

INTEGRATING FUNDAMENTAL ANALYSIS WITH MACHINE LEARNING: A PREDICTIVE MODEL FOR STOCK INVESTMENT DECISIONS

Tanishq Manoj Jarsodiwala, SVKM's IOT Dhule
 Parvej Usman Khatik, SVKM's IOT Dhule
 Shravani Vinod Lakkad, SVKM's IOT Dhule
 Makarand Shahade, SVKM's IOT Dhule
 Umakant Mandawkar, SVKM's IOT Dhule

ABSTRACT

Abstract—The stock market's volatility and complexity present substantial hurdles for investors seeking to make informed decisions. This research proposes a ML-based strategy for predicting stock purchasing decisions using financial variables such as Price to Earnings, Price to Book, Return on Equity, Return on Assets, and Debt to Equity. The project processes and analyzes historical stock data using advanced algorithms such as XGBoost, finding relevant patterns for predictive accuracy. The approach overcomes significant constraints of existing methods by incorporating robust data preprocessing, feature engineering, and model evaluation using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). A user-friendly interface guarantees that investors may gain actionable insights and make informed decisions. This study not only closes the difference between conventional financial analysis and cutting-edge Machine Learning, but it also provides the framework for scalable applications in financial consulting, education, and real-time stock market forecasting. Future developments will include real-time data processing, portfolio optimization, and expanded feature sets to improve prediction skills.

Keywords: Machine Learning, Stock Market Prediction, Extreme Gradient Boosting, Financial Ratios, Price to Earnings Ratio (P/E), Price to Book Ratio (P/B), Return on Equity (ROE), Return on Assets (ROA), Debt to Equity Ratio (D/E), Predictive Analytics, Investment Decision-Making, FinTech

INTRODUCTION

The share market has long been a foundation of financial structures, providing possibilities for people and institutions to generate wealth and invest. However, given its inherent volatility and reliance on a variety of factors, predicting stock movements is difficult. Investors frequently rely on traditional approaches such as fundamental and technical analysis, which, while useful in some cases, are susceptible to human error, subjectivity, and inefficiency when dealing with huge datasets. Investors have significant hurdles in processing and analyzing complicated financial data, including gauges like P/E, P/B, ROE, ROA, and D/E ratios. Existing solutions frequently rely on manual analysis or rudimentary models, which may fail to capture the complex linkages between these variables and market trends. Furthermore, many traditional systems lack scalability, real-time capabilities, and accessibility for non-expert users, leaving investors with insufficient and ineffective decision-making tools Ali, I., & Naushad, M. (2022).

To address these problems, this study proposes a ML-based strategy that uses XGBoost to anticipate stock purchase decisions with more accuracy and efficiency. To effectively manage big datasets, the proposed

system combines robust data preparation approaches as well as advanced feature engineering. By examining historical stock data and financial ratios, the model discovers trends and predicting features, providing a data-driven alternative to traditional methods. The system's performance is measured using measures such as Mean Absolute Error and Root Mean Squared Error, which ensures high reliability and accuracy.

This strategy offers numerous advantages. By automating the study of financial statistics, the technology reduces manual labor and removes biases, enabling investors to make informed decisions quickly. The addition of a user-friendly interface broadens access to advanced predictive algorithms, making them accessible to both rookie and expert investors. The system's scalability also enables the future incorporation of real-time data and additional feature sets, paving the path for dynamic applications in financial advisory, portfolio management, and education.

This paper describes the proposed system's methodology, implementation, and evaluation, with a focus on how it bridges the gap between traditional financial analysis and modern machine learning technology. It emphasizes how the system not only empowers investors but also benefits the larger financial ecosystem by providing accurate, dependable, and actionable insights. Future work will look into real-time data processing, portfolio optimization, and the integration of macroeconomic variables to expand the system's capabilities and impact.

LITERATURE SURVEY

Several studies have investigated the use of machine learning and computational technology to address the issues of stock market forecasting. This section discusses the significant contributions in this domain:

Shoban Dinesh et al.,(2021), [3] This study focuses on projecting stock price movements by combining technical analysis tools like moving averages with regression-based machine learning approaches. The suggested model solves lagging indicator latency by forecasting trend reversals, which improves trade signal timeliness and accuracy.

Vaibhav Gaur et al.,(2020), This research focuses on the move from classical time-series forecasting to machine-based models for stock prediction. By examining the performance of algorithms

Sahil Vazirani et al., (2020), This paper undertakes a thorough comparison of existing stock prediction models and presents a hybrid technique that integrates linear regression models sequentially. The findings show a considerable reduction in errors, emphasizing the need of novel model architectures for boosting predictive performance.

Sangeeta Kumari et al. (2020), This work uses CUDA parallel computing to improve the performance of stock prediction systems. By combining K-means clustering with real-time data processing, the model divides equities into buy, sell, or hold clusters, considerably lowering computation time and boosting intraday trading decision-making. The use of parallel computing promotes speedier processing, allowing investors to respond quickly to market swings. Agustinus Bimo Gumelar et al.,(2020), method for improving the accuracy of stock market predictions by using XGBoost and Long Short-Term Memory (LSTM) algorithms. The study uses historical data from the Indonesia Stock Exchange (IDX) for 25 companies. The results show that XGBoost outperforms LSTM with a 99 percent prediction accuracy, suggesting it's a promising tool for trading strategies. The authors also highlight the importance of using appropriate data, robust algorithms, and regular tuning for optimal results. Tran Phuoc et al.,(2024), It investigates the effectiveness of Long Short-Term Memory (LSTM) algorithms, combined with technical analysis indicators, in predicting stock price trends in Vietnam's stock market. The results show that the LSTM model achieves high accuracy for most of the stocks studied,

demonstrating its suitability for analyzing and forecasting stock price movements. This finding supports the use of machine learning in predicting stock prices and suggests that technical analysis indicators can contribute to improved forecasting accuracy.

Jaydip Sen et al.,(2023) Overall of the offered regression models findings show that, while both CNN and LSTM- based regression models are very accurate in forecasting NIFTY 50 open values, the CNN model that uses the previous week's data as input is the fastest in execution. On the other hand, the encoder-decoder convolutional LSTM model, which uses the preceding two weeks' data as input, has been shown to be the most accurate in anticipating results. Sidra Mehtab et al., (2020) It indicate that all the models are quite accurate in forecasting the NIFTY 50 open values, but the univariate encoder-decoder convolutional LSTM with the previous two weeks' data as input is the most accurate model. In terms of execution speed, a univariate CNN model containing the previous week's data is determined to be the fastest (Bhalla, P, et al. (2018).

Machine learning requires manual intervention, is faster to train, more interpretable, and usually performs better on data sets with simpler structures and smaller sizes. Deep learning requires fewer resources for feature extraction and discovery Bhattacharyya, S. S., & Thakre, S. (2021), has higher expression and fitting capabilities, and can better capture data correlation in larger and more complex sets. Dini Anggreini Khairunnisa et al., (2024)

[6] It contributes valuable insights to the growing body of research in forecasting Islamic stock market indices, shedding light on market dynamics and the applicability of predictive modeling techniques (Bryła, P, et al. (2022). Continued research and refinement of predictive models will further enhance our understanding of market behavior and guide investment strategies in Islamic finance. Malti Bansal et al., (2022) It Shows that data comprehensive research work have been presented and it has been inferred that the DL algorithm outperforms all the other algorithms for stock price or time series prediction and provides results with extensive accuracy. W Khan et al., (2022), (Chawla, U., et al. (2023)

It uses algorithms on social media and financial news data to discover the impact of this data on stock market prediction accuracy for ten subsequent days. For better prediction performance and quality, feature selection and spam tweet reduction are applied to the data sets. Furthermore, it conducts tests to identify stock markets that are difficult to predict and are heavily influenced by social media and financial news. It compares the results of various methods to identify a consistent classifier. Deepak Kumar et al., (2022) Stock market predictions use mathematical strategies and learning tools it provides a complete overview of 30 research papers recommending methods that include calculation methods, ML algorithms, performance parameters, and outstanding journals. (Deb, K. Tammi, K. Kalita and P. Mahanta, (2019) It employs machine learning to make stock market predictions more accessible and accurate. Stock market predictions have been made using a variety of machine learning methodologies. The purpose of this study is to conduct a review of relevant works on machine learning techniques to stock market prediction. Aryendra Singh et al., (2021) This paper used a variety of approaches, including linear regression, K-means clustering, K closest neighbour, LSTM, and so on. The employment of algorithms in stock prediction has proven to be crucial, resulting in its inclusion in robust market plans. These studies highlight machine learning's potential to transform stock market analysis. Building on these findings, this study incorporates sophisticated approaches such as XGBoost to solve restrictions and create a scalable, accurate, and user-friendly solution for stock prediction.

LIMITATIONS OF EXISTING SYSTEM

Current stock market prediction algorithms have various challenges that reduce their usefulness,

precision, and convenience of use. These restrictions include:

1. Relying solely on historical data: Many platforms overlook crucial financial measures like P/E, P/B, ROE, ROA, and D/E when analyzing company performance. This approach produces forecasts that may overlook the underlying value of equities.
2. Inability to Recognize Complex Interrelationships: Many traditional systems struggle to understand the intricate links between financial statistics and market developments, resulting in inaccurate predictions.
3. Overemphasis on Technical Analysis: Many solutions prioritize Technical gauges (like moving averages and RSI) above fundamental analysis, ignoring important components of a company's financial health.
4. Limited Algorithmic Variety: Some platforms use out-dated models, such as linear regression, that may not manage non-linear financial data.
5. Scalability Issues: Current models may not scale across large datasets or sectors, making them unsuitable for institutional investors and global markets.
6. Lack of Real-Time Analysis: Many systems lack the ability to deliver quick predictions or updates, which is crucial for making timely investment decisions.
7. High Error Margins: Inadequate data preparation, feature selection, and model optimization can result in high error margins in forecasts from existing systems.

By tackling these limitations, the proposed project aims to deliver a more robust, accurate, and user-friendly solution for stock market prediction. It combines financial ratios with machine learning techniques to enhance decision-making and provide actionable insights.

METHODOLOGY

The methodology for this research encompasses several systematic steps designed to ensure accurate and reliable stock prediction outcomes. Below is a detailed breakdown:

- **Data Collection:** The first step involves gathering historical stock market data and financial ratios. The data sources include publicly available financial APIs such as Yahoo Finance, Money Control, or Bloomberg Terminal. Key ratios like P/E, P/B, ROE, ROA, and D/E function as characteristics for input to ML algorithms.
- **Data Processing:** The obtained data is cleaned by dealing with values that are missing, eliminating redundant information, and normalizing numerical features. Relevant attributes are selected to enhance prediction accuracy. Financial ratios are scaled to ensure consistency. To guarantee objective model evaluation, this data set is divided into development, verification, and evaluation sets.
- **System Development:** XGBoost algorithm is chosen for its ability to handle non-linear relationships and large datasets effectively. **Model Training:** The algorithm is trained on the processed data using key financial ratios as input features. Hyperparameter tuning is conducted to optimize performance.
- **Model Evaluation:** Model correctness and dependability are evaluated using metrics like MAE as well as RMSE. The model is evaluated on the validation dataset to determine its generalizability.
- **Deployment:** A straightforward interface is developed to allow investors to enter stock information and receive recommendation. The interface displays key financial ratios and actionable insights. The trained model is integrated into the application to ensure smooth and efficient prediction outputs.
- **Real-Time Processing:** The system architecture is designed to incorporate real-time data integration in the future.

This will enable instant predictions and facilitate intraday trading decisions.

By following this structured methodology, the research ensures that the developed system is robust, scalable, and capable of providing reliable stock predictions.

Working of the Project

The project uses machine learning to create a stock market recommendation system that helps investors make informed decisions and teaches novice investors about stock market ideas. It begins with the collection and preparation of financial data, which includes key ratios like P/E, P/B, ROE, ROA as well as D/E. These indicators are examined with the XGBoost algorithm to accurately recommend stocks.

The system includes a easily navigable interface for investors to enter stock information and view projections, as well as a filtered AI chatbot that answers questions, teaches users about financial strategies, and simplifies difficult stock market ideas. Furthermore, the system is intended for real-time data integration and portfolio optimization, with individualized investment recommendations.

The system is managed by administrators, who ensure that activities run smoothly, data is accurate, and algorithms are updated. The project combines advanced machine learning techniques with financial research

to provide a scalable, accessible tool that assists both experienced and novice investors in their decision-making processes.

Use case diagram

The use case diagram of system for recommending stocks highlights the interaction between two primary user roles, Investor and Admin, with the system. Investors can utilize key functionalities such as predicting stock performance by inputting stock details, analyzing financial ratios like P/E, P/B, ROE, ROA, D/E for insights, learning stock market concepts and strategies via a filtered AI chatbot, and optimizing their portfolio with personalized suggestions. Admins, on the other hand, manage system operations, ensuring smooth functionality, real-time updates, and maintaining the prediction models and chatbot services. This diagram emphasizes a user-centric design, providing investors with intuitive tools for financial analysis and education while enabling admins to oversee system integrity and performance effectively.

CONCLUSION

This research demonstrates the efficacy of leveraging ML algorithms, specifically XGBoost, to predict stock purchasing decisions. By analyzing critical financial ratios such as P/E, P/B, ROE, ROA as well as D/E, the system provides a data-driven and reliable alternative to traditional stock market analysis methods.

The proposed system offers a user-friendly interface that simplifies the recommendation process for both novice and experienced investors, promoting informed decision-making. The inclusion of robust preprocessing, feature engineering, and rigorous evaluation ensures the preciseness as well as reliability of recommendations, while the scalability of the model facilitates potential future enhancements. Additionally, a filtered AI chatbot educates new investors by answering questions on various stock market topics, making complex financial concepts more accessible and empowering users with knowledge. The chatbot further supports educational goals by providing real-time answers to queries such as stock calculations, market mechanisms, and investment strategies.

One of the system's benefit is its ability to close the gap between contemporary ML technologies and conventional financial analysis. By automating manual tasks, it reduces human error and biases, providing investors with actionable insights tailored to their needs. This scalability and efficiency enable broader applications in financial advisory, Actual analysis of data, and education.

The inclusion of the AI chatbot also establishes a comprehensive educational framework for novice investors, addressing their knowledge gaps and building confidence in their investment decisions. With this feature, the system not only predicts stock market trends but also serves as a valuable learning resource, enabling users to grasp complex market dynamics easily.

Future work will explore real-time data integration to enhance the immediacy of predictions, advanced portfolio optimization techniques, and the inclusion of additional macroeconomic indicators for greater predictive depth. Such advancements would position this system as a comprehensive tool for understanding the complexity of investments in this field.

In conclusion, the result of this research underscore the total transformative power of ML in analysis of stock market, offering innovative solutions to improve investment strategies, reduce risk, and empower stakeholders in the financial ecosystem. By combining predictive analytics with educational resources such as an AI chatbot, the system contributes to a more inclusive, efficient, and informed investment landscape.

REFERENCES

- Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock market prediction with high accuracy using machine learning techniques. *Procedia Computer Science*, 215, 247-265.
- Dinesh, S., Rao, N., & Anusha, S. P. (2021, April). Prediction of trends in stock market using moving averages and machine learning. In *2021 6th International conference for convergence in technology (I2CT)* (pp. 1-5). IEEE.
- Gaur, V., Sood, S., Uppal, L., & Kaur, M. (2020, December). Revitalizing Stock Predictions with Machine Learning Algorithms—An Empirical Study. In *2020 IEEE 17th India Council International Conference (INDICON)* (pp. 1-6). IEEE.
- Gumelar, A. B., Setyorini, H., Adi, D. P., Nilowardono, S., Widodo, A., Wibowo, A. T., ... & Christine, E. (2020, September). Boosting the accuracy of stock market prediction using xgboost and long short-term memory. In *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 609-613). IEEE.
- Khairunnisa, D. A., & Ahmad Rodoni, A. R. Predicting Shariah Stock Market Indices with Machine Learning: A Cross-Country Case Study Prediksi Indeks Pasar Modal Syariah dengan Machine Learning: Studi Kasus Antar Negara.
- Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1-24.
- Kumar, D., Sarangi, P. K., & Verma, R. (2022). A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings*, 49, 3187-3191.
- Kumari, S., Patil, N., Nankar, P., & Kulkarni, M. (2020, September). CUDA parallel computing framework for stock market prediction using K-means clustering. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 467-473). IEEE.
- Mehtab, S., & Sen, J. (2020, November). Stock price prediction using CNN and LSTM-based deep learning models. In *2020 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 447-453). IEEE.
- Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Computer Science*, 216, 96-102.
- Phuoc, T., Anh, P. T. K., Tam, P. H., & 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(1), 1-18.
- Sen, J., Mehtab, S., & Nath, G. (2020). Stock price prediction using deep learning models. *Lattice: The Machine Learning Journal*, 1(3), 34-40.
- Singh, A., Gupta, P., & Thakur, N. (2021). An empirical research and comprehensive analysis of stock market prediction using machine learning and deep learning techniques. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012098). IOP Publishing.
- Vazirani, S., Sharma, A., & Sharma, P. (2020, October). Analysis of various machine learning algorithm and hybrid model for stock market prediction using python. In *2020 International conference on smart technologies in computing, electrical and electronics (ICSTCEE)* (pp. 203-207). IEEE.

Received: 27-Dec-2024, Manuscript No. AEJ-24-15665; **Editor assigned:** 29- Dec-2024, PreQC No. AEJ-24-15665 (PQ); **Reviewed:** 13-Jan-2025, QC No. AEJ-24-15665; **Revised:** 21- Jan-2025, Manuscript No. AEJ-24-15665 (R); **Published:** 28- Jan-2025