STOCK MARKET PREDICTION USING MACHINE LEARNING

Prof. Jogi John[1]

Aatish Kumar[2], Ankit Abhishek[3], Tejas A. Dhule[4], Amit Roy[5], Abhishek Jha[6]

[1,2,3,4,5,6] *Dept. of Computer Technology, Priyadarshini College Of Engineering, Nagpur, Maharashtra, India*

Abstract - *Stock market is a daily emerging topic everyone has on their mouths heard from news channels or read in newspapers. It's strongly related to the profit and loss beared by any company or big organization based on daily market ups and downs. Hence it makes very important for investors to know the product well and take a right decision of whether they should invest in the shares of a particular company or not. Here comes the major role of **Machine Learning!** Technological advancements are now growing in the field of AI and Machine Learning leading a Computer Systems to take decisions for humans after studying patterns in their daily behaviors and actions. Thus it is a very smart decision to use this talent of Machine to find similar patterns in changing behavior of market and accordingly let the machine think of the next similar pattern to predict the future stock prices. In this paper we have taken a raw data based on the past results of changing stocks of a company and organized it by forming a Regression Model. Each day more than 5000 trading companies noted in Bombay stock exchange (BSE) provides an average of more than 210000000 stocks, making an approximate of 2000 to 3000+INRs in investment. Thus it's been an treasure hunt for investors to crack the market by predicting the best future value. Designing an application which refines these much amount of data to find hidden **patterns** and using it to develop a model predicting a next recurring pattern in BSE Stock Market is necessary. It's a key requirement for the machine to represent this data graphically to the users or investors with approximate nodal value somewhere in the graph so that it is easy to understand the past behavior of that company. In this paper, we have proposed an application of Machine Learning and used **Support Vector Machine (SVM)** algorithm implemented in Python Programming Language to predict upcoming Stock Prices at some extent which will surely help an investor or a user to make it's decision when he/she is completely new in the topic of Stock Market. SVM can be truly used to train the data set obtained free of cost from different web platforms using **Web Scrapping**.*

Key Words: *Machine Learning, Bombay Stock Exchange, Support Vector Machine (SVM), Python, Web Scrapping.*

1. INTRODUCTION -

Stock Market always carry a common topic of discussion among the treads, investors, and end users which sells and buys their amount based on the market status. It's Never happens that the current stock prices of company staggered for a day, it's continuously changing and differentiating from the previous days results. In this situation it is very difficult for the share trader to find out hidden fluctuations in the changing market and to take a proper move to purchase the sells of the company. Other users always sticks to the daily news to see if any incident relevant to the company happened where they have invested their money and accordingly move up to withdraw it before price goes down. Considering the discontinuities and scattered behavior, Stock Price prediction became a major topic of discussion among big researchers and investors. This model of Machine Learning will focus on the above non Linear trends and discontinued changing stocks (prices) over the fixed segment of a day. It is for sure that this model will Never going to provide exact nodal fixation or amount to be invested at each successive prediction. In this we have tried to improve the prediction accuracy by making a Stock Classification Model, performing Web Scrapping, Data Mining, SVM Forecasting and optimization to be considered where predictive method will be multiple and linear.

2. WEB SCRAPPING -

Here Scrapping refers to the collection of data ethically without violating the general norms set by the website from which we are collecting data. We are collecting data from the web platforms available to all for use hence it is known as Web Scrapping. Now we can see the general examples of web Scrapping in day to day life such as copying any important educational data from educational website for study and research, copying the lyrics of any song you like from website and many more. But still as discussed all of these data should be legal to use. Similar concept is carried when we talk about Machine Learning. As to predict the future stock prices, we need to consider the past data of the company. Like to train our machine based on the past data set, a machine learning library called **Pandas_Datareader** must be imported in header of python code. We are Scrapping data from the website called **Yahoo Finance** (http://in.finance.yahoo.com) and storing the complete dataset into an CSV (Comma-Seperated Value) file. To understand the concept, let's look at the command to Scrape the data of (ex. Apple.Inc),

```
Data = of.read_csv("AAPL.csv")
```

```
Data.head()
```

Using the above commands, we have Scrapped the data of Apple.Inc from 1 Jan 2018 to 20 April 2019 and imported it in a 2-D Data Frame format which has 6 fields i.e. Open, High, Low, Close, Volume, Adj Clone.
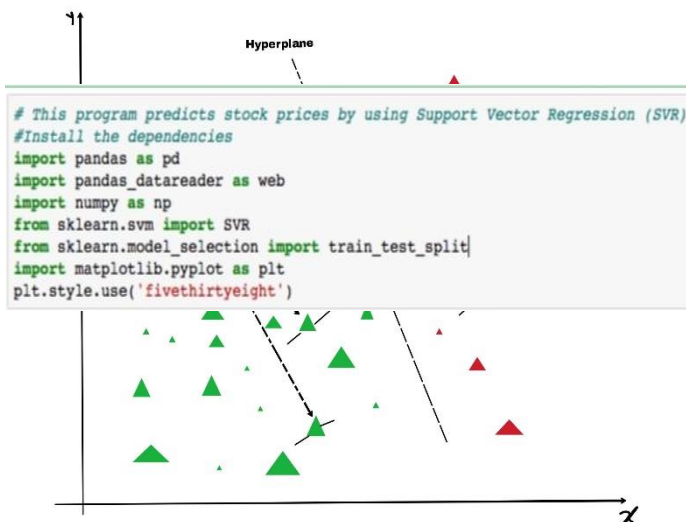
3. SUPPORT VECTOR MACHINE (SVM) -

Support Vector Machine is a standard **Supervised Learning** Algorithm in Machine Learning used for Classification and Regression problems. As we know the Supervised Learning Algorithm works on predefined labeled dataset where the machine has to distinguish the dataset in different classes such that if we have provided the combined dataset of two classes then the Algorithm will automatically learn the similarities in dataset and separate it based on similar points. Same Concept

happens with SVM where we have to plot different data points (each data point as a feature) on a n- dimensional plane then the classification is performed on the given dataset such that we proceed with finding the **Hyperplane** that separates the dataset into two different classes based on the similarities in data.

Support Vectors are the best possible coordinates of each individual class repressing the limiting boundaries.

4. TOOLS/PACKAGES AND PLATFORM USED -


```
# This program predicts stock prices by using Support Vector Regression (SVR)
#Install the dependencies
import pandas as pd
import pandas_datareader as web
import numpy as np
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

To begin with the project, we have implementation language as **Python 3** and some of dependencies as,

Here, **math** is used for mathematical operations, **pandas** for Data manipulation and analysis, **numpy** for core scientific computations, **sklearn** is a machine learning library to import SVM, **matplotlib** for 2-D plots of array.

5. PREDICTION OF STOCK MARKET BY USING SVM -

Following are the general steps considered for Stock Market Prediction using SVM:

STEP 1:- Data Extraction (Web Harvesting)-

The initial step is very important as all of our prediction is based on it i.e. previous dataset. To obtain it we have used the platform of **Yahoo Finance** (http://in.finance.yahoo.com). Below is the snapshot of python code which shows the process of Web Scrapping for obtaining the dataset of **Apple Inc. (AAPL)** from Yahoo Finance.

In above image, we have used an Python Library called **Pandas** for Scrapping the data of Apple Inc. from 1st Jan 2013 to 30th Dec 2019. Pandas automatically converts the web harvested data into structured 2D Frames, so that it can be easily loaded into the algorithm.

STEP 2:- Plotting The Data To Be Train -

Training is the most crucial process in Machine Learning because overall prediction will only depend upon the past experiences the Algorithm acquired during the probationary period. It can be thought as the training given to the soldiers to make them prepare for the real battle. Which training should be included in probation period and which should be neglected so that soldiers will not be confused during taking any decision at battle is a very delicate task. In terms of computation it is known as **Data Filtering.** Data Filtering allows to choose only the essential part of the dataset on which further training of Algorithm will depend.

In the below snapshot, we have chosen only the **Closing Price** history of Apple Inc. and will further going to train our model of **Support Vector Regression** using only this Closing Price of Stock. To represent the filtered data graphically, we have plot it for **Closing Price (in $)** against **Date** of each day using Machine Learning Library called **MatPlotLib.**

STEP 3:- Classification of Dependent and Independent Dataset -

As we have previously discussed that the Support Vector Machine (SVM) algorithm is a model of **Classification,** hence we have to provide the dataset as input categorizing it into **Depend** (Plotting on X-axis) and **Independent** (Plotting on Y-axis) data sets.

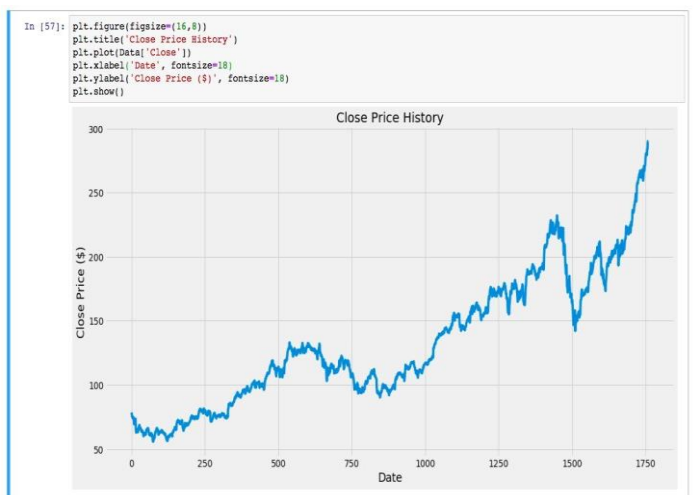### STEP 4:- Training The Model of SVM -

Now, it is the most important task in the this whole project i.e. Training Phase. The data captured by forecasting will now going to be used in training of SVM.

In general, 80% of dataset (Dependent & Independent) is utilized in training while 20% is used for end Testing. As seen above, we have trained our dataset on **rbf (Radial Basis Function) kernal** which has the highest accuracy among the rest of two SVM models i.e. **Linear Model** & **Polynomial Model.**

STEP 5:- Testing Model -

The overall time required for the Testing process is completely depends upon the type of dataset used, Market Fluctuations, and the amount of past data feed to the model.

After successful training, the final step before end outcome is to test the **Accuracy Level** provided by the model. In terms of Machine Learning, it is given by the term, **SVM Confidence** which is the best possible Score the machine

```
In [57]: plt.figure(figsize=(16,8))
         plt.title('Close Price History')
         plt.plot(Data['Close'])
         plt.xlabel('Date', fontsize=18)
         plt.ylabel('Close Price ($)', fontsize=18)
         plt.show()
```


acquired after the training phase. It can be also understand as the confidence of a soldier to win the battle after successful training in probationary period.

The **Ideal** value of Accuracy/SVM Confidence is always 1.0 (100%) which can not be achievable by any of the Machine Learning Algorithm. But we can try for the best nearby value.
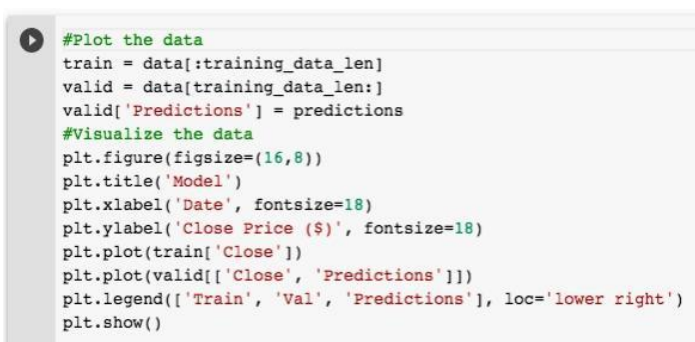
STEP 6:- End Prediction -

Finally after the huge stuff, we have been on our end outcome. As seen in below snapshot, all the 80% of training and 20% testing data is now linked with the **svr_rbf** model and x-forecast (Stock Prices) is thrown out by **.predict()** function for the future 30 days.

```
[257.49940778 215.05304989 240.48165486 294.94297702 284.02945574
 281.65598935 237.87790854 291.49366068 195.74322274 210.7502518
 292.93822424 278.58028342 281.26518921 262.99464896 165.71718508
 220.43091083 183.3456025  165.5771938  164.50604115 163.79828365
 163.79722929 163.79722928 163.79722929 163.79722929 163.7972293
 163.79722928 163.79722928 163.79722928 163.79722928 163.79722928]
```

:

We can also visualize the complete process of Training, Testing and Prediction through the graph drawn using the Machine Learning Library called **MatPlotLib.pyplot** as shown below,

```
#Plot the data
train = data[:training_data_len]
valid = data[training_data_len:]
valid['Predictions'] = predictions
#Visualize the data
plt.figure(figsize=(16,8))
plt.title('Model')
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price ($)', fontsize=18)
plt.plot(train['Close'])
plt.plot(valid[['Close', 'Predictions']])
plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
plt.show()
```

In below Model, we can see the **Train** curve by **Blue** Color, **Testing Value** curve by **Red** Color & **Prediction Value** curve by **Yellow** Color.

Date	Close	Predictions
2018-08-07	207.110001	216.218246
2018-08-08	207.250000	218.583771
2018-08-09	208.880005	220.522186
2018-08-10	207.529999	222.191788
2018-08-13	208.869995	223.404221
...	...	...
2019-12-23	284.000000	300.934845
2019-12-24	284.269989	302.366058
2019-12-26	289.910004	303.754059
2019-12-27	289.799988	305.584259
2019-12-30	291.519989	307.457428

352 rows × 2 columns

6. CONCLUSION -

In this project, we have tried to give an effective & useful application of Machine Learning i.e. **Prediction**. We have seen the ability of machine to self optimize itself based on the Supervision of developer. Stock Market was always an very interesting topic for everyone because with the help of little knowledge, one can maximize it's business profits. That's Why we have proposed a simple model which can help the investors to take decisions regarding whether to invest their money in the company/organization they wish or not.

Our end conclusion is depend on the following aspects we have learned so far:

SVM is a great algorithmic choice to implement this project because it can handle a large amount of data and can be very fast to train the model based on it.

RBF (Radial Basis Function) kernal is the strongest of all the three models of Support Regression.

Efficiency/Confidence value of SVM increases with the increasing experience obtain during Training Period.

Future efficiency can be greatly improved by implementing future aspects like **Sentiment Analysis** performed on people relating to the organization whose Predictions are made.

Thus future project modifications will be surely surprise all of us at next level when Machine not only know the past of the organization but also what can happen next in the future.

7. REFERENCES -

[1]. http://en.wikipedia.org/wiki/Stock_market

[2]. http://biz.yahoo.com/p/

[3]. http://en.wikipedia.org/wiki/MATLAB

[4]. Zhen Hu, Jibe Zhu, and Ken Tar – Stock Market Prediction Using SVM, 6th International Conference on Information Management, Innovation Management and Industrial Engineering,2013

[5]. C.J.C. Burges, - A Tutorial on Support Vector Machine for pattern Recognition, Data Mining and Knowledge Discovery Volume 2, pp. 1-43, 1998

[6]. Mr. Sachin Sampat Patio, Prof. Kailash Patidar, Assistant Prof. Meghan Jain, SSSIST,Sehore,Madhya Pradesh,India – Stock Market Prediction Using Support Vector Machine, Volume:02,Issue:01 (JAN-FEB,2016)