iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Stock Market Price Trend Forecasting using Machine Learning

Dr. Pankaj Agarwal[1], Shikhar Bajpai[2], Anjali Pathak[3], Rohit Angira[4]
[1]*HOD(CSE) ,* [2,3,4]*Students, IMSEC*

Abstract: support vector machine is a machine learning technique used in to forecast stock prices. Stock market price forecasting is the process of determining the future value of a company's stock. This study uses daily closing prices for many technology stocks to calculate price volatility and momentum for individual stocks and for the overall sector. These are basically used as parameters to the svm model. This model attempts to predict whether a stock price sometime in the future will be higher or lower than it is on a given day. We find little predictive ability in the short-run, but definite predictive ability in the long-run.

I. INTRODUCTION

Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques. Financial institutions and active traders have created various proprietary models to beat the market for themselves or their clients, but rarely did anyone achieve consistently higher than the average returns on investment. The challenge of stock market price forecasting is so appealing because an improvement of just a few points of percentage can increase the profit by millions of dollars. This paper discusses the application of Support Vector Machines and Linear Regression in detail along with the pros and cons of the given methods. The paper introduces the parameters and variables which can be used to recognize the patterns in stock prices which can be helpful in future stock prediction and how boosting can be integrated with various other machine learning algorithms to improve the accuracy of our prediction systems. [10]

Fundamental and technical analyses were the first two methods which were used to forecast stock prices. Artificial Neural networks (ANNs) is the most commonly used technique. In most of the cases ANNs suffer from over-fitting problem because of the large number of parameters to fix in the analyzed problem. Also, Support vector machines (SVMs) have been developed as an alternative which avoids such limitation. Its practical success can be attributed to solid theoretical foundation. SVM computes globally optimal solutions, unlike those obtained by ANNs, which tend to fall into local minima.

II. BACKGROUND

A. *Stock Prediction in Detail*

In general, there are 2 Stock value Prediction Methodologies:

1) *Fundamental Analysis:* This method is performed by the Fundamental Analysts and is concerned more with the company instead of the actual stock. The analysts take decisions based on the past performance of the company. [1]
2) *Technical Analysis:* This method is performed by the Technical Analyst. Using time-series analysis, it deals with the stock price based on the past patterns of the stock. We do a technical Analysis when applying Machine Learning algorithms to stock data to check if our algorithm can learn the underlying patterns in the stock accurately. Machine Learning can also play an important role in evaluating and predicting the performance of the company. The most successful stock prediction system use some kind of a hybrid model involving both Fundamental and Technical Analysis.

B. *Stock Market Efficiency*

Much economic research has been conducted into the EMH (Efficient Market Hypothesis) theory, which depicts that stock prices reflect all available information and are unpredictable. According to the Efficient Market Hypothesis (EMH), stock prices will only respond to new information and hence will follow a random walk. If they only respond to new information, they cannot be predicted. The stocks which follow a random walk is basically a sign of market efficiency [2]. Particularly, there are three variants of this theory- weak, semi-strong, and strong. Research has concluded that the semi-strong version of the theory holds true. This version claims that all publicly available information is reflected by the stock prices, but the private information can be used to

predict profits.

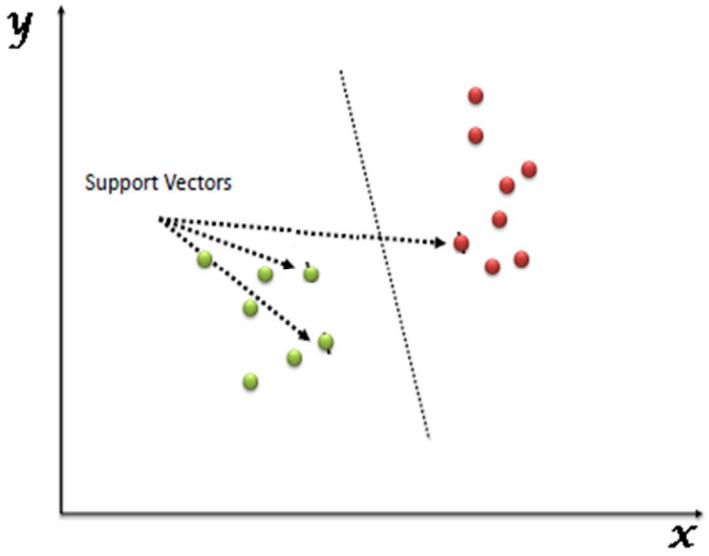
The Random Walk Hypothesis states that stock prices does not depend on past values and therefore patterns cannot be exploited since trends does not exist. Using the more powerful computing infrastructure which includes hardware and software, trading companies build efficient trading systems which can exploit the underlying patterns if a huge amount of data points are made available to them. Clearly with huge datasets, Machine Learning Techniques can seriously challenge the EMH.

C. General Machine Learning

Generally, there are two classes of machine learning techniques, the first is supervised learning in which the training data is a series of labelled examples. Each example is a collection of features that is labelled with the correct output corresponding to that feature set. This means that the algorithm is given features and outputs for a particular data set. It then applies what it learns from this particular dataset to predict the outputs for some another data set that is the test data. Unsupervised learning includes examples where the feature set is unlabelled. The algorithms try to cluster the data into distinct groups. [3] Supervised learning can be further divided into classification and regression problems. In classification, there is a set number of outputs that a feature set can be labelled as, while the output can take on continuous values in regression problems.

D. Support Vector Machines

Support Vector Machine (SVM) is a Supervised machine learning algorithm which can be used for both classification and regression problems. However, it is most commonly used in classification problems. In this, we plot each data item as a point in n-dimensional space where n is number of features, with the value of each feature be the value of a particular coordinate. We perform classification by finding the hyper-plane which differentiate the two classes. [1]



Support Vector Machine is one of the best binary classifiers. It creates a decision boundary in such a way that most points in one category fall on one side of the boundary while most of the points in the other category fall on the other side of the boundary.

III. IMPLEMENTATION

A. Data Collection

The data that we have used in the project is from Quandl which gives the statistical data in a csv format. We also referred to the data collected from yahoo finance for over 5 years. We choose to go with S&P 500 stock market to apply our method. The data consists of annual earnings, key statistics as well as Quarterly earnings. We have chosen 35 features upon which we will perform our supervised machine learning model .Then from which we can know which are the underperforming stocks and over performing stocks.[2][5]

Stock	S&P 500

Features	Debt equity Ratio, Trailing P/E, Price/Sales, Price/Book, Profit Margin, Operating Margin, Return on Assets, Return on Equity, Revenue Per Share, Market Capital, Enterprise Value , Forward P/E, PEG Ratio, Enterprise Value/Revenue, Enterprise Value/EBITDA, Revenue, Gross Profit, EBITDA, Net Income Annual to Common , Diluted EPS, Earnings Growth, Revenue Growth, Total Cash, Total Cash Per Share, Total Debt, Current Ratio, Book Value Per Share, Cash Flow, Beta, Held by Insiders, Held by Institutions, Shares Short (as of, Short Ratio, Short % of Float, Shares Short [2][5]
Data Source	Quandl , Yahoo Finance

B. Scaling and Normalising Features

Scaling and normalizing data in machine learning plays an important role for productive and systematic results. Data pre-processing relate all methods of processing applied on raw data to prepare it for *another* processing task [9] [10]. This involves many more procedures rather than just scaling and normalizing your data. The process of pre-processing can even include the use of unsupervised machine learning to reduce the amount of features in total to increase speed and efficiency. To maintain the correctness of our result we are randomising the data, we're going to shuffle our data for the process of learning. Some of the problems we encounter right now is that we're training on, for example, 80% of the data. Another problem is that our data setting is in alphabetical order of stocks in the stock market. Hence it means that, when we're testing, we're testing on stocks that begin with A-P, for example, and then leaving the Q-Z stocks for testing.

Accuracy = Correct count/test cases

Accuracy	55% ~ 65%

Our focus is to normalise the incorrect predictions. Our main goal is that our "loser" stocks that we pick are not very bad. Consider for example:

Machine Learning Algorithm is 85% accurate, and chooses 100 stocks to invest in:

1) 15 stocks perform an average -88% compared to market.
2) 85 stocks perform an average +2% compared to market.
3) This equates to an average of -11.5% in the end compared to market.

Next, instead, consider

Machine Learning Algorithm is 52% accurate, and chooses 100 stocks to invest in:

4) 48 stocks perform an average -2% compared to market.
5) 52 stocks perform an average +2% compared to market.
6) This equates to an average of +8% in the end compared to market.

This is an overstate example, of course, but it tells a major critical condition within the field of investing or trading. People often get hyper-focused on accuracy, rather than its performance. When people focus on the performance, they tend to focus there as well, ignoring the importance of risk.

C. Back-Testing Strategy

We finally finish up finding out our strategy is more than 60% accurate in its predictions, but, given our peculiar example, our main concern is to know the "degree" of our accuracy. We need to know how our out-performing predictions actually perform in an investing scenario, and how poorly or bad our incorrect predictions perform. So now we will be building a document that will actually consider a trading and investing scenario, using our method of machine learning algorithm. From the earlier datasets, we just trained with most of the samples and also trained against some in the last that is simply what we are going to do here. [7]

Our strategy for trading or investing is very simple, and can be improved easily, but the very basics of the algorithm go as follows: Using this machine learning algorithm, we make the predictions of the well performing companies. The learning was just to find major companies that were likely to out-perform "in the market" (S&P 500), over the course of one year. Hence, to do back testing,

we first train on a set of data, then we test against another set of data that is unlike the data we particularly trained on. So we go through some data, and then we find companies that the machine learning algorithm suggests us to buy. We then note down those companies, and evaluate their performances one-year later [11]. After that we add all of the outcomes of the stocks, as if we really had invested, and then we see how effective our algorithm is.

IV. CONCLUSION

In this project, we applied supervised learning technique i.e. Support Vector Machines in predicting the stock price trend of S&P 500 stock market. Our finds can be summarized into three aspects:

A. We found that SVM model can provide the predicting accuracy of (65%), We predict the stock price such that our "loser" stocks that we pick are not extremely bad.

B. Our feature selection analysis indicates that when use all of the 32 features. That's because the number of data points in the problem is much bigger than that of our features

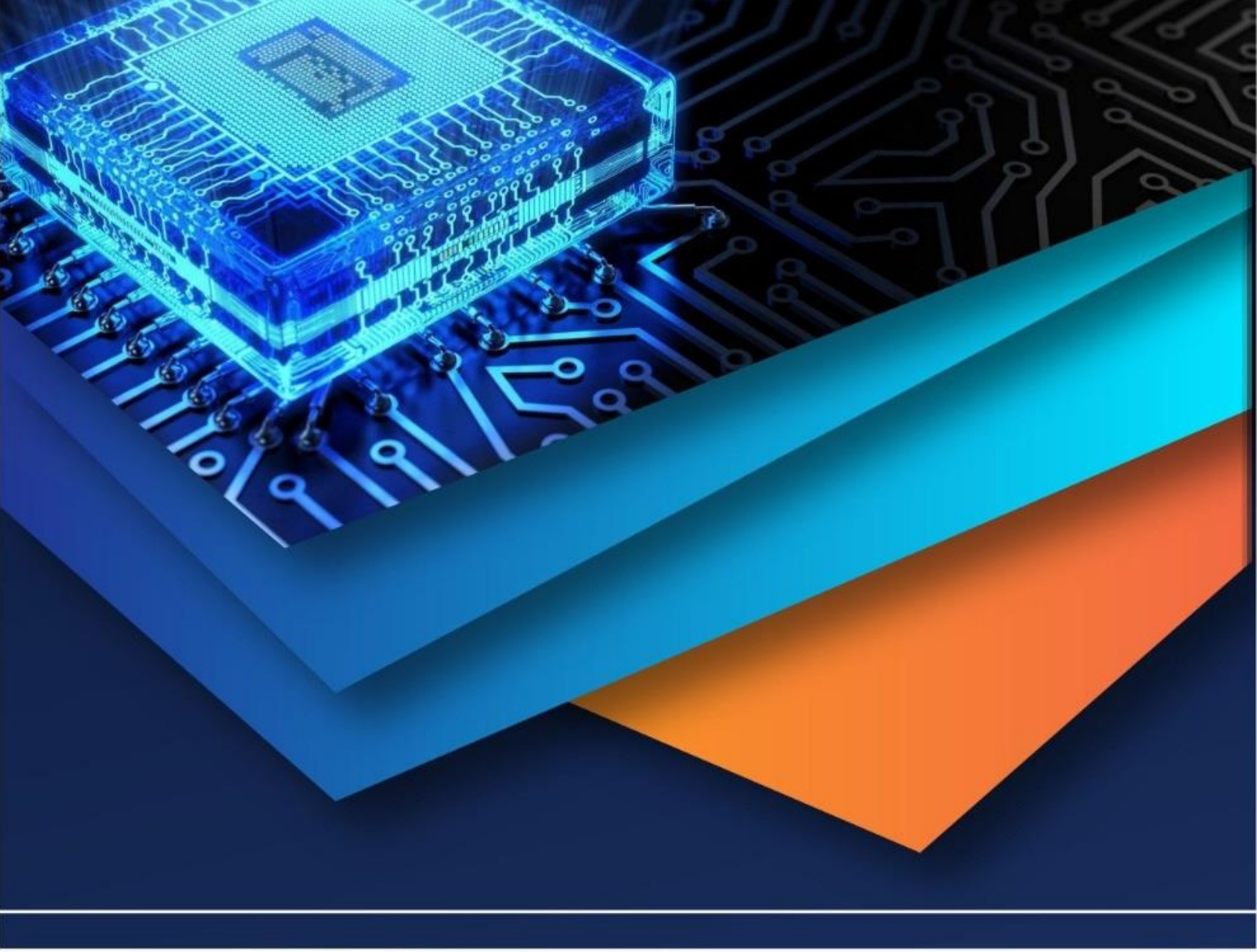
C. The trading strategy based on our prediction achieves very positive outputs by significantly outrunning the stock performance. For the future work, we believe that we can make the following improvements:

D. Test our predictor on different stocks to see its robustness. Try to develop a "more general" predictor for the stock market.

E. More improved back-strategy technique and feature selection process to diversify the risk

REFERENCES

[1] Abani, R., Kwek, S. & Japkowicz, N., 2005. support vector machines to classified datasets. s.l., Springer, pp. 39-50.
[2] Basky, R. B. & De Longe, J. B., 1992. Why there are variations in the stock market ?, s.l.: National Bureau of Economic Research.
[3] Hall, M. et al., 2009. The WEKA is a data mining application:
[4] Henderson, T. & Coulton, P. C., 2009. Automation, speed, and stock market quality: The NYSE's hybrid. Journal of Financial Markets, 14(4), pp. 167--424.
[5] Huang Ho, V., Nakamori, Y. & Wang, S.-Y., 2005. Forecasting stock market movement direction with the help of support vector machine. Computers & Operations Research, 202(10), pp. 2513-2421.
[6] Tsang, Z. et al., 2009. Credit rating analysis with regression anal
[7] ysis and neural networks: a market comparative study. Decision support systems, 37(4), pp. 245-788. M.J. Lee, Y.-C., 2007. Application of support vector machines to corporate world credit rating prediction. Expert Systems with the Applications, 33(1), pp. 67-74.
[8] Min, k. H. & Sung, Y.-C., 2005. Bankruptcy forecasting using support vector machine with preferable choice of main function parameters. Expert system with applications, 28(4), pp. 603-614.
[9] Phua, S. K. H., Zhu, X. & Koh, F. H., 2003. Forecasting stock index increases using artificial neural networks with cluster region methods. s.l., IEEE, pp. 260-265.
[10] Schumaker, R. P. & Chen, H., 2009. analysis of stock market prediction using financial news: The AKFin text system. ACM Transactions on Information Systems (TOIS), 47(2).
[11] Tsong, C. & Wang, S., 2009. Stock price prediction by intelligent machine learning techniques. s.l., s.n.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY