Stock Market Prediction Using Machine Learning Techniques

Rushikesh M. Khamkar
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering
rushi.khamkar5678@gmail.com

Rushikesh P. Kadam
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering
rpkadam27@gmail.com

Moushmi R. Jain
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering
jainmoushmi@gmail.com

Ashwin Gadupudi
Department of Computer Engineering,
Smt. Kashibai Navale College of Engineering
ashwingg97@gmail.com

Abstract

The Stock Market prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction using different Support Vector Machine(SVM), Single Level Perceptron, Multi -Level Perceptron, Radial Bias Function.

General Terms

Support vector machine, radial basis function, multi-level perceptron, single level perceptron, Machine learning.

Keywords:

Stock Market, Stock prediction, Machine learning, Classification of Stocks.

I.INTRODUCTION

For a new investor, the share market can feel a lot like legalized gambling. Randomly choose a share based on gut instinct. If the value of your share goes up you're in profit else you're in loss. The share market can be intimidating, but the more you learn about shares, and the more you understand the true nature of stock market investment, the better and smarter you'll manage your money.

Terms:

• A stock of a company constitutes the equity stake of all shareholders.

• A share of stock is literally a share in the ownership of a company[1]. When investor purchases a share of stock, the investor is entitled to a small fraction of the assets and profits of that company.

• Assets include everything the company owns (real estate, equipment, inventory)

• Shares in publicly traded companies are bought and sold at a stock market or a stock exchange.

These are some examples of popular stock exchanges:

• NYSE - New York Stock Exchange

•NASDAQ - National Association of Securities Dealers

• NSE – National Stock Exchange (India)

• BSE – Bombay Stock Exchange

There is no way to predict the accurate trends in stock market. Many factors affect rises the share prices of different companies[1]. The best way to understand stock markets is to analyze and study how the market movements have been in the past[2].

Share market trends tend to repeat themselves overtime. After you study the cycle of a particular stock, you can make predictions about how it will change over the course of time[3]. Some stocks might be truly arbitrary in which case the

movement is random but in most of the cases there is a particular trend that repeats itself. Recognizing these patters will enable you to predict the future trend[1].

The project goal is to build a system where the machine learning algorithms try to predict the prices of stocks based on their previous closing prices and other attributes that influence its price like Interest rates, Foreign exchange and Commodity prices[4].

II.MOTIVATION

Stock market movements make headlines every day. In India, 3.23 crore individual investors trade stocks. Maharashtra alone accounts for one-fifth of these investors. However, a report from Trade Brains shows that 90% of these investors lose money in due to various reasons like insufficient research, speculation, trading with emotions etc.

Higher inflation rate and lower interest rate makes it ineffective to put one's money into savings account or fixed deposits.[5][6] Thus, many people look up to stock market to keep up with the inflation. In this process of multiplying their money many investors have made a fortune while, some have lost a lot of money due to unawareness or lack of time to research about a stock.

There are lots of contradicting opinions in the news and an individual may not have the time or may not know how to research about a stock. Most importantly, it is very difficult to manually predict the stocks prices based on their previous performance of that stock. Due to these factors many investors lose a lot of money every year[6].

A system that could predict the stock prices accurately is highly in demand. Individuals can know the predicted stock prices upfront and this may prevent them from investing in a bad stock. This would also mean a lot of saved time for many of the investors who are figuring out wheather a particular stock is good or not.

III.LITERATURE SURVEY

1. Comparative analysis of data mining techniques for financial data using parallel processing[1]

[2014] [IEEE] Do the comparative analysis of several data mining classification techniques on the basis of parameters accuracy, execution time, types of datasets and applications. Simple Regression and multivariate analysis used, Regression analysis on attributes is used. No use of machine learning. Does not provide the algorithm used.

2. Stock market prices do not follow random walks: Evidence from a simple specification test[2]

[2015] [IEEE]Test the random walk hypothesis for weekly stock market returns by comparing variance estimators derived from data sampled at different frequencies. Simple trading rules extraction and Extraction of Trading Rules from Charts and Trading Rules. No alternative provided for human investing. Show only the flaws on manual investments.

3. A Machine Learning Model for Stock Market Prediction[3]

[2017] [IJAERD] Support Vector Machine with Regression Technology (SVR), Recurrent Neural networks (RNN). Regression analysis on attributes using simple Regression and multivariate analysis used. It is not tested in real market. Shows how social media affects share prices. Does not account for other factors.

4. Twitter mood predicts the stock market[4]

[2010] [IEEE] Analyze the text content of daily Twitter feeds by two mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States. These results are strongly indicative of a predictive correlation between measurements of the public mood states from Twitter feeds. Difficult to scan each every text extraction from large set of data, difficult Text mining.

5.Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets[5]

[2017] [Research] Propose a generic framework employing Long Short-Term Memory (LSTM) and convolutional neural network (CNN)for adversarial training to forecast high frequency stock market. This model achieves prediction ability superior to other benchmark methods by means of adversarial training, minimizing direction prediction loss, and

forecast error loss. It Can’t predict Multi scale Conditions and live data

6. Stock Market Prediction Using Machine Learning[6]

[2016] [IEEE] Uses different modules and give different models and give best accuracy using live streaming data. Predict Real Market Data and calculate Live data using single and multilevel perspective, SVM, Radial Bias. It Couldn’t work Textual Data form different Browsing Data (Web Crawling)

7. Stock Market Prediction by Using Artificial Neural Networks[7]

[2014] [IEEE] This model takes help of Artificial Intelligence and uses only neural networks to predict the data. Predicting data using single and multi-level perceptron. It uses 10 hidden layers with the learning rate of 0.4, momentum constant at 0.75 and Max Epochs of 1000. This model doesn’t use machine learning algorithms like SVM and radial basis function to determine their accuracy.

8. Price trend prediction Using Data Mining Algorithm[8]

[2015] [IEEE] This paper presented a data mining approach to predict the long-term trend of the stock market. The proposed model detects anomalies in data according to the volume of a stock to accurately predict the trend of the stock. This paper only provides long term predictions and does not give predictions to the immediate trends.

IV. GAP ANALYSIS

Sr.	Paper Title	Summary	Advantages	Disadvantages
1	Comparative analysis of data mining techniques for financial data using parallel processing [2014] [IEEE]	Do the comparative analysis of several data mining classification techniques on the basis of parameters accuracy, execution time, types of datasets and applications.	Simple Regression and multivariate analysis used, Regression analysis on attributes is used	No use of machine learning. Does not provide the algorithm used.
2	Stock market prices do not follow random walks: Evidence from a simple specification test [2015] [IEEE]	Test the random walk hypothesis for weekly stock market returns by comparing variance estimators derived from data sampled at different frequencies	Simple trading rules extraction and Extraction of Trading Rules from Charts and Trading Rules	No alternative provided for human investing. Show only the flaws on manual investments.
3	A Machine Learning Model for Stock Market Prediction [2017] [IJAERD]	Support Vector Machine with Regression Technology (SVR), Recurrent Neural networks (RNN)	Regression analysis on attributes using simple Regression and multivariate analysis used	It is not tested in real market. Shows how social media affects share prices. Does not account for other factors.
4	Twitter mood predicts the stock market [2010] [IEEE]	Analysed the text content of daily Twitter feeds by two mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States	These results are strongly indicative of a predictive correlation between measurements of the public mood states from Twitter feeds	Difficult to scan each every text extraction from large set of data, difficult Text mining
5	Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets [2017] [Research]	Propose a generic framework employing Long Short-Term Memory (LSTM) and convolutional neural network (CNN)for adversarial training to forecast high frequency stock market	This model achieves prediction ability superior to other benchmark methods by means of adversarial training, minimizing direction prediction loss, and forecast error loss	Can't predict Multi scale Conditions and live data
6	Stock Market Prediction Using Machine Learning	Uses different modules and give different models and give best accuracy using live streaming data.	Predict Real Market Data and calculate Live data using single and multilevel perspective, SVM, Radial Bias	Couldn't work Textual Data form different Browsing Data (Web Crawling)

V.PROPOSED WORK

Stock Market Prediction Using Machine Learning can be a challenging task. The process of determining which indicators and input data will be used and gathering enough training data to training the system appropriately is not obvious. The input data may be raw data on volume, price, or daily change, but also it may include derived data such as technical indicators (moving average, trend-line indicators, etc.)[5] or fundamental indicators (intrinsic share value, economic environment, etc.). It is crucial to understand what data can be useful to capture the underlying patterns and integrate into the machine learning system. The methodology used in this work consists on applying Machine Learning systems, with special emphasis on Genetic Programming. GP has been considered one of the most successful existing computational intelligence methods and capable to obtain competitive results on a very large set of real-life application against other methods. Section Different Algorithms used in algorithm[1].

Tools and Technologies Used

- Python
- Usage of libraries like – OpenCV, scikit, pandas, numpy
- Machine Learning techniques -classifiers
- Linear regression techniques
- Jupyter IDE

VI. METHODOLOGY

In this project we tried to predict the stock market prices using four different types of SVM and Artificial Neural Networks Algorithms.

5.1 Support Vector Machine (SVM)

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples[7], each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on[6].

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high dimensional feature spaces.

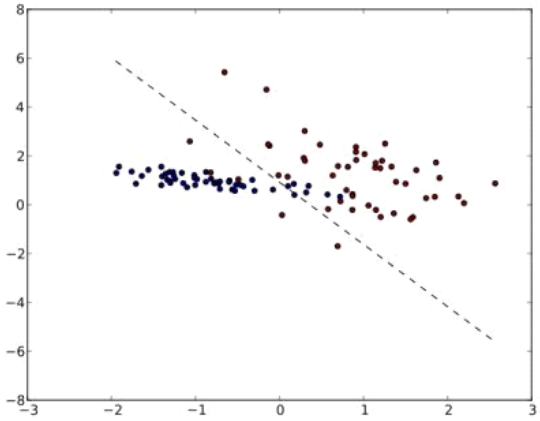


Figure 3. 1: demonstration of SVM

Linear discriminant function:

$$\mathbf{f(x) = w.x+b}$$

In this function, **x** refers to a training dataset vector, **w** is referred to as the weight vector and the value **b** as the bias term. The term **w.x** refers to the dot product (inner product, scalar product), which calculates the sum of the products of vector components.

Classification hyper-plane equations:
Positive margin hyper-plane equation: $\mathbf{w.x - b = 1}$
Negative margin hyper-plane equation: $\mathbf{w.x - b = -1}$
Middle optimum hyper-plane equation: $\mathbf{w.x - b = 0}$

5.2 Radial Bias

Radial basis function network is an artificial neural network which uses radial basis functions as activation functions. These networks are feed forward networks which can be trained using supervised training algorithms. These networks are used for function approximation in regression, classification and time series predictions[5]. Radial basis function networks are three layered networks where the input layer units does no processing, the hidden layer units implement a radial activation function and the output layer units implement a weighted sum of the hidden unit outputs. Nonlinearly separable data can easily be modeled

by radial basis function networks. To use the radial basis function networks we have to specify the type of radial basis activation function, the number of units in the hidden layer and the algorithms for finding the parameters of the network[3].

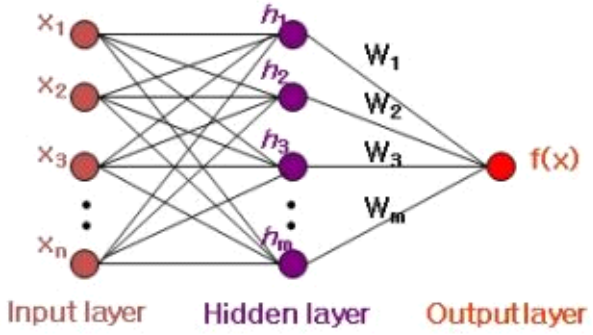


Figure 3.2: An demonstration of Radial Bias

$$\mathbf{h(x) = \Phi((x - c)^T R^{-1} (x - c))}$$

Where **Φ** is the function used, **c** is the center and **R** is the metric. The term $\mathbf{(x - c)^T R^{-1} (x - c)}$ is the distance between the input **x** and the center **c** in the metric defined by **R**. There are several common types of functions used such as Gaussian $\mathbf{\Phi(z) = e^{-z}}$, the multi-quadratic $\mathbf{\Phi(z)=(1+z)^{1/2}}$, the inverse multi-quadratic $\mathbf{\Phi(z) = (1+z)^{-1/2}}$ and the Cauchy $\mathbf{\Phi(z) = (1+z)^{-1}}$.

5.3 Single Layer and Multi-layer Perceptron

A single layer perceptron (SLP) is a feed-forward network based on a threshold transfer function. SLP is the simplest type of artificial neural networks and can only classify linearly separable cases with a binary target (1, 0)[1]. The single layer perceptron does not have a priori knowledge, so the initial weights are assigned randomly. SLP sums all the weighted inputs and if the sum is above the threshold (some predetermined value), SLP is said to be activated (output=1). The input values are presented to the perceptron, and if the predicted output is the same as the desired output, then the performance is considered satisfactory and no changes to the weights are made. However, if the output does not match the desired output, then the weights need to be changed to reduce the error[8].

A multi-layer perceptron (MLP) has the same structure of a single layer perceptron with one or more hidden layers. The backpropagation algorithm consists of two phases: the forward phase where the activations are propagated from the input to the output layer, and the backward phase, where the error between the observed actual and the requested nominal value in the output layer is propagated backwards to modify the weights and bias values[5]. 2 Propagation: Forward and Backward

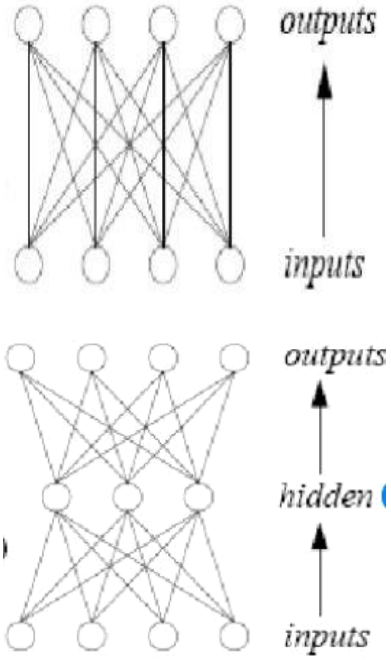
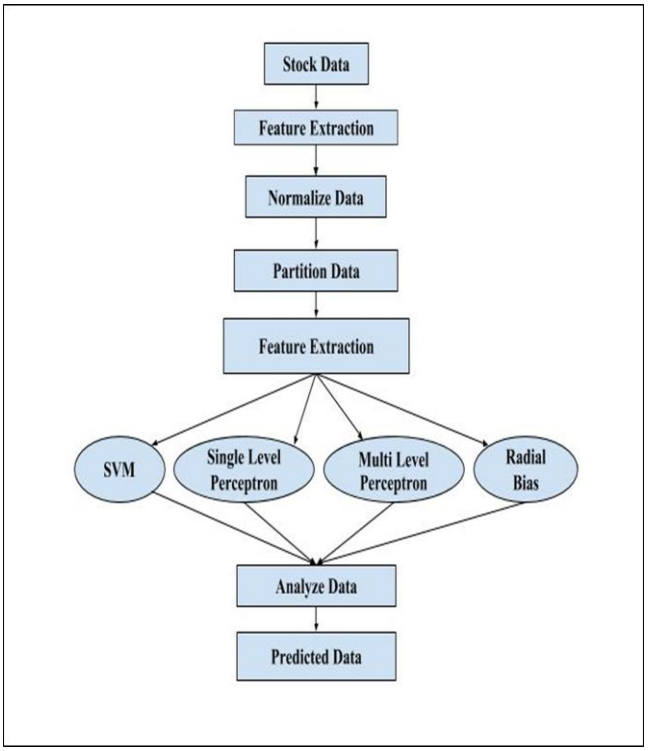


Figure 3.3: An demonstration of Single Level And Multi Level Perceptron

Single and multi-level perceptrons have multiple inputs and a single output. Consider **x1,x2,…xn** be input vectors and **w1,w2,…wn** be the weights associated with them[7].

Output $\mathbf{a = x1.w1 + x2.w2 + \ldots xn.wn}$

VII. SYSTEM ARCHITECHTURE



VIII. CONCLUSION

In this thesis, we looked at the problem of forecasting stock performance. Although a substantial volume of research exists on the topic, very little is aimed at long term forecasting while making use of machine learning methods and textual data sources. We prepared over ten year worth of stock data and proposed a solution which combines features from textual yearly and quarterly filings with fundamental factors for long term stock performance forecasting. Additionally, we developed a new method of extracting features from text for the purpose of performance forecasting and applied feature selection aided by a novel evaluation function. Problems Overcome[5]. To produce effective models, there were two main problems we were faced with and had to overcome. The first was that of market efficiency, which places theoretical limits on how patterns can be found in the stock markets for the purpose of forecasting. This property can become a concrete problem by patterns being exhibited in the data which are useless or even detrimental for predicting future values. The way we tried to deal with this was by carefully splitting our data into training, validation, and testing data with expanding windows so as to make maximum use of it while trying to avoid accidental overfitting. The second way we dealt with this was by using a tailored model performance metrics, which aimed to ensure good test performance of models by not only maximizing model validation, but also minimizing the variation across validation years of this value[7]. The third way we dealt with market efficiency was by performing feature selection using the Algorithm, so as to remove those features which performed poorly or unreliably. The second set of problems came from putting together a dataset to use for experimentation and testing. Due to the large volume of the data, care had to be taken when cleaning and preparing it, and the inevitable mistakes along the way required reprocessing of the data[4]. Using expert knowledge, we determined how to deal with the various problems in the data and ended up using mean substitution and feature deletion.

IX. FUTURE WORK

1. Model Updating Frequency:

They are trained once and then used for predicting stock performances over the span of a year. Since we use a return duration of 120 trading days, there is a necessary wait of half a year before data can be used to train models, which means that models end up making predictions using data which is over a year old. One way to make use of data as soon as it become available is to completely retrain the model every week (or less). A faster way to improve model performance may be through updating using incremental machine learning algorithms, which can update model parameters without re-training on all data[6].

2. Explore More Algorithms:

Although many different models were considered in this thesis, including various linear regression methods, gradient boosting, random forests, and neural networks, there is always more room to explore.

3. Improve Feature Extraction:

In this thesis, a few methods for extracting features from filings with textual data were explored. The problem of extracting features from text and determining text sentiment in particular are well studied, and other natural language processing methods may perform better. Our approach of using

autoencoders to extract features may also benefit from further exploration. In particular, when using the auxiliary loss, a more accurate method for estimating the financial effect corresponding to a given filing would be useful.

4. Utilize Time Series Information.:

Similar to the idea of updating model frequency, another area for exploration includes utilizing the time series aspect of the data. Our current models are not aware that the samples occur in any temporal order, and thus are not able to spot patterns in stock performance that depend on knowing the order of samples. One type of model that is often used to find and make use of these type of patterns are recurrent neural networks[9].

X.REFERENCES

[1] Raut Sushrut Deepak, Shinde Isha Uday ,Dr. D. Malathi, "Machine Learning Approach in stock market prediction"-2015 International Journal of Pure and Applied Mathematics Volume 115 No. 8 2017, 71-77.

[2] Tao Xing, Yuan Sun, Qian Wang, Guo Yu, "The Analysis and Prediction of Stock Price," 2013 IEEE International Conference on Granular Computing.

[3] A. W. Lo, & A. C. MacKinlay, "Stock market prices do not follow random walks: Evidence from a simple specification test," Review of financial studies, vol. 1, no. 1, pp. 41-66, 1988.

[4] Yash Omer , Nitesh Kumar Singh, "Stock Prediction using Machine Learning", 2018 International Journal on Future Revolution in Computer Science & Communication Engineering.

[5] Ritu Sharma, Mr. Shiv Kumar, Mr. Rohit Maheshwari "Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets", 2015 International Journal of Computer Science and Mobile Computing.

[6] Osma Hegazy, Omar S. Soliman, "A Machine Learning Model for Stock Market Prediction", International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013].

[7] S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee, "Stock Market Prediction using Machine Learning", International Journal of Advance Engineering and Research Development, vol. 4 2017.

[8] Xingyu Zhou , Zhisong Pan , Guyu Hu , Siqi Tang,and Cheng Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets", Mathematical Problems in Engineering, Volume 2018.

[9] J. Bollen, H. Mao, & X. Zeng, "Twitter mood predicts the stock market. Journal of Computational Science," vol. 2, no. 1, pp. 1-8, 2011.

[10] Ibrahim M. Hamed Ashraf S. Hussein, "An Intelligent Model for Stock Market Prediction", The 2011 International Conference on Computer Engineering & Systems