Stock Market Price Prediction using Machine Learning Based Technical Analysis

Pallabi Das[1], Madhurendra Singh[2], Aditya Anand[2], Devesh Kumar[2], Nainshi Kumari[2], Ravi Kumar[2], and Adrija Saha[2]

[1] Assistant Professor, Department of Computer Science Engineering, Guru Nanak Institute of Technology, West Bengal, India

[2] Student, Department of Computer Science Engineering, Guru Nanak Institute of Technology, West Bengal, India

Abstract. Accurately predicting stock market values has long been a goal of financial experts and practitioners. Machine learning techniques have supplemented traditional methods like statistical modeling and technical analysis, opening up new possibilities for increasing the forecast accuracy. In this paper, a novel approach to stock market price prediction by leveraging support and resistance levels derived from financial data is presented. The study involves a methodical procedure that includes collecting data, preprocessing, extracting features, training the model, and evaluating it. Several datasets are being used in this instance, including the 1-hour time frame of XAUUSD (gold versus US Dollar), XAGUSD (Silver Versus US Dollar), BTCUSD(Bitcoin-USD), and the US30 Index(Dow Jones 30). The temporal data format is converted and divided into training and testing sets during the preprocessing step. Next, using a look-back period of 50, support and resistance levels are calculated. Significance is determined by requiring a minimum of three touches. Next, using the preprocessed data, multiple machine learning (ML) models like Linear Regression, Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), and Support Vector Machine (SVM) are trained using the previously established support and resistance levels as predictive characteristics. On a different test set, the model's performance is assessed using Mean Absolute Error (MAE). By incorporating support and resistance levels into the modeling process, we achieve enhanced performance in forecasting stock market prices, particularly in capturing short-term price movements and market sentiment shifts. Furthermore, our study highlights the potential applicability of this approach to other financial time series datasets beyond gold prices, suggesting its broader utility in diverse financial markets and asset classes.

Keywords: Machine learning, Mean Absolute Error, Statistical Modeling and Stock Market Price.

1 Introduction

Predicting stock market prices accurately has long been a pursuit of researchers and practitioners in finance. To effectively manage risks and make well-informed decisions, traders, investors, and financial analysts need to be able to forecast price movements. Traditional methods such as technical analysis and statistical modeling have been augmented by the advent of machine learning techniques, offering new avenues for improving prediction accuracy. In this research paper, we introduce a novel approach that combines traditional statistical modelling with the incorporation of support and resistance levels derived from financial data to enhance the prediction of stock market prices.

Robust methodologies like as deep learning and machine learning have revolutionized the field of artificial intelligence. Aspiring data scientists and AI enthusiasts must comprehend the foundations of these methods as well as the widely utilized algorithms. In predictive modeling, regression is a basic idea that is essential to understanding and forecasting continuous variables. We can unlock tremendous potential in a variety of domains and develop and progress several sectors by utilizing the possibilities of these algorithms and methodologies.

This study was motivated by the realization of how difficult it is to predict stock market prices, which are influenced by a variety of factors like geopolitical events, market sentiment, and economic indicators. The integration of domain-specific insights like support and resistance levels has not got much attention in the literature, despite the fact that machine learning algorithms have demonstrated promise in capturing complex patterns in financial data. Support and resistance levels represent key psychological and technical thresholds in financial markets, reflecting points where buying and selling pressures converge. By integrating these levels into our predictive model, we aim to capture the inherent dynamics of market behavior and improve the robustness of our forecasts.

The remaining of the paper is designed in the following manner: section 2 discuss the review of the related work, section 3 discuss the proposed methodology. Section 4 shows the experimental result and novelty of the methodology. Section 5 concludes the paper.

2 Related Work

Previous research in the domain of financial time series forecasting has encompassed a wide array of techniques, ranging from traditional statistical methods to advanced machine learning and deep learning models. These methodologies have been employed with the aim of accurately predicting price movements in various financial markets. In [1] D. Kumar et al. reviewed multiple research papers based on Machine Learning based techniques for stock market prediction. This study made the assumption that stock market forecasting being a comprehensive process and unique criteria for predicting the stock market should be regarded as more accurate. In [2] authors highlighted on combining multiple datasets into a single data block. The database's characteristics were the charge for each day and the expiration date. Multiple functions were employed to train the device on an arbitrary version of time bar and forecast the item variable, which is the cost for a specific day. By employing traditional ML algorithms they achieved an accuracy of 0.8. In [3] A. Moghar et al. highlighted on the precision of a ML algorithm's prediction and how much the epochs can improve the model. In[4] Priyanka Srivastav et al. analyzed future stock prices using data-frame closing prices, built up and trained the LSTM model, and have taken a data set sample to generate stock forecasts and computed additional RMSE for correctness and effectiveness. G. Bathla et al.[5] investigated whether deep learning can predict high variations in stock prices in a specific time slot and built a new neural network based model. Pramod BS et al. [6] provided RNN-LSTM based comparison to currently available stock price predictor algorithms. The network was trained and evaluated with various sizes of input data to urge the graphical outcomes. J sen et al. [7] proposed a ML and deep learning-based predictive model for predicting the NIFTY 50 stock price movement in NSE of India. Vivek Varadharajan et al. [8] used Recurrent Neural Networks (RNN) with LSTM to predict the daily closing price of the Amazon Inc. stock. H. N. Bhandari et al.[9] used LSTM to predict the next-day closing price of the S and P 500 index.

However, despite the breadth of research in this field, few studies have delved into the potential benefits of incorporating support and resistance levels in predictive models. Machine learning algorithms, including deep learning models like LSTM, have garnered significant attention for their ability to capture complex patterns and temporal dependencies in financial data. These models offer promising avenues for improving prediction accuracy and have been applied to a diverse range of financial forecasting tasks. Statistical methods, on the other hand, provide robust frameworks for analyzing and modeling time series data, often incorporating econometric techniques to account for underlying market dynamics. The incorporation of support and resistance levels, derived from historical price data, represents a novel approach to enhancing prediction accuracy in financial markets. These levels serve as key psychological and technical thresholds, reflecting areas of potential price reversals or continuation patterns. By integrating support and resistance analysis into predictive modeling frameworks, researchers aim to capture the inherent dynamics of market behavior and improve the robustness of forecasts.

Despite the potential benefits of incorporating support and resistance levels, the literature on this topic remains sparse. Existing studies have primarily focused on other predictive features and techniques, with limited exploration of support and resistance analysis in predictive modeling. This research paper seeks to address this gap by presenting a novel approach that combines linear regression with support and resistance level identification to improve prediction accuracy.

Overall, while previous research has explored various techniques for predicting financial time series data, including machine learning algorithms and statistical methods, few studies have investigated the potential of incorporating support and resistance levels in predictive models. This study aims to fill this gap by presenting a novel approach that combines linear regression with support and resistance level identification, offering new insights into enhancing prediction accuracy in financial markets.

3 Proposed Work

This section introduces the proposed approach for predicting stock market prices by incorporating support and resistance levels derived from financial data. By synthesizing insights from previous research on financial time series forecasting techniques, including machine learning algorithms, deep learning models, and statistical methods, this study aims to contribute to the growing body of knowledge in the field. By highlighting the potential benefits of incorporating support and resistance levels in predictive models, this research endeavors to pave the way for future advancements in financial forecasting methodologies.

This approach combines various ML models with the identification of significant support and resistance levels, aiming to enhance prediction accuracy. A systematic framework encompassing data collection, preprocessing, feature engineering, model training, and evaluation have been used in this work. We begin by collecting historical price index data in the form of 1-hour candlesticks, focusing on the period from January 1, 2023, to January 1, 2024. The data is then cleaned and preprocessed, with the 'Local time' column converted to date time format for further analysis. Support and resistance levels are calculated using historical price data, utilizing a lookback period of 50 and a minimum threshold for significance determination. Subsequently, multiple ML models such as linear regression, Artificial neural network(ANN), LSTM are trained on the preprocessed data, leveraging the identified support and resistance levels as predictive features. The performance of the model is evaluated using Mean Absolute Error (MAE) on a separate test set, providing quantitative metrics to assess its effectiveness. Figure 1 shows the detailed workflow of the proposed methodology.

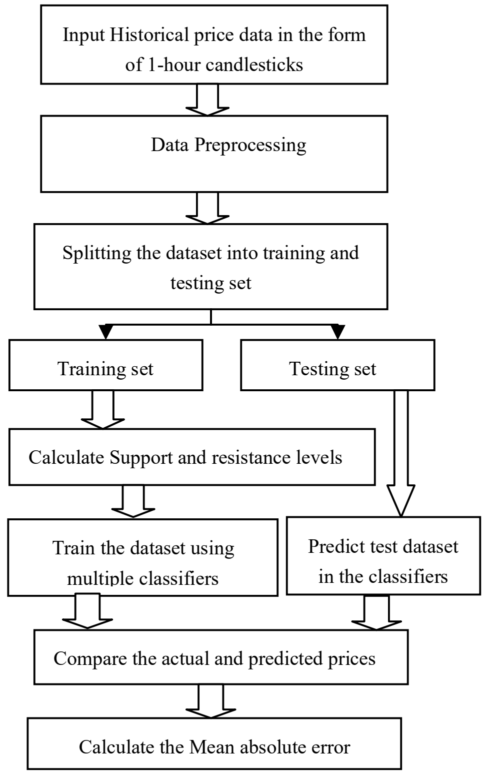


Fig.1. Workflow of the Proposed Model for the prediction of stock market prices.

4 Methodology

Dataset Collection Historical prices of XAUUSD, XAGUSD, BTCUSD and US30 Index data in the form of 1-hour candlesticks is collected for the experimental setup. The dataset spans from January 1, 2023, to January 1, 2024, providing a sufficient time frame for analysis.

Data Preprocessing The collected data undergoes preprocessing to ensure its quality and compatibility with the model. This includes cleaning the data and converting the 'Local time' column to datetime format. Preprocessing also involves splitting the data into training and testing sets, which are essential for model development and evaluation.

Data Splitting Data splitting is a crucial step in machine learning model development to ensure unbiased evaluation of the model's performance.The dataset is typically divided into two subsets: a training set and a testing set. The training set is used to train the model, while the testing set is used to evaluate its performance. It's important to shuffle the data before splitting to ensure that the distribution of data points is random in both the training and testing sets, which helps prevent any biases.

Support and Resistance Level Calculation The calculation of support and resistance levels is a crucial aspect of the experimental setup. These levels are determined based on specific criteria, including the minimum and maximum values of prices within defined ranges, as well as the cardinality of certain sets of price values. The parameters, such as the lookback period (L) and threshold value (T) are carefully chosen to ensure the effectiveness of the support and resistance level identification processes.

Feature Engineering and model training Support and resistance levels are calculated using historical price data. A lookback period of 50 is utilized for this calculation, with a minimum of three touches required for a level to be considered significant. These support and resistance levels serve as features for the subsequent model training phase. Multiple ML classification models are trained on the preprocessed data using the identified support and resistance levels as features. The training process involves fitting the model to the training data, enabling it to learn patterns and relationships between the input features and the target variable. During training, ML models learn the coefficients (weights) for each feature and the intercept term that best fits the relationship between the input features (historical prices) and the target variable (future price).

Testing and Predictions Once the model is trained, it is used to predict the future prices for the testing set. For each data point in the testing set, the script prepares the feature vector ('Open', 'High', 'Low', 'Close') and feeds it into the trained ML classification model to generate a prediction for the next time step's closing price.

Accuracy Assessment Accuracy assessment involves quantifying the performance of the model using appropriate evaluation metrics. Common metrics for evaluation tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (coefficient of determination).MAE represents the average absolute difference between the predicted values and the actual values. MSE and RMSE provide a measure of the average squared difference between predicted and actual values, with RMSE being more interpretable as it's in the same units as the target variable. R-squared measures the proportion of the variance in the dependent variable that is predictable from the independent variables. An R-squared value closer to 1 indicates a better fit of the model to the data. The formula is depicted in the equation (i-iv).

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}| \tag{1}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{2}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (3)$$

$$(\mathrm{R})^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2} \quad (4)$$

Where, N is the total no. of samples. y_i is the actual observed values in range of i=1 to N, $\hat{y}$ is predicted value of y and $\bar{y}$ is the mean value of y.

In addition to presenting our methodology and results, we provide a comprehensive review of relevant literature, highlighting previous research on predictive modelling techniques for financial time series data. While various approaches have been explored, few studies have examined the potential benefits of incorporating support and resistance levels into predictive models. In this study, we assess four machine learning models—linear regression, LSTM, ANN, and SVM—for their effectiveness in predicting stock prices. Our aim is to identify the most suitable model considering factors such as prediction accuracy, computational efficiency, and interpretability.Through this research, we aim to contribute to the growing body of knowledge in financial forecasting and provide insights that can inform decision-making processes in the financial industry.

5 Result & Analysis

Stock price prediction is crucial in financial markets, guiding investment decisions and risk management strategies. In this study, we compare the performance of four machine learning models—linear regression, Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), and Support Vector Machine (SVM)—in predicting stock prices. Using historical stock price data, we evaluate the models based on metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). Our findings demonstrate that the linear regression model outperforms the other models in terms of prediction accuracy, interpretability, and computational efficiency. We discuss the strengths and weaknesses of each model, offering insights for investment strategies and decision-making in financial markets.

Historical prices of Gold (XAUUSD), Silver (XAGUSD), US30 Index data in the form of 1-hour candlesticks is collected for the experimental setup. The dataset spans from January 1, 2023, to January 1, 2024, providing a sufficient time frame for analysis. The predicted values are compared with the actual values in the said time frame. Figure 2 shows a sample of the comparison of actual and predicted prices of Gold dataset within the given time frame.

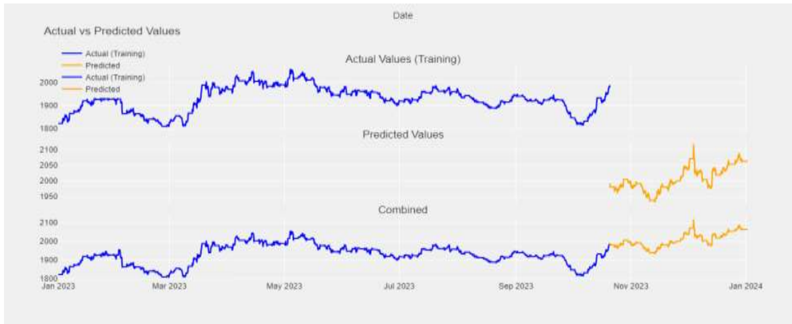


Fig.2. Line Chart of Actual and Predicted Prices of Gold

The absolute error is then calculated using MAE and MSE and then compared between the actual and predicted values. All the datasets are compared in the same way. The sample graphical presentation for the Gold dataset is shown in figure 3.

Fig.3. Graphical representation of Absolute Error between Actual and Predicted close prices of Gold

Each model linear regression, LSTM, ANN, and SVM is trained using the training set and hyper parameters are tuned. Predictions are generated using each model on the testing set, and performance is evaluated using MAE, RMSE, and R^2. Statistical tests are conducted to assess significant performance differences between the models .Figure 4 shows the comparison of the models in form of graph and chart for the different types of models.

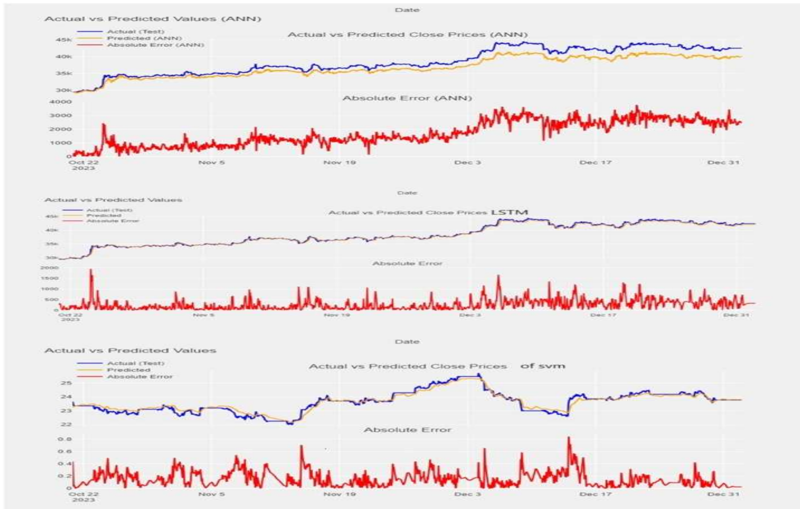
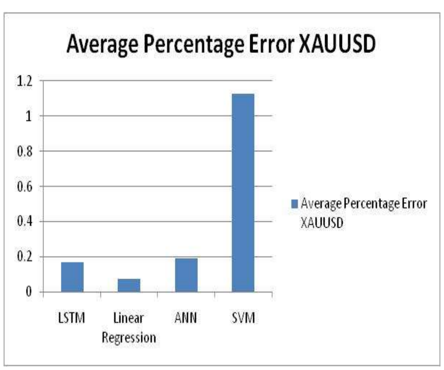
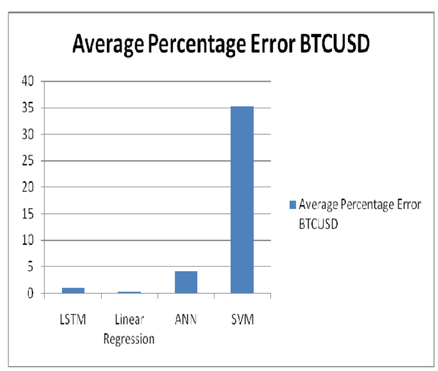


Fig.4. Comparison of the ANN, LSTM and SVM models

Finally, we have calculated the absolute error between the actual and the predictive values for the closing prices for the Gold price data. Literature highlights the strengths and weaknesses of each model in stock price prediction. In our experiment, LSTM and ANN are excellent at capturing temporal dependencies. But, in some situations their complexity might be more of a drawback than a benefit. The effectiveness of SVM in capturing non-linear relationships limits its versatility. The performance of all these models in calculating average percentage error is compared on XAUUSD, BTCUSD, XAGUSD and US30 in figure 5. It clearly reveals that linear regression, with its simplicity and transparency, provides actionable insights for investors and financial institutions, facilitating informed decision-making.





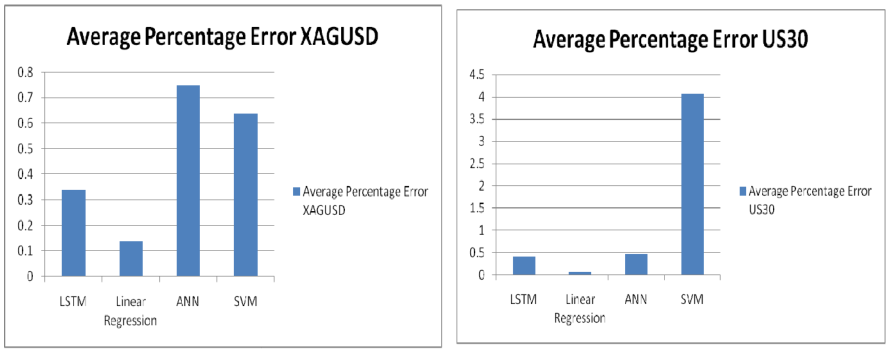


Fig.5. Comparison of absolute error of different models for the four datasets

6 Comparison with State-of-the-art methods

The proposed model, which combines linear regression with the identification of significant support and resistance levels, demonstrates better prediction accuracy when compared to traditional linear regression models. The model performs better at predicting stock market prices by utilizing support and resistance levels that are taken from past price data.

The study also emphasizes how this method might be applied to other financial time series datasets, indicating that its uses are not limited to the particular context of gold price prediction. This suggests that the approach might be expanded to include a wider range of financial instruments and markets, which could result in improved forecasting abilities in a variety of financial industry domains.

By comparing the two models, it is also possible to see how much better linear regression is than LSTM models in terms of model stability, data availability, simplicity, interpretability, and efficiency. While LSTM models are renowned for their ability to capture complex temporal dependencies, the research suggests that linear regression may be more suitable for short-term price prediction tasks in financial markets, where rapid decision-making and transparency are paramount.

7 Conclusion

Our study demonstrates the effectiveness of combining linear regression with support and resistance levels for enhancing stock price prediction accuracy. By leveraging historical price data and technical analysis indicators, our approach offers valuable insights into market dynamics and trends, empowering investors and financial institutions to make more informed decisions. Furthermore, the simplicity and interpretability of linear regression models make them accessible to a wide range of users, highlighting their potential as a practical tool for short-term price prediction in financial markets.

References

1. Kumar, Deepak, Pradeepta Kumar Sarangi, and Rajit Verma. "A systematic review of stock market prediction using machine learning and statistical techniques." Materials Today: Proceedings 49 (2022): 3187-3191.
2. Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. Journal of King Saud University-Science, 34(4), 101940.
3. Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. Procedia computer science, 170, 1168-1173..
4. Srivastava, P., & Mishra, P. K. (2021, October). Stock market prediction using RNN LSTM. In 2021 2nd Global Conference for Advancement in Technology (GCAT) (pp. 1-5). IEEE..
5. Bathla, G., Rani, R., & Aggarwal, H. (2023). Stocks of year 2020: prediction of high variations in stock prices using LSTM. Multimedia Tools and Applications, 82(7), 9727-9743.
6. Pramod, B. S., & Pm, M. S. (2021). Stock price prediction using LSTM. Test Engineering and Management, 83, 5246-51.
7. Sen, J., Mehtab, S., & Dutta, A. (2023). Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. Authorea Preprints.
8. Varadharajan, V., Smith, N., Kalla, D., Samaah, F., Polimetla, K., & Kumar, G. R. (2024). Stock Closing Price and Trend Prediction with LSTM-RNN. Journal of Artificial Intelligence and Big Data, 4, 877.
9. Varadharajan, V., Smith, N., Kalla, D., Samaah, F., Polimetla, K., & Kumar, G. R. (2024). Stock Closing Price and Trend Prediction with LSTM-RNN. Journal of Artificial Intelligence and Big Data, 4, 877.
10. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320.
11. Uddin, R., Alam, F. I., Das, A., & Sharmin, S. (2022, February). Multi-Variate Regression Analysis for Stock Market price prediction using Stacked LSTM. In 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET) (pp. 474-479). IEEE.
12. Hindrayani, K. M., Fahrudin, T. M., Aji, R. P., & Safitri, E. M. (2020, December). Indonesian stock price prediction including covid19 era using decision tree regression. In 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 344-347). IEEE.
13. Rajeswari, K., Vispute, S. R., Gangatirkar, S., Dixit, K., Chaudhari, S., & Bhole, S. (2022, December). Comparative Analysis of various Machine Learning Algorithms for Stock Price Prediction. In 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) (pp. 1-6). IEEE.
14. Sanapala, S., Reddy, V. A., Choudhury, S. S., Akshaya, V. V., & Varma, V. M. (2023, November). Optimising Trading Strategies using Linear Regression on Stock Prices.

12

In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-6). IEEE.