Stock Price Prediction Using Linear Regression- A Machine Learning Algorithm

*Mukesh Raj\*, Ravi Vishwakarma, Ritik Sharma, Sandeep Sharma, Shivam Mishra*

Department of Computer Science and Engineering

JSS Academy of Technical Education, Noida, 201301, India

**Abstract**

Stock price prediction is a crucial aspect of financial markets, impacting investors, traders, and policymakers. This review paper aims to provide a comprehensive overview of the application of linear regression models in predicting stock prices. We examine the historical progression of research in this field, discuss the fundamental concepts of linear regression, explore data sources and feature selection techniques, evaluate model performance metrics, and analyze challenges and limitations. Additionally, we highlight recent advances and innovations, compare linear regression with other predictive modeling techniques, and propose future directions for research. To solve this problem we are developing one stock price prediction website using Python and Linear Regression which is one of the best Machine Learning statistical method for predictive analysis. We are using historical Data for the prediction. Finding a method to use linear regression models to obtain more precise values is the major goal. To acquire more precise results from the linear regression models, it is possible to change the dataset that will be used to train the models. The purpose of this paper is to demonstrate that linear regression is the most appropriate and effective method for forecasting stock market analysis

**Keywords***.* Machine Learning, Linear Regression, Python, Django framework, Yahoo Finance.

**Introduction**

For many business analysts and researchers, forecasting the stock market price is always a challenge. Stock market prices estimation is not only an interesting but also challenging area of research. Predicting the stock market with full accuracy is very difficult as external entities such as social, psychological, political and economic have a great and substantial influence on it.

Investors must predict the future stock value of companies in order to obtain high profits. Various prediction techniques have been developed to do predictions on the stock market accurately. There were two methods widely known as conventional methods at the time when there were no computational methods for risk analysis.

Two methods that are widely used in general are namely Fundamental Analysis and Technical Analysis.

**Fundamental Analysis:** To determine accurate product value, reliable and accurate information on the financial report of the company, it is necessary to have competitive strength and economic conditions in which they are interested. For the long-term predictions, Fundamental analysis is useful and the advantages are due to their systematic approach and their ability to predict changes.

**Technical Analysis:** “The idea behind technical analysis is that investors’ constantly changing attributes in response to different forces/factors make stock prices trends/movements”. Different technical factors of quantitative parameters can be used for analysis, such as trend indicators, lowest and highest daily values, indices, daily ups and downs, stock volume, etc.

**Literature Review**

Here are summaries of some research papers on stock price prediction using linear regression algorithm as a machine learning approach:

* 1. The purpose of this paper is to demonstrate that linear regression is the most appropriate and effective method for forecasting stock market analysis. The primary objective of this project was to predict stock market prices at a higher accuracy rates and it had been achieved at an approximate accuracy ranging in between 75% to 85 %.
  2. In this research paper, the effectiveness of the linear regression model was examined for used car price prediction. The linear regression model was applied to the data set that includes the features and price information of vehicles in Turkey as the year 2020. As a result, when we selected 1/3 of the data set as the test data, it was observed that the R2 score for the prediction success of model was 73%. To improve the effectiveness of the results the dataset could be extend or preprocessing part be detailed.
  3. This paper will use the concept of machine learning to predict the stock prices of three listed companies based on three different regression models (i.e., OLS, Ridge and XGBoost). According to the analysis, the OLS model and the Ridge model are very accurate in predicting stock prices, especially in the low and middle price ranges. In contrast to these typical linear regression models, the XGBoost model is not as accurate in predicting stock prices and even has a significant prediction bias in the high price range. These results will enable subsequent research to make better choices when selecting models for forecasting, especially for data sets with different characteristics.
  4. In this paper, the stock price prediction of regression has applied SVM and compared them with different models using algorithms such as random forest and multilayer perception. Based on the result of experiment 90 percent of training set gained the most accurate outcome that the value of mean absolute error was 0.0005 and with the same training set proportion of 70%, SVM got the highest accuracy among all the six models which has the relative absolute error of 0.94%.
  5. In this paper many regression-based models are compared based on their R- squared value. In regression modeling, the R-squared (R2) value represents the percentage of difference explained by the independent variable(s). A higher (R2) value near to 1 indicates better performance. Obtained R-Squared values 0.832, 0.832, 0.574, 0.838, 0.825, and 0.815 for SVR, RF, KNN, LSTM, GRU, and LSTM with GRU, respectively. Consequently, the most effective model for correctly predicting stock market closing prices is the LSTM learning model, which had the greatest R-Squared value of 0.838.
  6. This paper compares the degree of accuracy of ANN and ARIMA, which are considered to be the most commonly used and accurate models in stock price forecasting for the past 20 years, in predicting Nvidia's stock price. The time period chosen for the prediction is from June 2020 to June 2021, when Nvidia's stock price rises sharply, and the data is obtained from Kaggle and Yahoo Finance. According to the analysis, the ANN model is clearly capable of predicting the sharp fluctuations in Nvidia's stock price between 2020 and 2021, which also indicates the potential of artificial neural network-based models for stock price predictions.
  7. In this paper, a CNN-BiLSTM-Attention-based model is proposed to boost the accuracy of predicting stock prices and indices. First, the temporal features of sequence data are extracted using a convolutional neural network (CNN) and bi-directional long and short-term memory (BiLSTM) network. Comparing this method with the LSTM, CNN-LSTM, and CNN-LSTM-Attention models, it is found that the accuracy of stock price prediction is highest using the CNN-BiLSTM-Attention model in almost all cases
  8. In this paper, we introduce various techniques used to predict share price in the stock market and find R-square value for different stocks and then figured out that LSTM has highest value of R-square value.
  9. This paper first introduces four popular machine learning algorithms used for stock price prediction which are linear regression, support vector machines, artificial neural networks and long short-term memory and then compares them and finds that long short-term technique for very much suitable technique for predicting the stock prices.
  10. This research paper by Indu Kumar, Chetna Utreja etc. (2018): This study assessed the effectiveness of different machine learning techniques, including Forestry and Support vector machines.
  11. This research paper by Rashi Jaiswal, Brijendra Singh (2022) proposed a hybrid convolutional recurrent model architecture for stock price prediction based on exploitation of two different deep learning models properties into a single model. Hybrid Convolutional Recurrent Model consists of 1D-CNN and GRU where 1D-CNN is leveraged as feature extraction task and GRU for temporal regression. The proposed CNN-GRU Model outperform for stock price prediction. The performance evaluation and comparison analysis between the existing hybrid models and proposed hybrid model have also been done with experiments to justify.
  12. This research paper by A.Bhavani, A.Venkata Ramana etc. (2022) focuses on comparative analysis between LSTM and GRU in stock price prediction. LSTM and GRU are comparable, although GRU has fewer characteristics. GRU outperforms LSTM in terms of performance.
  13. This research paper by Amirali Rayegan, Ali Shri etc. (2022) has applied several machine learning techniques to develop a valid forecast consisting of linear models and various artificial neural networks.
  14. This research paper by Ranganath Kanakam, Dadi Ramesh etc. (2022) discussed about what is stock market prediction, how to predict with using Multiple Linear Regression and Support Vector Machine, analysed on some set of data and present results of our analysis.
  15. This research paper by Md. Ebtidaul Karim, Md. Foysal etc. (2022) proposed a novel hybrid deep learning model employing the bidirectional long short term memory (BiLSTM) and gated recurrent unit (GRU) network. Individually the long short memory (LSTM), Bi-LSTM, GRU and traditional neural network (NN) modules are implemented to forecast the stock price.
  16. This research paper by Atharva Joshi, Dr. Pradeep K. Deshmukh etc. (2022) compares the performance of Vanilla LSTM with the other LSTM variant called Peephole LSTM. The experiments revealed that Peephole LSTM outperforms Vanilla LSTM by a significant margin.
  17. This research paper by Ranjan Kumar Roy, Koyel Ghosh etc. (2021) tried to incorporate the issue when someone implements it as per the model outcome. It cannot give the proper result when someone implements it in real life since capital market data is very sensitive and news-driven. To avoid such a situation, we use the hedging concept when implemented.
  18. This research paper by Bipin Aasi, Syeda Aniqa Imtiaz etc. (2021), a Multivariate Multistep Output Long-Short-Term-Memory (MMLSTM) model is proposed to provide a one-week prediction on the stock close value for the technology company, “Apple Inc.” with the stock name “AAPL”.
  19. This research paper by Xinrong Zhou (2021) aim to compare both CNN and LSTM on the stock price prediction problem. Combined LSTM-CNN model is proposed to achieve a better performance, which avoids the layback of LSTM and increase the robustness of CNN.
  20. This research paper by Ramkrishna Patel, Vikas Chaudhary etc. (2021) discusses various machine learning approaches (Supervised or Unsupervised) and methods through which the investors get to know the stock prices increase or decrease.

**Methodology**

The other tools and technology used in the given website are html and CSS which are used for frontend. The library named scikit learn is used as it has all machine learning algorithms and functions inbuilt in it. The framework Django which comes under Python is used at the backend to access web services and resources which is inbuilt in python programming language. To fetch the data for the website which should be reliable and accurate, Yahoo finance site is used and connecting it with API, So the prediction of stocks of all the companies which are there in yahoo finance can be done through our website.

In the given project, stock prices are predicted using Linear regression algorithm in machine learning and data implementation using python tools and libraries like Scikit-learn, Numpy are done.

**Linear Regression:**

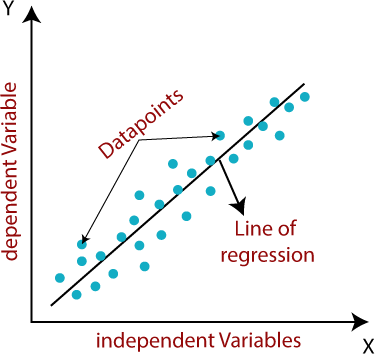
A linear regression contains the labelled data (supervised learning) which generates the relationship between independent and dependent variables using simple mathematical equations and thereby calculate the best fit line or line of minimum reluctance. This line can be used to calculate stock predictions using graph or curve analysis.

Linear regression is considered to be better than most of the other techniques as it is easier to Implement and contains basic mathematical and computational theory. It involves fitting a straight line to the given data points of independent variable(x) and dependent variable(y) whose slope is (m) and error is (e). This line is implemented by reducing the sum of squared differences between actual values understood in the above equations and diagrams:

y = mx + c + e (1)

here, ‘c’ is the intercept formed on dependent axis y. for multiple data sets with slopes m1, m2…mk

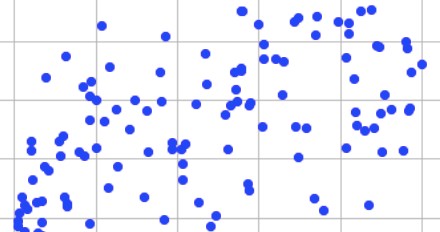
we can use-:

y = m1x + m2x …. mkx + c + e ( 2)

**Fig. 1.** Representation of best fit line of regression

Steps for implementation of Linear regression includes-:

* + 1. Importing necessary computing packages and creating the data from input files using Pandas library.
    2. Filter out the essential variables and generate a numpy array from Pandas data frame to store continuous set of data.
    3. Assigning the input variable (x, y) and performing recursive feature normalization, where we will compress our input variables to magnified values for faster speed and accuracy.
    4. Plotting the data sets using matplotlib and plotly tools through normalized data and generate the graph as shown below-:



**Fig. 2.** Normalized stock market data

* + 1. Training the algorithm and computing the hypothesis (approximate target value-y) for stock price vs time duration curve.
    2. Calculating the hypothesis hx and SSE (sum of squared error) cθ using the below equations-:

hx = θ0 x0 + θ1x1 (3)

where θ0 and θ1 have minimal error or deviation variables for linear fit and x0 tends to 1. SSE which is difference between hx and actual values can be calculated using-:

𝑛

cθ = ½ m∑

θ=0

(hθx – y)2 (4)

by reducing the SSE, we can optimize our results and best fit line.

* + 1. calculating the gradient descend and compute optimal θ0 , θ1 which are required for plotting best fit line using below equation-:

θ1 = θ0 –α Ω cθ (5)

Where θ0 and θ1 are current and next optimal positions, α is small step , Ω is direction and c(θ) is SSE..

* + 1. model testing, plotting best fit line and graph, deployment.

**Implementation Flow**

In order to develop GUI (graphical user interface) to make this predictor tool accessible to users, predictor is converted to a website which will show real time stock market prediction.

The first step in this is to collect and fetch the data from reliable source site Yahoo- finance and creating a DBMS (data base management system) using SQLite. Data preprocessing is yet another important step in which inaccuracy or inconsistencies in data had been removed and data is transformed into user understandable format using pandas library in python. In the backend part numPy, plotly, matplotlib are used for multiple operation execution. The prediction part is performed using linear regression technique of machine learning. To access the machine learning tools and resources for the website, scikit learn library and Django framework are used respectively. For UI/UX part Html5 and CSS3 versions are used. The website is maintained at regular intervals and fetch new and current status of data every time.



Collect Data

Data Pre-processing

Build Database

Choose ML Algorithm

Connecting to Web Server and API

Create Frontend and Backend

Update and Maintenance

Model’s Testing

Training Algorithm

**Conclusion**

Many people desire to forecast future stock prices in order to increase their fortune because the stock market is continuously changing. However, due to the volatile nature of the stock market, current solutions that employ cutting-edge technology like Deep Learning, AI, and Neural Networks have not proven successful in making precise forecasts. As a result, the analysis of the current methods for stock market prediction using a linear regression algorithm is the main goal of this review work. A supervised machine learning approach called linear regression establishes a linear relationship between independent and dependent variables. The choice of an appropriate dataset is crucial for successful stock market prediction using linear regression, according to a review of research papers. According to the experiments, the linear regression method performed better in terms of accuracy than other machine learning techniques. However, a lot of experts also stated that they planned to investigate neural networks' potential for stock market prediction in the future.

**References**

1. Sonali Antad, Saloni Khandelwal, Anushka Khandelwal, Rohan Khandare, Prathamesh Khandave, Dhawal Khangar, Raj Khanke “Stock Price Prediction Website Using Linear Regression - A Machine Learning Algorithm” Department of Engineering, Sciences and Humanities (DESH). Published in August 2023.
2. Sümeyra MUTİ1 , Kazım YILDIZ2 “Using Linear Regression For Used Car Price Prediction” International Journal of Computational and Experimental Science and Engineering (IJCESEN). Published in February 2023.
3. Yuhan Li “Stock Price Prediction based on Multiple Regression Models”. Beijing University of Technology, Beijing 100083, China. Published in April 2023.
4. Chuchu Qi “Stock price prediction of Chinese company using support vector machine”. Shanghai University, Shanghai, China. Published in March 2023
5. Muhammad Rehman , Muhammad Fuzail, Muhammad Kamran Abid, Naeem Aslam “Financial Prices Prediction of Stock Market using Supervised Machine Learning Models”, Department of Computer Science, of Engineering and Technology, Multan, Punjab, Pakistan. Published in May 2023.
6. Bowen Ma , Yanchi Yang , Jiaming Zhang and Keli Zhang “A Comparison of Stock Price Prediction with ANN and ARIMA”. Published in March 2023
7. Jilin Zhang, Lishi Ye and Yongzeng Lai “Stock Price Prediction Using CNN-BiLSTM-Attention Model”. Published in April 2023
8. S. S Zalte, P. N. Patil, S. N. Deshmukh “Study of Machine Learning Algorithms For Stock Price Prediction”. Published in February 2023
9. Yang Shi “Research on the Stock Price Prediction Using Machine Learning”, Weatherhead School of Management, Case Western Reserve University, Cleveland OH 44106 United States. Published in September 2023.
10. Sulochana Sonkamble , Vaibhav Vyas , Prathamesh Shimpi , Aniket Mule, Mihir Sonawane “Stock Price Prediction System”. Department of Computer Engineering, JSPM’s Rajarshi Shahu School of Engineering and Research, Pune, Maharashtra, India. Published in April 2023.
11. R. Jaiswal and B. Singh, "A Hybrid Convolutional Recurrent (CNN-GRU) Model for Stock Price Prediction," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 2022, pp. 299-304, doi: 10.1109/CSNT54456.2022.9787651.
12. A. Bhavani, A. V. Ramana and A. S. N. Chakravarthy, "Comparative Analysis between LSTM and GRU in Stock Price Prediction," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 532-537, doi: 10.1109/ICECAA55415.2022.9936434.
13. A. Rayegan, A. Shiri and B. Bahrak, "A comparative study of machine learning techniques for stock price prediction," 2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), Behshahr, Iran, Islamic Republic of, 2022, pp. 1-7, doi: 10.1109/ICSPIS56952.2022.10044075.
14. Kanakam, Ranganath & Dadi, Ramesh & Mohmmad, Sallauddin & Shabana, & Prakash, T.. (2022). Stock price prediction using multiple linear regression and support vector machine (regression). AIP Conference Proceedings. 2418. 020046. 10.1063/5.0082035.
15. Karim, Md & Foysal, Md & Das, Sunanda. (2022). Stock Price Prediction Using Bi-LSTM and GRU-Based Hybrid Deep Learning Approach. 701-711. 10.1007/978-981-19-3148-2\_60.
16. Joshi, Atharva & Deshmukh, Pradeep & Lohokare, Jay. (2022). Comparative analysis of Vanilla LSTM and Peephole LSTM for stock market price prediction. 1-6. 10.1109/IC3SIS54991.2022.9885528.
17. Roy, Ranjan & Ghosh, Koyel & Senapati, Apurbalal. (2021). Stock Price Prediction: LSTM Based Model. 143-151. 10.21467/proceedings.115.19.
18. Aasi, Bipin & Imtiaz, Syeda & Qadeer, Hamzah & Singarajah, Magdalean & Kashef, Rasha. (2021). Stock Price Prediction Using a Multivariate Multistep LSTM: A Sentiment and Public Engagement Analysis Model. 1-8. 10.1109/IEMTRONICS52119.2021.9422526.
19. Zhou, Xinrong. (2021). Stock Price Prediction using Combined LSTM-CNN Model. 67-71. 10.1109/MLBDBI54094.2021.00020.
20. Patel, Ramkrishna & Choudhary, Vikas & Saxena, Deepika & Singh, Ashutosh. (2021). REVIEW OF STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES. 10.1109/ICOEI51242.2021.9453099.