Stock Market Predication Using A Linear Regression

Dinesh Bhuriya
Govt. Women's Polytechnic, Indore
dineshbhuriya2006@gmail.com

Girish Kaushal
Shri Vaishnav Polytechnic, Indore
kaushalgirish@gmail.com

Ashish Sharma
HOD Govt. Women's Polytechnic, Indore
sharmaashish_gwpci@yahoo.co.in

Upendra Singh
upendrasingh49@gmail.com

Abstract: It is a serious challenge for investors and corporate stockholders to forecast the daily behavior of stock market which helps them to invest with more confidence by taking risks and fluctuations into consideration. In this paper, by applying linear regression for forecasting behavior of TCS data set, we prove that our proposed method is best to compare the other regression techniquemethod and the stockholders can invest confidentially based on that.

Keywords: Stock Market, Forecast, Data Mining, linear regression, Data set.

I. INTRODUCTION:

In stock market the most growing topic in research is "Forecasting"because of its importance and popularity among the masses and also small and large companies due to financial benefits and its low risk. Despite the risk of falling too much value per share due to market fluctuations rarely happens, but again, the risk is there. These fluctuations which effect on stock price and trading volume have some difficulties in forecasting. The fluctuations effect on the behavior of people in terms of capital savings or investment, the stock price and the increase or decrease of risk for investors. Therefore, in general, forecasting the stock market behavior through techniques and various methods is a useful tool to assist investors to act with greater certainty and taking the risks and volatility of an investment into consideration and know when to buy the cheapest price and when to sell to highest price [2].

In various fields such as economy, policy and engineering Data mining techniques have a more successful rate compare to traditional statistical method by discovering hidden knowledge of data [3, 4, and 5].Experience has shown that machine learning techniques could be successful in forecasting daily stock price and its trading volume [6]. In this paper, at first review the prior researches done for forecasting stock market and then we describe the importance of trading volume. By using linear regression, we forecastTCS data set [7] behavior and at the end we compared and evaluated the result of our proposed method with other approaches

II MOTIVATION

Stock market price prediction is a problem that has the potential to be worth billions of dollars and is actively researched by the largest financial corporations in the world. It is a significant problem because it has no clear solution, although attempts can be made at approximation using many different machine learning techniques. The project allows techniques for real-world machine learning applications including acquiring and analyzing a large data set and using a variety of techniques to train the program and predict potential outcomes.

II RELATED WORKS

Nowadays the stock market has been called for research in many fields due to its effects on financial challenging and capacity of forecasting its various

aspects through different scientific methods such as genetic algorithm, Artificial Neural Network (ANN) and other Meta heuristic algorithms.Many institution and academic researchers are trying to propose a method for forecasting next day behaviors of stock indexes in order to be better than the other methods, like a research that Majhi and other friends [8] did via applying bacterial foraging optimization technique for forecasting stock market and S&P500 indexes in short and long terms, and they made a linear combiner model which its weights updated by BFO and comparing it with Multi-Layer Perceptron (MLP) based method showed that Majhi and other friend's method has less calculative complexity and more precision to MLP method. Another forecasting system [9] in which counting of complex keyword topples and its transformation to predict stock market behavior periodically and doing real-time forecasting on web has been done. Some researchers used text mining approach [10], their findings investigate effects of financial news in forecasting stock market. Increasing social networks and their popularity among people have been led into new ideas of investigating of effect of the popularity and application of these social networks that can have on stock market behavior.

Like a work about effect of emotions like hope, fear and worry have on increasing or decreasing amount of Dow Jones on the next day [11] or investigating effects of Facebook [12] on stock market. The relation between the tendencies of investors and activities of stock market found by using a new time scale which operates on updated mood of about 100 million American Facebook users between the periods of 10/09/2007 to 10/09/2010. In this paper, forecasting of trading volume is considered which is similar to the introduced methods.

III PREDICATION METHOD

Regression predicts a numerical value [16]. Regression performs operations on a dataset where the target values have been defined already. And the result can be extended by adding new information [14]. The relations which regression establishes between predictor and target values can make a pattern. This pattern can be used on other datasets which their target values are not known. Therefore the data needed for regression are 2 part, first section for defining model and the other for testing model. In this section we choose linear regression for our analysis. First, we divide the data into two parts of training and testing. Then we use the training section for starting analysis and defining the model. 80% data used for training purpose and 20% data used for testing purpose.

Data Representation

The dataset that was used was collected from the TCS Stock Database [2] as a collection of comma-separated values where each row consisted of a stock on a specific day along with data on the volume, shares out, closing price, and other features for that day in time. The Python scientific computing library numpy was used along with the data analysis library pandas in order to convert these CSV files into pandas DataFrames that were indexed by date. Each specific stock is a view of the master DataFrame that is filtered based on that stock's ticker. This allowed efficient access to stocks of interest and convenient access to date ranges. These stock DataFrame views are then used as the data to be fed into our regression black boxes.

Table [1] Dataset Sample (Tcs stock dataset)

Date	Open price	High price	Low price	Close price	Number of trend
8/25/2004	269	270	244	246	26145236
8/26/2004	248	249	249	244	8572064
8/27/2004	245	245	245	239	6013268
8/30/2004	241	247	247	241	4499276

Open price,High price, Low price and Number of trend consider as input independent variable and Close price is consider as target dependent variable and date is used as index variable.

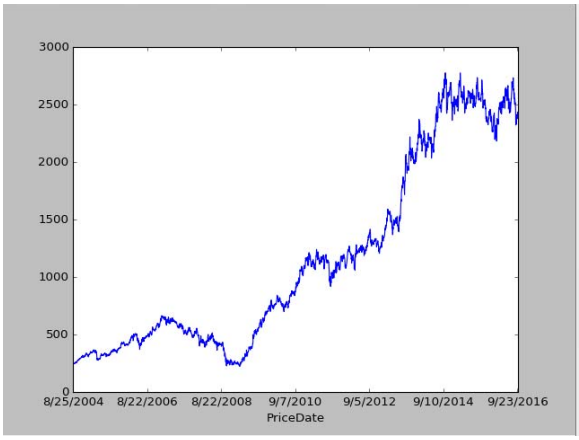

Fig.1 Time series plot of trading value

We implement linear, polynomial and RBF regression model on python 2.7 using machine learning libraries. Results of these methods are compared on bases of Confidence value.

Confidence values are percent ratio right predicated values v/s total test values.

Result analysis of various regression methods

Regression Method	Confidence value
Linear regression	0.9774
Polynomial	0.468
RBF	0.5652

Table 2. Comparison of various regression methods.

On the basis of this compression we analyzed that linear regression model provide best result compared to polynomial and RBF regression.

Table 3 Predicted Result

Date	Close price
11/17/2016	2142
11/18/2016	2165
11/19/2016	2178
11/20/2016	2135
11/21/2016	2171

V CONCLUSIONS

The aim of our research study is to help the stock brokers and investors for investing money in the stock market. The prediction plays a very important role in stock market business which is very complicated and challenging process due to dynamic nature of the stock market. As per the discussed works above our system, predicts the stock prices based on Linearregression, polynomial and RBF regression approach using 5 variables. We compare these methods on bases of confidence value and analyzed that linear regression provide betst result compare another method.

REFERENCES

[1] Enke, D., & Thawornwong, S. (2005), The use of data mining and neural networks for forecasting stock market returns, *Expert Systems with Applications*, *29*(4), 927-940.

[2] oone, L., Giorno, C., & Richardson, P. (1998), *Stock market fluctuations and consumption behaviour: some recent evidence* (No. 208). OECD Publishing.

[3] Gharehchopogh, F.S., Mohammadi, P., &Hakimi, P. (2012). Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study. International Journal of Computer Applications, 52(6), 21-26.

[4] Gharehchopogh, F. S. (2011). Approach and Developing Data Mining Method for Spatial Applications. In *Proceedings of International Conference on Intelligent Systems & Data Processing (ICISD), India* (pp. 342-345).

[5] Gharehchopogh, F.S., &Khaze, S.R. (2012), Data Mining Application for Cyber Space Users Tendency in Blog Writing: A Case Study. *International Journal of Computer Applications*, *47*(18), 40-46.

[6] Berry, M. J., &Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Wiley. com.

[7] http://au.finance.yahoo.com/q/hp?s=ASX.AX, Last available: 28/06/2013

[8] Majhi, R., Panda, G., Sahoo, G., Dash, P. K., & Das, D. P. (2007, September). Stock market forecast of S&P 500 and DJIA using bacterial foraging optimization technique. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on* (pp. 2569-2575). IEEE.

[9] Wuthrich, B., Cho, V., Leung, S., Permunetilleke, D., Sankaran, K., & Zhang, J. (1998, October). Daily stock market forecast from textual web data. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on* (Vol. 3, pp. 2720-2725). IEEE.

[10] Nikfarjam, A., Emadzadeh, E., &Muthaiyah, S. (2010). Text mining approaches for stock market forecast. In *Computer and Automation Engineering (ICCAE), 2010 the 2nd International Conference on* (Vol. 4, pp. 256-260). IEEE.

[11] Zhang, X., Fuehres, H., &Gloor, P. A. (2011). *Forecastingstock market indicators through twitter "I hope it is*

not as bad as I fear". Procedia-Social and Behavioral Sciences, 26, 55-62.

[12] Karabulut, Y. (2011). Can Facebook predict stock market activity? *Available at http://bus.miami.edu/umbfc/_common/files/papers/Karabulut.pdf* [last Available 02.07.2013].

13. Multiple Regression analysis from http://www.google.com.

14. M.Ray, Harswarup Sharma "A Reference book on Mathematical Statistics" 4TH Edition, 2006.

15. Introduction to multiple regression downloaded pdf file from http://google.com.

16. Data sources from http://yahoofinance.com