STOCK PREDICTION USING MACHINE LEARNING ALGORITHM

Manthan Nayani*1

*1Student, Department Of Computer Engineering, Pacific School Of Engineering, Surat, Gujarat, India.

ABSTRACT

The projection of the Stock market has always caught the attention of many analysts and researchers. Prediction of any stock price is an arduous task as there are lots of variables involved to understand the particular situations in the market. Artificial Intelligence is supposed to be a predictive approach that analyzes the stock market a wide range of not only fundamentals and technical patterns but it has the ability to perceive the financial ratio of the company price prediction required tremendous computational power. This paper aims at outlining the research work for Stock Market Prediction with a special focus on Machine Learning algorithms such as Support Vector Machine (SVM) and Linear Regression by which one can get an output of price prediction.

Keywords: Artificial Intelligence, Machine Learning, Support Vector Machine, Stock Trend, Stock Market Predication, Trading Pattern.

I.   INTRODUCTION

The stock market is the place where various facilities are provided to the Individual Investor, Mutual Fund, Financial Institutions, and Foreign Institutional Investors (FII's) to sell and buy any stock, commodities, and any other assets in the market. To put it another way, Stock Market is a system for trading various securities and bonds without any hurdle. In India, if anyone wants to list their company in the stock market then they have mainly three options Bombay Stock Exchange (BSE), National Stock Exchange (NSE), and Calcutta Stock Exchange (CSE). There are two popular indexes namely Sensex and Nifty Fifty which represent an overall trend of the market. The trend in the stock market affects the profitability of the stockholder because firstly if the value of any stock is going upward then the stockholder makes the profit as they have bought the stock at a lower rate and now it is going up. Secondly, if the value of any stock is going down then stockholders have to bear the losses as they have bought the stock at a higher rate and now it is below the buying price. So, every participant wants to buy any stock at a lower rate and sell it at a higher price. Investors have to do a top-notch analysis of any company which involves fundamentals analysis and technical analysis. Fundamental analysis, it a method company's intrinsic value of particular will be evaluated. It also considers many financial ratios such as Price-to-earnings(P/E), debt-to-equity ratio(D/E), price-to-book value (P/BV), , return on equity (ROE), operating profit margin (OPM) all of this will eventually improve the process of making money in the long run. Technical analysis is the method used by traders to forecast the future price of a stock by analyzing past trading activity. The objective of the analysis is to reach a number corresponding to the actual price of the stock to assure whether the actual price is undervalued or overvalued. Chart patterns and statistical numbers are used extensively by technical analysts and this makes technical analysis strenuous and more complex computation. Here AI plays a crucial role in the prediction of stock price. Advance Artificial Intelligence (AI) is supporting humans to a great extent with its intelligence. Various algorithms are available in AI which is used to predict the future.[1] Stock market prediction is indeed a tough task however, bow out with profit from the stock market is practicable with the advancement in AI. Many researchers take advantage of AI and implement a different mechanism to build a good decision-making system with the help of available past data. Fearful, greediness, intimidation, and many more human sentiments will be totally removed while predicting the stock price in the case of AI. The computative method will have a higher chance of predicting accurate prices because it eliminates all human emotion and wholly works on computational methods and historical data. They are helpful tools for people navigating the investment and risk assessment decision-making process.[2] So, algorithmic trading is much more efficient and productive that's why humans are more dependent on this type of method to make more money. A trend that has been increasingly explored is the use of complex techniques for preprocessing the input data, which facilitates the execution of Machine Learning algorithms and increases its accuracy, since the noisy data tend to be eliminated, leaving only the most relevant data. Machine Learning is one type of AI which allows the software application to become more precise at predicting output without being explicitly programmed to perform. Generally, Machine Learning uses historical data as input to provide

output as a new predicted value. A Machine Learning procedure that will be trained from the historically available information and then uses the appropriate knowledge that it has gained for the end result which is to predict the accurate output. In this study, we will use an ML technique called Support Vector Machine (SVM) and Linear Regression.

II. METHODOLOGY

This paper, analyzing the stock market variables using linear regression and support vector machine, the technical research of stock market operations. It is a more advantageous decision-making process using algorithms to evaluate different statistics. figure 1 shows all the processes step by step to predict the stock market price. These steps can be used in almost every possible method used to predict stock price.
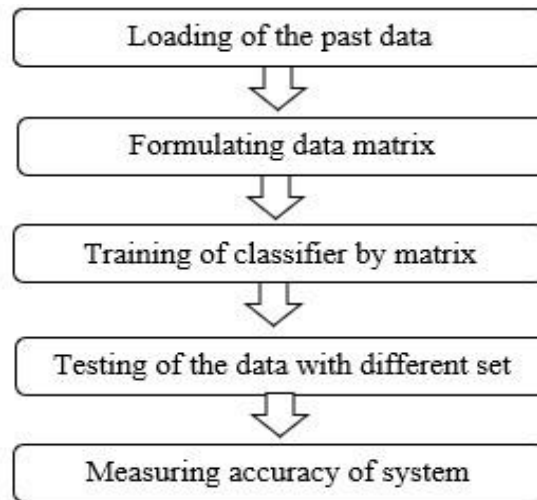

Figure 1: The flow chart used in stock market price prediction.

2.1 Support Vector Machine: -

SVM algorithm manifests the data by bringing out output separating hyperplane. SVM is considered to be one of the most suitable algorithms available for time series prediction. This algorithm can be connected to the presumption of the stock trade by taking advantage of the absolute parameters of any corporation such as cost per profit propagation of stock, net income by stock. At the initial stage, the model trained with a practice training data set where there is a classification of a particular category for each record. The SVM model is considered suitable for arranging non-straight exemplification by developing a strong connection between the information variables and class names. It has been seen that the model proposed is very reliable and more efficient if extra variables are included. As shown in figure 2 The data points that are nearest to the hyperplane are called support vectors. Basically, SVM is based on the idea of separating a dataset into two classes with the hyperplane. If the support vector data point is removed then it will reform the position of the dividing hyperplane. Because of this property, they can be pointed as the critical elements of a data set.
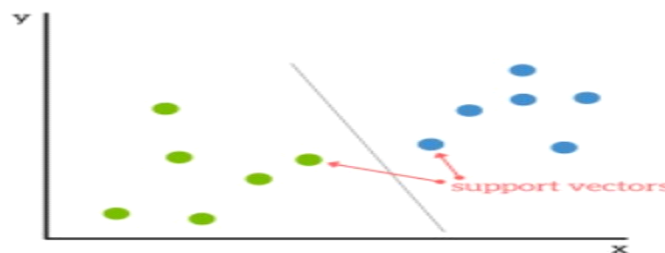

Figure 2: Example of Support Vector.

Here, a hyperplane in the n-dimensional Euclidean space is flat, and of that space n-1, dimensional subset divides the space into two disconnected parts. As a simple example, for the task of classification with only two features, you can think of a hyperplane as a line that linearly separates and classifies a set of data. Innately, as long as possible space between the hyperplane and data points the result is convinced that they have been classified correctly. Therefore, we wish data points to be far away from the hyperplane, at the same time, still

stand on the correct side of it. So, at any time, every bit of new experimental data is enumerated, it can land on anyway two sides of the hyperplane that will decide the class that we assign to it. The distance between the hyperplane and the nearest data point from either set is known as the margin [3]. Here, the aim is to excerpt a hyperplane with the huge possible margin between the hyperplane and any point within the trial dataset, giving a higher expectation of new data being classified correctly.
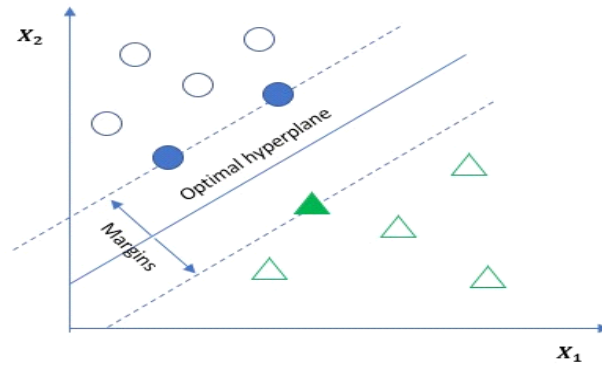

Figure 3: Example Optimal hyperplane and Margins.

2.1 Linear Regression: -

Linear regression has abundant usage in the finance sector with lots of applications.[4] Regression evaluation is based on the study of the relationship between one variable, generally known as the dependent variable, and several other variables broadly called the independent variables. These are among the most popular data-driven model because of their easy functions and very well-known techniques. These models start from linear to nonlinear and parametric to nonparametric models. This model endeavors to form the interconnection between two variables by applying the linear equations to ascertained data. At any moment, if the linear regression model is applicable to a certain data then one thing that should be carefully observed is the range of the data. While using the regression equations to anticipate values outside of this range is largely disproportionate, and may give incredible answers. This practice is known as extrapolation.
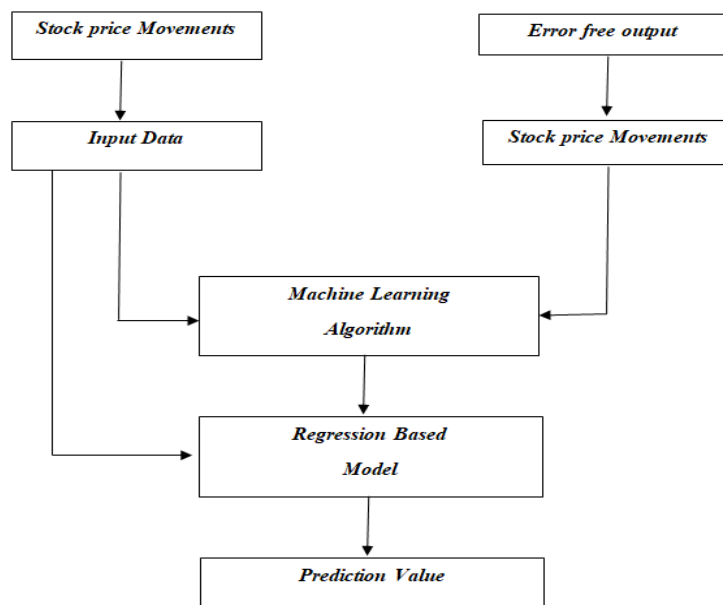

Figure 4: Flow chart for Regression Based Model.

The line is an equation in the linear regression that is accountable for representing historical data to predict future stock values. There are two possible scenarios in the stock market, first is a stock may be overvalued when it is above the line of linear regression, and secondly, the stock is probably undervalued when its value is below the linear regression line. Typically, many investors evaluate the stock regression line with essential historical stock data. In order to forecast the future price of the stock, one needs to procure historical stock market data for a particular time period. Historical stock data are easily available on financial websites, and

there many convenient mobile applications available to access that information. The two major portions of information required are the closing stock price and the associated closing date.

The equation y = mx + b is the formula for the linear regression line. In this calculation, "y" is the dependent variable, "m" is the slope of the line, "x" is the given independent variable and "b" is the y-intercept. The calculation of the y-intercept, subtract the mean of all the stock prices from the mean of all the dates. Eventually, put the values back into the formula. For instance, if you calculate a slope of 1.5 and a y-intercept of 20, the absolute linear regression equation for the stock prediction is y=1.5x+20.

III.   COMPARISONS IN TECHNIQUES FOR MARKET STOCK PRICE PREDICTION

S.no	Techniques	Advantages	Disadvantages	Parameter Used
1	**Stock trend prediction by using Regression Analysis**	Moving Average reduces the Fluctuation	It requires large datasets to achieve a stable and good accuracy. [5]	Last closing price and date from daily and weekly activities whose data Extracted from Stock exchange
		It does not require the input features to be scaled and is simple and very efficient.		
		The output is more informative and explains the contributing factors	It cannot be applied to problems that cannot be linearly separated.	
2	**Support vector Machine for Stock Prediction**	The risk of overfitting is less as they provide a good generalization of the data. [6]	Performs poorly when the dataset is noisy.	Price to earnings ratio of stocks, net revenue, net income price, consumer spending, unemployment rate.
		Don't lose much accuracy when applying to a Sample from Outside a training sample	It lacks transparency in its output which is caused due to the high number of dimensions. [6]	

IV.   RESULTS AND DISCUSSION

A support vector machine as stated in this literature plots points on a hyperplane such that data points belonging to two different classes are separated by Support Vectors by the largest gap possible. It is, therefore, conclusive that Support Vector Machine performs better than Linear Regression. The resultant plots are shown below in figures Fig. 5 and Fig.6.
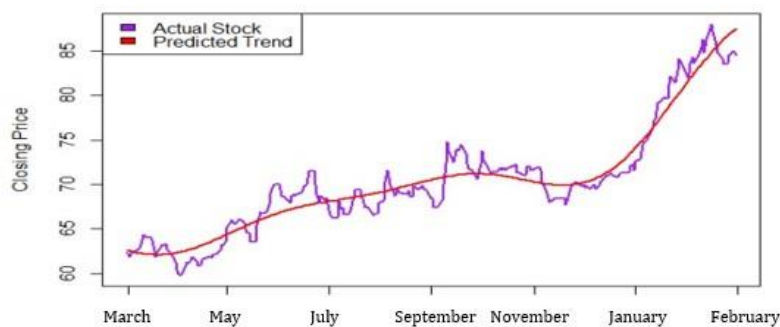

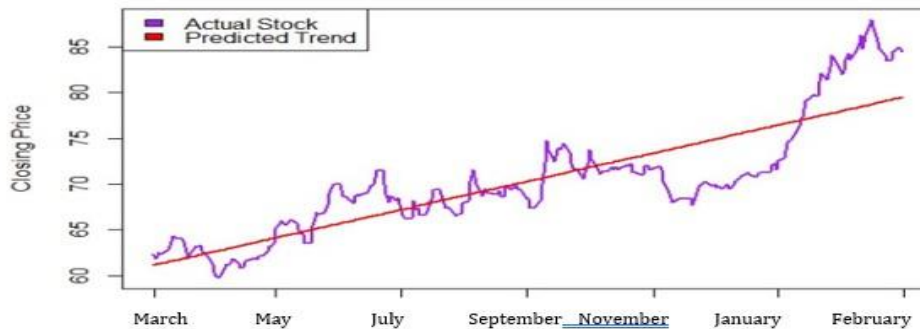**Figure 5:** Stock prediction of ONGC using SVM

Figure 6: Stock prediction of ONGC using LR

V.   CONCLUSION

Stock market prophecy will help individuals, organizations and many shareholders to track market tendency. It also helps to figure out whether to buy or sell any particular stock so that risk can be reduced and get massive profit. In this paper, I have studied the two techniques that are used for the prediction of t the stock market. The risk factor can be analyzed on the basis of historical data and previous business trends. From this research, it is to be concluded that the execution of the Support Vector Machine is more advantageous than Linear Regression and it also gives faultless stock prediction result. Therefore, large number stock exchange analyst has started using SVM method extensively to make predictions of the stock market.

VI.   REFERENCES

[1] Eunsuk Chong et at. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, Expert Systems with Applications,8 (2017): 187-205.

[2] "sigmoidal.io," LLC, Sigmoidal.

[3] KDnuggets(2016). Support Vector Machines: A Simple Explanation[online] Available at: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

[4] Programmingforfinance(2018).linearregression.[online]Available at:

[5] https://programmingforfinance.com/2018/01/predicting-stock-prices-with-linear-regression/

[6] N. Singh Pahwa and N. Khalfay, "Stock Prediction using Machine Learning a Review Paper," Int. J. Comput. Appl., vol. 163, no. 5, pp. 975–8887, 2017.

[7] L. Auria and R. a Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis," Discuss. Pap. Dtsch. Inst. Wirtschaftsforsch., no. August, 2008.