e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 5, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.54

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506085 |

Stock Market Prediction Using Classification

Jeffy Pannuel Raj PS, Nantha Kumar S, Vijayapandiyan P, Balachander K

U.G Scholar, Department of CSE, Velammal Institute of Technology, Chennai, Tamil Nadu, India

U.G Scholar, Department of CSE, Velammal Institute of Technology, Chennai, Tamil Nadu, India

U.G Scholar, Department of CSE, Velammal Institute of Technology, Chennai, Tamil Nadu, India

Assistant Professor, Department of CSE, Velammal Institute of Technology, Chennai, Tamil Nadu, India

ABSTRACT: The goal of this paper is to review totally different techniques to predict stock worth movement victimisation the sentiment analysis from social media, data processing. During this paper we are going to realize economical technique which may predict stock movement additional accurately. Social media offers a robust outlet for people's thoughts and feelings it's a fast-ever-growing supply of texts starting from everyday observations to concerned discussions. This paper contributes to the sphere of sentiment analysis that aims to extract emotions and opinions from text. A basic goal is to classify text as expressing either positive or negative feeling. Sentiment classifiers are designed for social media text like product reviews, blog posts, and even email corpus messages. With increasing complexness of text sources and topics, it's time to re-examine the quality sentiment extraction approaches, and probably to re-outline and enrich the definition of sentiment. Next, in contrast to sentiment analysis up to now, we have a tendency to examine sentiment expression and polarity classification inside and across varied social media streams by building topical datasets inside every stream. Totally different data processing ways area unit accustomed predict market additional with efficiency in conjunction with varied hybrid approaches. We have a tendency to conclude that stock prediction is incredibly advanced task and varied factors ought to be thought of for prognostication the market additional accurately and with efficiency.

I. INTRODUCTION

Stock market prediction is the act of trying to determine the future value of a stock from social media Social media offers a robust outlet for people thoughts and feelings Analysis of social media is strongly related to sentiment analysis This is used to extract emotions and opinions from text Data mining methodologies like NLP, Random forest, Neural network is used for analyzing social network content and improves the average accuracy Recent analysis reveals the existence of attention-grabbing communication patterns among completely different participants of various social network platforms. These patterns are shown to be helpful in predicting product sales and stock costs . Compared to a social network, which may be thought of as representing connections among folks within the public, a company network connects solely staff in a very huge corporation. While participants of a social network will specific opinions on any problems with interest, members of a company communication network area unit expected to chiefly say company-specific business. If human communication patterns will be discovered within the social networks to predict product sales or stock performance, one might surprise if such patterns additionally exist among members in company communication network to permit constant to be done. in contrast to social networks, in a very company communication network, e-mails have long been used as a tool for interorganizational and interorganizational data exchange. within the same means, a social network platform is ready to capture participants' behaviour and their opinions concerning varied problems and events. Thus, we tend to argue that a company communication network within the sort of Associate in Nursing e-mail scheme additionally contains perceptive data, like structure stability and hardiness, a couple of company's developments. we tend to believe our argument is in line with company communications, that suggests that "employee communications will mean the success or failure of any major amendment program" ensuing from a merger, acquisition, new venture, new method improvement approach, or alternative management problems. In alternative words, worker communication will serve a crucial "business operate that drives performance and contributes to a company's financial success". Based on these broad company communication theories, we tend to anticipate that each company has its own communication approach with identifiable patterns. we tend to believe that these communication patterns will reflect however a company manages major company activities

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506085 |

(such as mergers, acquisitions, new ventures, new method improvement approaches, going considerations, or bankruptcy) which will afterwards influence the company's performance within the exchange.

II. LITERATURE SURVEY

1. **Title:** Public sentiment analysis in Twitter data for prediction of a Company's stock price movements.

Description:

There has recently been some effort to mine social media for public sentiment analysis. Studies have suggested that public emotions shown through speaker unit could well be related with the Dow-Jones Industrial Average Industrial Average. However, will public sentiment be analyzed to predict the movements of the stock value of a specific company? If thus, is it attainable for the stock value of 1 company to be a lot of predictable than that of another company? Is there a particular quite firm whose stock value area unit a lot of predictable supported analyzing public sentiments as mirrored in Twitter data? During this article, we tend to propose a technique to mine Twitter knowledge for answers to those queries. Specifically, we propose to use an information mining algorithmic program to see if the price of a range of thirty firms listed in data system and therefore the New York securities market will truly be foreseen by the given fifteen million records of tweets (i.e., Twitter messages). We do thus by extracting ambiguous matter tweet knowledge through IP techniques to outline public sentiment, then build use of an information mining technique to get patterns between public sentiment and real stock value movements. With the projected algorithm, we tend to manage to get that it's attainable for the stock value of some firms to be foreseen with a median accuracy as high as seventy six.12%. During this paper, we tend to describe the info mining algorithmic program that we tend to use and discuss the key findings in relation to the queries display.

2. **Title:** Prediction of rainfall using back propagation neural network model.

Description:

Majority of Indian farmers depend upon downfall for agriculture. Thus, in a farming country like Asian nation, rainfall prediction becomes important. This paper presents comparative study of neural network architectures specifically back Propagation Neural Network (BPNN), Generalized Regression Neural Network (GRNN) and Radial Basis operate Neural Network (RBNN) to predict downfall in Thanjavur district of southern province Madras, India. The various models area unit trained mistreatment the coaching information set and are tested for accuracy on accessible take a look at information. MATLAB has been used for model development. When coaching all networks and testing them We found that RBNN provides best result for prediction.

III. EXISTING SYSTEM

In existing system, we tend to propose that a company's performance, in terms of its stock worth movement, is foreseen by internal communication patterns. to get early warning signals, we tend to believe that it's vital for patterns in company communication networks to be detected earlier for the pre- diction of serious stock worth movement to avoid attainable adversities that an organization could face within the securities market in order that stakeholders' interests is protected the maximum amount as attainable. Despite the potential importance of such data regarding corporate communication, very little work has been tired this vital direction. We attempt to bridge these research gaps by employing a data-mining method to examine the linkage between a firm's communication data and its share price. As Enron Corporation's e-mail messages constitute the only corpus available to the public, we make use of Enron's e-mail corpus as the training and testing data for our proposed algorithm.

IV. SYSTEM DESIGN

PROPOSED SYSTEM

In this Proposed System, opinions can now be found almost everywhere - blogs, social networking sites like Face book and Email, news portals, ecommerce sites, etc. While these opinions are meant to be helpful, the vast availability of such opinions becomes overwhelming to users when there is just too much to digest. Over the last few years, this special task of summarizing opinions has stirred tremendous interest amongst the Natural Language

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

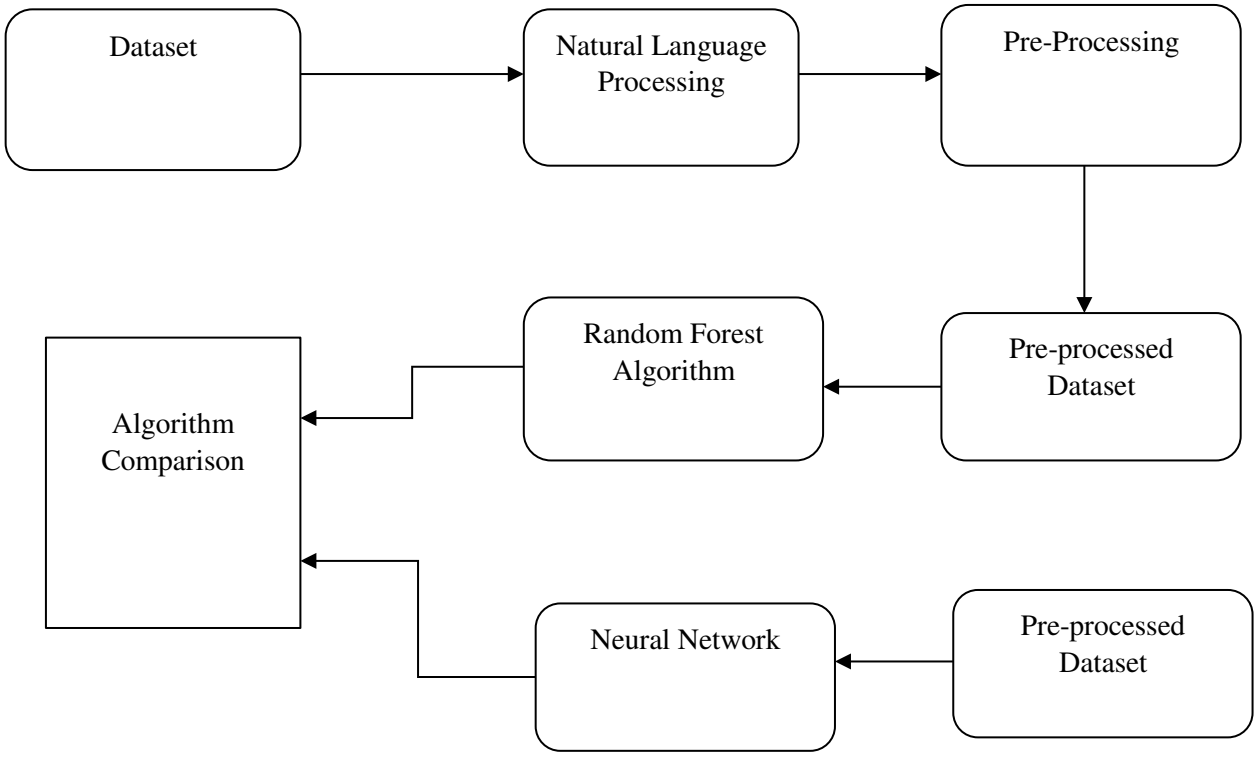
| DOI:10.15680/IJMRSET.2022.0506085 |

Processing (NLP) and Random Forest communities. „Opinions" mainly include opinionated text data such as blog/review articles, and associated numerical data like aspect rating is also included. While different groups have different notions of what an opinion summary should be, we consider any study that attempts to generate a concise and digestible summary of a large number of opinions as the study of Opinion Summarization. For example, sentiment prediction on reviews of a product can give a very general notion of what the users feel about the product. If the user needs more specifics, then the topic-based summaries or textual summaries may be more useful. Regardless of the summary formats, the goal of opinion summarization is to help users digest the vast availability of opinions in an easy manner. The approaches utilized to address this summarization task vary greatly and touch different areas of research including text clustering, sentiment prediction, Random Forest, NLP analysis, and so on. Some of these approaches rely on simple heuristics, while others use robust statistical models.

Advantages

- We will classify the approaches in various ways and describe the techniques used in an intuitive manner.
- We will also provide various aspects of evaluation in opinion summarization, which was not covered by other previous surveys.
- Finally, we will provide insights into the weaknesses of the approaches and describe the challenges that remain to be solved in this area.

ARCHITECTURE DIAGRAM



MODULES

DATA COLLECTION

Our data consists of mainly daily email about the stock market. Initially we are going to try to analyze the NYSE then upon perfection we will be using our model with the most accuracy to data from Bangladesh and implement a system to help investors of Bangladesh. We started off with finding a tagged data; unfortunately there were no free sources to find such. What we did then was built our own web script to scrap chat from the email API. We run this script every

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506085 |

day for 8 hours and we collect a huge amount of chat daily. Then we store it in JavaScript Object Notation (JSON) to our file storage and parse it.

PREPROCESSING

These noisy words will interfere with our learning algorithm. Also the biasness seems to increase as a result of including these words. If we include these words then our learning algorithm seems to look for quantifiers like "a", "an", "this" etc. Also auxiliary verbs seem to be of no interest to us. So we need only words that give a sense of "good", "bad" or "neutrality". The following list of things we had to do in order to pre-process our data:

1. Delete all hyper-text links from the tweet data. For example: "http://" or "https://t.co/3k7Bai5crQ" is removed from the tweet feed.
2. Change all word blocks from the email corpus data to lower case. This increases 14 uniformity and changes helps us to remove repetitions if present.
3. Removing white spaces from the tweet data. We keep the emoticons because they provide helpful insights about the tweet.
4. We remove punctuations marks like commas, full stop etc.
5. Clear out any tag to any person in the email data. Tags starting with "@" are removed.

We keep hashtags because sometimes tags like "#Stock #Crash" helps us better understand the mood.

6. Do take corpus chat in our data email. So remove tweets with "RT" from the data set.

DATA SCORING

Our approach for scoring a tweet was simple and effective. The first problems with tweets in the CSV files were that they contain a lot of noisy words which have little 15 to no significance to the actual context. We must remove them to get words of interest to us. Considering this objective, we first collected a list of positive, negative and neutral words in the dictionary. Then what we did was scored based on the unique positive, negative and neutral words on the tweets and our list of words. Suppose, the tweet consists of n words. Now considering the score for positive, negative and neutral score be Scorepos, Scoreneg and Scoreneu respectively and notate the set of all positive, negative and neutral words as listpos, listneg and listneu and frequency of positive, negative and neutral words as frequencypos, frequencies and frequencyneu respectively we come up with the following formula for scoring the data.

V. RESULT WITH ANALYSIS

The experimental setup consists of two components. The first one is collecting data and scoring it. We collect data from Twitter feed and score it. The second component being the learning models. Some of the models are simple and some cutting edge. Based on the learning models we have worked on previously we have come up with interesting results. We used rapid Table and meta-charts to make visual representations. We have used three learning models on data that we have collected from a dedicated machine for 3 months. We have divided seventy percent of it for training set and the rest 30% for test set.

CONCLUSION

The findings and theoretical implications from this paper are 2-fold. On one hand, we tend to capture the communications among nodes in Enron's major company communication network and identified employees' communication patterns. This paper demonstrates that a company e-mail scheme contains significant data concerning employees' communication patterns. notwithstanding we tend to solely concentrate on the communication frequency, an organization (Enron in our case) has identifiable patterns of e-mail exchange. Such identifiable patterns will reveal necessary data concerning major company activities and structure stability which will later influence the focal company's performance within the securities market.Therefore, join forces communication patterns will function an honest proxy to predict a company's stock performance. Our experimental results incontestable the existence of dependence between e-mail communication network and stock worth for Enron the contributions of the planned algorithmic program may be characterized as below. First, we tend to use "adjusted residual" to find the connection between e-mail communication frequency and stock worth. The "adjusted residual" live is probabilistic, thus it will work effectively even once the periods of the stock knowledge covered are unequal with missing or inaccurate values.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54|

| Volume 5, Issue 6, June 2022 |

| DOI:10.15680/IJMRSET.2022.0506085 |

REFERENCES

1. A.Timmermann, ‘‘Elusive return predictability,’’ Int. J. Forecasting, vol. 24, no. 1, pp. 1– 18, Jan./Mar. 2008.
2. R. D. McLean and J. Pontiff, ‘‘Does academic research destroy stock return predictability?’’ J.
3. Finance, vol. 71, no. 1, pp. 5–32, Feb. 2016.
4. M.-W. Hsu, S. Lessmann, M.-C. Sung, T. Ma, and J. E. V. Johnson, ‘‘Bridging the divide in financial market forecasting: Machine learners vs. Financial economists,’’ Expert Syst. Appl., vol. 61, pp. 215–234, Nov. 2016.
5. J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, The Econometrics of Financial Markets.
6. Princeton, NJ, USA: Princeton Univ. Press, 1997, ch. 1, sec. 1.5, pp. 20–25.
7. J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, ‘‘The predictability of asset returns,’’ in The Econometrics of Financial Markets. Princeton, NJ, USA: Princeton Univ. Press, 1997, ch. 2, pp. 27–82.
8. W. Lo and A. C. MacKinlay, ‘‘Stock market prices do not follow random walks: Evidence from a simple specification test,’’ Rev. Financial Stud., vol. 1, no. 1, pp. 41–66, Jan. 1988.
9. H. Jang and J. Lee, ‘‘An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information,’’ IEEE Access, vol. 6, pp. 5427– 5437, 2018.

BIOGRAPHY

Mr. K. Balachander, Assistant Professor in the department of Computer Science and Engineering from Velammal Institute of Technology, Panchetti.

Nanthakumar S is a B.E. final year student in the department of Computer Science and Engineering from Velammal Institute of Technology, Panchetti. His current research focuses on stock price pridiction.

Jeffypannuel raj PS is a B.E. final year student in the department of Computer Science and Engineering from Velammal Institute of Technology, Panchetti. His current research focuses on stock price pridiction.

Vijayapandiyan P is a B.E. final year student in the department of Computer Science and Engineering from Velammal Institute of Technology, Panchetti. His current research focuses on stock price pridiction.

Impact Factor
7.54

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com