JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

STOCK PRICE PREDICTION USING LINEAR REGRESSION: A MACHINE LEARNING APPROACH

[1]Pavana,[2]Prof.Dikshit M B,[3]Prema Sharadi T, [4]Shabari

[1]Final year B.E. Student, [2] Assistant Professor, CSE, [3]Final year B.E. Student, [4]Final year B.E. Student
[1]Department of Computer Science & Engineering,
[1]Srinivas Institute of Technology, Mangaluru, India.

Abstract: This project explores the application of linear regression, a fundamental machine learning algorithm, for stock price prediction. The study involves collecting and preprocessing historical stock data, selecting relevant features, and training a linear regression model. The model's performance is evaluated using metrics such as Mean Squared Error (MSE) and R-squared, and its accuracy is compared with other simpler baseline methods. The findings highlight the computational efficiency of linear regression but also reveal its limitations in capturing the complex and non-linear dynamics of the stock market. This research provides a foundational understanding of linear regression's applicability to stock price prediction and offers insights for future exploration of more sophisticated machine learning models to enhance prediction accuracy.

Keywords: Machine Learning, CNN, OpenMP, VGG16.

I. INTRODUCTION

Stock price prediction is a crucial yet challenging task in financial markets due to the volatile and complex nature of stock prices. Traditional approaches often rely on subjective analysis, which may lead to inconsistent results. This project aims to address these challenges by leveraging machine learning, specifically linear regression, to predict stock prices. Linear regression, being a fundamental and computationally efficient model, provides a baseline to understand the dynamics of stock prices. By analyzing historical data, identifying key features, and evaluating the model's performance, this research explores the applicability and limitations of linear regression in stock prediction. The insights gained will serve as a foundation for developing more sophisticated predictive models in the future. Accurately predicting stock prices is a significant challenge due to the volatile, complex, and non-linear nature of financial markets. Existing methods, ranging from simple technical indicators to advanced machine learning models, often fail to deliver consistent and reliable predictions. This project aims to address this issue by evaluating the applicability of a linear regression model for stock price prediction, identifying its strengths and limitations, and comparing its performance against other simpler predictive methods. The goal is to understand the feasibility of using linear regression as a baseline model for this complex task and to explore areas where it falls short in capturing market dynamics.

This project proposes the use of linear regression as a baseline approach to stock price prediction. The solution begins with collecting and preprocessing historical stock data, ensuring it is clean, normalized, and free of missing values. Feature selection is then performed to identify critical variables, such as historical prices, trading volume, and market indicators, that significantly influence stock prices. A linear regression model is trained using these selected features, leveraging its simplicity and computational efficiency to predict future prices. The model's performance is evaluated using metrics like Mean Squared Error (MSE) and R-squared to assess its accuracy and reliability. Additionally, the results are compared with other simpler predictive methods, such as moving averages, to understand the strengths and limitations of linear regression in capturing stock market dynamics. This approach provides a foundation for understanding the feasibility of linear regression in stock price prediction and paves the way for exploring more advanced machine learning models in the future.

This project focuses on developing and evaluating a linear regression model for stock price prediction. It covers the processes of data collection, cleaning, normalization, and feature selection to prepare high-quality inputs for the model. The study also includes training the model, measuring its performance using standard metrics like MSE and R-squared, and comparing its effectiveness with other basic prediction techniques. By understanding the strengths and limitations of linear regression in this context, the project lays the groundwork for future exploration of more sophisticated machine learning approaches to improve prediction accuracy in stock market applications.

The implementation of a stock price prediction system involves integrating various components such as hardware resources, software tools, and a robust machine learning algorithm. This project aims to leverage historical stock price data to predict future trends using Linear Regression, a widely used supervised learning algorithm. Accurate predictions of stock prices can help investors, analysts, and financial organizations make informed decisions, minimizing risks and maximizing returns.

The Linear Regression algorithm is the backbone of the stock price prediction system. It is a supervised machine learning algorithm used to predict a continuous output, making it ideal for estimating future stock prices based on historical data.

Linear regression works by establishing a relationship between the input features (e.g., previous stock prices, moving averages) and the output variable (future stock price). The algorithm assumes that there is a linear relationship between these variables, meaning that changes in the input data correspond proportionally to changes in the predicted output.

In this project, the historical stock price data serves as the input for training the model. The algorithm "learns" patterns from this data, identifying trends and relationships that influence stock prices. For example, the closing prices of previous days might show a strong correlation with the next day's price, and linear regression uses this relationship to make predictions.

The training process involves analysing the input data and adjusting the model's parameters to minimize the error between the predicted stock prices and the actual prices. Once the training is complete, the model is tested on unseen data to evaluate its accuracy and generalization capabilities. The simplicity and efficiency of linear regression make it a preferred choice for predicting stock prices, particularly when working with time-series data.

One of the key advantages of linear regression is its interpretability. Unlike complex machine learning models, the relationships identified by linear regression are easy to understand and visualize. This makes it an excellent starting point for building prediction models and analysing how different input features contribute to stock price movements.

By leveraging the linear regression algorithm, the project can analyse historical stock trends and make accurate predictions about future stock prices, providing valuable insights to investors and decision-makers.

II METHODOLOGY

In the context of a stock price prediction system using linear regression, it involves defining how the system will collect, process, and analyze stock market data, as well as how it will present predictions to users. A well-structured design ensures efficient data handling, accurate predictions, and a user-friendly interface.

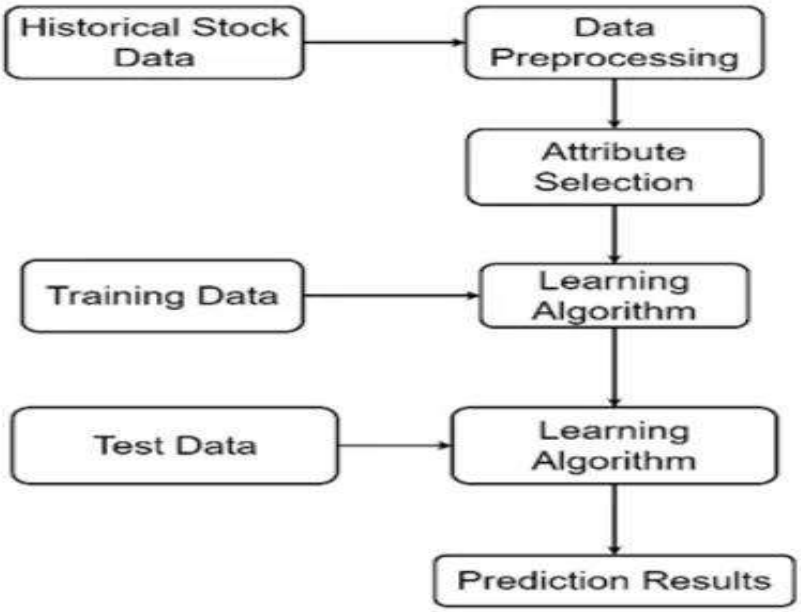


Fig.1: Linear regression prediction flow chart

The flowchart illustrates the process of building and implementing a stock price prediction system using historical stock data. The process begins with Historical Stock Data, which serves as the raw input for the system. This data includes stock prices, volumes, and other relevant market indicators.

The figure 2 Shows the flow diagram of model operations.

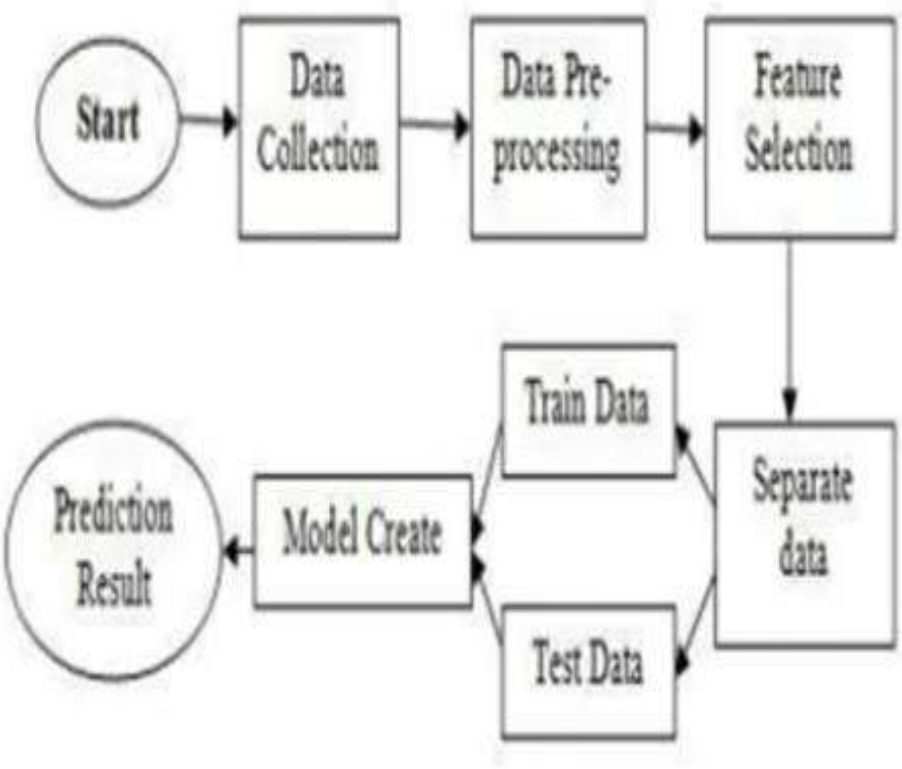


Fig.2: Methodology.

The flowchart outlines the methodology for a stock price prediction system, beginning with a clear sequence of steps thatensure systematic data processing, model training, and result generation. The process starts at the start stage and progresses to Data Collection , where relevant historical stock market data such as stock prices , volume , and time-based information is gathered from reliable sources. This collected data serves as the foundation for building the prediction model.

The stock price prediction system is designed through a series of well-structured steps aimed at building a reliable and accurate forecasting model. It starts with data collection from trusted financial platforms such as Yahoo Finance, Quandl, or Alpha Vantage, gathering essential information like historical prices, volumes, and timestamps. This raw data is then processed through data preprocessing, where missing values are handled, data is normalized to a common scale, and outliers or noise are removed to ensure data consistency and accuracy. Following preprocessing, feature selection is carried out to extract meaningful indicators such as moving averages, volatility, and lagged price features. These selected features help improve the model's predictive power by reducing redundancy and focusing on relevant variables. The refined data is then divided into training and testing sets—training data is used to build the model, while testing data evaluates its performance.

Using linear regression, the system identifies relationships between the independent variables (features) and the dependent variable (stock price). The model learns by minimizing the error between actual and predicted values. After training, the model is evaluated using performance metrics like Root Mean Square Error (RMSE) and R^2 score to determine its accuracy and reliability.

Finally, the trained model is used to generate stock price predictions, offering insights into future trends. These predictions assist investors and traders in making informed financial decisions, highlighting the practical value of the system in real-world market analysis.

IV. RESULTS AND DISCUSSION

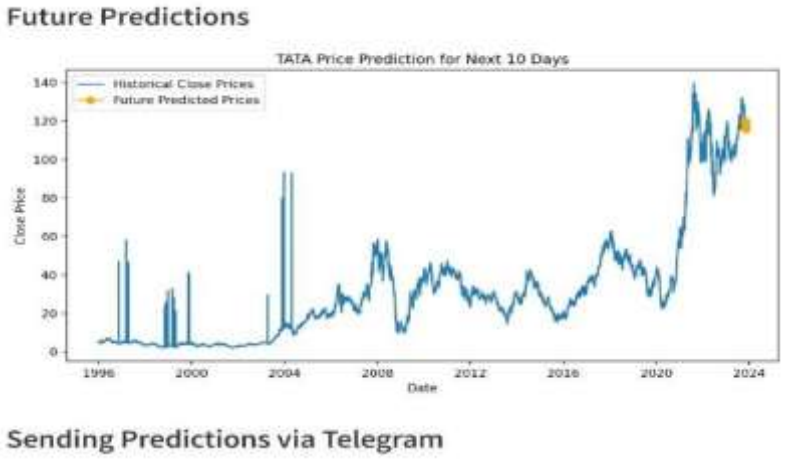
Linear regression is a widely used statistical method for predicting a dependent variable based on one or more independent variables. For this analysis, we used historical stock price data for each of the five companies over a defined period. The goal was to predict the future stock price of each company, using the available data to train the linear regression model.

Tata (TATA Motors)

Historical Data Analysis: The historical stock price data for Tata Motors was used to model its future price movements. The company experienced fluctuations due to factors such as market trends, production performance, and external economic factors.

Fig. 3: Historical Data

Prediction Results: The linear regression model for Tata Motors produced a predicted future price that closely tracked the general trend seen in the historical data, but with some deviations. The model captured the upward or downward trends to a certain degree but struggled to predict sharp price movements driven by market events.



Wipro

Historical Data Analysis: Wipro's stock data exhibited a more stable growth pattern compared to the others, reflecting its strong position in the IT sector.

Fig.4: Historical Data

Prediction Results: Linear regression was relatively successful in capturing the overall price trend for Wipro, showing a consistent prediction in line with historical movements. However, sudden market shifts led to some prediction errors.

V. CONCLUSION

Many individuals seek to predict future stock prices in hopes of increasing their wealth, driven by the ever-changing dynamics of the stock market. However, the volatile nature of the market has made it challenging for even the most advanced technologies, such as Deep Learning, AI, and Neural Networks, to deliver precise forecasts consistently. As a result, this review aims to delve into the current methods of stock market prediction, focusing specifically on the application of linear regression. Linear regression, a powerful supervised machine learning technique, identifies a linear relationship between independent and dependent variables. Research highlights that selecting the right dataset is critical for successful stock market predictions using linear regression. Experiments have shown that linear regression often outperforms other machine learning models in terms of accuracy. Nevertheless, many experts are still keen to explore the potential of neural networks for enhancing stock market predictions, as they may offer deeper insights and capture more complex patterns

VI. REFERENCES

[1] S. Shukla and B. Shah's "Stock Price Prediction Using Multiple Linear Regression Models" International Journal of Engineering Science Invention (IJESI). Published in October 2018.

[2] N. Rouf and S. Singh's "Stock Price Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions" MDPI. Published in November 2021.

[3] V. Gururaj, S. V R and A. Karthick's "Stock Price Prediction Using Linear Regression and Support Vector Machines". International Journal of Applied Engineering Research. Published in 2019.

[4] N. Karlsson's "Comparison of Linear Regression and Neural Network for Stock Price Prediction".

[5] P. Soni, Y Tewari, D. Krishnan- "Machine Learning approaches in Stock Price Prediction: A Systematic Review", Journal of Physics Conference Series. Published in 2022.

[6] M. Vijh, D. Chandola's "Stock Closing Price Prediction using Machine Learning Techniques". International Conference on Computational Intelligence and Data Science (ICCIDS).

[7] K. Daniel's "A Study on Stock Price Prediction Using Linear Regression and Random Forest Regression" (2022).

[8] O. E. Orsel, S. S. Yamada- "Comparative Study of Machine Learning Models for Stock Price Prediction" (2022).

[9] J. Sen- "Stock Price Prediction Using Machine Learning and Deep Learning Framework", 6th International Conference on Business Analytics and Intelligence (ICBAI). Published in December 2018.

[10] I. Kumar, C. Utreja's "A Comparative Study of Supervised Machine Learning Algorithms for Stock Market Trend Prediction". IEEE Conference. (2018).