2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)
Department of Electrical & Electronics Engineering, Galgotias College of Engineering and Technology, Gr. Noida, India

2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-7281-7741-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACITE51222.2021.9404733

Stock Market Prediction Using Linear Regression and SVM

Bhawna Panwar
School of Computer Science and Engineering Galgotias University, Greater Noida, India
panwarbhawna99@gmail.com

Gaurav Dhuriya
School of Computer Science and Engineering Galgotias University, Greater Noida, India
proudygarv@gmail.com

Prashant Johri
School of Computer Science and Engineering Galgotias University, Greater Noida, India
johri.prashant@gmail.com

Sudeept Singh Yadav
School of Computer Science and Engineering Galgotias University, Greater Noida, India
sudeept999@gmail.com

Nitin Gaur
School of Computer Science and Engineering Galgotias University, Greater Noida, India
nit.gaur1982@gmail.com

***Abstract*—In Stock Market is the financial epitome of financial business and trading since it came into existence it has shown the impact of hits low and similarly when it is high. The stock market crash in 2008 showed the world that the business hit the low when the Dow Jones Industrial Average fell 777.68%. Several machine learning algorithms have shown that these stock prices can be predicted and these algorithms can be implemented using the approach of supervised learning. In Supervised Learning, we have test data using this we train the models. Although the results obtained after training the model may differ from the actual but it has been observed that in many cases accuracy is satisfactory. In this paper, the first task is to use web scrapping to collect datasets from stock data. Then we plot the data on the graph, from the graph we can analyze the stock prices going high or low. After this, we will predict stock prices using SVM and Linear Regression, that Linear Regression for stock market analysis is better than the SVM for the same.**

***Index Terms*—SVM(State Vector Machine), SVR(State Vector Regression), Linear Regression(LR)**

I. INTRODUCTION

Machine Learning in the field of computer science based on training the machine to analyze and work on problems by learning through experience, it is one of the parts of Artificial Intelligence. It is widely used for making things easier like filtering emails, credit card fraud detection, computer vision, etc. The main reason behind using machine learning algorithms is to reduce the risks of tasks that have failed in the past to not occur in the future using the past information, this involves analyzing past data for discrepancies to monitor any such thing in the future. Just figure out the burden people have to go through each entry every time new data is to be analyzed and categorized for future use. The base of any machine learning is the dataset which in common use we say the training set. The training set contains the past data which consist info regarding the work of system what were the data recorded for a period also a time bar to determine that system has data in form duration, the data should be in a properly labelled manner as it may cause in results that are not accurate. The data is continuously in running state like new data is being added and it is categorized.[4] Also, in case of stocks data is very dynamic and depends on a lot of factors. There are a lot of machine learning algorithms to choose from that have different advantages over the other in different situation. For this paper, we have analyzed two of such algorithms on their comparisons analyzing the more accurate one for the stock market prediction. The machine learning algorithms are Linear Regression and SVM.[2] The dataset is obtained by web scraping the data from the yahoo finance site. The data is then formulated in the form of a table to the user for plotting the graphs of companies like Apple, Amazon, Microsoft, Google, etc. To plot we are using a function under plotly candlestick which is very beneficial to get data on a particular point. The graphs can be analyzed for visually analyzing the stock going up and down. After that, we are using these data to check the confidence percentage or accuracy of the two mentioned machine learning algorithms.[10]

II. DATA SET

For stock market prediction I used web scraping technique to get the data, the site was yahoo finance for this purpose I used the python module 'beautiful Soup'. It is an efficient tool for web scrapping, it comes in handy if you are looking for only the required data. The data received is then used to plot the graph b/w the closing rate of stocks vs the days. It results of the plot can easily help a person analyze that the stocks are rising or falling on a particular interval. Below figure 1 shows the plot of stocks of Amazon from 1st Oct to 31st Dec 2019.

Fig. 1: Amazon stock plot

III. SVM

SVM is a machine learning algorithm that is used in solving some of the regression and prediction problems.[1] It is primarily based on statistical Learning approach. In SVM we to predict results we find out the plane between which the values might range for the input. Let us take an example of a plot that shows examples of some circles and triangles[fig2].[8]

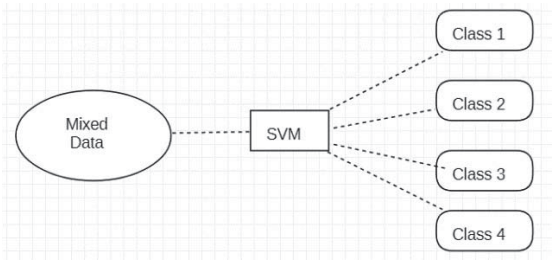


Fig. 2: A SVM Categorization

For SVM we first draw the planes intersecting values of the circles and rectangles and using those we develop the median of both the planes this results in another name for SVM that is the widest street approach.[3] The figure below describes hyperplane[fig3]

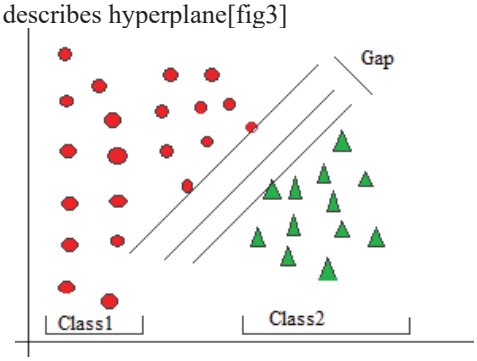


Fig. 3: Hyperplane dividing classes

A. SVM Architecture

In SVM we analyze the two classes for dividing them with a hyperplane, it is a line that splits the data of the two classes such that value between the last points of the classes is maximum it is called maximum margin.[5] Now using the hyperplane and the support vectors we achieve the widest street, which is in further enhance towards SVM approach [fig4].

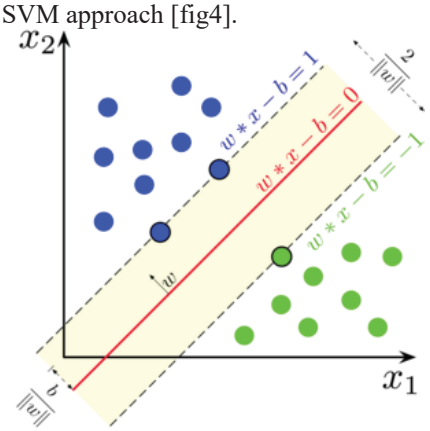


Fig. 4: SVM Architecture In the above figure

w = vectors x= variables
b= biased term(wo)
$w * x - b = 1$
$w * x - b = 0$
$w * x + b = 1$

The above equations help to analyze the new data and the plane nearest to it determines the class of the data. Also, these equations help find the maximum margin for the dataset, which in turn helps predict the class of data with accuracy.

B. How SVM works?

The SVM computer model learns by assigning labels to objects. For example, to detect fraud and scams in the credit card services, it analyzes hundreds and loads of successful and unsuccessful transactions of credit card reports. The hyperplane is made such that the characteristics can be divided of the separator[fig6] for the categories[fig5] of data that it separates are drawable on the hyperplane[fig7], using this the new data is analyzed for which group it belongs.[7]

Let us take an example:

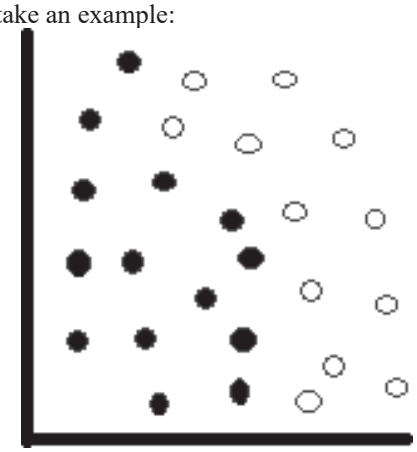


Fig. 5: The above data contains two class of data

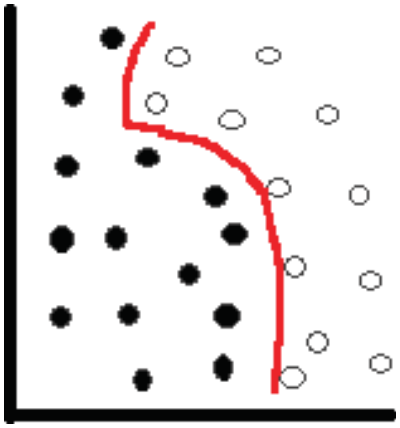


Fig. 6: The imaginary separator separates the data

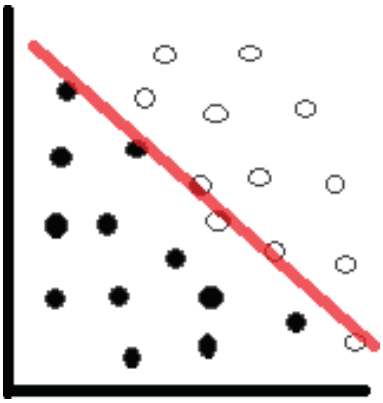


Fig. 7: The hyperplane is formed with help of boundary.

C. Application of SVM

Credit Card Fraud Detection: Loads of information of past information on stocks is analyzed and model is trained to avoid any future repetition. Also, different organization use SVM for cybersecurity frauds.[9]

Recognize Handwritten Alphabets or digits: SVM can learn to recognize handwritten alphabets or digits which is achieved by training the function with pictures of handwritten alphabets and digits which works as a training dataset.[11]

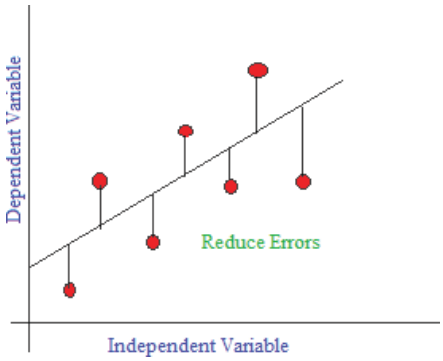
Prediction of Stock Prices: In various scenarios SVM has been accurate in predicting stock prices. The main way of doing is just making a regressor model out of the dataset. The predictions may vary but in different cases, it works differently.

D. Problems with SVR

It works not well when the data is noisy and unlabeled data. Data should be in format to represent on the hyperplane.

IV. LINEAR REGRESSION

Linear Regression is a model that displays the relationship between the dependent and independent variables. Also, the variations involved in the dependent variable on changes being observed in the independent variable.[6]



Linear Regression is also used for predictive analysis, predict the needs of the customer based on its shopping list. Also, it is used to optimize business workspace like the right conditions for the model to excel without any more inclusions[fig8].

A. Linear Regression Architecture

$$y = bo + b1 * x1$$

In the above equation, y is the dependent variable x is the independent variable bo is constant b1 is coefficient.

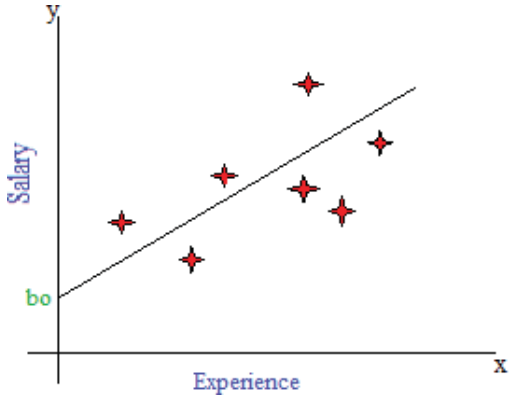


Fig. 9: LR on Salary Experience Relationship

B. How Linear Regression works

Let us take one simple example of salary experience relationship to explain[fig9]-

$$\text{sallary} = \text{bo} + \text{b1} * \text{experience}$$

C. Linear Regression Applications

Predict Weather: Based upon the humidity, rise and fall in temperature we can predict the weather for upcoming days. This application would be very useful in areas with less rainfall, in advance people can line up strategies to store rainwater.

Academic Report: The linear regression can further be used to predict the performance of the student based upon the activities of the students.

Forecasting Result: Adding time series analysis, the data can be predicted as in time period basis.

V. RESULT

On analyzing the prediction results of SVR and Linear Regression the Amazon stock data Linear Regression aligned to an accuracy of 98.76% and SVR accuracy is 94.32%. So, it can be inferred that in stock market prediction Linear Regression is better than the State Vector Regression[fig.10].

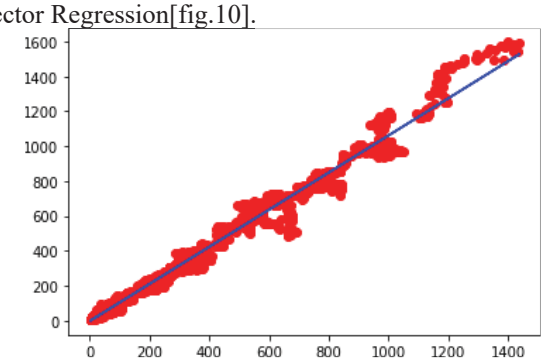


Fig. 10: Linear Regression Prediction

VI. CONCLUSION

On observing the accuracy of both the algorithms on same the stock data it was observed linear regression predicted more accuracy as compared to the SVM model. Thus, it is proved that Linear Regression for stock data or stock market analysis is better than the SVM for the same.

VII. REFERENCE

[1] HaiqinYang, Laiwan Chan, and Irwin King "Support Vector Machine Regression for Volatile Stock Market Prediction".

[2] Mo Chun Yuen, Lam Man Yiu, Tang Kai Man "COMP 3211 Final Project Report Stock Market Forecasting using Machine Learning".

[3] Corinna Cortes, Vladmir Vapnik."Support-Vector Networks".

[4] Tejinder Singh, Er. Mohit "Implementation of Stock Market Prediction".

[5] V Kranthi Sai Reddy. "Stock Market Prediction Using Machine Learning".

[6] Kelly H. Zou, PhD Kemal Tuncali, MD Stuart G. Silverman, MD. "Correlation and Simple Linear Regression".

[7] Yanhui Guo, Siming Han " An Adapive SVR For High Frequency Stock Price For3ecasting".

[8] Yomsheng Ding, Linjun Cheng, Witold Pedrycz, and Kuangrong Hao "Global Nonlinear Kernel Prediction for Large Data Set With a Partial Swarm-Optimized Internal Support Vector Regression".

[9] Yoram Baram " Partial classification : The Benefit of Deferred Decision".

[10] Mortaba Nabipour, Pooyan Nayyeri, Hamed Jabani "Predicting stock market trends using machine learning deep learning algorithms via continuous and binary data ; a comparative analysis on the Tehran stock exchange".

[11] L.j.Cao, Francis E.H. Tay " Support Vector Machine With Adaptive Parameters in Financial Time