ANALYSIS OF STOCK MARKET FROM INVESTMENT PERSPECTIVE

Gorantla Venkata Praneeth Babu¹, Yerra Lakshmi Venkata Sai Sourabh², Guntupalli Manoj Kumar³
¹Department of Computer Science, and Engineering, SRMIST, Kattankulathur, Chennai, India, Gg7033@srmist.edu.in
²Department of Computer Science and Engineering, SRMIST, Kattankulathur, Chennai, India, Yy1795@srmist.edu.in
³Department of Computer Science and Engineering, SRMIST, Kattankulathur, Chennai, India, manojkun1@srmist.edu.in

ABSTRACT:

Stock market prediction regarding stock of a company for a long term. The opinion about stack market is high risk and returns. Also number of investors who having potential to invest in stock market is high; but investors count is less in number. The successful way of prediction can only makes more profit. Recent trend in stock market is predicting value for future in accordance with previous values using machine learning as key factor. This proposed method presenting support vector machine (SVM), decision tree and random forest algorithm. To perform prediction individually and to examine which one is better by calculating efficiency is carried out in this study.

Keywords: Support Vector Machine (SVM), Decision Tree, Random Forest Algorithm

I. INTRODUCTION

Trading stocks invest money in any one of the company and earn a particular amount of money from that is followed from initial stage. A difficulty in manual method of prediction is day-by-day the stock value of every company will vary. Apart from that financial experts are required to predict some stocks where it may be raise or down. In recent years, predicting stock market before going to invest is more popular. The stock market is a volatile place in which markets price values tend to vary on a daily basis. The necessity to predict this volatility has been increasing significantly during the course of years. Machine learning plays a vital role in predicting closing/opening values more accurate and also dealt with finding what will be the next step after opening (i.e. increase or decrease in stock). The main resource for performing stock prediction through machine learning is whatever the data we provided. The proposed work regards with analysing current value of stock market with values taken from few years ago as input. By providing that data as input, three differential control algorithm is progressed. From that, the operating performance and level of prediction is estimated.

II. RELATED WORK

Mahendra Reddy et.al, exploring and discussing in [1] about long term stocks of companies. Depending upon exchange of products and raw materials, the stock price of a company is calculated. Mainly machine learning based techniques focus on time domain framing. Some techniques may fail in accuracy. To resolve this, SVM and long short term memory (LSTM) are preferred. LSTM looks like a memory storage unit and it stores data for future reference. Other than sequential time series data also it handles single valid data. Thus neural network is chosen to perform above said analysis in an effective manner. In that neural network, recurrent neutral network is particularly chosen to obtain information about stocks in time basis. The contextual information’s of input is transmitted from one stop to others by connecting loop with it. The input value range from 0 up to variable rate; so that it is scaled from 0 to 1. Further X and Y train is created. Every X train can accommodate values for last 60 days in present Y data. Presence of dropout process, time and memory allocation is comparatively low. These things are performed in improving accuracy in comparison with conventional stock market prediction methods. By following the above said methodology, stock analysis upon 7 Companies for 10 days is carried out.
Nikita Agrawal et.al, and Swati Chedwal et.al in [2, 3] revise whatever the data gathered from financial news and other things available in social media to find a strategy to predict loss and raise in a minimal time duration. But strategies used in it only differ. A well-known Random forest classification algorithm is preferred to look upon large data sets and this one manages the time needed to perform prediction.

Uma Devi et.al, focus on CNX Realty, BANK NIFTY and MIDCAP 50 risk factors in those three indices. The stock market is said to be non-linear dynamic system, variable in accordance with time and complex. The investors who invest in stocks are said to be less in number because of several risk factors. Although earning is more, even most of them facing issues in it. To avoid such that severe circumstances, the proposed work is taken into account with collecting data from past three years (i.e. from January 2009 to September 2011). Rise and fall in it is clearly defined by those data analysis. Tools namely beta and correlation are employed to define stock and risk behind it. The relation between index and company individuality is demonstrated through correlation tool; risk in stocks is noted through beta tool. From analysis and above case study about stocks describing that investing in banking index is profitable one [4].

Ashwini Pathak et.al, proposing various algorithms in machine learning to focus upon data sets of stock market and it would make able to predict in a effective manner. The proposed algorithm is SVM, random forest, KNN and finally logistic regression [5]. A successful prediction will be the only way to maximize profit. The essentiality of this case study is to find a possible solution from above four methodologies. This tries to bring positive impacts on stock market among investors. SVM functions by mapping data to enlarged dimensional future space and adopts separator which classifies cases. Hyper-planes are boundary that differ data points. The major drawback in this algorithm is it may fail in large data set. Random forest classifier does classification and regression and it possess functioning capability on large data sets. Multiple decision trees are combined and from that results are obtained. More number of trees in it makes slower. KNN responds only when there is a request arrived or else it won’t function upon data sets.

As well as above two methods, this method also takes more time when size of data is large. Logistic recursion performs well in both binary digits and multiclass data classification. Among four algorithms, random forest classification only attained this much efficiency (80.7%) It is more reliable.

Kunal Pahwa et.al, proposes SVM method to resolve risk factors in stock market prediction. Nearly 14 years data is chosen as raw data for analysis. A clear view upon attributes is given in this paper [6]. Further testing and resulting on future stock price is visualized graphically.

Ishita Parmar et.al, in [7] conveyed LSTM and regression to predict stocks. The stocks bid prices are classified as low, high, volume, close and open. Other than volume, remaining is progressed at variable time factor and with their names the bid rate is prescribed; volume denotes number of shares carried out by a shareholder from another. The data sets are collected from yahoo finance and stock prediction is carried out. It is having record of nine lakhs; in that record stock data and other related values are included. After that the same progress followed up in [1] is repeated.

Neha Bhardwaj et.al, in [8] provides detailed study of following algorithms: logistics recursion, random forest, KNN, ARIMA algorithm and this work compares individuality functioning of algorithms. Raw data of few companies are chosen and efficiency is calculated. Companies namely APPL, TCS, J&J, VZ and INFY details are taken from February 2015 to February 2019.

Sahil Vazirani et.al, proposed a novel method to resolve impacts in handling large volume of data. New Linear regression models are selected to work on above said impacts [9]. Error reduction is moreover possible in this case. The output taken from first regression model is fed to secondary system for improving the data analysis.

Rohan Taneja et.al in [10] dealt with regression and classification. In regression function, closing rate of company’s stock is judged and in next stage what are all the changes will takes place in closure value is determined.

From above survey it is evident that random forest classification algorithm is more viable and it is well suited for large data classification. Time requirement is low in this method in comparison with predecessors.
III. PROPOSED TOPOLOGY

![Flowchart of proposed system](image)

The flow analysis of proposed method and process is visualized in above block diagram.

**Data set:**

The first and foremost case is to collect data whether generated or developed data is available in internet. A data set is a collection of date, close/open, rise/fall, and volume at that particular duration. In most case, data are available in a single database table; in that each value prescribing specific term. The following dataset represents data of Amazon.com Inc from end of September 2015 to February 2016. The obtained data is available in raw format and it is not able to take it directly for further process. Dealing with opening and closing price, classes are arranged and binary values are assigned. If opening price is lower than closing price then it should be denoted as 1; in opposition with above statement 0 is mentioned.

```
10 rows x 6 columns
```

**Data pre-processing:**

SVM: A supervised learning algorithm which performs classification and given data is stored in classes. It finds space between differential data sets. A boundary line that can separate datasets is called hyper-planes. Depending upon hyper-planes the dataset position allocation is performed. Kernel, gamma and regularisation parameters are used in tuning SVM parameters.
In a simpler structure the whole process present in it is presented below:

**Decision tree algorithm:** It is well suited for classification and regression problems. In Most case it is preferred for classification of datasets. Its appearance looks like tree structure; an internal node stands for features of datasets, branches for decision rules ad each leaf nodes for outcome.
Random forest algorithm: It is also similar to decision tree. A combination of multiple trees is present in to generate outcome. As same as decision tree, this should have hyper-parameters. But its speed of response is said to be high on large data sets. The below structure describes about steps involved in it.

IV. RESULT AND DISCUSSION

The analysis and resulting of different topology is progressed by taking data for actual completed time period with what are all the output obtained from those three proposed topologies. For this analysis closing price for February 2021 is taken and it is presented in figure 7.
Figure 8 denotes values of closing price at the second week of February 2021 and following this structure above pictorial representation is presented.

Fig.8.values representing actual closing price for February 2021

The enclosure attached in below figure 9 describing about future predicted value. The closing value for 2021 is predicted and visualized graphically.

Fig.9.predicted closing price for February 2021

Other than this, efficiency calculation is progressed with the below waveforms taken from software.

(a) SVM algorithm
Fig. 10. Efficiency analysis

From figure 10, the efficiency is plotted and it ranges from 95.6%, 93.8% and 94.6%. Thus SVM method is more relevant to actual value method. As per the efficiency values represented in it, identifies SVM will be a better choice for future prediction.

V. CONCLUSION

To predict before performing investing is the only way to invest in profitable way. From this paper, how to resolve risk factors and promoting stock market prediction in a right way is done. The prediction result obtained from proposed method is higher than pre-existing method. From above discussion and results obtained from proposed differential methods, SVM can have more capability to trace the actual value going to reflect in closing price is experimentally verified.

REFERENCE

2. “Stock Market Prediction Approach: An Analysis” Nikita Agrawal; (IJERT); July-2019
9. “Analysis of various machine learning algorithm and hybrid model for stock market prediction using python” Sahil Vazirani, Abhishek Sharma, Pavika Sharma; 2020 (ICSTCEE)
10. “STOCK MARKET PREDICTION USING REGRESSION” Rohan Taneja, Vaibhav, May-2018
11. “Efficacy of News Sentiment for Stock Market Prediction” Sens Kalra, Jay Shankar prasad
12. “Literature Review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support tools” Muhammad Firdaus, Swelandiah
13. Endah Pratiwi, Dionysia Kowanda, Anastocia Kowanda
15. “Predicting the Effects of News Sentiments on the Stock Market” Dev Shah, Haruna Isah, Farhana Zulkernine
17. “Indices prediction of Bangladeshi stock by using time series forecasting and performance...”
22. "A Decision Support Approach for Online Stock Forum Sentiment Analysis" Desheng Dash Wu, Lijuan Zheng, and David L. Olson
23. "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms; Srinath Ravikumar, Prasad Saraf
25. Weerithyaset, Saratose Tritilnunt