# Stock Market Analysis Using Supervised Machine Learning

**Hasna V N [1], Shemitha P A [2], Dr. G Kiruthiga[3]**

[1]*Department of Computer Science &Engineering, I E S College of Engineering*
[2]*Assistant Professor ,Department of CSE & I E S College Of Engineering*
[3] *CSE Dept . HOD & I E S College Of Engineering*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Stock Market is one of the most versatile sectors in the financial system, and Stock Market plays an important role in economic development. Small ownerships, brokerage corporations, banking sector, all depend on this very body to make revenue and divide risks; a very complicated model. However, this paper proposes to use machine learning algorithm to predict the future stock price for exchange by using open source libraries. The prediction Methods using Machine learning algorithms are SVR & Linear regression.*

**Key Words: *Data Analysis, Fundamental, Implementation, Linear Regression, Support vector Regression (SVR), Stock Market, Supervised Machine Learning***

## 1. INTRODUCTION

The concept of stock markets came to India in 1875, when Bombay Stock Exchange (BSE) was established as 'The Native Share and Stockbrokers Association' a voluntary non-profit making association. We all know it, the Bhaji (Sabji) market in your neighborhood is a place where vegetables are bought and sold. Like Bhaji (Sabji) market, a stock market as a place where stocks shares are bought and sold. The stock market determines the day's price for a stock through a process of bid and offer. You have right to bid and buy a stock shares and offer to sell the stock shares at a valuable price. Buyers compete with each other for the best bid and got their highest price quoted to purchase a particular Stock Market Shares. Similarly, sellers compete with each other for the lowest price quoted to sell the stock. When a match is made between the best bid and the best offer a trade is executed. In automated exchanges high speed computers do this entire job. Stocks of various companies are listed on stock exchanges. Presently there are 23 stock markets In
India. The Bombay Stock Exchange (BSE), the National Stock Exchange (NSE) and the Calcutta Stock Exchange (CSE) are the three large stock exchanges. There are many small regional exchanges located in state capitals and other major cities.

Stock markets are also expected to play a major role in disciplining company's managements. In India, Equity market development received emphasis since the very first phase of liberalization in the early 'eighties. Additional emphasis followed after the liberalization process got deepened and widened in 1991 as development of capital markets was made an integral part of the restructuring strategy. Today, Indian markets conform to international standards both in terms of structure and in terms of operating efficiency.

Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. In machine learning we have adapted the same algorithm where we use the features to train the classifier which then predicts the value of the label with certain accuracy which can be checked while training and testing of the classifier. For a classifier to be accurate you must select the right features and have enough data to train your classifier. The accuracy of your classifier is directly proportional to the amount of data provided to the classifier and the attributes selected.

The proposed system using a machine learning algorithm called Support Vector Regression (SVR) & Linear Regression. The program will read in Google (GOOG) stock data and make a prediction of the open price based on the day. Support Vector Regression(SVR) is quite different than other Regression models. It uses the Support Vector Machine(SVM, a classification algorithm) algorithm to predict a continuous variable. While other linear regression models try to minimize the error between the predicted and the actual value, Support Vector Regression tries to fit the best line within a predefined or threshold error value. What SVR does in this sense, it tries to classify all the prediction lines in two types, ones that pass through the error boundary( space separated by two parallel lines) and ones that don't. Those lines which do not pass the error boundary are not considered as the difference between the predicted value and the actual value has exceeded the error threshold, $\varepsilon$(epsilon).And Linear regression a simple linear regression model will predict the dependent variable(denoted y) values as a function of the independent variables(denoted x). Basically, this function generates a straight line in the cartesian

coordinate system. This straight line is the prediction line for the simple linear regression model which tries to predict the dependent values as accurately as possible.

**Support Vector Machine Pros:**

- It is effective in high dimensional spaces.

- It works well with clear margin of separation.

- It is effective in cases where number of dimensions is greater than the number of samples.

**Support Vector Machine Regression Cons:**

- It does not perform well, when we have large data set.

- Low performance if the data set is noisy ( a large amount of additional meaningless information).

**Types Of Kernel:**

1. linear

2. polynomial

3. radial basis function (rbf)

4. sigmoid

**Linear regression** is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

**Linear Regression Pros:**

a) Simple to implement.

b) Used to predict numeric values.

**Linear Regression Cons:**

i. Prone to overfitting.

ii. Cannot be used when the relation between independent and dependent variables are nonlinear

## 2. PREDICTION MODEL

### A. Data Analysis Stage

In this stage we shall look at the dataset.. Load the Google (GOOG) stock data that I got from finance.yahoo.com into a variable called 'df' short for data frame. The actual data used during the time of this writing can be found here. Then I will print the first 7 rows of data. Create the variables that will be used as the independent and dependent data sets by setting them equal to empty lists. Get all of the rows from the Date column store it into a variable called 'df_dates' and get all of the rows from the Open column and store the data into a variable called 'df_open'.

We have successfully analyzed the data and extracted the useful information that we shall be needing for the classifier. This is a very crucial step and shall be treated with extreme care. A miss of information or small error in deriving useful information will lead to a fail prediction model and a very inefficient classifier. Also, the features extracted are very specific to the subject used and will definitely vary from subject to subject. Generalization is possible if and only if, the data of the other subject is collected with the same coherence as the earlier subject.

### A. Training And Testing Stage

In this stage we extended  from our data and imlement in our machine learning model . Create a function that uses 3 different Support Vector Regression (SVR)models with three different kernels to see which one performs the best. The function will have three parameters, the dates, prices, and the day that we want to do the prediction on to get the price.To do this, first I will create the 3 SVR models with 3 different kernels (linear, polynomial, radial basis function ). Also add in the linear regression model. Next we will train each model with the dates and prices data. Last but not least I will plot the models on a graph to see which has the best fit and return the prediction of the day.

### B. Results
Now we can start making my stock price prediction. Recalling the last row of data that was left out of the original data set, the date was 06–28–2019, so the day is 28. This will be the input of the model to predict the open price

which is ₹1076.390015. So now we will predict the open price giving the models a value or day of 28. The best model from the graph below seems to be the RBF which is a Support Vector Regression model that uses a kernel called radial basis function. That model predicted the price for day 28 to be ₹1091.7058004910944, which is pretty close to the actual price of ₹1076.390015.
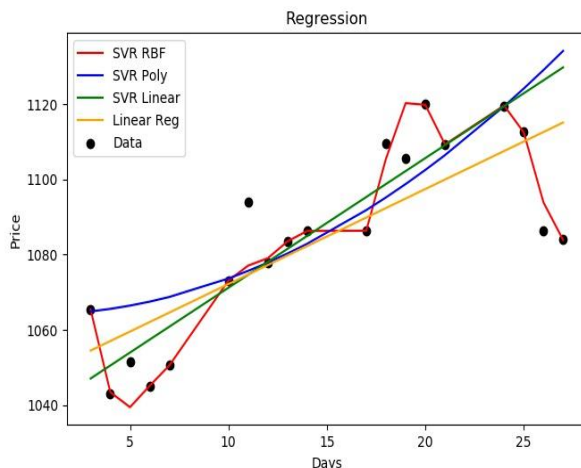


**Chart -1**: *Predict the price of GOOG on day 28*

## 3. CONCLUSION AND FUTURE WORKS

In this paper, we proposed several methods to analyze GOOGLE stock datasets. We also presented stock prediction methods using algorithms to apply a supervised learning of the dataset.

For future work, we would like to extend this study to use other datasets such as Amazon dataset or facebook dataset and use different feature selection methods. Furthermore, we may apply ANN algorithms to predict stock prices using various tools such as Python and R or R studio, Statistical Analysis System (SAS), and Stata; then we will evaluate the performance of our work with some of these tools.

## ACKNOWLEDGEMENT

## REFERENCES

[1]. Kunal Pahwa,Neha Agarwal (2018), "Stock Market Analysis Using Supervised Machine Learning" ,

[2]. Agarwal,"Introduction To The Stock Market (july 14,2017).

[3]. Montgomery, D.C, Peck ,E.A , And Vining , "Introduction To Linear Regression Analysis", (2012)

[4]. Yadav , Sameer, "Stock Market Volatility – A Study Of Indian Stock Market " (2012)

[5]. Google Developers, Oct 2018, "Decending into ML: Linear Regression", Google LLC, https://developers.google.com/machinelearning/ crash-course/descending-into-ml/linear-regression.

[6]. Hurwitz, E. and Marwala, T., 2012. Common mistakes when applying computational intelligence and machine learning to stock market modelling. arXiv preprint arXiv:1208.4429.

[7]. Draper, N.R.; Smith, H. (1998). Applied Regression Analysis (3rd ed.). John Wiley. ISBN 0-471-17082-8.

[8] .https://en.wikipedia.org/wiki/Stock_marke

[9]. https://en.wikipedia.org/wiki/Regression_analys is.