# A MACHINE LEARNING PERSPECTIVE ON FORECASTING STOCK PRICES - A REVIEW

 P. Prudhvi Raj

# 1. ABSTRACT:

Stocks represent a form of security that provides shareholders with a degree of ownership in a corporation, along with rights to the company's profits and assets. As a result of the company's changing valuation, stock prices also vary, making the prediction of stock values a difficult undertaking due to the volatile and nonlinear characteristics of financial markets. This is where Advanced Machine Learning comes into play; Machine Learning, which is a branch of Artificial Intelligence, employs algorithms to enable computers to learn from data without being explicitly programmed. In traditional approaches, the model is often overwhelmed by noise or hindered by rigid labeling methods. So by using Machine Learning algorithms and techniques, stock price predictions become more accessible and accurate, as they can identify patterns and trends in financial data. In order to fix the issue of noise, there are various labeling methods to be found in Machine Learning, including but not limited to N-Period Min-Max labeling, along with advanced techniques like Variational Mode Decomposition, Model-agnostic meta-learning, and Long short-term memory. These approaches work together to enhance the precision of stock predictions in complex financial markets. This study presents a new framework for predicting stock prices that tries to turn these very challenges into opportunities. The study aims to provide reliable, interpretable stock price forecasts

**2. KEYWORDS:** *Machine Learning Algorithms, Artificial Intelligence, N-Period Min-Max labeling, Variational mode decomposition, model-agnostic meta leaning, long short-term memory network*

# 3. INTRODUCTION:

# Forecasting stock market trends remains a critical area of interest in financial research, as the inherent volatility and complex data dependencies of financial markets pose challenges for accurate prediction. Traditional methods, such as fundamental and technical analysis, offer valuable insights into market behavior by examining company-specific and historical price information. Yet, these methods often struggle with the scale and rapid evolution of modern financial data, which limits their ability to capture intricate market patterns and respond to changes swiftly. This challenge has directed research towards advanced, data-driven solutions, particularly within machine learning (ML), to improve the adaptability and accuracy of stock price forecasts.

# Machine learning models have proven highly effective in processing large datasets, identifying hidden patterns, and enhancing prediction precision across various financial applications. Techniques like Support Vector Machines (SVM), XGBoost, and Logistic Regression have shown strong performance in market trend prediction, especially when paired with advanced labeling techniques. For instance, Han, Kim, and Enke (2023) have shown that employing XGBoost with N-Period Min-Max labeling offers advantages by effectively handling noise and addressing short-term fluctuations, which results in enhanced model precision. Furthermore, Long Short-Term Memory networks, which are tailored for processing sequential data, excel at recognizing long-term dependencies, a crucial factor for forecasting in time-series analysis.

# Recent developments have expanded ML’s applicability in finance, with advanced techniques like meta-learning, variational mode decomposition (VMD), and explainable AI making strides in robustness and interpretability. Liu et al. (2022) explored the potential of meta-learning in combination with VMD, which helps address the issues of non-stationarity and concept drift in financial time-series data by adapting model learning as data patterns evolve. Additionally, XAI methods, including Belief Rule-Based Expert Systems (BRBES) and Local Interpretable Model-Agnostic Explanations (LIME), increase transparency in model predictions, providing insights into the reasoning behind model outputs—an essential feature for building trust in high-stakes financial decisions (Rudin, 2019).

# Combining these advanced ML techniques with sentiment analysis further broadens the approach to stock prediction by capturing external factors that influence market dynamics. This paper offers a comparison of machine learning models, including XGBoost, meta-learning, and VMD, combined with various labeling techniques, to evaluate their effectiveness in dynamic financial environments. By analyzing the predictive power and interpretability of these approaches, this study seeks to identify the strengths and limitations of different models for robust, adaptable, and transparent stock price forecasting.

# 4. RELATED WORK:

Machine learning has proven instrumental in advancing stock price prediction by addressing challenges such as noise, adaptability, and model interpretability. Han, Kim, and Enke [1] proposed a trading system using XGBoost combined with N-period Min-Max labeling to reduce data noise, thereby improving prediction accuracy in the fluctuating stock market. This method highlights the impact of strategic labeling in refining data for more precise forecasting.

Liu et al. [2] introduced a framework combining meta-learning with variational mode decomposition (VMD) to manage non-stationary data and adapt to market concept drift. Meta-learning enhances the model’s adaptability, while VMD stabilizes data by breaking it into subseries, thereby boosting prediction accuracy in variable financial condition.

Yun, Yoon, and Won [3] addressed the need for model interpretability by integrating a genetic algorithm for feature selection with machine learning regressions. This approach streamlines the model, allowing for clearer insights into the decision-making process, which is essential for transparency in financial predictions.

Chen et al. [4] conducted a comparative study of machine learning techniques, including graphic signal recognition, assessing their effectiveness in stock prediction. This comparison provides valuable insights into how different algorithms perform in handling complex financial data patterns, guiding algorithm selection for various prediction scenarios.

Lastly, Akhtar et al. [5] analyzed statistical data-based machine learning models, comparing their performance for stock prediction tasks. Their findings underscore the importance of selecting suitable algorithms based on data characteristics, as model effectiveness can vary widely in financial forecasting**.**

Hossain et al. [6] explored the use of Belief Rule-Based Expert Systems (BRBES) integrated with machine learning for stock price prediction. Their study demonstrates how BRBES, which incorporates belief rules and fuzzy logic, can enhance accuracy in forecasting by effectively managing uncertainty and ambiguity in financial data, especially in volatile markets.

Vijh, Chandola, and Tikkiwal [7] investigated various machine learning techniques for predicting stock closing prices. Their work provides a comparative analysis of different algorithms and identifies those that yield higher accuracy, emphasizing the role of algorithm selection in stock prediction and the need for robust methods that adapt to changing market dynamics.

Bansal, Goyal, and Choudhary [8] focused on achieving high accuracy in stock market prediction by using machine learning techniques. Their study highlights advancements in model performance through hyperparameter tuning and model optimization, showcasing how careful calibration can lead to significant improvements in predictive accuracy.

Maqbool et al. [9] integrated sentiment analysis with a Multi-Layer Perceptron (MLP) regressor to improve stock prediction accuracy. By incorporating sentiment scores derived from financial news, the study demonstrates how external market sentiment can be combined with machine learning models to provide a more comprehensive understanding of market trends.

L. M and Gnanasekaran [10] examined the application of classification algorithms for stock price prediction. Their work underscores the potential of classification approaches in identifying stock trends, particularly by comparing different machine learning classifiers and highlighting which models are better suited for classification tasks within financial forecasting.

Çelik, İcan, and Bulut [11] extended machine learning models in financial prediction by incorporating Explainable AI (XAI). Their research underlines the importance of transparency and interpretability in financial forecasting, showing how XAI techniques can enhance trust in machine learning predictions by providing clearer insights into model decision-making processes.

Mehta, Malhar, and Shankarmani [12] used a combination of machine learning and sentiment analysis to predict stock prices. The study emphasizes the effectiveness of combining traditional data-driven approaches with sentiment analysis, enabling a more dynamic prediction that accounts for the influence of public sentiment on market movement.

Dinesh et al. [13] utilized moving averages and machine learning models to predict stock market trends. Their approach showcases how blending technical indicators with machine learning can improve forecasting, particularly in capturing trend directions and making short-term predictions in fluctuating markets.

K. C et al. [14] focused on predicting stock prices through various machine learning techniques, comparing the effectiveness of different algorithms. Their study provides insight into the relative strengths of each method, guiding model selection based on prediction accuracy across multiple datasets.

Kadu and Bamnote [15] performed a comparative study of stock price prediction using machine learning. By analyzing the performance of various algorithms, the research highlights which models are most effective in different market scenarios, reinforcing the value of comparative analysis in selecting optimal models for financial forecasting.

# 5. METHODOLOGY:

5.1 Problem Definition:

The primary aim is to advance stock price prediction by improving model accuracy and adaptability using machine learning techniques and optimized data labeling. Traditional methods often struggle with financial data’s complexity, including noise and high volatility. This research explores machine learning models, such as XGBoost, SVM, and LSTM, alongside labeling techniques like N-Period Min-Max and Variational Mode Decomposition (VMD), to reduce noise and stabilize predictions. Meta-learning is employed to enhance model adaptability to shifting market conditions, ensuring model predictions are transparent and actionable. Additionally, sentiment analysis is incorporated to capture external factors influencing market trends, creating a more well-rounded prediction model.

5.2 Data Collection and Preprocessing:

Data for this study was sourced from financial platforms such as Yahoo Finance, covering daily stock prices, trading volumes, and sentiment scores derived from financial news sources. Preprocessing steps included addressing missing values, normalizing data through Z-score scaling to ensure consistency, and applying advanced labeling techniques such as N-period Min-Max to enhance data structure for accurate prediction. Additionally, feature engineering was used to create relevant financial indicators (e.g., moving averages, volatility measures) to help capture essential patterns within the data. Outlier detection and handling were also conducted to minimize any skewing effects on the predictions, resulting in a robust and reliable dataset for model training.

5.3 Machine Learning Models:

XGBoost is a powerful ensemble learning method that builds sequential models to improve predictive accuracy, particularly effective in handling noisy and high-dimensional data. When combined with Min-Max labeling, XGBoost excels in capturing short-term price fluctuations by prioritizing significant data points, which is crucial for financial data where sudden changes can impact overall trends. This method is especially useful in settings where short-term forecasting is critical, such as intraday trading and rapid response strategies.

Support Vector Machines is a versatile supervised learning model with strengths in both classification and regression, making it well-suited for financial forecasting. In the context of stock prices, SVM’s strength lies in its capacity to construct hyperplanes that maximize the margin between data classes, effectively separating different market trends and potential reversals. When combined with preprocessing methods like Variational Mode Decomposition (VMD), which decomposes data into stable, trend-isolating components, SVM becomes more resilient to noise. This synergy enables SVM to detect and predict underlying trends in various market conditions with enhanced accuracy

Belief Rule-Based Expert System provides a unique advantage in financial prediction by combining rule-based reasoning with probabilistic modeling, thus offering interpretability alongside adaptability. In stock price forecasting, BRBES employs belief rules that integrate expert knowledge, helping the model manage uncertainty and respond dynamically to market changes. Unlike black-box approaches, BRBES yields interpretable outputs, which is crucial in financial settings where model transparency directly influences decision-making confidence. This interpretability, paired with BRBES’s ability to dynamically adapt to new data patterns, makes it a reliable choice for predictions in volatile market conditions. The below figure outlines a three-phase process for algorithmic trading. Phase 1 involves extracting technical indicators and labeling data. Phase 2 trains an XGBoost model on this labeled data. In Phase 3, the trained model generates trading signals (buy, hold, sell) and evaluates its performance.



Fig-1: XGBoost-based Trading Strategy Pipeline

Labeling Techniques:

N-Period Min-Max labeling is a data-labeling strategy that assigns significance to the minimum and maximum values observed within a designated period, helping models focus on prominent price changes while ignoring minor noise. This labeling technique is particularly beneficial in high-volatility markets, where isolating major price shifts enhances the model’s ability to identify relevant trends. By using N-Period Min-Max labeling, algorithms like XGBoost and SVM can better prioritize impactful data points, ultimately refining their predictive accuracy.

This labels as Up(0) at the maximum point and Down at the minimum point and the Ct is the closing price of the N-term , Nmax is the maximum closing price of the N-term, and Nmin is the minimum closing price of the N-term, the directions of the closing price in the future is Lt+1

Lt+1 = $\left\{\begin{array}{c}0, if Ct == Nmax\\1 if Ct == Nmin\end{array}\right.$



Fig -2: Trading signal generated by the NPMM Labeling

5.4 Model Training and Evaluation:

The process of training and evaluating models followed a systematic method, dividing the dataset into training, validation, and test sets to ensure a thorough performance evaluation. Hyperparameter tuning was performed using techniques like grid search and Bayesian optimization to enhance the accuracy of each model, concentrating on methods such as XGBoost, Support Vector Machine (SVM), and Belief Rule-Based Expert Systems. Metrics for evaluation, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared, were utilized to measure predictive accuracy, while cross-validation was applied to reduce overfitting and improve model generalization. A comparative analysis of each model's performance revealed that XGBoost's capability of capturing intricate, non-linear relationships was especially beneficial, whereas the Belief Rule-Based Systems offered further understanding of feature importance. The final model was chosen based on these metrics and the trade-offs between accuracy and interpretability, highlighting the necessity of tuning and metric-based evaluation in stock prediction models. Figure Fig-3 below presents a comparison of three machine learning models—XGBoost, SVM, and BRBES—using their performance metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared. In conclusion, XGBoost stands out as the most precise model according to both MSE and RMSE. Its lower error values indicate that it can deliver more accurate predictions when compared to SVM and BRBES.



Fig-3: Different models performance metrics

5.5 Challenges and Limitations:

Implementing machine learning for stock price prediction faces several challenges. Stock market data is highly volatile and noisy, which makes it difficult to build stable, accurate models. Techniques like N-Period Min-Max and VMD help reduce noise but add complexity to data processing. Models often have trouble adjusting to sudden shifts in the market, even with methods like meta-learning designed to help them adapt. Additionally, understanding how these models make decisions can be difficult. Techniques like XGBoost and LSTM are often hard to interpret, and while tools like LIME can provide some insight, it remains challenging to fully clarify how the models come to their conclusions. Lastly, relying on sentiment analysis introduces potential biases due to inconsistent textual data, affecting overall prediction accuracy. These limitations underscore the need for ongoing model improvements to handle the complexities of financial forecasting effectively.

**6. RESULTS AND DISCUSSION:**

6.1 Results:

The comparative analysis of machine learning models revealed distinct advantages and limitations associated with each approach. Table 1 summarizes the performance of the models based on the evaluation metrics—Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared—calculated over the test dataset. The results indicate that XGBoost outperformed other models, delivering the lowest MSE and RMSE values, as well as the highest R-squared, showcasing its superior ability to capture nonlinear patterns in stock price movements.

Table 6.1.1: Performance Metrics of Machine Learning Models



6.2 Discussion:

The success of XGBoost can be attributed to its ability to handle complex interactions in the dataset, leveraging techniques like N-Period Min-Max labeling to prioritize significant price movements. This approach is particularly effective in financial markets characterized by high volatility and noise. However, despite its high accuracy, XGBoost lacks interpretability, which is critical in high-stakes financial applications. Integrating techniques such as Local Interpretable Model-Agnostic Explanations (LIME) could mitigate this challenge by providing post-hoc explanations for the model's predictions.

The Support Vector Machine (SVM) model demonstrated moderate performance, with its strength lying in its capacity to separate data trends using hyperplanes. When combined with Variational Mode Decomposition (VMD), SVM effectively reduced noise, making it suitable for scenarios requiring robust noise handling. However, its computational complexity and sensitivity to hyperparameter tuning presented challenges, particularly in real-time applications.

**6. CONCLUSION:**

# This research emphasizes the potential of machine learning methodologies to enhance the accuracy of stock price predictions by tackling issues like noise, volatility, and adaptability. By incorporating sophisticated techniques such as XGBoost, SVM, and LSTM, along with labeling strategies and sentiment analysis, we can improve the precision and dependability of stock market forecasts. Although challenges remain, including how to adjust to swift market fluctuations and enhance model interpretability, the blend of these advanced tools appears promising in making stock price predictions more resilient and informative. Future investigations should aim to refine these models and seek methods to increase their transparency and adaptability to changing market circumstances.

# 7. REFERENCES:

[1] Yechan Han, Jaeyun Kim, David Enke (2023), “A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost”, Expert Systems with Applications, Volume 211,118581, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2022.118581>.

[2] Teng Teng Liu, Xiang Ma, Shuo Li, Xuemei Li, Caiming Zhang(2022), “A stock price prediction method based on meta-learning and variational mode decomposition”, Knowledge-Based Systems, Volume 252, 109324, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2022.109324>

[3] Kyung Keun Yun, Sang Won Yoon, Daehan Won (2023), “Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection”, Expert Systems with Applications, Volume 213, Part A, 118803, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2022.118803>

[4] Junde Chen, Yuxin Wen, Y.A. Nanehkaran, M.D. Suzauddola, Weirong Chen, Defu Zhang(2023), “Machine learning techniques for stock price prediction and graphic signal recognition, Engineering Applications of Artificial Intelligence”, Volume 121, 106038, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2023.106038>

[5] Md. Mobin Akhtar, Abu Sarwar Zamani, Shakir Khan, Abdallah Saleh Ali Shatat, Sara Dilshad, Faizan Samdani(2022), Stock market prediction based on statistical data using machine learning algorithms, Journal of King Saud University - Science, Volume 34, Issue 4, 101940, ISSN 1018-3647, <https://doi.org/10.1016/j.jksus.2022.101940>.

[6] Emam Hossain, Mohammad Shahadat Hossain, Pär-Ola Zander, Karl Andersson(2022), “Machine learning with Belief Rule-Based Expert Systems to predict stock price movements”, Expert Systems with Applications, Volume 206, 117706, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2022.117706>.

[7] Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal(2020), “Stock Closing Price Prediction using Machine Learning Techniques”, Procedia Computer Science, Volume 167, Pages 599-606, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.326>.

[8] Malti Bansal, Apoorva Goyal, Apoorva Choudhary (2022), ”Stock Market Prediction with High Accuracy using Machine Learning Techniques”, Procedia Computer Science, Volume 215, Pages 247-265, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2022.12.028>.

[9] Junaid Maqbool, Preeti Aggarwal, Ravreet Kaur, Ajay Mittal, Ishfaq Ali Ganaie(2023), “Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach”, Procedia Computer Science, Volume 218, Pages 1067-1078, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2023.01.086>.

[10] L. M and P. Gnanasekaran(2023), "Prediction of Stock Price Using Machine Learning (Classification) Algorithms," International Conference on Advances in Computing, Communication and Applied Informatics, Chennai, India, 2023, pp. 1-5, doi: 10.1109/ACCAI58221.2023.10199307.

[11] Taha Buğra Çeli̇k, Özgür İcan, Elif Bulut(2023), ”Extending machine learning prediction capabilities by explainable AI in financial time series prediction”, Applied Soft Computing, Volume 132, 109876, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2022.109876>.

[12] Y. Mehta, A. Malhar and R. Shankarmani(2021), "Stock Price Prediction using Machine Learning and Sentiment Analysis”, 2nd International Conference for Emerging Technology, Belagavi, India, 2021, pp. 1-4, doi: 10.1109/INCET51464.2021.9456376.

[13] S. Dinesh, N. R. R, S. P. Anusha and S. R(2021), "Prediction of Trends in Stock Market using Moving Averages and Machine Learning”, 6th International Conference for Convergence in Technology, Maharashtra, India, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418097.

[14] K. C, S. A. N. A, A. P, P. R, M. D, and V. B. D(2021), "Predicting Stock Prices Using Machine Learning Techniques“, 6th International Conference on Inventive Computation Technologies, Coimbatore, India, pp. 1-5, doi: 10.1109/ICICT50816.2021.9358537.

[15] P. P. Kadu and G. R. Bamnote(2021), "Comparative Study of Stock Price Prediction using Machine Learning”, 6th International Conference on Communication and Electronics Systems, Coimbatore, India, pp. 1200-1204, doi: 10.1109/ICCES51350.2021.9489170.