Stock Price Trend Forecasting and Stock Selection using Supervised Learning Methods

1Priyanshi Bhola  
Computer science and engineering  
Amity University, U.P, India  
priyanshi.bhola5@gmail.com

2Garima Ahlawat  
Computer science and engineering  
Amity University, U.P, India  
garimaahlawat.10@gmail.com

3Dev Kumar Chaudhary  
Computer science and engineering  
Amity University, U.P., India  
dkchaudhary@amity.edu

ABSTRACT

In this paper, we are going to present and review a more feasible method to predict the stock movement with higher accuracy. The first thing we have taken into account is the dataset of the stock market prices from the previous year. The dataset was pre-processed and tuned up for real analysis.

Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper, we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

KEYWORDS - Stock Market, Machine Learning, Predictions, Support Vector Machine, Data Pre-processing, Data Mining, Dataset, Stock, Stock Market.

1. INTRODUCTION

Quantitative traders with much money from stock markets buy stocks derivatives and equities at a low price and later on selling them at a high price. The trend in a stock market prediction is not a new thing, and yet this issue is kept being discussed by various organizations.
The stock market is an aggregation of various buyers and sellers of stock. A stock in general represents ownership claims on business by a particular individual or a group of people. The attempt [3] to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient. The system must work according to the real-life scenarios and should be well suited to real-world settings. The system is also expected to take into account all the variables that might affect the stock's value and performance. There are various methods and ways of implementing the prediction system like Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and Time-series aspect structuring. With the advancement of the digital era, the prediction has moved up into the technological realm. The most prominent and [3] promising technique involves the use of Artificial Neural Networks, Recurrent Neural Networks, that is the implementation of machine learning. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being scheduled time and again.

Traditional methods of prediction in machine learning use algorithms like Backward Propagation, also known as Backpropagation errors. Lately, many researchers are using more of ensemble learning techniques. It would use low price and time [3] lags to predict future highs while another network would use lagged highs to predict future highs.

The probable stock market prediction target can be the future stock price or the volatility of the prices or market trend. In the prediction, there are two types like a dummy and a real-time prediction which is used in stock market prediction system. In Dummy prediction, they have defined some set of rules and predict the future price of shares by calculating the average price. In the real-time prediction, compulsory used the internet and saw the current price of shares of the company.

Computational advances have led to the introduction of machine learning techniques for predictive systems in financial markets. In this paper, we are using a Machine Learning technique, i.e., Support Vector Machine (SVM) to predict the stock market, and we are using Python language for programming.

2. Methodology
Classification is an instance of supervised learning where a set is analyzed and categorized based on a common attribute. From the values or the data are given, classification draws some conclusion from the observed value. If more than one input is given, then classification will try to predict one or more outcomes for the same. A classifier that are used here for the stock market prediction includes SVM classifier. A Support Vector Machine (SVM) is a discriminative classifier that formally defined by the separating hyperplane. In other words, the given labelled training data (supervised learning), the algorithm outputs the optimal hyperplane which categorizes new examples. In the two-dimensional space this hyperplane is a line dividing a plane into two parts where in each class lay in either side.

Support Vector Machine (SVM) is considered to be as one of the most suitable algorithms available for the time series prediction. The supervised algorithm can be used in both regression and classification. The SVM involves in plotting of data as a point in the space of n dimensions. These dimensions are the attributes that are plotted on particular co-ordinates. SVM algorithm draws a boundary over the data set called the hyper-plane, which separates the data into two classes, as shown in Fig 1.

The hyper-plane is a decision boundary which is later extended or maximized on either side between the data points. Considering the same figure, if μ is some unknown data point and w is a vector which is perpendicular to the hyperplane, then the SVM decision rule will be
Radial Basis Function (RBF)

In the machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in the various kernelized learning algorithms. In particular, it is most commonly used in support vector machine classification.

A radial basis function is a real-valued function whose value depends only on the distance from the origin, so that; alternatively on the distance from some other point, called a centre, so that. Any function which satisfies the property is a radial function.

**RBF = Local Response Function**

The RBF Kernel is nothing more than a low-band pass filter, which is well known in Signal Processing as a tool to smooth images. RBF Kernel acts as the prior that selects out smooth solutions.

The Radial basis function kernel is also called as the RBF kernel, or Gaussian kernel is a kernel that is in the form of a radial basis function (more specifically, a Gaussian function). The RBF kernel is defined as

\[ KRBF (\mathbf{x}, \mathbf{x}') = \exp \left[ -y \|\mathbf{x} - \mathbf{x}'\|^2 \right] \]

Where \( y \) is the parameter that sets "spread" of the kernel.
The RBF units provide a new basis set for synthesizing the output function. The radial basis functions are not orthogonal and are over complete.

3. Result

Steps for Stock Market Prediction

Steps for Stock Market Prediction

Step 1: This step is essential for the download data from the net. We are predicting the financial market value of any stock. So that the share value up to the closing date is download from the site.

Step 2: In the next step, the data value of any stock that can be converted into the CSV file (Comma Separated Value) so that it will easily load into the algorithm.

Step 3: In the next step in which GUI is open, and when we click on the SVM button, it will show the window from which we select the stock dataset value file.

Step 4: After selecting the stock dataset file from the folder it will show graph Stock before mapping and stock after mapping.

Step 5: The next step algorithm calculated the log2c and log2g value for minimizing error. So, it will predict the graph for the dataset value efficiently.

Step 6: In the final step algorithm display the predicted value graph of select stock, which shows the original value and the predicted value of the stock.

Our proposed system is to predict the price of the stock by analyzing its historical data.
Figure 3: Flow Diagram Design

4. Discussion

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Stock price volatility. This is an average over the past $n$ days of percent change in the given stock’s price per day.</td>
<td>$\sum_{i-t-n+1}^{t} \frac{C_i - C_{i-1}}{C_{i-1}}$</td>
</tr>
<tr>
<td>Stock Momentum</td>
<td>This is an average of the given stock’s momentum over the past $n$ days. Each day is labeled 1 if closing price that day is higher than the day before, and $-1$ if the price is lower than the day before.</td>
<td>$\sum_{i-t-n+1}^{t} \frac{y}{n}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Index volatility. This is an average over the past $n$ days of percent change in the index’s price per day.</td>
<td>$\sum_{i-t-n+1}^{t} \frac{I_i - I_{i-1}}{I_{i-1}}$</td>
</tr>
<tr>
<td>Index Momentum</td>
<td>This is an average of the index’s momentum over the past $n$ days. Each day it is labeled 1 if closing price that day is higher than the day before, and $-1$ if the price is lower than the day before</td>
<td>$\sum_{i-t-n+1}^{t} \frac{d}{n}$</td>
</tr>
</tbody>
</table>

Figure 4: Formula Implemented
The .xlsx file contains the raw data based on which we are going to publish our findings. There are eleven columns of eleven attributes that describe the rise and fall in stock prices.

<table>
<thead>
<tr>
<th>DATE</th>
<th>TRADING CODE</th>
<th>LTP</th>
<th>HIGH</th>
<th>LOW</th>
<th>OPENP</th>
<th>CLOSEP</th>
<th>YCP</th>
<th>TRADE</th>
<th>VALUE (mn)</th>
<th>VOLUME</th>
</tr>
</thead>
<tbody>
<tr>
<td>28-12-2017</td>
<td>1JANATAMF</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>6.4</td>
<td>6.4</td>
<td>6.5</td>
<td>79</td>
<td>1.888</td>
<td>2,94,720</td>
</tr>
<tr>
<td>27-12-2017</td>
<td>1JANATAMF</td>
<td>6.5</td>
<td>6.5</td>
<td>6.4</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
<td>73</td>
<td>1.295</td>
<td>2,00,062</td>
</tr>
<tr>
<td>26-12-2017</td>
<td>1JANATAMF</td>
<td>6.5</td>
<td>6.6</td>
<td>6.4</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
<td>103</td>
<td>4.119</td>
<td>6,30,548</td>
</tr>
<tr>
<td>24-12-2017</td>
<td>1JANATAMF</td>
<td>6.6</td>
<td>6.6</td>
<td>6.4</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
<td>46</td>
<td>0.654</td>
<td>1,01,104</td>
</tr>
<tr>
<td>21-12-2017</td>
<td>1JANATAMF</td>
<td>6.6</td>
<td>6.6</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>6.4</td>
<td>24</td>
<td>0.241</td>
<td>37,098</td>
</tr>
<tr>
<td>20-12-2017</td>
<td>1JANATAMF</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>6.4</td>
<td>6.4</td>
<td>6.4</td>
<td>37</td>
<td>0.296</td>
<td>45,885</td>
</tr>
<tr>
<td>19-12-2017</td>
<td>1JANATAMF</td>
<td>6.4</td>
<td>6.6</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>6.5</td>
<td>55</td>
<td>1.387</td>
<td>2,16,529</td>
</tr>
<tr>
<td>18-12-2017</td>
<td>1JANATAMF</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>6.4</td>
<td>6.5</td>
<td>6.4</td>
<td>36</td>
<td>0.141</td>
<td>21,817</td>
</tr>
<tr>
<td>17-12-2017</td>
<td>1JANATAMF</td>
<td>6.5</td>
<td>6.6</td>
<td>6.5</td>
<td>6.4</td>
<td>6.5</td>
<td>6.6</td>
<td>118</td>
<td>2.904</td>
<td>4,52,125</td>
</tr>
<tr>
<td>16-12-2017</td>
<td>1JANATAMF</td>
<td>6.5</td>
<td>6.6</td>
<td>6.5</td>
<td>6.6</td>
<td>6.6</td>
<td>6.6</td>
<td>36</td>
<td>0.596</td>
<td>90,597</td>
</tr>
</tbody>
</table>

Figure 5: Raw Data

This is a pictorial representation of the data present in our .xlsx file. This particular file contains 121608 such records. There are more than ten different trading codes available in the dataset, and some of the records do not have relevant information that can help us train the machine, so the logical step is to process the raw data. Thus we obtain a more refined dataset which can now be used to train the machine.

<table>
<thead>
<tr>
<th>DATE</th>
<th>TRADING CODE</th>
<th>LTP</th>
<th>HIGH</th>
<th>LOW</th>
<th>OPENP</th>
<th>CLOSEP</th>
<th>YCP</th>
<th>TRADE</th>
<th>VALUE (mn)</th>
<th>VOLUME</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-08-16</td>
<td>1JANATAMF</td>
<td>6.2</td>
<td>6.3</td>
<td>6.1</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>56</td>
<td>0.757</td>
<td>122,741</td>
</tr>
<tr>
<td>2018-08-16</td>
<td>1STPRIMFMF</td>
<td>11.2</td>
<td>11.2</td>
<td>10.9</td>
<td>11.0</td>
<td>11.1</td>
<td>10.9</td>
<td>145</td>
<td>2.640</td>
<td>238,810</td>
</tr>
<tr>
<td>2018-08-16</td>
<td>AAMRANET</td>
<td>80.1</td>
<td>80.4</td>
<td>78.5</td>
<td>78.5</td>
<td>79.7</td>
<td>78.3</td>
<td>545</td>
<td>15.488</td>
<td>195,035</td>
</tr>
<tr>
<td>2018-08-16</td>
<td>AAMRATECH</td>
<td>30.8</td>
<td>31.6</td>
<td>30.7</td>
<td>31.0</td>
<td>30.9</td>
<td>31.0</td>
<td>195</td>
<td>5.100</td>
<td>164,899</td>
</tr>
<tr>
<td>2018-08-16</td>
<td>ABB1STMF</td>
<td>6.1</td>
<td>6.1</td>
<td>5.9</td>
<td>6.0</td>
<td>6.1</td>
<td>6.0</td>
<td>109</td>
<td>11.214</td>
<td>185,7588</td>
</tr>
</tbody>
</table>

Figure 6: Calculate Head

This is the result of using the head(). Since we are using the pandas library to analyze the data, it returns the first five rows. Here five is the default value of the number of rows it returns unless stated otherwise.
Figure 7: Time series plot of GP

Figure 8: Histogram of CLOSEP-OPENP

Figure 9: Histogram of HIGH-LOW
5. Conclusion

In the project, we proposed the use of the data collected from different global financial markets with machine learning algorithms to predict the stock index movements. SVM algorithm works on the considerable dataset value, which is collected from different global financial markets. Also, SVM does not give a problem of overfitting. Various machine learning-based models are proposed for predicting the daily trend of Market stocks. Numerical results suggest high efficiency. The practical trading models built upon our well-trained predictor. The model generates higher profit compared to the selected benchmarks.

By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the SVM algorithm. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.
6. References


