Stock Market Analysis and Prediction Applying Machine Learning

Varalakshmi K, Prashanth P, Mounika B, PavanKumar V, Yugandhar K, PavanKumar K

1,2,3,4,5Student, 6Professor
Department of CSE
Raghu Engineering College, Dakamarri, Visakhapatnam

Abstract: The Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange and explains the prediction of a stock applying Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market applying machine learning is Python. Here, 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. We will use machine learning techniques called Support Vector Machine (SVM) and Naive Bayes to predict stock prices for the large and small capitalizations, employing prices with both daily and up-to-the-minute frequencies. The support vector machine (SVM) is a data classification technique that has been recently proven to perform better than other machine learning techniques especially in stock market prediction. SVM try to build a model using a set of training examples given to it. Each training data instance is marked as belonging to one of two categories. The SVM will attempt to classify the data instances into those two categories. The trained SVM model can then be tested with new data instances to predict which category they belong to based on the training performance. Bayesian networks are also used for prediction. The Naive Bayes algorithm is a Bayesian Network technique used for the Bayesian Network construction using the historical data. The algorithm is called Naive because it assumes that the features in a class are unrelated to the other features and all of them independently contribute to the probability calculation. The final prediction result is used to decide the algorithm which gives the best prediction of stocks based on the data.

Index Terms: Stock Market, Support Vector Machine, Naive Bayes, Prediction.

I. INTRODUCTION

The stock market is a complex system where shares of publicly-traded companies are issued, bought and sold. To some it is a nebulous, dark chasm where people gamble. Actually, it is not gambling at all. Why? Let’s say you put $100 on one roll of the dice. If you win, you win $X. If you lose, you lose the entire $100. When you invest in stocks, you will win $X or lose $Y. It’s rare to lose it all, unless of course you invest in a company that goes bust. You could say that the stock market is a group of people pitting their expertise against one another. The Stock Market is an Adversarial System of Trading. The stock market is a collection of millions of investors with diametrically opposing views. This is because when one investor sells a particular security, someone else must be willing to buy it. Since both investors cannot be correct, it is an adversarial system. In short, one investor will profit and the other will suffer loss. Therefore, it’s important to become well versed on the investment you are considering. There are many factors that determine whether stock prices rise or fall. These include the media, the opinions of well-known investors, natural disasters, political and social unrest, risk, supply and demand, and the lack of or abundance of suitable alternatives. The compilation of these factors, plus all relevant information that has been disseminated, creates a certain type of sentiment (i.e. bullish and bearish) and a corresponding number of buyers and sellers. If there are more sellers than buyers, stock prices will tend to fall. Conversely, when there are more buyers than sellers, stock prices tend to rise. Benjamin Graham, the father of value investing, once said, “The buyer of common stocks must assure himself that he is not making his purchase at a time when the general market level is a definitely high one, as judged by established standards of common-stock values.”

STOCK MARKET PREDICTION: Stock market price prediction for short time windows appears to be a random process. The stock price movement over a long period of time usually develops a linear curve. People tend to buy those stocks whose prices are expected to rise in the near future. The uncertainty in the stock market refrain people from investing in stocks. Thus, there is a need to accurately predict the stock market which can be used in a real-life scenario. The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modelling and predicting the variable stock market. The datasets of the stock market prediction model include details like the closing price opening price, the data and various other variables that are needed to predict the object variable which is the price in a given day. The previous model used traditional methods of prediction like multivariate analysis with a prediction time series model. Stock market prediction outperforms when it is treated as a regression problem but performs well when treated as a classification.

The aim is to design a model that gains from the market information utilizing machine learning strategies and gauge the future patterns in stock value development. The Support Vector Machine (SVM) can be used for both classification and regression. It has been observed that SVMs are more used in classification based problem like ours. The SVM technique, we plot every single data...
component as a point in n-dimensional space (where n is the number of features of the dataset available) with the value of feature being the value of a particular coordinate and, hence classification is performed by finding the hyperplane that differentiates the two classes explicitly.

Naïve Bayes algorithm is a classification technique which generates Bayesian Networks for a given dataset based on Bayes theorem. It assumes that the given dataset contains a particular feature in a class which is unrelated to any other feature. For example, an object is considered to be A because of some features. These features presence may depend on each other or on other features but all of the features presence independently contribute to the probability that this object is A. and that is the reason it is known as Naïve. Advantages of Naïve Bayes algorithm are it is easy to build and useful for very large datasets and even known to outperform highly sophisticated classification techniques.

II. RELATED WORK

The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction. Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect it. The technique that was employed in this instance was regression. Since financial stock marks generate enormous amounts of data at any given time a great volume of data needs to undergo analysis before a prediction can be made. Each of the techniques listed under regression has its own advantages and limitations over its other counterparts. One of the noteworthy techniques that were mentioned was linear regression. The way linear regression models work is that they are often fitted by applying the least squares approach, but they may alternatively be also be fitted in other ways, such as by diminishing the “lack of fit” in some other norm, or by diminishing a handicapped version of the least squares loss function. Conversely, the least squares approach can be utilized to fit nonlinear models. [1]

2. Stock Market Prediction Using Historical Data Analysis
The stock market prediction process is filled with uncertainty and can be influenced by multiple factors. Therefore, the stock market plays an important role in business and finance. The technical and fundamental analysis is done by sentimental analysis process. Social media data has a high impact due to it’s be helpful in predicting the trend of the stock market. Technical analysis is done [6] by applying machine learning algorithms on historical data of stock prices. The method usually involves gathering various social media data, news to extract sentiments expressed by individuals. Other data like previous year stock prices are also considered. The relationship between various data points is considered, and a prediction is made on these data points. The model was able to make predictions about future stock values.

3. A Survey on Stock Market Prediction Using SVM
The recent studies provide a well-grounded proof that most of the predictive regression models are inefficient in out of sample predictability test. The reason for this inefficiency was parameter instability and model uncertainty. The studies also concluded the traditional strategies that promise to solve this problem. Support vector machine commonly known as SVM provides with the kernel, decision function, and sparsely of the solution. It is used to learn polynomial radial basis function and the multi-layer perception classifier. It is a training algorithm for classification and regression, which works on a larger dataset. There are many algorithms in the market but SVM provides with better efficiency and accuracy. The correlation analysis between SVM and stock market indicates strong interconnection between the stock prices and the market index.

4. Machine Learning Approach In Stock Market Prediction
The vast majority of the stockbrokers while making the prediction utilized the specialized, fundamental or the time series analysis. Overall, these techniques couldn't be trusted completely, so there emerged the need to give a strong strategy to financial exchange prediction. To find the best accurate result, the methodology chose to be implemented as machine learning and AI along with supervised classifier. Results were tried on the binary classification utilizing SVM classifier with an alternate set of a feature list. The greater part of the Machine Learning approach for taking care of business [2] issues had their benefit over factual techniques that did exclude AI, despite the fact that there was an ideal procedure for specific issues. Swarm Intelligence [2] optimization method named Cuckoo search was most easy to accommodate the parameters of SVM. The proposed hybrid CS-SVM strategy exhibited the performance to create increasingly exact outcomes in contrast with ANN. Likewise, the CS-SVM display [2] performed better in the forecasting of the stock value prediction. Prediction stock cost utilized parse records to compute the predicted, send it to the user, and autonomously perform tasks like buying and selling shares utilizing automation concept. Naïve Bayes Algorithm was utilized. [8]

III. METHODOLOGY

1. NAIVE BAYES
Naïve Bayes is a classification algorithm for binary and multi-class classification problems. The technique is easiest to understand when described by applying binary or categorical input values. It assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naïve Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naïve Bayes is known to outperform even highly sophisticated classification methods. The in representation for Naïve Bayes is probabilities.

In the Eq. 1 below Bayes theorem provides a way calculating posterior probability $p (c \mid x)$ from $p(c)$, $p(x)$ and $p (x \mid c)$. 

\[ p(c \mid x) = \frac{p(x \mid c) p(c)}{p(x)} \]
Here \( P(C|X) \) is the posterior probability which tells us the probability of hypothesis C is true.

\[ P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \]  

(1)

\( P(X|C) \) is a conditional probability of occurrence of event X given hypothesis C is true.

It can be estimated from the training data. The frequency counts of the words are stored in hash tables during the training phase.

The following are the important steps to be performed in the algorithm:

1. The given dataset is to be converted into a frequency table.
2. Calculate probabilities of the events and using the probabilities create table.
3. By applying the Naïve Bayesian equation, calculate the posterior probability for all classes.
4. The class with the highest posterior probability is the outcome of the prediction.

In the implementation, the predicted prices are very much similar to the actual values that denote the algorithm capability to predict the values accurately.

2. SUPPORT VECTOR MACHINE

The main of Support Vector Machine algorithm is to create the best lines or decision boundary that can segregate N dimensional spaces into classes so that we can easily put the new data point in the correct category in the future. Support Vector Machine is considered to be one of the most suitable algorithms available for fine series prediction. Support Vector Machine chooses the extreme points that help in creating the hyperplanes. More formally, a support-vector machine constructs a hyperplane or et of hyperplanes in a high dimensional space, which can be used for classification, regression, or other tasks like outlier's detection.

Here we are going to use Radial Basis Function kernel. It is a popular kernel function is used in SVM. RBF is used to check the predictability of the prices against the actual stock prices. It does not forecast the price but rather compares the actual data points to the predicted data points. This kernel function returns the inner product between two points in a suitable feature space. We use RBF when there is no prior knowledge about the data.

\[ K(x_i, x_j) = \exp\left( -\pi||x_i - x_j||^2 \right) \]  

(2)

Where \( ||x_i - x_j|| \) is the Euclidean distance between \( x_i \) and \( x_j \).

By applying the distance in the original space we calculate the dot product of \( x_i \) and \( x_j \).

C is inverse of the strength of regular.

This system helps us in predicting the stock price and also helps to control the over fitting problem in the stock market prediction. This system has various advantages like best feature subset is selected, the subset which contain only those features which are highly related to the output function so that it help to increase the accuracy.

IV. IMPLEMENTATION

In this paper we focus on predicting the Stock Market by applying Machine Learning model i.e., Support Vector Machine (SVM) by RBF kernel and Naive Bayes.

Fig 1: Implementation

Steps for Stock Market Prediction
Step 1: In this step we are predicting the financial market value of any stock. So that the share value up to the closing date are 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 Separate Value) so that it will easily load into the algorithm.

Step 3: In the next step the CSV file of the dataset is taken to perform data preprocessing where the unsupervised data is converted into supervised data which is fit for prediction and split into training set and testing set.

Step 4: In the next step in which GUI is open and when we click on the SVM button it will show the window of predicted value of Stock based on the training set applying SVM algorithm.

Step 5: In this step algorithm display the predicted value graph of select stock which shows the original value and predicted value of the stock.

Step 6: The steps 4 and 5 are repeated when we click on the Naive bayes button and finally we will conclude the best Machine learning algorithm by observing the accuracy of both the algorithms.

V. RESULTS

<table>
<thead>
<tr>
<th>Date</th>
<th>Open Price</th>
<th>High Price</th>
<th>Low Price</th>
<th>Last Trade</th>
<th>Close Price</th>
<th>Total Trad Turnover (in Lakhs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/31/2021</td>
<td>502.59</td>
<td>506.61</td>
<td>500.29</td>
<td>502.57</td>
<td>502.9</td>
<td>4799358</td>
</tr>
<tr>
<td>12/30/2021</td>
<td>499.91</td>
<td>501.5</td>
<td>496.98</td>
<td>499.34</td>
<td>499.34</td>
<td>7007324</td>
</tr>
<tr>
<td>12/29/2021</td>
<td>497.96</td>
<td>501.14</td>
<td>497.01</td>
<td>498.77</td>
<td>498.1</td>
<td>5382146</td>
</tr>
<tr>
<td>12/28/2021</td>
<td>501.76</td>
<td>503.37</td>
<td>495.08</td>
<td>497.06</td>
<td>497.77</td>
<td>4262283</td>
</tr>
<tr>
<td>12/27/2021</td>
<td>502.71</td>
<td>507.7</td>
<td>499.31</td>
<td>500.14</td>
<td>501.26</td>
<td>2782404</td>
</tr>
<tr>
<td>12/24/2021</td>
<td>500.19</td>
<td>505.66</td>
<td>498.77</td>
<td>503.3</td>
<td>503.37</td>
<td>4548917</td>
</tr>
<tr>
<td>12/23/2021</td>
<td>502.47</td>
<td>505.75</td>
<td>499.01</td>
<td>504.11</td>
<td>503.23</td>
<td>7020886</td>
</tr>
<tr>
<td>12/22/2021</td>
<td>509.6</td>
<td>510.64</td>
<td>497.98</td>
<td>501.43</td>
<td>501.04</td>
<td>7281803</td>
</tr>
<tr>
<td>12/21/2021</td>
<td>503.47</td>
<td>510.59</td>
<td>502.69</td>
<td>509.79</td>
<td>509.76</td>
<td>7742619</td>
</tr>
<tr>
<td>12/20/2021</td>
<td>496.39</td>
<td>507.32</td>
<td>496.2</td>
<td>502.59</td>
<td>501.97</td>
<td>7631144</td>
</tr>
<tr>
<td>12/16/2021</td>
<td>502.42</td>
<td>503.52</td>
<td>494.04</td>
<td>502.45</td>
<td>502.02</td>
<td>7473637</td>
</tr>
<tr>
<td>12/15/2021</td>
<td>500.67</td>
<td>505.42</td>
<td>492.11</td>
<td>500.17</td>
<td>500.62</td>
<td>10315322</td>
</tr>
<tr>
<td>12/14/2021</td>
<td>495.44</td>
<td>503.52</td>
<td>494.96</td>
<td>501.43</td>
<td>501.26</td>
<td>9663328</td>
</tr>
<tr>
<td>12/13/2021</td>
<td>486.44</td>
<td>496.86</td>
<td>497.77</td>
<td>494.92</td>
<td>494.85</td>
<td>8100982</td>
</tr>
</tbody>
</table>

Fig 2: Stock dataset for TC1 Reliance Inc. In CSV file

![SVM](image)

Fig 3: Predicted output by SVM for TC1 Reliance Inc.
VI.

CONCLUSION

Stock Market follows the random walk, which implies that the best prediction you can have about tomorrow’s value is today’s value. The stock market indices are highly fluctuating and it effects the investor’s belief. Stock prices are considered to be a very dynamic and susceptible to quick changes because of underlying nature of the financial domain and in part because of the mix of a known parameter and unknown parameters. This paper will develop a financial data predictor program in which there will be a dataset storing all historical stock prices and data will be treated as training sets for the program. The main purpose of the prediction is to reduce uncertainty associated to investment decision making. We used two Machine Learning models called SVM (Support Vector Machine) and Navie Bayes to predict the future price of a stock. We have shown the comparison between the predicted price and actual price in Fig 3 and Fig 4. As it clearly visible from the graph that, SVM model gives accurate predictions. This method of predicting the return on investment will help in a great way to financial institutions and stock brokers to predict the future price in such uncertain conditions.
VII. FUTURE SCOPE

Although stock market prediction is extremely difficult, it provides an interesting and potentially rewarding area of research. Our methods provide a starting point for future market analysis on which there are several possible areas for improvement. We found the success of our relatively simple regressive techniques to be surprising, the extremely limited features that they examine, and the possibility that stock prices are independent from one another. It is possible that these methods were over-fit to our training examples, and would not be generalizable to other stocks. One possible means of addressing this issue would be combining linear regressive methods with other machine learning techniques such as neural networks or support vector machines. Regression is clearly the strongest method, showing higher predictive accuracy and an ability to actually make money in a real-world situation.

REFERENCES