International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195361

Stock Market Prediction Using Machine Learning

Prof. Shailendra Gaur[1*], Rishabh Bhardwaj[2], Vinay Bansal[2], Nidhi Kumari[2], Shalley Gupta[2]
[*1]Assistant Professor, Department of Information Technology, BPIT, Delhi, India
[2]Department of Information Technology, BPIT, Delhi, India

ABSTRACT

Stock price prediction is one of the most complex machine learning problems. It depends on a large number of factors which contribute to changes in the supply and demand. In this paper, we propose a stock prediction analysis using machine learning based on support vector machines (SVM), linear regression and reinforcement learning. SVM are favored in applications where text mining is used for market prediction. SVMs can be used for both linearly and non-linearly separable data sets. when the data is linearly separable, SVMs construct a hyperplane on the feature space to distinguish the training tuples in the data such that the margin between the support vectors is maximized. Correlation is used between stock prices of different companies to predict the price of a stock by using technical indicator of highly correlated stocks, not only stock to be predicted.

Keywords : Support Vector Machines, Prediction Model, Linear Regression, Prediction Using Decision Stumps, Expert Weighting, Online Learning

I. INTRODUCTION

1 An informal Introduction to Stock Market Prediction

Prediction of the stock market has always been a bane and goal of investors since its existence. Everyday a handsome amount of dollars is traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Entire companies rise and fall regularly on the basis of the behavior of the market. To make correct choices, investors have to judge based on technical analysis, such as company's charts, stock market indices and information from newspapers and microblogs. However, it is difficult for investors to analyze and forecast the market by churning all this information. In this prediction model, different types of attributes are selected on the basis of different literature review. Therefore, to predict the trends automatically, many Machine learning techniques have been investigated

2. Motivation

The main objective and purpose of the project is to study and apply Machine Learning Algorithms as possible on a dataset involving a particular domain, namely the Stock Market, as opposed to coming up with a newer (and/or better) algorithm that is more efficient in predicting the price of a stock.[2]In this paper, we discuss the Machine Learning techniques which have been applied for stock trading to guess the rise and fall of stock prices before the actual event of an increase or decrease in the stock price occurs. In particular the paper comprises the application of Support Vector Machines, Linear Regression, Prediction using Decision Stumps, Expert Weighting and Online Learning in detail along with the benefits and pitfalls of each method. The paper focuses on the parameters and variables that plays an important role in order to recognize the patterns in stock prices which can be helpful in the future prediction of the company's stocks and how Boosting can be combined

with other learning algorithms to improve the accuracy of such prediction systems.

II. RELATED RESEARCH

Numerous machine learning techniques have been used over the past decade to forecast the stock market. Machine learning techniques are by far the most widely used technique.[3][4]Support vector Machines have been used in for stock market trend prediction. Over the last few years, the use of SVMs for stock market forecasting has made significant progress. SVMs were initially used by Tay & Cao for the purpose of financial time series forecasting. Kim developed an algorithm to predict the stock market direction by using technical analysis indicators as input to SVMs. Studies have given a comparative analysis of SVM with Back Propagation Neural Networks (BPN). The experimental results showed that SVM outperformed BPN most often though there are some markets for which BPN have been found to be better. These results may play an important role to the fact that the SVM implements the structural risk minimization principle and this leads to better generalization than Neural Networks, which implement the empirical risk minimization.

III. SUPPORT VECTOR

1. What is Support Vector Machine?

Support Vector Machine (SVM) is a administered machine learning algorithm that can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot the facts and figures (data) as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. [5]

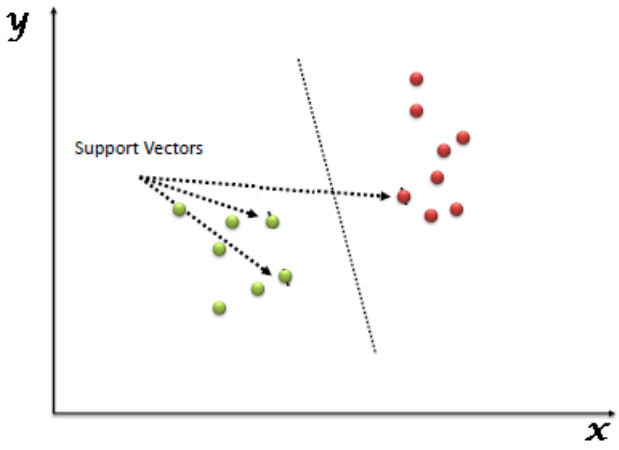


Fig 1

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best isolates the two classes (hyper-plane/ line).

2. Implementation of SVM in python

In Python, scikit-learn is a extensively used library for applying machine learning algorithms, SVM is also available in scikit-learn library and follow the same structure (Import library, object creation, fitting model and prediction). Let's look at the below code:

```
#Import Library
from sklearn import svm
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create SVM classification object
model = svm.svc(kernel='linear', c=1, gamma=1)
# there is various option linked with it, like changing kernel, gamma and C value. Will converse further about it in next section.Train the model by means of the training sets and check score
model.fit(X, y)
model.score(X, y)
#Predict Output
predicted= model.predict(x_test)
```

3. Pros and Cons associated with SVM

Pros:

- It works really well with clear margin of separation
- It is effective in high dimensional spaces.
- It is effective in cases where number 0of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.[5][6]

Cons:

- It doesn't perform well, when we have large data set because the required training time is higher
- It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping
- SVM doesn't directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. It is connected SVC method of Python scikit-learn library.

IV. LINEAR REGRESSION

Linear regression is widely used throughout Finance in a plethora of applications.[7] Linear regression is a method used to model a relationship between a dependent variable (y), and an independent variable (x). With simple linear regression, there will solitary be one self-governing variable x. There can be many independent variables which would fall under the category of multiple linear regression. The equation for linear regression can be written as:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here, x_1, x_2,….x_n represent the independent variables while the coefficients θ_1, θ_2, …. θ_n represent the weights.

1. Assumptions

Standard linear regression models with outmoded approximation practices make a variety of assumptions about the predictor variables, the response variables and the relationship between them. Frequent allowances have been developed that allow each of these conventions to be relaxed (i.e. reduced to a weaker form), and in some cases eradicated completely. Generally, these extensions cause the estimation procedure to be rather composite and laborious, and may also require even more data in order to produce an accurate model.

2. Interpretation

A fitted linear regression model can be used to recognize the connection between a single predictor variable xj and the response variable y when all the other predictor variables in the model are constant. Explicitly, the understanding of βj is the anticipated change in y for a one-unit change in xj when the other covariates are held fixed, that is, the expected value of the fractional derivative of y with respect to xj. This is from time to time called the unique effect of xj on y. In contrast, the marginal effect of xj on y can be evaluated using a correlation coefficient or simple linear regression model relating only xj to y; this effect is the total derivative of y with respect to xj.

V. MONTE CARLO METHOD

Monte Carlo methods are a wide-ranging period of computational algorithms that depends on repeated random sampling to get the statistical or numeric results. Their indispensable idea is using uncertain values to crack complications that might be deterministic in principle. Monte Carlo methods are mainly used in three problem classes: numerical integration, optimization, and generating draws from a probability distribution. In principle, Monte Carlo methods can be used to resolve any difficult having a probabilistic analysis. By the law of large numbers, integrals defined by the expected value of some random variable can be estimated by taking the empirical mean (a.k.a. the sample mean) of self-governing samples of the variable.

Monte Carlo imitation is commonly used to evaluate the threat and improbability that would affect the result of dissimilar decision options. Monte Carlo simulation allows the business risk predictor to incorporate the total effects of uncertainty in variables like sales volume, commodity and labor prices, interest and exchange rates, as well as the consequence of distinct risk events like the withdrawal of a contract or the change of a tax law.

Monte Carlo methods in finance are often used to calculate funds in projects at a business unit or corporate level, or to estimate monetary derivatives. They can be used to model project programs, where imitations combined guesstimates for worst-case, best-case, and most likely periods for each task to define consequences for the overall project. Monte Carlo methods are also used in option pricing, defaulting risk analysis.

VI. SOME RESULTS AND FINDINGS

1 AAPL(Apple) Stock Data Frame to get a feel for the data.

Table 1

```
In [4]: AAPL.head()
```

Out[4]:

	Open	High	Low	Close	Volume
Date					
2016-10-27	115.39	115.86	114.10	114.48	34562045
2016-10-28	113.87	115.21	113.45	113.72	37861662
2016-10-31	113.65	114.23	113.20	113.54	26419398
2016-11-01	113.46	113.77	110.53	111.49	43825812
2016-11-02	111.40	112.35	111.23	111.59	28331709

2 Summary stats of Apple Stocks

Table 2

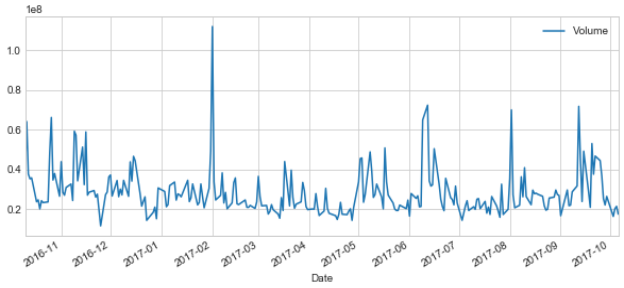
```
In [5]: # Summery stats for Apple Stock
        AAPL.describe()
```

Out[5]:

	Open	High	Low	Close	Volume
count	251.000000	251.000000	251.000000	251.000000	2.510000e+02
mean	139.381474	140.304622	138.494263	139.488327	2.805794e+07
std	17.106701	17.101638	16.891555	16.951448	1.193381e+07
min	106.570000	107.680000	104.080000	105.710000	1.147592e+07
25%	120.435000	121.100000	120.025000	120.715000	2.082378e+07
50%	143.720000	144.500000	143.100000	143.700000	2.559729e+07
75%	153.880000	154.450000	152.900000	153.805000	3.195864e+07
max	164.800000	164.940000	163.630000	164.050000	1.119850e+08

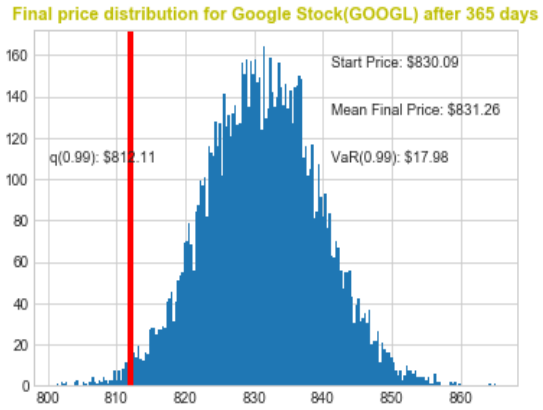
3 Total volume of stocks being traded each day over pass years

```
<matplotlib.axes._subplots.AxesSubplot at 0x23b90cf27f0>
```



4. Final Price Distribution for Google

```
<matplotlib.text.Text at 0x12999daf4a8>
```



VII. CONCLUSION

Stock Market is the alleviation of risk through the spreading of investments across multiple establishments, which is achieved by the combining a number of small investments into a large bucket. Stock Market is an appropriate investment for the common man as it offers an opportunity to invest in an expanded, professionally managed portfolio at a relatively low cost. With the help of this program, the user can analyze the current market conditions and make an informed decision based on the statistics. Using these statistical depictions, the investor can mitigate the risk and can achieve better returns. Predicting the stock market however is not an easy task and no one can do provide cent percent assurance, but with the use of such a program, the risk can slightly be depicted.

VIII. REFERENCES

[1]. Wei Huanga;b, Yoshiteru Nakamoria, Shou-Yang Wangb.Forecasting stock market movement direction with support vector machine. Computers & Operations Research 32 (2005) ,2513 – 2522. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.4127&rep=rep1&type=pdf

[2]. Vatsal H.shah.Machine learning techniques for stock prediction.[online]Available at: https://www.scribd.com/document/78997928/Machine-Learning-Techniques-for-Stock-Prediction

[3]. Shunrong Shen,Haomiao Jiang,Tongda Zhang. Stock Market Forecasting Using Machine Learning Algorithms(2012).Available at: https://cs229.stanford.edu/proj2012/ShenJiangZhangStockMarketForecastingusingMachineLearningAlgorithms.pdf

[4]. Rohit Choudhry, Kumkum Garg.A Hybrid Machine Learning System For Stock Market.world Academy Of Science, Engineering and Technology 39 (2008) Available at: https://www.researchgate.net/publication/238747905_A_Hybrid_Machine_Learning_System_for_Stock_Market_Forecasting

[5]. AnalyticsVidhya(2017).SVM.[online]Available at: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

[6]. Datascience904.SVM.[online]Available at: https://datascience904.wordpress.com/support-vector-machine/

[7]. Programmingforfinance(2018).linear-regression.[online]Available at: https://programmingforfinance.com/2018/01/predicting-stock-prices-with-linear-regression/

[8]. Investopedia(2019).montecarlo.[online]Available at: https://www.investopedia.com/articles/07/montecarlo.asp

Cite this article as :

Prof. Shailendra Gaur, Rishabh Bhardwaj, Vinay Bansal, Nidhi Kumari, Shalley Gupta, "Stock Market Prediction Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 3, pp. 206-210, May-June 2019. Available at doi : https://doi.org/10.32628/CSEIT195361

Journal URL : http://ijsrcseit.com/CSEIT195361