International Journal of Advance Research in Science and Engineering
Vol. No.9, Issue No. 07, July 2020
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Stocks Analysis and Price Prediction

Snehal.A.Yadav[1], Amruta.N.Shisode[2],
Snehal.S.Yadav[3], Prof. Deepak S. Khachane[4]

1(Department of Computer Engineering, New Horizon Institute of Technology and Management, Thane(W), University of Mumbai, India)

2(Department of Computer Engineering, New Horizon Institute of Technology and Management, Thane(W), University of Mumbai, India)

3(Department of Computer Engineering, New Horizon Institute of Technology and Management, Thane(W), University of Mumbai, India)

4(Department of Computer Engineering, New Horizon Institute of Technology and Management, Thane(W), University of Mumbai, India)

ABSTRACT

The objective of this paper is to construct a model to predict stock value movement using the opinion mining and clustering method to predict stock prices. We have used domain specific approach to predict the stocks.Each domain we have taken has some shares that has maximum capitalization. The proposed method is not at all like past methodologies, here sentiments of the particular subjects of the organization or sector are fused into the stock prediction model. The topics and related opinion of shareholders are automatically extracted from the writings in a message board by utilizing our proposed strategy alongside isolating clusters of comparable sorts of stocks from others using clustering algorithms. The proposed methodology will give us two output set i.e. one from sentiment analysis and another from clustering-based prediction with respect to some specialized parameters of stock exchange. By examining both the results an efficient prediction is produced. In this paper stocks with maximum capitalization within all the important sectors are taken into consideration for analysis.Also, Big data analytics are essentially used in many fields forprecise prediction and study of the large datasets. They give us the significant information fromlarge data sets which is sometimes hidden. The ApacheHadoop big-data framework allows to managebig datasets through distributed storage and processing,stocks fromthe National Stock Exchange are selected and it is split into training and test data set to predict the stocks using Machine Learning and Sentiment Analysis.

Keywords- *Big data analytics, big data, machine learning, sentiment analysis, stock market.*

I. INTRODUCTION

Stock Market prediction and analysis is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over certain period of time. The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decrease

The National Stock Exchange of India Limited (NSE) is the leading stock exchange of India, located in Mumbai. The NSE was established in 1992 as the first demutualized electronic exchange inthe country. NSE was the first exchange in the country to provide a modern, fully automated screen-based electronic trading system which offered easy trading facility to the investors spread across the length and breadth of the country. Due to involvement of many numbers of industries and companies, it contains very large sets of data from which it is difficult to extract information and analyse their trend of work manually. The application developed in this project, not only helps in prediction the future movement if the stock in the market, but also automate the data retrieval, trend analysis, predictive analysis and insights generation of a stock, just at the click of a button. Stock market analysis and prediction will reveal the market patterns and predict the time to purchase stock.

The successful prediction of a stock's future price could yield significant profit. This is done using large historic market data of 12 months in this project, to represent varying conditions and confirming that the time series patterns have statistically significant predictive power for high probability of profitable trades and high profitable returns for the competitive business investment. Herding behaviour is common among investors, all investors do not get all information at the same time and the time it takes to evaluate information before they act differs between investors. Many investors do not show rational behaviour. Greed and fear are strong feelings and may result in panic sales and stock market bubbles. Hence, to regulate the stock market to obtain maximum profit or achieve a certain objective in general without falling prey to inconsistencies, predicting stock behaviour is a pressing requirement.

II. PROPOSED DESIGN

Overall,we built a system which analyses stocks to predict daily gains based on the real time data from NSE (National Stock Exchange). They are picked up and their daily gain data are divided into training and test data set to predict the stocks with high daily gains. Based on our analysis we propose a robust Hadoop based data pipeline to perform this analyses for any type and scale of data. The big data analytics are used for efficient stock market analysis and prediction. Generally, stock market is a domain that uncertainty and inability to accurately predict the stock values may result in huge financial losses. Through our work we were able to

propose a approach to help us identify stocks with positive everyday return margins, which can be suggested to be the potential stocks for enhanced trading. Such approach will act as a Hadoop based pipeline to learn from past data and make decisions based on streaming updates which the stocks are profitable to trade in.

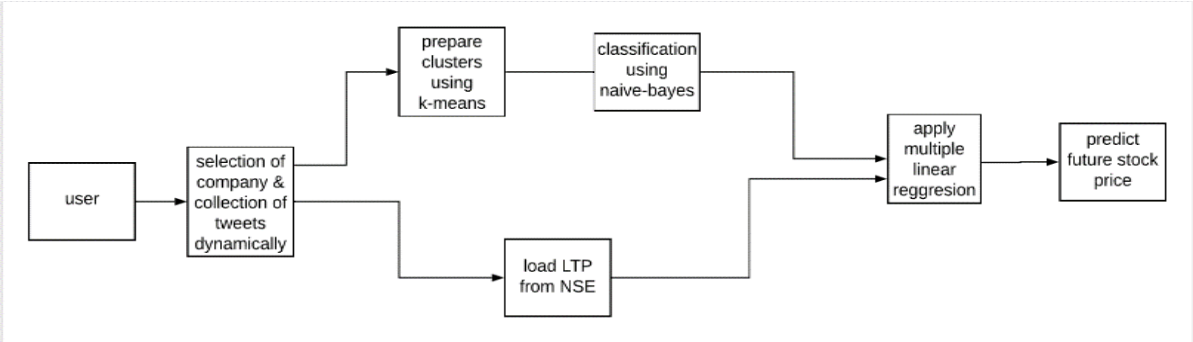

Fig: 2.1

The above block diagram shows the working of our project. User will select the company from which he would buy the shares.Accordingly, system will fetch the tweets & last traded price(LTP) of the company stocks from NSE. The data is stored in Hadoop for making the processing fast. Also, we will use Sentiment analysis on fetched tweets and rank them from 0-2 .Our System will apply algorithms on this stored information. Clustering of tweets will be done on the basis of company related tweets & company unrelated tweets using the K-means algorithm. After that Naive Bayes algorithm will segregate the positive, negative & neutral tweets. Lastly MultipleLinearRegression will be applied on segregated tweets & on last traded price of stocks. Hence the future stock price will be predicted.

III. ALGORITHMS-

3.1 Naive Bayes Algorithm:

A supervised learning probabilistic classifier in which given dataset is classified and considers each of these features to contribute independently to the probability.It is a classification technique which generates Bayesian Networks for a given dataset based on Bayes theorem. It assumes that the given dataset contains a particular feature in a class which is unrelated to any other feature. For example, an object is considered to be A because of some features. These features presence may depend on each other or on other features but all of the features presence independently contribute to the probability that this object is A. and that is the reason it is known as Naïve.

3.2 K-Means:

An unsupervised learning in which clusters are formed of given dataset K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data (i.e., data without defined categories or

groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. This algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

- where,

'||xi - vj||' is the Euclidean distance between xi and vj.

'ci' is the number of data points in ith cluster.

'c' is the number of cluster centers.

The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data.
2. Labels for the training data (each data point is assigned to a single cluster).

3.3 Multiple Linear Regression:

MLR also known simply as multiple regression is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. Simple logistic regression analysis refers to the regression application with one dichotomous outcome and one independent variable; multiple logistic regression analysis applies when there is a single dichotomous outcome and more than one independent variable. The outcome in logistic regression analysis is often coded as 0 or 1, where 1 indicates that the outcome of interest is present, and 0 indicates that the outcome of interest is absent. If we define p as the probability that the outcome is 1, the multiplelogistic regression model can be written as follows:

$$\hat{p} = \frac{\exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p)}{1 + \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p)}$$

International Journal of Advance Research in Science and Engineering
Vol. No.9, Issue No. 07, July 2020
www.ijarse.com
IJARSE
ISSN 2319 - 8354

stockname	Current Price	Prediced Pr...	Status	difference
VRLLOG	259.8	469.86563	High	210.06563
APTECHT	146.25	369.7586	High	223.5086
SBIN	327.5	381.20203	High	53.702029...
ARVIND	38.1	548.9818	High	510.8818
ASIANPAINT	1873.95	1711.0717	Low	-162.87830...
TCS	1766	398.04782	Low	-1367.95218

Fig: 3.1

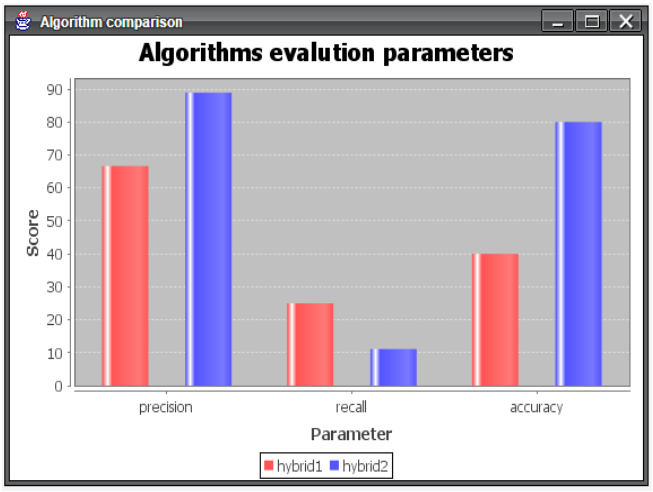

Fig: 3.2

CONCLUSION

[1] There are various ups and downs in Indian stock market. In order to invest money in stock market for purchasing the shares it is very essential for the investors to predict the stock market condition.

[2] We are giving comparative study on model comprising combination of three algorithm (K-means, Naïve Bayes, Multiple Linear Regression)

[3] K-means gives 2 clusters for classifying company related & unrelated feeds.

[4] Naive Bayes algorithm helps to give distinct and clear results of classification.

[5] MLR gives an accuracy of almost 80-84%.

ACKNOWLEGEMENT

We would like to take this opportunity to thank all who has helped us. It is a pleasure to express gratitude to our project guide, Mr. Deepak. S. Khachane for providing us with constructive and positive feedback during the preparation of this project. We would like to thank Ms. ArathiKamble, who constantly guided us and gave us valuable insights. We would like to share our acknowledgement to our head of Computer Department Dr.Sanjay.S.Sharma. At last but not the least, we are thankful to our friends for their encouragement and suggestion. We are also grateful to our parents for their constant support and best wishes along with other staff members for their support and coordination.

REFERENCES

Websites:

[1] NSE- National Stock Exhange (n.d). Retrieved from https://www.nseindia.com/

Books:

[2] White, T. (2011).Hadoop: the definitive guide. Sebastopol, CA:OReilly.

Journal Papers:

[3] EvangelosTriantaphyllou, C.-T.L., Development and evaluation of five fuzzy maldistributed decision-making methods. International Journal of Approximate Reasoning 1996. 14(4): p. 281-310.

[4] Angadi, M. C., & Kulkarni, A. P.(2015). Time Series Data Analysis for Stock Market Prediction Using Data Mining Techniques with R. International Journal of Advanced Research in Computer Science.

[5] Attigeri, G. V., MM, M. P., Pai, R. M., & Nayak, A. (2015). Stock MarketPrediction: A big data approach. In TENCON 2015-2015 IEEE Region 10 Conference (pp. 1-5)