STOCK MARKET PREDICTION USING BIG DATA

Dr C K Gomathy, Assistant Professor, SCSVMV Deemed to be University, India

Ms T. Lalitha Sagari, Ms R.V.N. Rutvika, Ms K. Sai Lakshmi, Ms S. Kavya Sree

UG Scholars- SCSVMV Deemed to be University, India.

ABSTRACT

Big data is a new and emerging buss word in today's times. Stock market is an up and ever evolving, volatile, uncertain and intriguingly potential niche, which is an important extension in finance and business growth and prediction. Stock market has to deal with a large amount of vast and distinct data to function and draw meaningful conclusions. Stock market trends depend broadly on two analyses; technical and fundamental. Technical analysis is carried out using historical trends and market values. On the other hand, fundamental analysis is done based on the sentiments, values and social media data and responses. Since large, complex and complicated and exponentially growing data is involved, we use big data analysis to help assist in the prediction and drawing accurate business decisions and profitable investments.

Keywords: Big data, prediction, Stock Market, Machine Learning.

1. INTRODUCTION

The first function of a financial exchange is to encourage the procedure for the organizations by methods for which they can exchange. The second step is to organize and manage the environment in which exchange can take place. Contributing to and benefiting from the market has never been easy, owing to the market's obvious vulnerability and highly unpredictable nature, in which shares/values can rapidly rise and fall in value. Instability is a true proportion of the dispersion of profits for a specific security or market file. Generally, the higher the unpredictability, the riskier the security. The instability of genuine prices of basic stocks is referred to as recorded instability. They have proven to be the most challenging, yet rewarding and beneficial. Big data analytics put together proves to be extremely beneficial.

Many research groups are investigating the use of social media analytics to predict stock market trends. To determine the polarity

There are several methods for each tweet/news.

1. Creating your own dictionary with semi-supervised learning

2. A dictionary-based approach tailored to the domain.

3.The semi-supervised learning approach is used to build dictionary, which takes time because of the initial level of manual labour Words are added after some threshold values are set

to either the positive or negative dictionary This approach is suitable for real-time analytics

4.Various open- source tools are used to analyze various websites.

based on Hadoop They have solely relied on manual labour. These takes time requires adhere too

II. METHODOLOGY

This section gives a description of one stock market forecast methodology. One of the novel methods suggested in the literature for event-based supervised learning stock market prediction is deciding on the major event criterion and then selecting the relevant news based on that decision. Then, based on the connected event, assign each news item the proper label, and use the tagged tweets to train a classifier. collect tweet sentiments and forecast the tone of upcoming news. And last, based on the net collected sentiment, place a long or short position.

A. Data Collection

Data collection occurs in the stock market. Two sets of data are used for this purpose: the data from the earnings calendar and the daily stock market information Various websites can be used to gather daily stock market information

B. Feature Selection

Many numerical properties can be defined from the large data set of stock prices and profits figures that has been collected. For each company and each amount of earnings The Surprise factor, earnings per stock, and the difference between the previous ESP, Market Cap, Earning Jump, as well as some operations on EPS and Market Cap, are among those features, and they are among the most crucial ones.

III. IMPLEMENTATION

Linear Regression Algorithm:

In order to forecast values, LR is utilized to determine the relationship between independent and dependent labels. LR is involved with numerous independent labels. To investigate the correlations between the independent and dependent labels, we used multiple linear regression. Assuming that labels a and b are either independent or dependent, the regression equation is as follows.

$$A = nb + e$$

In LR, a similar idea can be applied to determine the precise value for Spark. The supervised machine is necessary for the LR model. It projects what the stock price will be. The model sets values as targets based on independent or dependent changing values. The LR model makes predictions about prices based on independent values. This model can be used to anticipate future values of using datasets from various companies.

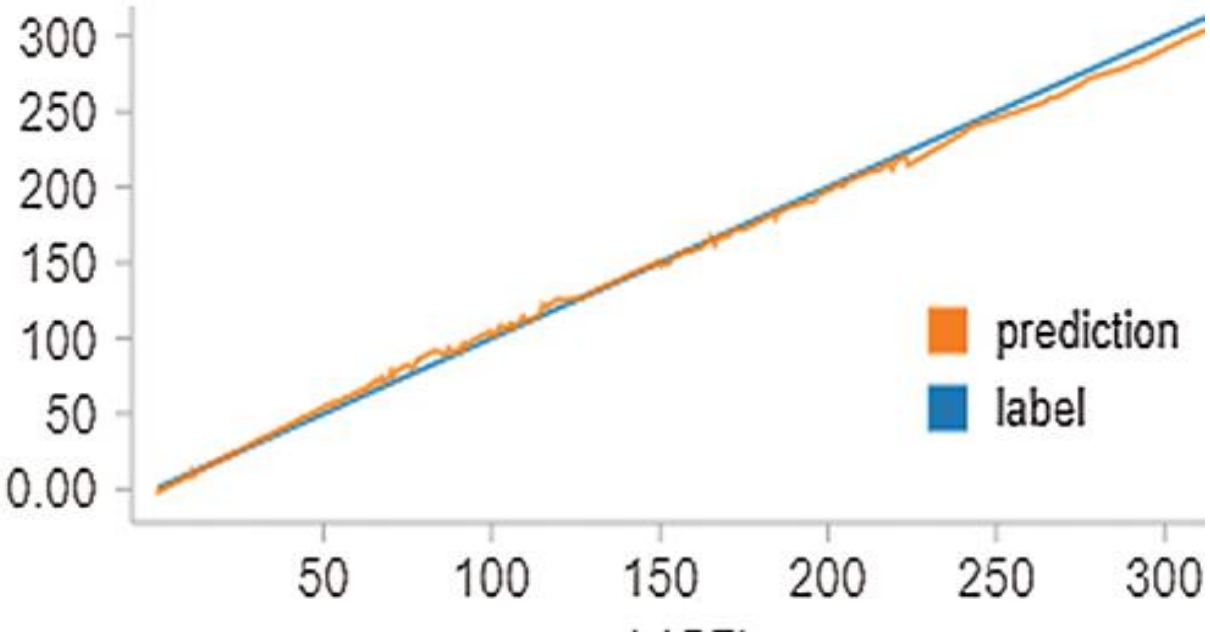

Fig 1: Prediction label in Linear Regression Algorithm

Decision Tree Algorithm:

The decision tree (DT) model has also been employed. Algorithms for supervised machine learning are necessary for this model. We divided the data into many classes and features for this model, one for each dataset.

The supervised decision tree approach is effective for both classification and regression applications. It can't perform better than the random forest. This model and Spark are used to prepare the data for analysis.

The outcomes that AAPL projected using this model are shown in Fig.

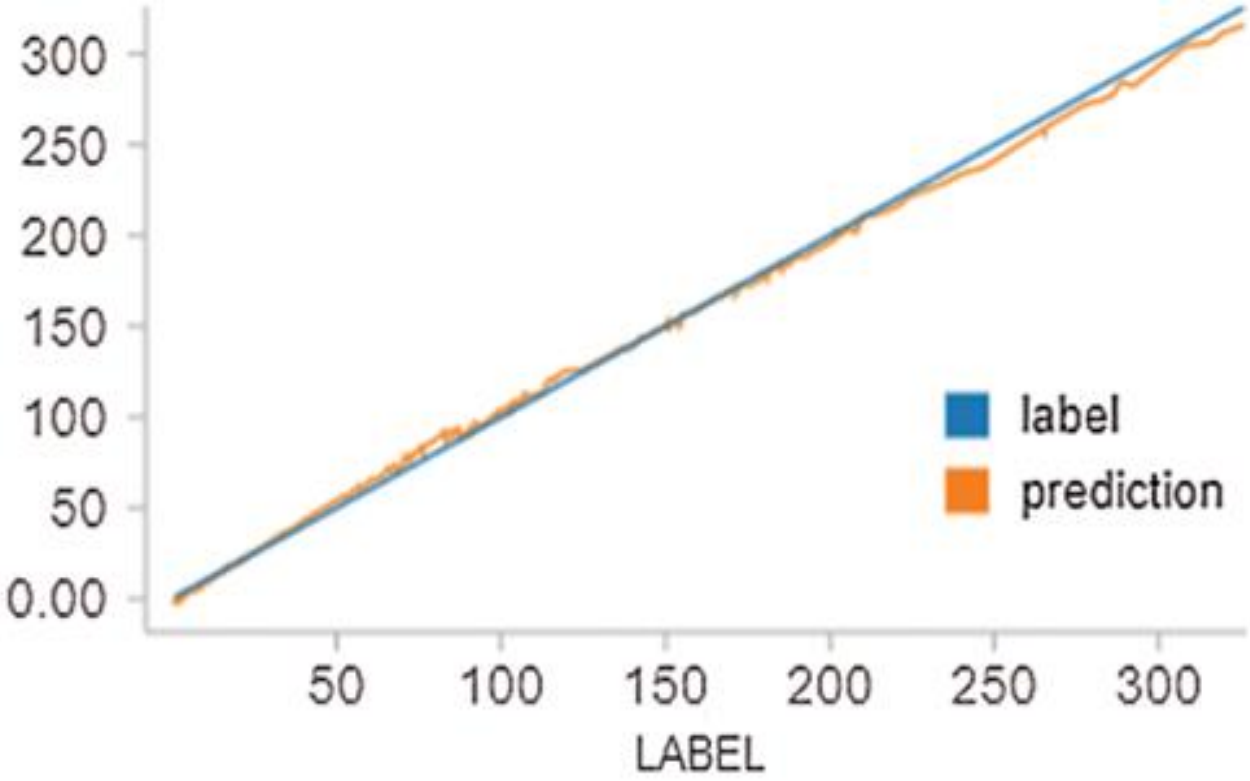

Fig 2: Label prediction in Decision Tree Algorithm

Random Forest Algorithm

Algorithms developed by supervised machines make up random forest (RF) models. The decision tree model and the RF model are comparable (DM). However, it can measure numerous trees using the same information and get the predicted value for each individual tree. The anticipated outcomes for the Apple (AAPL) stock using the RF model are displayed in Fig. Compared to the results generated by the DT model, the findings of this model are more reliable at forecasting changes in stock price.

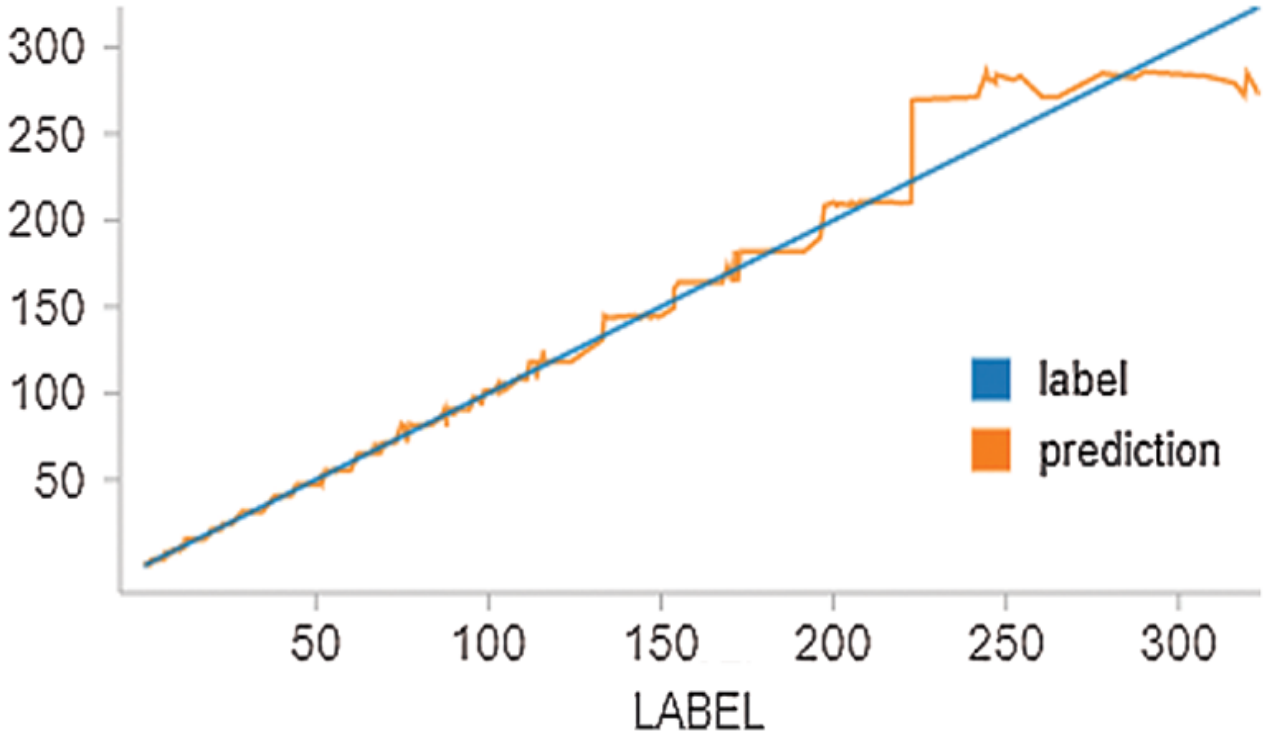

Fig 3: Label prediction in Random Forest.

IV.RESULTS:

We compare the results of all models and highlight the model that produced the most accurate results in predicting the future values of stock prices in this section. We employed a number of machine learning models to forecast stock price movements using the Spark big data platform. Using Spark ML lib, we predicted shifts in stock prices. On historical data, we used machine learning libraries for ten different companies. According to the results, generalized linear regression, random forest, and linear regression all produced more accurate results than the decision tree model. The accuracy ratios are between 77% and 80% when naive Bayes and logistic regression are applied to the texture of the data. We recommend utilizing deep learning models via LSTM for subsequent studies.

V.CONCLUSION

Big data analytics are effectively applied in this study's stock market analysis and forecasting. Generally speaking, the stock market is an area where uncertainty and the incapability to precisely estimate stock values can lead to significant financial losses. Through our research, we were able to recommend a method for locating equities with positive everyday return margins that may be suitable for increased trading. Such a strategy will function as a Hadoop-based pipeline to draw lessons from the past and decide which US equities are profitable to trade based on streaming updates. We also look for areas where our study could be strengthened in the future. In order to advance our research, we plan to automate the analyzing procedures.

VI. REFERENCES

[1] DR.C.K.Gomathy , V.Geetha , S.Madhumitha , S.Sangeetha , R.Vishnupriya Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.

[2] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[3] Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018

[4] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April ’2017

[5] Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018.

[6] C K Gomathy and V Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA,ISSN No: 0975-8887

[7] C K Gomathy and V Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2):14-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887

[8] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3,Issue IV, April 2014,P.No:1302-1306, ,ISSN:2278-1323.

[9] C. K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, doi: 10.1109/ICCTET.2014.6966260.

[10] Dr.C K Gomathy, V Geetha ,T N V Siddartha, M Sandeep , B Srinivasa Srujay Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.

[11] C.K.Gomathy.(2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X .

[12] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[13] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[14] Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April ’2017

[15] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-

2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN:978-9-3510-7216-4.

[16] Parameshwari, R. & Gomathy, C K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications. 115. 16-20. 10.5120/20163-2253.

[17] C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET'14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.

[18] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM) ,Volume 1,Issue IV, August 2011,P.No:50-53,ISSN : 2231-1009

AUTHOR'S PROFILE:

1. Ms. T Lalitha Sagari, Student, B.E Computer Science and Engineering, Sri Chandrasekharendra Saraswati Viswa MahaVidyalaya, Enathur, Kanchipuram, Tamil Nadu, India.
   Area of Interest: Data Science and Data Analytics
2. Ms.R.V.N. Rutvika, Student, B.E Computer Science and Engineering, Sri Chandrasekharendra Saraswati Viswa MahaVidyalaya, Enathur, Kanchipuram, Tamil Nadu, India.
   Area of Interest: Data Science and Data Analytics.
3. Ms. K Sai Lakshmi Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswati Viswa MahaVidyalaya, Enathur, Kanchipuram, Tamil Nadu, India.
   Area of Interest: Data Science and Data Analytics.
4. Ms.S. Kavya Sree, Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswati Viswa MahaVidyalaya, Enathur, Kanchipuram, Tamil Nadu, India.
   Area of Interest: Data Science and Data Analytics.
5. Dr.C.K.Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her area of interest is Software Engineering, Web Services, Knowledge Management and Big data analytics.