CLASSIFICATION OF SOCIAL MEDIA PRODUCT REVIEWS BY USING VARIOUS MACHINE LEARNING CLASSIFIER

RATHOD DIPEEKA KANHIRAM[1], ASST.PROF. V. KARWANDE[2]

ME Student, Department of Computer Science & Engineering, EESGOI , India[1]

HOD, Assistant Professor, Department of Computer Science and Engineering, EESGOI,India. [2]

Abstract: The Web site is now a day full of information about people's opinions through apps such as social media, micro blogging sites, review websites, personal blogs, etc. Sentimental analysis is an area where people's opinions may be evaluated and categorized as positive, negative or neutral in text mining. In this paper, the tweets or review feelings published on the Twitter are recognized by searching the specific term in tweets and then the polarity of the tweets is assessed as positive and negative. The tweet emotions are tweeted on a Twitter depending on the selection of the features of each score. Naive Bayes Classifier (NBC) is used for training and testing word features, as well as assessing the feeling polarity of each tweet. to choose the best features. Parameters such as accuracy, precision and time are taken into account in assessment of performance compared to three machine classifications, namely the Random Forest, Naive Bays and Support Vector Machines (SVM).

Keywords: Support Vector Machine (SVM); Naive Bayes Classifier (NBC); Twitter Analytics;Sentiment analytics.

I INTRODUCTION

The rise of social network platforms has allowed individuals to produce and consume a lot of content on the Internet. In the past, information was obtained via portal sites. A vast number of websites offer a comprehensive array of subjects from politics to entertainment. These conventional web sources of information are helpful but less efficient, as they often contain duplicated information. As social networking sites online arrive, however, individuals seek to obtain information from these platforms because of their rapid and effective functionality. These platforms are accessible for consumers to select the source of information they want. And many social network sites like Twitter, Google+ and Facebook also give users with information. Twitter is the world's best-known microblogging network. It is also the fastest expanding social network platform and dominates the field of microblogging. Over 500 million registered users post 340 million twitter posts every day, sharing their views and everyday activities. Twitter messages are substantially shorter than conventional microblogging sites. Only 140 characters or fewer may be posted in one Twitter message. This function helps Twitter to obtain the primary message out of the large volume of internet material. Twitter users may follow the people and information source they like, depending on the needs of the users. Twitter has therefore become a formidable medium with all the above-

mentioned advantages, from international news to home-buying items. [1] [2]

In the previous several years, the Twitter information streams have seen the popularity of this social network expand unbelievably. The users have a huge quantity of knowledge on several issues. All the information is not nonetheless valuable to users and each user has its own interests and preferences. Users need customized services urgently. Currently, the consumers benefit from more and more tailored services. This tailored service is needed to make your life more productive. Users share a vast volume of information on the Twitter platform every day. These statistics relate to user behavior and hence Twitter and this information gathering are the target of numerous research projects. One of Twitter's research areas is user modelling. To offer a tailored service, researchers have begun to study the Twitter rating and recommendations of web sites. A wide range of studies focused on estimating user interests based on tweets provided by individuals. [5]

Regardless of the content of the tweets and possible usage of Twitter, researchers have also shown that tweets often carry relevant user emotional information. This has made emotional analysis on Twitter a major study problem in the domain of micro blogging. Most emotional research focuses on the categorization of feelings on Twitter. Several characteristics and approaches for training sentiment classification on Twitter platform have been investigated

with variable outcomes in recent years. There are also some additional Twitter emotional analysis study investigations. One of the experiments in this field is to gather feedback on products by extracting the emotion of clients on the Twitter network. In addition, prior research have focused on analyzing public opinions through the removal of emotions from Twitter postings. Fake news is an essential social media concern in our everyday lives. The greater the number of criminal actions in the world, the more fake news, it causes human life defects to avoid and prevent criminal actions using these approaches. Our approach takes crowd signals into consideration for the detection of false news and is prompted by Twitter's latest facilities that allow users to report false news. Since emotion plays a critical part on Twitter, I may anticipate a more effective technique to build a user profile if I consider user feelings. It isn't a completely novel notion to combine emotion with user modelling. Some earlier research already focused on this combination and proven to increase the quality of the user profile by combining emotion and user modelling. However, most of these studies feature an interactive portion in which users may collect the emotional reaction and nobody combines emotion and user modelling on the Twitter platform. The major purpose of the study is to investigate the emotional aspects in Twitter and to incorporate emotional characteristics to user modelling tactics.

Social media has become a leading networking platform for individuals in the form of reviews, tweets, postings, debates etc. The broad usage of social media platforms has also resulted in a wide range of opinion and emotive messages. The exponential rise in digital unstructured subjective and emotional textual information on resources publicly available on Twitter means that it is essential to create strategies to get important information. Tweet emotional analysis is the application of automated approaches to predict the feelings content and emotional status of a tweet. Even with massive computer capabilities, recognising the feelings and emotions included in the processed text remains one of Twitter's difficult problems. This drives you to provide emotional predictions via tweets. [6]

II LITERATURE SURVEY

In this research, online reviews have a huge influence on company and company today. The decision to buy things online primarily rests on customer reviews. Opportunistic people or organisations are thus trying to alter product reviews for their own objectives. This research presents semi-supervised and monitored text mining algorithms for detecting false internet reviews and evaluates the efficiency of both strategies on hotel reviews data sets. In this research,

Ottet algold's standard dataset is employed in our assessments. The dataset includes 1600 text-based evaluations of 20 hotels in the Chicago region, USA. We have 800 bogus reviews and 800 real reviews here. A tag of '0' refers to faulty reviews in the evaluations, whereas '1' refers to authentic reviews. 400 are written with negative emotional polarity in the dataset while 400 include positive emotional polarity. Similarly, 400 include favorable evaluations, while 400 rest reviews offer negative emotional polarity. These reviews have been collected from different sources. Discreet reviews were made utilizing Amazon Mechanical Turk (AMT) and other websites like Yelp, TripAdvisor, Expedia and Hotels.com have got the results [1].

In this research, Social media has now been expanding for a few days. Millions of users evaluate and evaluate tourist locations on tourist sites every day. Sentiment analysis may be carried out with these reviews that are beneficial to identify the popularity of tourism destinations. Based on the results of feeling analyses, tourists may simply determine which to visit. The examination of feelings in this research was done utilizing the machine learning methodology. The Dataset was obtained from many websites for tourism assessment. Here we conducted a comparative analysis of extraction techniques such as Count Vectorization and TFIDF Vectorization. In addition to Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest classification techniques (RF). Algorithms performance has been compared using several factors including accuracy, reminder, accuracy and f1-score. From the experiment, we showed that the extraction technique of TFIDFVjectorization enhances the accuracy of the classification method compared to Count Vectorization for a specific data set. TFIDF Vectorization+RF has delivered the maximum accuracy 86% for a study dataset utilised in the sentiment classification for tourism site reviews [2].

In this research, The developments in requirements engineering (RE) procedures and procedures developed in recent decades have shown a significant increase in the efficiency of various Machine Learning (ML) approaches to address multiple multidimensional RE problems. One of the challenges is the effective identification and classification of stack overflow (SO) software requirements for building quality systems.)e the adequacy of ML-based techniques to tackle this issue showed significant results that were far more effective than those produced by the usual NLP techniques. However, it is somewhat rare to have a thorough, systematic and deep understanding of these ML-based strategies.

Objective. To identify, detect and categorise the types of ML algorithms utilised mostly in SO for software needs. Methods.)is study reports A systematic review of literature (SLR) which gathers empirical evidence published by May 2020. The results of the SLR study demonstrated that (1) the Latent Dirichlet Assignment (LDA) theme modelling is one of the most generally used ML algorithms for chosen studies, and (2) the precision and recalls are one of the most generally used ways to measure the performance of these ML techniques. Conclusion. Our SLR study shows that ML algorithms are able to detect the software needs on SO, but that they still face several open problems/issues which will ultimately restrict their practical applicability and performance. Our SLR project requires a tight cooperation between RE and ML communities/researchers to address outstanding difficulties facing the development of machine-based quality-based learning systems in the actual world [3].

In this research, Social media has provided everyone with the opportunity to express opinions and speak to the masses, but it also serves to make people unpleasant and unpleasant. However, it is difficult or not to determine a comment or post, since most social media platforms still hunt for more effective ways to find an effective and moderate answer. Automating this will assist to identify abusive comments and to save websites, promote user safety and improve online dialogue. In this article, the toxic comment dataset of Kaggle is used to train and categorise the comments in the following categories: toxic, severe toxic, obscene, menacing, insulting and hateful identity. The dataset is trained with several profound learning strategies and analyses the deep learning model in the categorization of comment. Deep learning techniques are employed, such as the long-term memory cell with and without word GloVe embedding, and the Convolution Neural Network (CNN) with or without GloVe [4].

In this research, Due to its considerable processing capability in applications like speech, picture or text processing, deep learning (DL) drew growing attention. For the exponential expansion and broad availability of digital social media (SM), it is hard or uncomfortable to analyse this data using traditional techniques and technology. DL is a suitable answer to this situation. In this study, we address the practised DL structures with great interest through a taxonomic overview after the substantial efforts done towards SM analysis (SMA). This work however focuses on discussing the SMA-oriented difficulties with the DL-based remedies rather than the technical explanation. To this purpose, we also emphasise the limitations of DL research

(such as scalability, heterogeneity and multimodality) and future developments [5].

In this research, With social media such as Twitter and distribution software platforms such as app stores, people have different means of expressing their opinions on software goods. Popular software companies receive thousand times a day consumer comments. Research has proven this feedback to include vital information for the development of software teams, such as bug reports or feature requests and support requests. Since the user feedback analysis is difficult and complex to maintain, several academics and technology providers advise that automated analysis based on classic supervised machine learning algorithms be used. In this work, in categorizing user input in English and Italian into issue reports, queries and irrelevant outcomes of traditional machine learning and in-depth learning. Our results indicate that utilizing classic machine learning, while we have gathered thousands of labels, we can still get equivalent results to deep learning [6].

In this research, Because there are many social networking programmes that are significantly less time-consuming. Really users are so thrilled in interacting with this programme. There's plenty of blooming social networking apps like Facebook, Twitter, YouTube, etc. So users may not only access the provided information in this application but can also share what they feel when they see the material. There are many channels on YouTube which are growing every day and the channel management upload material according to their channel, thus they must evaluate the comments or feedback of the client. If these comments and feedback are assessed, then the channel manager will decide whether or not the consumer likes the material. If adjustments in content are required by reviewing feedback, they may readily modify. To analyse customer feedback, many categorization algorithms such as the Decision Tree, K Nearest Neighbors and Support Vector Machine have been developed. For example. The most accurate method is then used to develop a model that serves as the sentimental analysis model for other channel managers [7].

In this research, Identifying and classifying extremist tweets is a hot problem. Extremist gangs used social media platforms such as Facebook and Twitter to spread their ideas and recruit people. This paper proposes a paradigm for the study of information connected to terrorism that focuses on the classification of tweets into extremist and non-extremist classifications. On the basis of social media postings created by users on Twitter, we construct a tweet classification

system employing deep learning sentimental analysis techniques to categorise the tweets as extremist or nonextremist. The experimental results are encouraging and give future researchers with a doorway [8].

In this research, Prior to the creation of ICT, social connections developed inside local cultural limits, such as geospatial places. The current breakthroughs in communication technology have significantly exceeded the limits of traditional communications in terms of time and space. These social technologies have caused a revolution in information generated by users, online human networks, and rich data relating to human behaviour. The misuse of social technology such as social media (SM) platforms, however, has created a new type of online anger and violence. This research highlights a novel way of exhibiting aggressive conduct on SM websites. The motives for constructing light aggressive behaviour prediction models in SM are also addressed. We explore in-depth cyberbullying prediction models and discuss the key problems associated with building cyberbullying prediction models in SM. This study gives insights into the whole cyberbullying detection process and, most all, looks into the technique. Although data collecting and function engineering processes have been developed, the main focus is on functional selection algorithms, and subsequently on different machine learning techniques for cyberbullying behavior. Finally, the concerns and problems that are new research guidelines for researchers to examine have been underlined [10].

III. SYSTEMS ARCHITECTURE

The system model architecture the feelings of tweets or reviews of products published on twitter are detected by the search in tweets for the specific term and then assess the tweet polarity as positive and negative. The feelings of tweets posted on a twitter evaluated on the basis of each score word selection. Naive Bayes Classifier (NBC) is used to train and test the features of a word and also evaluate the polarity of the feeling of the tweets to choose the best features. Parameters such as accuracy, precision and time are taken into account in assessment of performance compared to three machine classifications, namely the Random Forest, Naive Bays and Support Vector Machines (SVM).
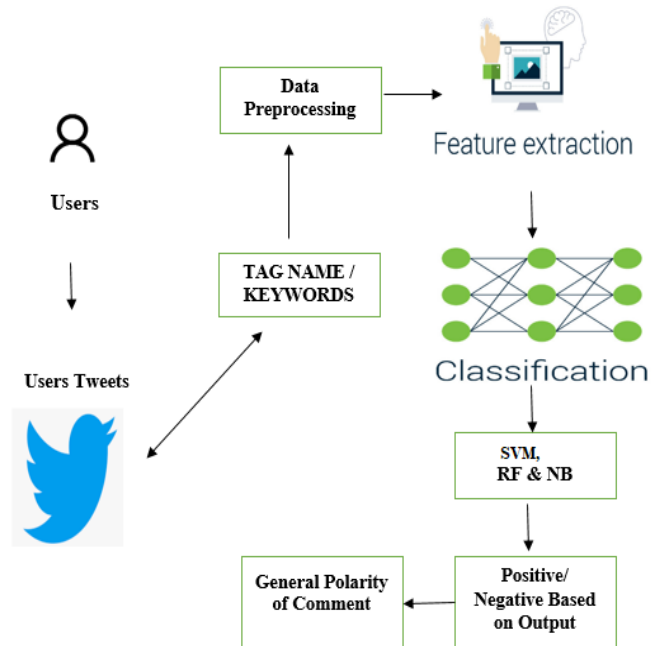

Figure No 3.1: System Architecture

IV EXPERIMENTAL RESULTS

The in creating the feature vector, the context-based method is superior than the list-based method. The support vector machine (SVM) approach is designed to find a hyperplane that best divides the data into two groups. However, we have three classes for sentiment analysis (positive, neutral, negative). How can we deal more than two kinds of datasets?

A multi-class classification issue to a binary classification problem has two approaches: one-vs-all and one-vs-one. One SVM classifier is constructed per class in a single-vs-all technique. This classification assumes that one class as the positive class and the other as the negative class. You create an SVM Classifier per pair of classes in the one-to-one technique. To simplify things, we picked a one-to-one strategy, where our challenge was broken down into three binary classifications. When you design your own analyzer, the optimum method does not fit all text data types. According to a study carried out by Stanford University, Naive Bayes provides support vehicles with brief sentiment classification tasks, whereas an SVM is a superior performer in evaluating feelings for lengthier assessments.
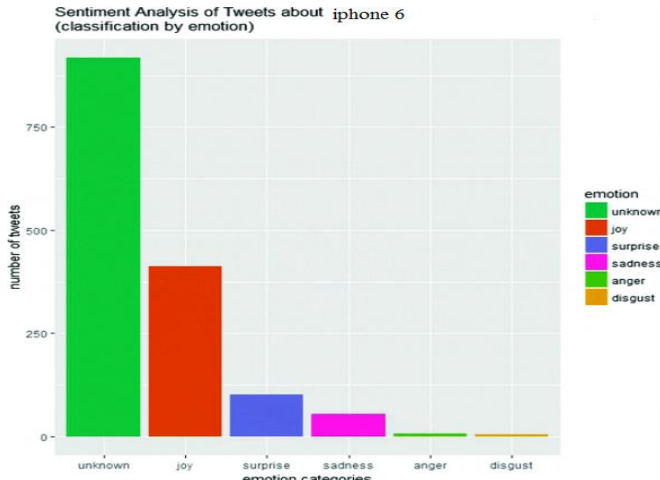
Figure No 4.1: Sentimental Analysis of Tweets about Iphone6.

V CONCLUSION

Sentiment analysis is used to uncover behind his statement the author's feeling. This proposed effort will extract tweets using a particular string search, which will be evaluated with RF, SVM, and NB classification for positive, neutral and negative classification. In the analysis, the RF, SVM and their correctness will be taken into consideration, taking into account three characteristics and thus, increasing the number of tweets. The RF, SVM and NB accuracy is measured by increasing the number of tweets

REFERENCES

1.Rakibul Hassan and Md. Rabiul Islam,"Detection of fake online reviews using semi-supervised and supervised learning",IEEE International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.

2.Apeksha Arun Wadhe and Shraddha S. Suratkar,"Tourist Place Reviews Sentiment Classification
Using Machine Learning Techniques ",IEEE International Conference on Industry 4.0 Technology (I4Tech) Vishwakarma Institute of Technology, Pune, India. Feb 13-15, 2020.

3.Arshad Ahmad,Chong Feng,Muzammil Khan,Asif Khan,Ayaz Ullah,Shah Nazir and Adnan Tahir,"A Systematic Literature Review on Using Machine Learning Algorithms for Software Requirements Identification on Stack Overflow",Hindawi Security and Communication Networks Volume,2020.

4.Mukul Anand and Dr.R.Eswari,"Classification of Abusive Comments in Social Media using Deep Learning",IEEE Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC),2019.

5.Malik Khizar Hayat, Ali Daud, Abdulrahman A. Alshdadi,Ameen Banjar,Rabeeh Ayaz Abbasi,Yukun Bao And Hussain Dawood,"Towards Deep Learning Prospects: Insights
for Social Media Analytics",IEEE Access, March 6,2019.

6.Christoph Stanik, Marlo Haering and Walid Maalej,"Classifying Multilingual User Feedback using Traditional Machine Learning and Deep Learning",arXiv,12 September 2019.

7.Ashutosh Bansal, Chunni Lal Gupta , A Muralidhar,"A Sentimental Analysis for YouTube Data using
Supervised Learning Approach",International Journal of Engineering and Advanced Technology (IJEAT), ISSN:2249-8958, Volume-8 Issue-5, June 2019.

8.Shakeel Ahmad, Muhammad Zubair Asghar,Fahad M. Alotaibi and Irfanullah Awan,"Detection and classifcation of social media-based extremist afliations using sentiment analysis techniques",Human Cent. Comput. Inf. Sci.,2019.

9.Rajalaxmi Hegde,Sandeep Kumar Hegde,Sanjana,Sapna Kotian and Shreya C Shetty,"Personality classification using Data mining approach",IJRAR March 2019, Volume 6, Issue 1,2019.

10.Mohammed Ali Al-Garadi, Mohammad Rashid Hussain, Nawsher Khan, Ghulam Murtaza, Henry Friday Nweke, Ihsan Ali, Ghulam Mujtaba, Haruna Chiroma, Hasan Ali Khattak And Abdullah Gani,"Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms:Review of Literature and Open Challenges",IEEE Access,May 22,2019.

11.Rohit Raturi,"Machine Learning Implementation for Identifying Fake Accounts in Social Network",International Journal of Pure and Applied Mathematics
Volume 118 No. 20,4785-4797,2018.

12.Mitulgiri H. Gauswami and Kiran R. Trivedi,"Implementation of Machine Learning for Gender Detection using CNN on Raspberry Pi Platform",IEEE Proceedings of the Second International Conference on Inventive Systems and Control(ICISC),2018.

13.Chad Crowe,"Predicting User Interaction on Social Media using Machine Learning",University of Nebraska at Omaha,11-2018.

14.Anu Sharma,Dr. M.K Sharma and Dr. R.K Dwivedi,"Literature Review and Challenges of Data Mining Techniques for Social Network Analysis",Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 5,pp. 1337-1354,2017.

15.Rita G. Guimaraes,Renata L. Rosa, Denise De Gaetano, Demostenes Z. Rodriguez and Grac¸a Bressan,"Age Groups Classification in Social Network Using Deep Learning",IEEE ACCESS, ACCESS-2017.