Stock Price Prediction Using Clustering and Regression: A Review

Article · January 2018

CITATION
1

READS
804

3 authors, including:

Mittal Darji
G H Patel College of Engineering and Technology (GCET)
8 PUBLICATIONS **10** CITATIONS

SEE PROFILE

International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Stock Price Prediction Using Clustering and Regression: A Review

Janki Patel, Prof. Miral Patel, Prof. Mittal Darji
Department of Information Technology G.H.Patel collage of engineering V.V.Nagar, Anand, Gujarat, India

ABSTRACT

Prediction of stock market has been an attractive topic to the stock brokers. In stock market the decision on when buying or selling stock is important in order to achieve profit. There are number of techniques that can be used to help investors in order to make a decision for financial gain. In this research work I have propose a prediction algorithm that will give the relation between the dependent factor like price and independent factors like opening price, closing price, high value of stock, low value of stock and volume of stocks bought. In this research I have explained development of stock price prediction with the use of regression analysis and clustering algorithm. Here clustering algorithm is used to partitioning the data and it also gives the high performance, and regression is used to predict the stock price of a company for a particular date.

Keywords: Prediction, Regression, Clustering

I. INTRODUCTION

1.1 BASIC INTRODUCTION

Prediction of stock market is an attractive topic to the stoke brokers. In stoke market the decision when buying or selling the stock is an important in order to achieve profit. As market fluctuating everyday it is difficult to predict the future stock price. In data mining, there are number techniques which is designed to overcome this uncertainty. (I.e. clustering, regression, SVM, neural network etc.) In this research work I am concentrating on clustering and regression. As data mining is classified in to predictive and descriptive method, clustering is descriptive method and regression is a predictive method. Clustering gives the high performance and it is used to partition the data. Regression analysis is used to predict the stock price of the particular company [2].

1.2 PURPOSE OF DATA MINING IN STOCK MARKET

Many researchers attempts to predict stock prices by applying statistical and charting approaches. But those methods lacks behind heavily due to human biased decisions on stock market based on day to day mind set of human behavior. By applying data mining in a suitable way hidden patterns can be uncovered which was not possible by traditional approaches. Also by applying business intelligence future price prediction with increased accuracy levels are possible with data mining techniques. The huge amount of data generated by stock Markets forced the researchers to apply data mining to make investment decisions [1].

There are some challenges of stock market can be effectively addressed by mining techniques like:

1. Future stock price prediction
2. To generate effective patterns of past data for further analysis.
3. To optimally utilize the capital of shareholders.
4. For the growth of country economy.
5. To bring more investors to stock market who are lacking in analysis.
6. To stabilize the market

7. To increase transparency in the market.
8. To check corruptive practices.
9. To bring more lazy and tech savvy investors into market.

Here, I had taken one of this challenge that is future stock price prediction.

1.3 BASIC INFORMATION OF STOCK MARKET

A stock market is a public market for trading of company stocks. Stock market prediction is the act of trying to determine the future price of a company stock. The value of the share depends on how many people want to buy it and how many people are selling it. If many people want to buy a stock, the price will go up. If there are more sellers than buyers, the price will go down. People usually buy/sell shares in stocks with the help of a broker. A broker also helps customers make good choices in stock [2].

The analysis of stock market or in simple terms figuring at what to buy and what to sell and at which price can be done through three methods fundamental, technical and time series [1][2] . Technical analysis uses the past traded price and volume data to predict future market price and movement. Technical analysis does not look into the actual nature of market, company, currency or commodity, it is based solely on historical price and volume information. This price data is plotted on chart and analyzed [2].

Fundamental analysis approach is finding out the true value of a stock and compares it with the current trading levels and recommends buying of stock which is traded lesser than its true value [1].

In case of Time series prediction linear flow prediction models are generated and historic patterns are traced [1].

1.4 CLUSTERING

Custering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters).

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. It is unsupervised learning technique which is used to group similar instances on basis of features [2].

For example: Suppose, you are the head of a rental store and wish to understand preferences of your costumers to scale up your business. Is it possible for you to look at details of each costumer? Definitely not. But, what you can do is to cluster all of your costumers into say 10 groups based on their purchasing habits and use a separate strategy for costumers in each of these 10 groups. And this is what we call clustering.

1.5 REGRESSION

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. Regression is widely used for prediction and forecasting. It is also gives the relationship between independent variable and dependent variable. More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables that is, average value of the dependent variable when the independent variables are fixed [6].

II. LITERATURE SURVEY

In [1] they have discussed several attempts made by researches for stock price prediction. These works show that data mining techniques can be applied for evaluation of past stock prices and acquire valuable information by estimating suitable financial indicators.

In this paper, an analysis system which helps the people to identify the more profitable companies using data mining approaches. The clustering and regression are the two techniques of data mining used here, Validation index is used for analyzing the performance of different clustering methods such as partitioning technique, hierarchical technique, model based technique and density based technique. And regression is used for prediction of stock price.

Future work : As future work stock prediction system can be created using Partitioning based or model based technique along with multiple regression technique.

In this paper they had investigate to predict the stock prices using auto regressive model. The auto regression model is used because of its simplicity and wide acceptability. They have also conducted a study on the effectiveness of auto regressive model.

In this paper, Researchers developed an automated approach which combines data mining, a field of computer science and technical analysis, a way of market analysis to help investors grow their money. Prediction of future trends in capital markets is based upon current data and past available data.

In this paper, They have proposed the system which will study the database of shares and given predictions according to it. With the help of study of neural networks the system is designed. For prediction particularly ARMA (autoregressive-moving-average) algorithm is used. Hence the system will be able to give highest probability predictions for particular shares.

Future work : As future work there is a need to improve the parameters accuracy and performance.
This paper they had done work on regression analysis as a data mining technique and and they had developed system for exploiting time series data in financial institution. The development and implementation of a stock price prediction application using machine learning algorithm and object oriented approach of software system development. The algorithm was used in training a set of market data collected for the period of one thousand, two hundred and three days.

This paper explains the framework predict the close prices of the stock precisely utilising the regression methods. As it can be seen from the charts, Support Vector Regression is the best among the three.

In this paper they have done the development and implementation of a stock price prediction is explained in this project and for this purpose regression algorithm and object oriented approach of software development is used.

The proposed system will make use of neural networks with back propagation for parallel and distributed computing and regression analysis to prevent system failure in case of sudden fluctuation in the stock price due to unprecedented events such as natural calamity, scams, etc.

This paper shows the impact of Demonetisation on Stock Market of India. Result from the Ordinary Least Square support that demonetisation or withdrawal of higher denomination currency has a significant impact on the Stock market for the Indian economy. The result reveals that average returns on most sectors have exhibited negative values. Public sector banking segment, Pharma Energy and IT has recorded a rise in returns.

III. MOTIVATION

As stock price fluctuating every day it is required intensive planning for making profit from it. Since the beginning of stock market, analyst have always struggled to predict the future stock price because of its complexity and profitability. The most reliable way to forecast the future is to try to understand the present but the amount of data available nowadays is

huge. Data analysis can be used to better understand the present scenario of the Stock market so as to understand and try to create a better future scope for investment. With Data analysis, we can add a degree of certainty to the unpredictable and volatile nature of stock prices. This certainty can go a long way to ensure that losses are minimized and profits are maximized. Though the predictions can never be fully accurate even a minute increase in accuracy of prediction can help a lot in terms of profitability.

IV. PROPOSED WORK

My object of this research work is to predict the stock price. For that I am going to use clustering which gives the high performance and Regression which is used for prediction of future stock price. Here I have taken database from BSE site of selected stock companies of date 1st September 2016 to 31st October 2017. After gathering database we have done pre-processing of data which contains daily price value of opening price, high price, low price and closing price of stock. This data also called time series data. After that I had found the mean of this four value which is used for prediction of next day's price of stock.

Table 1. details of graph

No.	Name	Colour
X1	Open	Blue
X2	High	Red
X3	Low	Yellow
X4	Close	Purple

V. IMPLEMENTATION

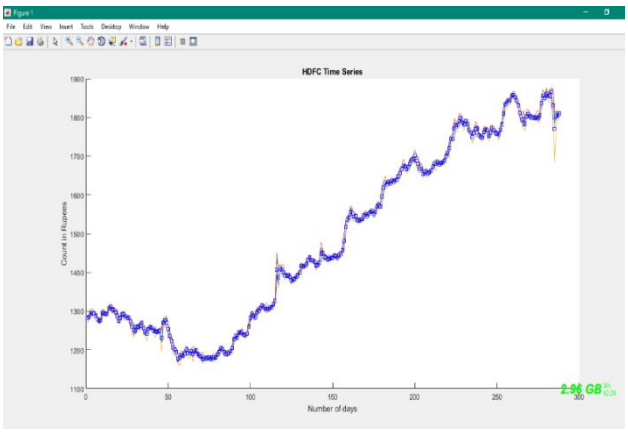

Figure 1. Time series graph of stock data for 300 days

Figure 1 shows the time series graph of HDFC stock for 300 days from date 1st September 2016 to 31st October 2017. There are 4 parameters were used which is opening price of stock, high price, low price and close price of stock of HDFC. After taking 4 parameters, after that I had found the mean of this four value which is used for prediction of next day's price of stock.

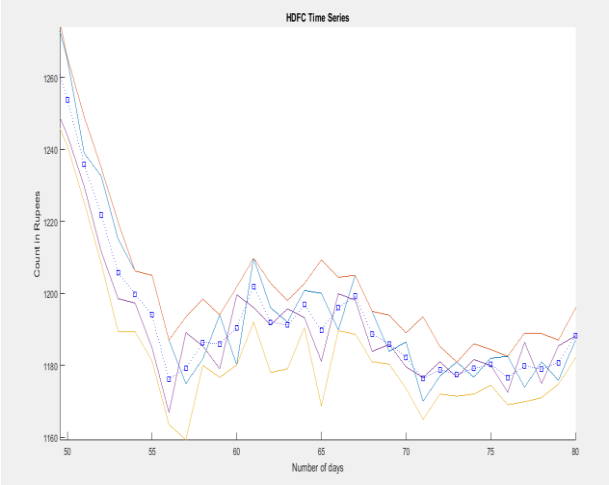

Figure 2. Detail graph of time series for mean value

Figure 4 shows the Detail graph of time series for mean value of HDFC company of per day price of opening price, closing price, high price, and low price.

VI. CONCLUSION AND FUTURE PLAN

I have referred many work related papers, it shows some of data mining techniques like clustering and regression for prediction of stock price. From this we

can conclude that the historical data shows the value depend on many factors which can help anyone to predict the value of the stock. Applying algorithm will increase more precise values. I have gathered the data of different 10 companies. And after the gathering of the data I had done pre-processing. After pre-processing I had found the mean of 4 parameters like opening price, high price, low price, and closing price of 10 companies.

Future Work Plan: In next , I will apply the partitioning technique of clustering which gives high performance which is used for partitioning of Mean data. Then I will apply regression which gives prediction of selected stock companies. Also I will add some real time features.

VII. REFERENCES

[1]. Prof S.Prasanna, Dr.D.Ezhilmaran, "An analysis on Stock Market Prediction using Data Mining Techniques", Feb 2013.

[2]. Bini B.S,Tessy Mathew, "CLUSTERING AND REGRESSION TECHNIQUES FOR STOCK PREDICTION",2015.

[3]. Dr.P.K.Sahoo, Mr.Krishna charlapally, "Stock Price Prediction Using Regression Analysis", 2015.

[4]. Pooja Taori1, Archana Nikose, "Prediction of Future Trends in Capital Markets using Mining Methodologies", February 2016.

[5]. G.S.Navale, Nishant Dudhwala, Kunal Jadhav, "Prediction of Stock Market using Data Mining and Artificial Intelligence", 2016.

[6]. Bhagyashree Nigade, Aishwarya Pawar at el.,"Stock Trend Prediction Using Regression Analysis-A Data Mining Approach", February 2017.

[7]. Bhagyashree Nigade at el.,"Comparative Study of Stock Prediction System using Regression Techniques", March - April 2017.

[8]. Mr.Pramod Mali, Hemangi Karchalkar at el.,"Open Price Prediction of Stock Market using Regression Analysis", May 2017.

[9]. Vinod Mehta at el.,"Stock Price Prediction Using Regression And Aritificial Neural Network", 2017"

[10]. Dr.P.Chellasamy, Anu.K.M, "Impact of Demonetisation on Indian Stock Market: With Special Reference to Sectoral Indices in National Stock Exchange of India", May.-June.2017