linear Regression and Gradient Boosting algorithm to forecast the stock prices of companies listed on the National Stock Exchange (NSE) and compare their performance. We employ a sliding window methodology for long-term forecasting and evaluate model performance using the Root Mean Square Error (RMSE) metric.

I. INTRODUCTION

Financial markets are complex and dynamic systems where participants engage in the buying and selling of assets such as stocks, currencies, and derivatives through various platforms, including exchanges and over-the-counter markets. These markets offer investors opportunities to generate returns, often with lower initial capital requirements compared to starting new businesses or pursuing high-salary careers. However, stock markets are influenced by numerous factors, leading to significant volatility and uncertainty.

While automated trading systems (ATS) leverage computer algorithms to execute trades with speed and efficiency beyond human capabilities, their performance necessitates the implementation of robust risk management strategies and safety measures, guided by human oversight. Developing effective ATS involves integrating various components, including trading strategies, complex mathematical models that capture stock behaviors, machine learning algorithms for forecasting future stock values, and analysis of pertinent news related to the stocks under consideration.

Time-series analysis is a prevalent technique in financial forecasting, utilizing historical data to identify patterns and predict future market movements. This method aids in discerning trends, cyclical fluctuations, and seasonal variations, thereby enhancing the accuracy of financial projections. However, it's important to note that forecasting financial markets is inherently challenging due to their non-stationary and noisy nature. Despite these challenges, time-series models remain valuable tools for investors and analysts seeking to make informed decisions based on historical data.

II. EXISTING SYSTEM

Time series forecasting and modeling are crucial in data analysis, particularly in econometrics and operations research. These techniques are widely applied in analytics and data science to predict future values based on historical data. Stock prices are inherently volatile, influenced by numerous factors, making accurate prediction challenging. Traditional methods like Autoregressive Moving Average (ARMA) models and Random Walk (RW) theory have been employed to forecast stock prices. ARMA models combine autoregressive and moving average components to capture linear dependencies in time series data. However, stock prices often exhibit non-linear patterns that these models may not fully capture. The Random Walk theory posits that stock prices follow a random path, suggesting that past

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

movements cannot predict future prices. This theory aligns with the efficient market hypothesis, which asserts that stock prices reflect all available information, making them unpredictable.

With advancements in machine learning, models like Long Short-Term Memory (LSTM) networks have been introduced for stock price prediction. LSTMs are a type of recurrent neural network capable of learning and remembering over long sequences, making them suitable for time-series forecasting. Studies have demonstrated that LSTM models can capture complex, non-linear relationships in stock price movements, potentially offering more accurate predictions than traditional methods.

III. PROPOSED SYSTEM

The proposed system employs machine learning techniques to predict stock prices by analyzing historical data obtained from Yahoo Finance. The dataset comprises approximately 900,000 records, each containing attributes such as date, stock symbol, opening price, closing price, lowest price, highest price, and trading volume, focusing on a single company's stock performance.

Data preprocessing involves reading the CSV file using Python's Pandas library, followed by normalization using scikit-learn's preprocessing module. The dataset is then partitioned into training and testing subsets, with 20% allocated for testing to evaluate model performance. For predictive modeling, machine learning algorithms like Random Forest and Support Vector Machines (SVM) are utilized to forecast stock prices. Random Forest is an ensemble method that constructs multiple decision trees during training and outputs the mean prediction for regression tasks, enhancing accuracy and controlling overfitting. SVM, on the other hand, is effective in high-dimensional spaces and is suitable for regression challenges by finding the hyperplane that best fits the data.

The system's efficacy is assessed using performance metrics such as Root Mean Square Error (RMSE), which measures the average magnitude of prediction errors, providing insight into the model's accuracy. By integrating these machine learning algorithms, the proposed system aims to deliver accurate stock price predictions, offering valuable insights for investors and stakeholders in the financial sector.

IV. MODULES

The system presented here composes of five modules:-

1. Input as Dataset

Historical stock data is sourced from the National Stock Exchange (NSE), encompassing daily records for a specific company. The dataset includes attributes such as date, stock symbol, opening price, closing price, lowest price, highest price, and trading volume.

2. Pre processing

Utilizing Python's Pandas library, the dataset is loaded and examined. Preprocessing steps involve applying Min-Max scaling to normalize the data, transforming feature values to a 0 to 1 range. This normalization enhances the performance of machine learning algorithms by ensuring uniformity in feature scales.

3. Data splitting

The preprocessed dataset is partitioned into training and testing subsets, with 80% allocated for training and 20% for testing. This division allows for model training and subsequent evaluation on unseen data, providing insights into the model's generalization capability.

4. Model Development and Training

Machine learning algorithms, specifically Linear Regression and Support Vector Machine (SVM) Classifier, are employed to model the relationship between historical stock data and future price movements. The training process involves learning from the training subset to identify patterns and relationships within the data.

5. Prediction and Evaluation

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803196

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Utilizing the trained models, predictions are made on the testing subset. Performance is assessed using metrics such as Root Mean Square Error (RMSE), which quantifies the average magnitude of prediction errors, offering a clear measure of model accuracy.

By integrating these modules, the system aims to provide accurate and reliable stock price predictions, aiding investors and stakeholders in making informed financial decisions.

V. RESULTS AND DISCUSSION

The application of machine learning techniques to stock price prediction has yielded varying degrees of success, with certain models demonstrating notable performance. For instance, studies have shown that the Random Forest model achieved accuracies of over 91% in predicting stock price movements. Similarly, XGBoost has been recognized for its ability to handle complex feature interactions, leading to improved prediction accuracy.

However, stock price prediction remains a formidable challenge due to the market's inherent volatility and the influence of external factors like news events and investor sentiment. Models are susceptible to overfitting, particularly when capturing noise rather than genuine patterns. To mitigate these issues, future research could explore the integration of diverse data sources and the application of advanced deep learning techniques, such as Long Short-Term Memory (LSTM) networks, which have shown promise in modeling temporal dependencies in financial data.

VI. KEY INSIGHTS

XGBoost: This ensemble learning algorithm, which builds decision trees sequentially to correct errors from previous trees, has demonstrated superior accuracy in stock price prediction tasks. Its robust handling of complex data patterns contributes to its high performance.
Random Forest: An ensemble of decision trees that splits data based on feature values, Random Forest effectively captures non-linear interactions within the data. Its structure provides a balance between predictive accuracy and model complexity.
Linear Regression: While simpler and computationally efficient, Linear Regression may not fully capture the intricate relationships present in stock market data, potentially leading to lower predictive accuracy.
Decision Trees: These models offer interpretability but are prone to overfitting, especially when not properly tuned, which can affect their reliability in dynamic market conditions.
Support Vector Machine (SVM): SVM models, known for their effectiveness in high-dimensional spaces, can be computationally intensive, which may limit their practicality for real-time stock price prediction.

VII. EXAMPLE IMAGE FOR STOCK

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. FUTURE WORK

Future enhancements to this project aim to incorporate additional parameters and factors, such as financial ratios and multiple data instances, to improve prediction accuracy. Integrating financial metrics like Return on Assets (ROA), Return on Equity (ROE), Debt-to-Equity Ratio, and Price-to-Earnings (P/E) Ratio has been shown to influence stock returns significantly. Beyond numerical data, analyzing textual content from public comments can reveal patterns in customer sentiment and corporate performance. Techniques such as sentiment analysis enable businesses to gauge public perception and respond effectively.

Employing traditional algorithms and data mining techniques can Expanding the analysis to encompass the entire company's stock performance, including various financial indicators and market trends, can provide a more comprehensive understanding. Developing systems that facilitate automated trading decisions based on predictive models could offer real-time investment strategies. Additionally, providing advisory services to assist new investors in navigating the stock market can further enhance the project's impact. further enhance the prediction of corporate performance. For instance, predictive analytics can anticipate customer needs, while clustering methods can identify distinct customer segments.

By integrating these diverse data sources and analytical techniques, the project can evolve to offer more accurate predictions and valuable insights into stock market dynamics.

IX. CONCLUSION

The proposed algorithm, trained on an extensive dataset of historical stock prices, offers significant potential for brokers and investors seeking to enhance investment strategies. By leveraging machine learning techniques, the model aims to provide more accurate stock value predictions compared to traditional methods. This advancement aligns with recent studies highlighting the effectiveness of machine learning in financial forecasting.

However, it's important to acknowledge that stock market predictions are inherently uncertain, and the model's performance should be evaluated carefully before making investment decisions.

REFERENCES

1. Chen, M., Han, J., & Yu, P. S. (2012). Data mining for the internet of things: Literature review and challenges. In International Conference on Internet of Things (pp. 1-9). IEEE.
2. James, G., Witten, D., Hastie, T., &Tibshirani, R. (2013). An introduction to statistical learning. New York: Springer.
3. Kimoto, T., Asakawa, K., Yoda, M., &Takeoka, M. (1990). Stock market prediction system with modular neural networks. In Proceedings of the International Joint Conference on Neural Networks (IJCNN) (Vol. 1, pp. 1-6). IEEE.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. New York: Springer.
5. Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019
6. Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. IEEE Transactions on Neural Networks, 12(4), 875-889.
7. Hull, J. C. (2015). Options, futures, and other derivatives. Pearson Education.
8. Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media.
9. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10) (pp. 807-814).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com