International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Predicting Stock Market Using Machine Learning Algorithms

*Vadla Uday Kumar Chary[1], M. Vasavi[2]

[1]M.Tech.-A.I. student, Department of CSE, CVR College of Engineering, Vastunagar, Mangalpally, Ibrahimpatnam, Telangana, India

[2]Assistant Professor, Department of CSE, CVR College of Engineering, Vastunagar, Mangalpally, Ibrahimpatnam,, Telangana, India

Article Info
Volume 9, Issue 5
Page Number : 160-167

Publication Issue :
September-October-2022

Article History
Accepted :09 Oct 2022
Published: 20 Oct 2022

ABSTRACT

Predicting how the stock market will move could be useful as a way for short-term investors to get early advice and as a way for long-term shareholders to get early warning of financial trouble. The most crucial consideration when choosing a forecasting approach is predicting accuracy. Since the past ten years, more research has been done to increase the forecasting models' accuracy. It can be very challenging to choose the right stocks that are ideal for investment. The primary objective of any investor should be to maximise returns. The goal of stock market forecasting is to estimate how stock prices on a given exchange will fluctuate in the future. If it were possible to accurately forecast the direction of stock prices, investors would be able to earn more. This research uses machine learning to greatly reduce the uncertainty of future trend predictions. We will improve the accuracy of stock market predictions by using the boosting models found in machine learning algorithms.

Keywords : Stock Market, Machine Learning Algorithm

I. INTRODUCTION

Predicting the performance of a stock market has historically been a difficult problem for statisticians and financial analysts. This forecast is based on the idea that investors should put their money into stocks that are expected to rise in value and liquidate those that are expected to decrease in value. There are typically two methods for forecasting the stock market. One such method is fundamental analysis, which looks at the strategy and basic data of a company, such as its market position, expenses, and annual growth rates. The second is called "technical analysis," and it uses data from past stock prices and values to predict future stock prices. In order to forecast the market, this study looks at past charts and patterns. In the past, financial experts could reliably foresee the direction of stock markets.

Stocks, also known as equity, are a type of security that represents a claim to a piece of a business. Accordingly, the stockholder is entitled to a portion of the company's assets and profits equal to the number of shares of stock owned by the stockholder. The plural of "shares" is "shares."

Forecasting the stock market is fraught with difficulty, and data scientists often run into roadblocks when attempting to create a predictive model. The correlation between investor psychology and market behaviour, as well as the volatility of the stock market, pose two major challenges: complexity and nonlinearity.

II. MOTIVATION

Since it could be worth billions, the world's largest financial institutions are investing heavily in studying stock market price prediction. It is a serious issue since there is no apparent solution, despite the fact that approximations can be made using a variety of machine learning approaches. A vast data set can be collected and analysed as part of the project, and a number of different methodologies can be used to train the programme and forecast possible results.

III. LITERATURE SURVEY

Due to its impact on financial difficulties and its ability to be predicted, the stock market is currently the subject of study in a wide range of disciplines. The ICECA 2017 Stock Market Forecasting International Conference on Electronics, Communication, and Aerospace Technology By employing Linear Regression, Reference: Dinesh Bhuriya et al. Our study's ultimate goal is to offer stock brokers and investors a hand up as they put their money to work in the stock market. Stock market business is a very complex and challenging process due to the dynamic nature of the stock market, so prediction plays a very important role in this business. Consistent with the aforementioned literature, our system makes accurate stock price predictions using a variety of regression techniques, including linear regression. After looking at the results of each method and comparing them based on the confidence values they gave, they decided that linear regression gave the best results.

Maqsood, H., et al. [2] Market fluctuations are affected not only by the nonlinearity of data and economic rules but also by public opinion and the state of the economy and government. They used linear regression, support vector regression, and deep learning to effectively predict stock prices based on Twitter data for events that happened between 2012 and 2016. The findings indicate that not all major events have a consequential effect on stock market forecasting. However, the effectiveness of prediction algorithms can be impacted by more significant local events. Some global events have been noted to affect stock markets in other countries as well. Predictions of the stock market now also incorporate data from news reports. The state of the country typically has a significant impact on the stock market, and the news can be a good place to learn about these developments. Machine learning methods can be used to anticipate trends in the news, which can then be applied to the stock market. One potential area of development is the application of such a massive Twitter dataset to the problem of predicting interest and exchange rates. Foreign exchange and interest rates are sensitive to market sentiment and can quickly change.

The idea of an improved AdaBoost algorithm model to pick outperforming stocks was proposed by Sun Yutong et al. [3]. This model analyses stocks in depth to select those with a return that is greater than the market average. These findings demonstrate the superior effectiveness and precision of the AdaBoost algorithm when choosing stocks. It has higher than average output reliability and return. When using this model, you won't have to worry about overriding or waiting around for hours while it trains. In reality, we typically only invest in high-performing stocks, so there is a wide range in the error costs associated with various categories. Therefore, in this paper, we give different kinds of errors different amounts of weight to improve precision and effectiveness. A dissatisfactory accuracy rate of 54.5% on the testing dataset led them to the conclusion that the sophisticated AdaBoost algorithm falls short. Also, they discovered that there is a significant influence of factors and that simple weak learner perform poorly towards a single factor, resulting in a slow decline in error rate. Although the factors are calculated in this model, they are not independent, which leads to fluctuating results and poor boosted classifier performance. In future projects, we will use methods like fundamental analysis and industry analysis to find stocks that look good.

According to J. B. Duarte et al [5], efficiency as a metric for judging the impact of investors' mentality on a simulated stock market. The goal of this research was to construct a scenario, grounded in the principles of behavioural finance that would allow for the representation of the behaviour of a market with multiple stock agents and the effect of the investor on the efficiency of that market. Changing the market's dispositions and the number of each type of agent can reduce the financial series' persistence level in the mixed cellular automaton model. It is possible to create a mixed learning cellular automaton model by instituting a system of punishments and incentives when the position taken by the behavioural agents is in response to the economic behaviour at the t+1 instance.

According to Mehar Vijh et al. [6], two main strategies have been proposed to forecast a company's stock price. When predicting the future price of a stock, technical analysis relies on information about the stock's past performance, such as its closing and opening price, volume traded, adjacent close values, etc. Qualitative analysis, on the other hand, is conducted by economic analysts based on external factors such as company profiles, market conditions, political and economic factors, and textual information in the form of financial news articles, social media, and even blogs. These days, stock price forecasts are made using cutting-edge artificial intelligence methods based on technical or fundamental analysis. In the case of stock market analysis in particular, the data size is enormous and the data structure is non-linear. A model that is both efficient and capable of discovering previously unnoticed patterns and intricate interrelationships in this massive dataset is required for processing data of this type. When compared to conventional methods, recent applications of machine learning in this field have been shown to increase productivity by an impressive 60–86%.

IV. Proposed Methodology

In the proposed system we have collected the dataset from the Tesla Stock Exchange from October 19, 2015 to October 16, 2020. The stock data obtained from Tesla contains the following parameters:

- Date: It has the date of the stock trading
- Open: The first price at which a security is traded during a business day is known as its "opening price."
- High: The period's highest closing price for a given stock
- Low: The period's lowest closing price for a given stock
- Close: The closing price is the last price at which an asset was traded on a given business day, as indicated by the word "close."
- Adj Close: When corporate actions are taken into consideration, the closing price of a stock is "adjusted" to reflect its true value
- Volume: The volume of a security is the number of shares that were bought and sold in a certain amount of time.

The data set is split into a training part, which has 90% of the data, and a testing part, which has 10% of the data.

Data Preprocessing

Data pre-processing is done by various methods in those we are using MinMaxScaler and it is imported from the scikit learn.

MinmaxScaler: Transform features by scaling each feature to a given range. This estimator scales and translates each feature individually such that it is in the given range on the training set, e.g. between zero and one.

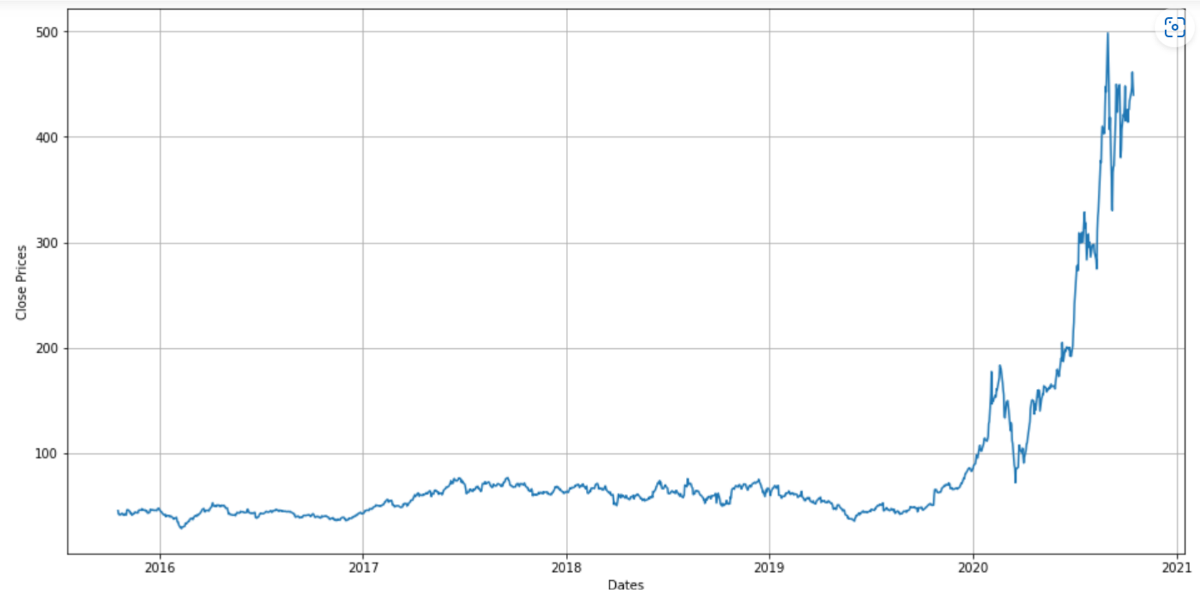

Figure 1: Visualisation of whole stock close data

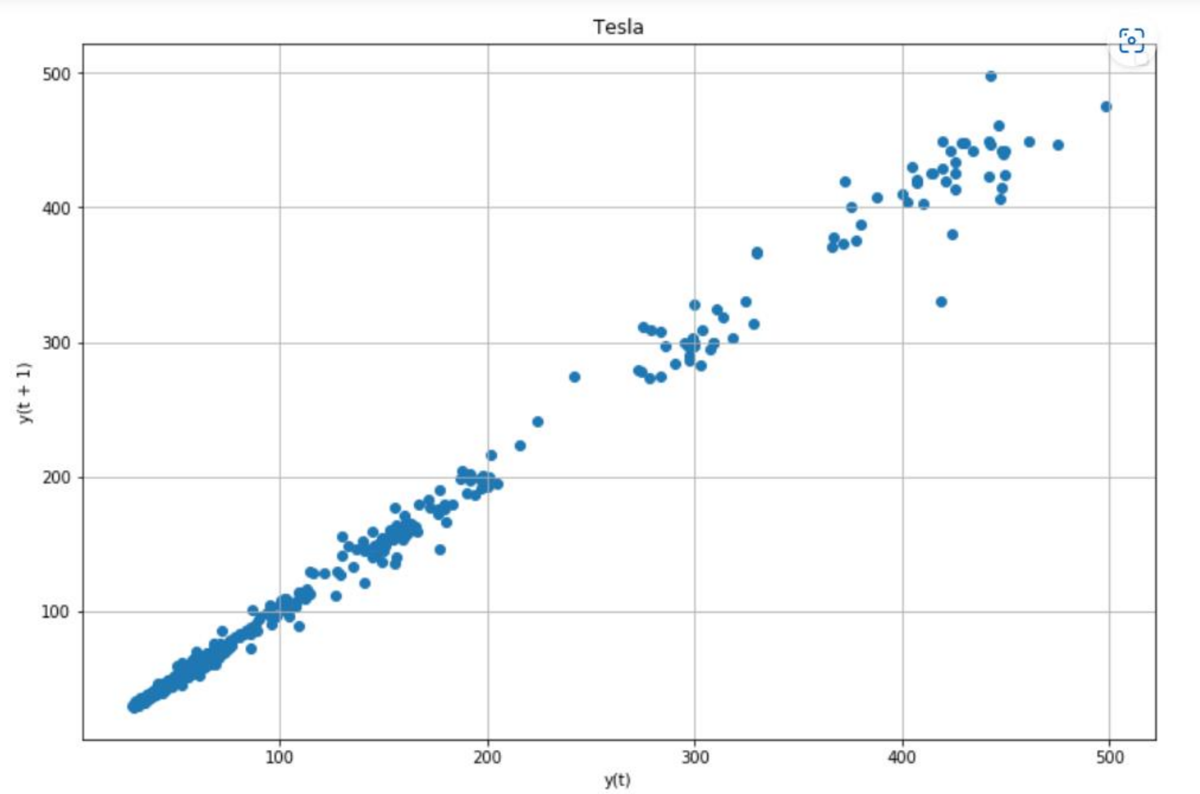

Figure 2: correlation of close coloumn in a Data Frame.

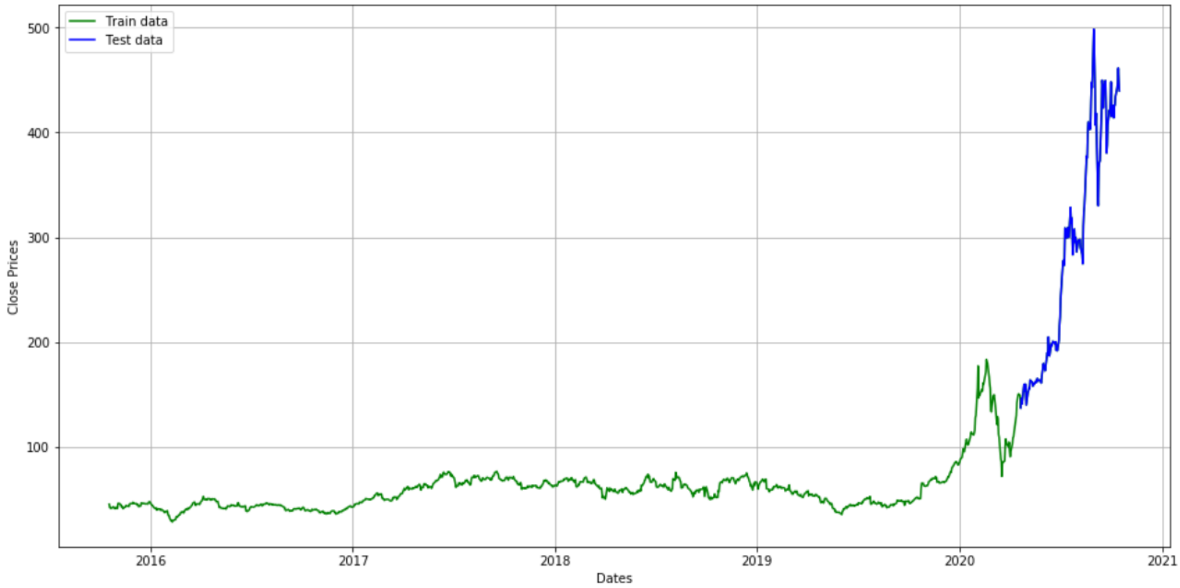

Figure 3: Visualization of training and testing data

V. PREDICTION MODELS

Some of the machine learning techniques that are used here are Linear Regression, Ridge Regression, AdaBoost, CatBoost, LightBoost, and XgBoost.

Regression model

For regression to work, a dataset must already have target values defined. Plus, the final product can be improved upon by including additional data. Regression can establish a pattern in the relationship between predictor and outcome values. Applying this pattern to other datasets where the desired values are unknown will yield similar results. For regression, you'll need two sets of information: training data and test data. Here, we conduct an analysis using linear regression. As a first step, we split the data into a training set and a test set. After that, we use the training phase to initiate analysis and establish the model's parameters. The data was split into a 90% training set and a 10% test set.

Linear Regression

The supervised learning technique known as linear regression has been implemented in machine learning. It's tasked with carrying out a regression analysis, which it does. Using independent variables, a target prediction value is modelled in a regression analysis. Most commonly, it is employed to investigate the connection between observables and future predictions.

Ridge Regression

Multiple regression Data with multi-collinearity can be analysed with a technique called ridge regression. It works best when there are more predictor variables than observations in the data set. As a result, the second-best case scenario is multi-collinearity in a set.

ADA Boost

The boosting method is a process that takes a group of weak learners and turns them into strong ones. To improve upon the predictions of every individual learning strategy, AdaBoost employs a specialised form of boosting known as an ensemble model. Boosting's main purpose is to teach less capable learners how to improve their accuracy by revisiting their old forecasts in a step-by-step process. This model is a meta-predictor, which means that it fits a model to the primary dataset and then uses that model to fit additional copies of itself to the primary dataset. By adjusting sample weights in response to the actual forecasting error, the training process helps the model zero in on the most challenging data points.

XG Boost

The decision tree-based ensemble model XG Boost is relatively new. The principles of boost for slow students are used here. The XG Boost model was introduced to outperform and outrun its tree-based counterparts in terms of performance and speed. The XG Boost method has many advantages over other similar approaches, including regularisation to prevent over-fitting, built-in cross-validation capability, expert handling of missing data, catch awareness, parallelized tree building, and tree pruning.

Light GBM

The model's efficiency is improved and memory consumption is decreased with Light GBM, a gradient boosting framework that uses decision trees. A One-Sided Sampling Method Using Gradients for Light-Weighted GBM The importance of various data instances in calculating information gain varies.

Cat Boost

Gradient-boosted decision trees are the foundation of Cat Boost. When trained, a series of decision trees are constructed in order. Loss is minimised with each new tree constructed. The initial settings determine how many trees will be created.

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section visualizes the performance of the model and communicates the report of the accuracy of the machine learning models.

The experiment were performed using Jupyter notebook. These experiments were supported on a system that is equipped with intel i5 10^{th} Generation processor having 8GB RAM and running on 64-bit Windows 10 operating system.

Dataset Description

In the proposed system we have collected the dataset from the Tesla Stock Exchange from October 19, 2015 to October 16, 2020. The total rows in the dataset has 1259. The stock data obtained from Tesla contains the following parameters:

- Date: It has the date of the stock trading
- Open: The opening price is the price from the first transaction of a business day
- High: The highest price at which a stock has traded during the time period
- Low: The highest price at which a stock has traded during the time period
- Close: The closing price is the price from the end transaction of a business day
- Adj Close: The adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions
- Volume: Volume is the number of shares of a security traded during a given period of time

The dataset is splitted in the form of training and testing, training consists of 90% and testing as 10%.

Table 1. Results of machine learning models

Methods	Accuracy	R Square
Linear Regression	98%	98%
Ridge Regression	89%	86%
AdaBoost	97%	97%
Xg Boost	99%	99%
Light Boost	84%	64%
Cat Boost	84%	86%

The XG boost Classifier technique has outperformed the other techniques and achieved a prediction accuracy of 99%.

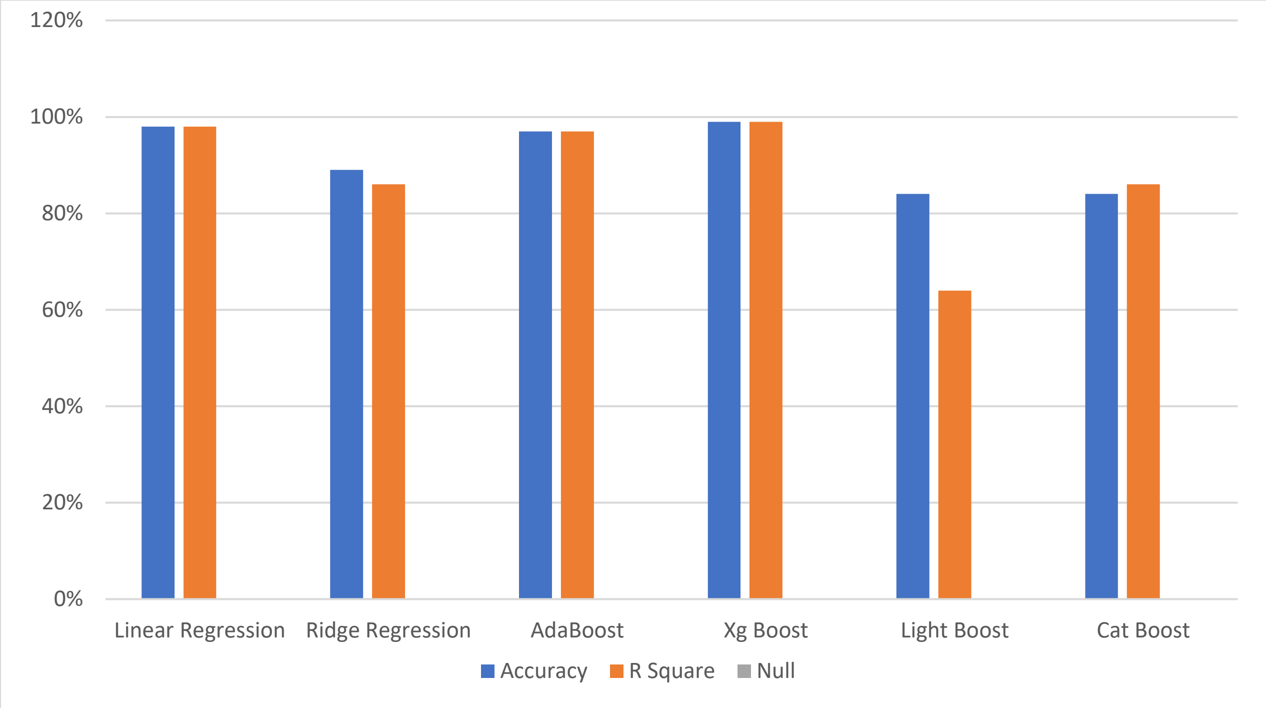

Figure 4: Accuracy and R squared chart visualization.

Table 2. Comparative analysis

Author and year	**Title**	**Techniques**	**Accuracy**
Dinesh Bhuriya, Girish Kaushal, Ashish Sharma and Upendra Singh,2017 [1]	Stock Market Predication Using a Linear Regression	Linear regression Polynomial RBF	97% 46% 56%
Sun Yutong, Hanqing Zhao, 2015 [3]	Stock Selection Model Based on Advanced AdaBoost Algorithm	AdaBoost	54%
Proposed work	Predicting Stock Market Using Machine Learning Algorithms	Linear Regression AdaBoost Ridge Regression Xg Boost Light Boost Cat Boost	**98%** **97%** 89% 99% 84% 84%

From the above existing papers by comparative analysis the Linear regression and AdaBoost algorithms are got better accuracy.

Dinesh Bhuriya et al. in [1] calculated the level of confidence in linear regression as 97%. Confidence defines the probability of the single input occurred in the form of a event. If a class has high probability then it has high confidence. In proposed methodology accuracy of Linear regression is predicted rather then the confidence and got better accuracy then existing once. Sun Yutong et al. in [3] used advanced adaptive boost algorithm. Factors calculated in this model are not independent which makes result unstable and boosted classifier perform badly. In proposed methodology adaboost algorithm used to predict and R square is calculated and achieved better results.

VII. CONCLUSION AND FUTURE SCOPE

To predict stock moment several machine learning methods are employed like Linear Regression, Ridge Regression, Ada boost, XG Boost, Cat Boost and Light Boost. Results show that boosting Classifier technique has outperformed the other techniques and achieved a prediction accuracy of 99%.This work can be extended by implementing some other Machine Learning models to data you may improve the findings and accuracy.

VIII. REFERENCES

[1]. Dinesh Bhuriya, Girish Kaushal, Ashish Sharma, Upendra Singh “Stock Market Predication Using a Linear Regression” International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

[2]. H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim, and K. Muhammad, “A local and global event sentiment based efficient stock exchange forecasting using deep learning” Int. J. Inf. Manage., vol. 50, pp. 432–451, Feb. 2020.

[3]. Sun Yutong, Hanqing Zhao “Stock Selection Model Based on Advanced AdaBoost Algorithm” 7th International Conference on Modelling, Identification and Control (ICMIC 2015).

[4]. M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, “Evaluating multiple classifiers for stock price direction prediction,” Expert Syst. Appl., vol. 42, no. 20, pp. 7046–7056, Nov. 2015

[5]. J. B. Duarte Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez, “Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency” Contaduriay Administracion, vol. 62, no. 4, pp. 1361–1376, Oct. 2017.

[6]. Mehar Vijh, Deeksha Chandola,Vinay Anand Tikkiwal,Arun Kumar “Stock Closing Price Prediction using Machine Learning Techniques” International Conference on Computational Intelligence and Data Science(ICCIDS 2019).

[7]. J. B. Duarte Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez, “Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency,” Contaduría y Administración, vol. 62, no. 4, pp. 1361–1376, Oct. 2017.

[8]. Y. Chen and Y. Hao, “A feature weighted support vector machine and Knearest neighbor algorithm for stock market indices prediction,” Expert Syst. Appl., vol. 80, pp. 340–355, Sep. 2017.

Cite this article as :

Vadla Uday Kumar Chary, M. Vasavi, "Predicting Stock Market Using Machine Learning Algorithms", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 5, pp. 160-167, September-October 2022.

Journal URL : https://ijsrset.com/IJSRSET229533