Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4

2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 978-1-6654-1960-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICV50876.2021.9388525

Analysis of Various Sentiment Analysis Techniques of NLP

Dr. Satyen M. Parikh
Prof. & Head AMPICS, DCS
Ganpat University
Ahmedabad, India

Mitali K. Shah
PhD Scholar DCS
Ganpat University
Kherva, India
shahmitali07@gmail.com

***Abstract*—Sentiment Analysis and opinion mining complement each other. This approach is concerned with classifying the opinions of public regrading a product using NLP. This approach considers sentiments and viewpoint of individuals about the occurrence of an episode. Assessment mining is generally helpful in different domains such as business brand surveys, web-based media study, and movie analyses, and so forth. The sentiment analysis is an important method in formation of frameworks based on recommendations. The people give the content audits like web-based reviews, feedback or remarks on the web-based media and web-based business sites. This textual content is an essential wellspring of client's conclusions. In sentiment analysis, the main aim is to classify the users' sentiments into positive, negative and unbiased. This type of analysis indicates the ubiquity or importance the product in the marketplace. Every human being has their different opinion, feelings, thoughts, and emotion for an event this can be known with the help of sentiment analysis. The extraction of features and classification are the steps in sentiment analysis. This work is centered around implementing a fresh approach to sentiment analysis. The sentiment analysis techniques have various phases which include pre-processing, feature extraction and classification. The various machine learning algorithms for sentiment analysis are reviewed in terms of certain parameters.**

Keywords—Sentiment Analysis; Machine Learning; NLP

I. INTRODUCTION

A. Sentiment Analysis

Over the past few years, an interesting and popular research area emerging lately is sentiment analysis. The opinions that are held by any numbers of individuals are reviewed and analyzed using sentiment analysis [1]. These opinions can be associated to an occasion, brand, people or item. Earlier, magazines, newspapers and other sources were used to express people views. However, with the advancement in technology the people have begun to express their feelings on different social networking and micro-blogging sites. In a beneficial way, opinions of people have been separated, studied and afterward evaluated via researchers.

Twitter has gained the highest popularity in comparison to all other micro blogging platforms in the past few years. It can be regarded as a substantial pointer for the sentiments of user. Different ways have been developed by several media organizations to mine the twitter information [2]. Any kinds of views related to different products and services can be generated using this information. Since there are several social media sites which are releasing their Application Programming Interfaces (APIs) today, there are been improvement in the process of data collection within research and development field. The users and developers are provided within various APIs through twitter. The three different API variants available in Twitter are RESTAPI, the Search API, and the Streaming API. The status data and client's information can be accumulated by the developers using REST API. The specific Twitter content can be queried using Search API. The content present in real-time is collected using Streaming API. Different kinds of sentiment analysis can be performed by these massive online data available today [3]. Tweets are known as the micro blogs posted on Twitter.

B. Data Extraction

The data source was outsourced from the tweets posted on Twitter. The messages posted on the twitter are extracted using its API. The "twitter auth" version of the public API is used and implemented in PHP. Either the web servers or local hosts can execute this directly or for the query few parameters are considered. An extensive set of filtering parameters are set during the extraction of tweets from Twitter such that they can match any specific criteria. API is used allows the query processing once it is generated. The output of this query corresponds to overall relevant twitter source data. The data is directly embedded within the MySQL database for being utilized in future. In every record that is generated though a tweet, information such as tweet id, text, client name and so on can be extracted. If any user publicizes its location, the data

relevant to place of message is generated in the form of latitude and longitude from the Twitter API. However, due to security concerns and client privacy, people have stopped sharing their locations since 2012. The site is used as a filtering metric in the principal query later this microblogging site. Thus, depending upon the steady set of places, the extraction of messages is carried out [6].

C. Sentiment Analysis Data Preprocessing

There is certain amount of irrelevant data available within the data collected from Twitter. Different kinds of illogical terms or useless information need to be removed from the data. The NLP tool is applied to remove this useless data. Any kinds of grammatical association present between the words of sentences are generated as outcome by this tool. Within the general natural language research, it is not useful to include certain advanced linguistics available in English language. Thus, there are 50 predefined relations called dependencies available in NLP which are listed and explained in this standard description [7]. The information analysts consider the key word associations as important although linguistics defines additional associations of words within a sentence, due to which these 50 dependencies have been defined in NLP. The most utilized conditions among these 50 are nsubj, amod, dobj. The messages carrying meaningful information can be identified through these associations. The outcomes are not at all aided by facilitating filtering along with supplementary associations. The associations amidst nouns and adjectives or verbs are discovered using nsubj relation within a noun sentence. Irrespective to accompanying a noun within a sentence or not, this is of great importance.

II. Literature Survey

A. Hybrid Approach for Sentiment Analysis

M anagha, et.Al (2015) introduced a hybrid technique for extracting sentiments from the reviews of Malayalam film [9]. The recommended scheme combined machine learning and fuzzy logic for sentiment analysis. The recommended approach was used for the tagging of input file. The tags were given by tnt tagger. The recommended approach used some specific rules for handling the exemptions. The sentiments were extracted using fuzzy membership functions. The tested outcomes demonstrated the efficiency of the recommended approach in decision making regarding a review. It was required to insert sentiment terms for novel fields. The future work would be based on the automatic production of sentiment lexicons. This phenomenon would decrease manual involvement in dictionary preservation.

Amira shoukry, et.Al (2015) proposed a hybrid technique for sentiment analysis of Egyptian language tweets [10]. The recommended scheme combined machine learning and SO (semantic orientation) approaches for this purpose. The tested results obtained showed a notable improvement in various performance parameters. Thus, these results confirmed the efficacy of the recommended scheme in the sentence level sentiment analysis. Some more efforts would be made in the nearby future for creating a more wide-ranging list of positive and negative sentiment terms. This work would enhance the overall performance of semantic orientation scheme for sentiment analysis.

Ping han, et.Al (2016)suggested approach was the mix of similarity-based and lexicon-based method.[11]. The sentiment polarity of small message was obtained using semantic rules and emoticons. Semantic rules based on Chinese language were used in this work. The sentiment based, and similarity-based techniques were combined for the extraction of sentiment terms. Also, emoticons revised the achieved results one more time. The examined results showed that suggested scheme effectively characterized dual sentiments. The words from micro blogs were used in this work for validating the accuracy of the recommended scheme. The recommended approach could be generalized and implemented easily on other small texts based on Chinese language.

Orestes appel, et.Al (2016) recommended a hybrid technique for analyzing sentiments. The recommended approach combined semantic rules, fuzzy sets, unsupervised machine learning algorithms and a sentiment lexicon modified by the senti-wordnet for sentiment analysis [12]. At first, a hybrid typical classification was performed. This approach was converted into a hybrid advanced approach through modification. This approach integrated linguistic classification of semantic polarity. Fuzzy sets were used here to model the polarity of sentiments. The suggested approach was once used for computing the polarity of a unique textual content and a preferred openly current dataset primarily based on movie review.

Mariam biltawi, et.Al (2017) recommended a hybrid approach for classifying the sentiments of Arabic dialect [13]. The recommended approach was the combination of lexicon based and corpus-based methods. The main objective of this approach was to present review for CBS (corpus-based approach) just like it represented in LBM (lexicon-based approach).In request to accomplish this point, the polarity phrases had been changed with their equal tag 'POS' (positive) or 'NEG' (negative) in the lexicon. Several classification models were compared in this work for performance evaluation. The examined results showed that suggested scheme carried out higher than CBA. The recommended scheme achieved maximum accuracy rate of 96.34%. However, the research work is still going on additional datasets for expanding lexicons and improving some classification model using just general words.

A. Haripriya, et.Al (2018) suggested a hybrid method to raise out real-time sentiment analysis of distinctly trending incidents on the foundation of location[14]. Several sentiment lexicons, unigram and bi-gram language paradigm with different machine learning algorithms were combined for study in this work. The evaluation of the achieved outcomes was carried out for verifying the performance of the recommended system based on several metrics. The recommended system outperformed the other location based

existing systems in terms of accuracy. The future work would be focused on extending the recommended system to multiple dialects with emoticons. Also, more lexicons could be discovered for classifying polarity.

Farkhundiqbal, et.Al (2019) recommended a hybrid system for sentiment analysis using genetic algorithm-based feature reduction [15]. The suggested approach coordinated lexicon- based strategies and AI techniques for improving exactness and versatility in sentiment investigation. The size of feature set could be reduced up to 42% without affecting the accuracy by the recommended hybrid system. The recommended approach was compared with other popular feature reduction approaches such as PCA (principal component analysis) and LSA (latent semantic analysis) in this work for performance analysis. The obtained results depicted that the recommended approach achieved enhanced accuracy rate of 15.4% and 40.2% over PCA and LSA approaches respectively. The future work would be focused on the expansion of this system for cyber-intelligence. This system would assist to generate suggestions for regulatory bodies based on users' views.

Vani kansal, et.Al (2019) recommended a hybrid approach for analyzing sentiments in financial matters based on artificial intelligence and cuckoo search [16]. Initially, the training of ANN (artificial neural network) based sentiment analysis model was done to get positive, negative and neutral sentiments. However, cuckoo search algorithm was recommended in this work due to some limitations of this classifier. Optimizing the fetched attributes with a new fit function was the main aim of this algorithm. The suggested calculation returned appropriate capabilities based on positive and negative conclusions. Every sentiment analysis model required an optimization method due to the occurrence of rough data. However, the recommended system used different pre-processing techniques for improving the data. The future work would involve different types of optimization techniques. In addition, more efforts would be made to deal with sarcastic tweets.

Korovkinas, et.Al, (2019) presented that ascent of informal organizations and spread of web-related splendid gadgets was followed by impact in information open for assortment and arranging, which offered authentic mechanical and computational difficulties alongside new energizing outcomes in assessment, division, and use of new and present information science and AI techniques [17]. In this work, a crossover method was used to improve the SVM precision of classification utilizing test preparing information and hyperparameter tuning. The proposed structure applies gathering to pick preparing information and boundary tuning to improve the suitability of the classifier. The paper audits that higher impacts had been gotten utilizing our proposed strategy in all examinations, interestingly with past impacts showed method's work.

Priyanka tyagi, et.Al, (2019)presented that any sentiment of a human through which the feelings, mentality, and thoughts can be passed on is perceived as assessment. [18]. Such data assessment which is refined from the news reports, customer overviews, online media refreshes, or microblogging locales are called sentiment analysis which is generally called assessment mining. The reviews of human beings in the direction of precise occasions, brands, object or enterprise can be recognized via sentiment analysis. Population responses have been collected and improved by experts performing the evaluation. Population responses have been collected and improved by experts performing the evaluation. The popularity of sentiment analysis is improving today as the number of comments shared publicly by small blogging sites is also increasing. All ideas can be grouped into three categories called positive, negative and neutral. Twitter, being the most frequent micro running a blog internet page, is used to collect data for testing. Tweepy is used to extricate the source data from twitter. Python language is used in this experiment to create an algorithm for the collected information. The author is responsible for using the n-gram modeling features that will be released. The idea is divided between positive, negative and neutral using a supervised learning algorithm known as the K-Nearest Neighbor.

B. Lexicon based for Sentiment Analysis

Endangwahyupamungkas, et.Al (2016) analyzed sentients in the language of Indonesia using a lexicon-based approach [19]. sentient data classification is performed on three senses. The experimental results showed that the suggested method obtained a 0.68 classification accuracy. In many cases, the recommended method has yielded positive results in the classification of sentiment. It was conceivable to utilize these challenges as a stage to complete more examination work. Originally, an unfamiliar dialect was found in the database. The following concern was identified with the disappointment of waiting episode. The ultimate problem was about imprecision.

Hunaidaawwad, et.Al (2017) proposed a hybrid stemming scheme for improving lexicon-oriented sentiment analysis (SA) in Arabian language [20]. To improve the chances of attaining the text of the sentiments, the suggested scheme utilized root-based degrees and light stemmers all in all. The implementation of recommended scheme was carried out on two datasets and four lexicons. The recommended scheme was compared by means of root-based stemmer and light stemmers independently. The achieved outcomes indicated that the recommended scheme performed superior to other approaches regarding accuracy when these approaches were used individually.

Rezvanehrezapour, et.Al (2017) recommended and assessed an improved model for SA (sentiment analysis) [21]. To enhance the accurateness of SA, the recommended model integrated informative hash tags into a lexicon. The tweets were collected regarding every contender of president post prior to year 2016 in general election held in USA to get the unity of all posts, the '# informed symbol LBA (dictionary-based approach) was utilized. Based on these tweets, the ranking was given to the contenders as per their reputation. The tested outcomes depicted that the intuitive hints about a contender's fame particularly on social media could be provided by assigning ranks to contenders because of

sentiments expressed in tweets. However, this work had several limitations. Firstly, sarcastic and symbolic texts were very frequent on twitter. The approaches employed in this work did not explain these effects in specific manner.

Annet john, et.Al (2019) recommended context-based sentiment analysis using hybrid lexicon. This work combined a wide-ranging lexicon and domain-dependent lexicon along with other context-based sentiment adjusting rules for improving the classification accuracy of opinion-based tweets' [22]. The main aim here was to handle the context-based polarity of tweet in that case, the past polarity of the word defined in the dictionary may differ from the one expressed in the tweet. In the results, it was demonstrated that new approach performed superior to existing approaches in terms of different performance metrics. This function will be expanded in the future by incorporating flexible rules.

C. Machine Learning based for Sentiment Analysis

Jamil hussain, et.Al, (2017) because to the huge growth of communication technology and the number of people the use of social media has been expanded[25]. As with the progression of innovation, mini computers and smart phones are easily available to the human pockets with ease to share and express new ideas using social media platform like face book, blogs, and twitter etc. In this paper, author presented the way of finding the sadness rate of an individual like its condition of low temperament and repugnance for a movement that can influence an individual's considerations, conduct, sentiments, and feeling of prosperity. It tends to be seen by separating feelings from the content, utilizing feelings speculations, AI procedures, and characteristic language handling strategy on various microblogging sites. In this research, they have compared different existing classifiers such as SVM, ME and NB to measure depression by sentence level sentimental analysis. The creator had received feature selection, voting model procedures and performed various investigations utilizing twitter and 20newsgroups dataset. The results showed that SVM is superior to naïve bayes in terms of accuracy.

Zahra rezaeiet.Al, (2017) large amount of data has been created using social networking sites and users share their ideas and small messages using micro blogging website. Micro blog sites are mainly used to share their opinions and to post their daily views, using micro-blogging service popularly known as twitter [26]. An overall review of a point has become another test as number of tweets increases on daily basis. Predicting the sentiments in tweets, algorithmic approaches will be used, and tweets are continually increasing and arriving at very high speed in limited span. Hoeffding tree algorithm has been used as popular tool in mining data streams and appropriate approach to obtain splitting attribute as data in twitter relies on data stream framework. To enhance the performance filtering and wrappers techniques were used. In twitter language common preprocessing task is used before applying any algorithms. Hoeffding tree has a high cycle time, to use this mcdiamid tree algorithm. Mcdiarmid tree performance is better than hoeffding tree but the accuracy of both are same. Due to large amount of twitter data, minimum processing time is required for the sentimental analysis therefore, mcdiarmid tree is better than hoeffding tree.

Mondherbouaziziet.Al,(2017) introduced another methodology for opinion mining that, notwithstanding the previously mentioned assignments of paired and ternary classifications, goes further in the order of writings gathered from Twitter and orders these writings into numerous sentiment classes [27]. While in that work, they limit their degree to seven distinctive sentiment classes, the proposed approach is versatile and can be hurried to arrange messages into more classes. They started by introducing SENTA, a tool designed to help users choose from a variety of items that are very similar to their system, to use them for classifying, through a simple to-utilize graphical UI. After this, SENTA is applied to perform several tests to classify tweets into many classes.

Mondherbouaziziet.Al, (2018)focused on the sentiment analysis and opinion mining [28]. In this work, people's sentiments were analyzed automatically on certain subjects through analysis of social media messages. Multi-level emotional analysis specifically, addresses on identifying specific emotions sent by the user rather than the quantity of the concept of his or her text or post. If so, they are delivering a different function than the standard multi-class categories, which apply to a set of data collected on twitter. This task was called "quantification". By the expression "quantification", it refers to the realization of overall emotions in online posts (e.g., Tweet) rather than providing a solo emotion carries out its labelling. This is the reason a method was suggested that automatically sets the dissimilar scores for all emotions in a message and select the scores with the maximum scoring they perceive as transmitted in the message. To arrive at this objective, in addition to their recently presented tools SENTA the important parts included running and performing an assignment .All through this work introduced the additional segments; study the attainability of evaluation and propose a way to deal with performing it on an informational collection made of tweets for 11 distinctive sentiment classes. The informational collection is transcribed, and the aftereffects of the mechanized analysis are checked with human remarks.

Mohammed H. Abd el-jawad, et.Al (2018) compared several machine learning and deep learning algorithms for performance analysis [29]. Moreover, a novel hybrid model was recommended in this work. This model used text mining and neural networks to classify sentiments. A dataset containing over 1 million tweets gathered from five fields was used in this work. The training and testing of the suggested model are done using 75% and 25% of the dataset, with the suggested model beat other existing methodologies as far as precision (83.7%). The future work would be focused on combining sentiments and text for analyzing sentiments. In addition, the hybrid classification process will be used soon for the sentiment analysis of arabic tweets.

Megharathi, et.Al (2018) classified emotions because of Twitter data. The ensemble machine learning algorithms were applied in this work for improving classification outcomes in terms of opinion mining [30]. This paper applied a hybrid

scheme for analyzing sentiments. This approach combined two classifier models called SVM and DT together. The ensemble algorithms improved the efficacy and trustworthiness of the recommended approach. The tested results depicted that the recommended scheme provided better classification outcomes in terms of different performance parameters.

Mondherbouaziziet.Al,(2019)introduced that majority of studies in the field of emotional analysis are concerned with classifying these messages into binary and ternary. The work of multi-class classification has gotten little consideration [31]. Classifying messages using this approach is a difficult work when you consider the difficulty of normal languages and the complexity of interpreting and measuring "quantifying" the way people show their emotions. In this movement, they gain proficiency with the capacity of multi-class characterization in online posts for Twitter clients, and afterward, show how it is conceivable to sort, and the limits and challenges of that task. Regardless, they propose a novel model to speak to the various sentiments and show how this model assists with seeing how sentiments are connected. The model is then used to examine the difficulties posed by the various categories and to highlight future improvements in the accuracy of the various categories.

III. RESEARCH GAP

1) The techniques which are designed previously are proposed on the small size data, when the size of the data gets increase parametric value gets reduced at steady rate. The accuracy parametric value also describes other values like precision, recall and f-measure which also get affected [14] [12].
2) The sentiment analysis approaches need to handle large quantity of historical information to train the model. While handling such large amount of data, the execution time is increased which affects performance [11][17].
3) In the previous years, much work is done on the various type of techniques for the feature extraction and classification which are too complex, it directly increases the execution of the model [13][16]
4) The technique of feature extraction like N-gram is applied for the feature extraction. The feature extraction technique can divide the data into certain parts. The N-gram method can be further enhanced to divide data randomly [14][17]
5) The dataset of the sentiment analysis which is collected from the twippy API has many attributes which can be reduced using the feature extraction algorithm [21][22]
6) The supervised classification algorithms require the target set for the classification of data into certain classes. The dataset which is collected does not have the target which needs to assign dynamically which affect the efficiency of the model [10][19]
7) The sentiment analysis can be done using the lexical analysis phase, the technique of lexical analysis can be done using the priority phases. The threshold value can assign by taking average of the priority values which affect efficiency of the proposed model[17][23]
8) The classification methods like SVM, KNN, tree methods are already applied for the sentiment analysis. The classification methods require various parametric values which needs to define for the sentiment analysis [20][25]

TABLE I.

Authors' Names	**Year**	**Research Process**	**Outcomes**
Anagha, M., Et al.	2015	Introduced a hybrid technique for extracting sentiments from the reviews of Malayalam film. The recommended scheme combined machine learning and fuzzy logic.	The tested outcomes demonstrated the efficiency of the recommended approach in decision making regarding a review. It was required to insert sentiment terms for novel fields.
Shoukry, Amira, and Ahmed rafea	2015	Proposed a hybrid technique for analyzing sentiments of Egyptian language tweets.	The achieved tested results showed noteworthy enhancement in different performance parameters.
Han, Ping, Shan li, and Yunfeijia	2016	Recommended a hybrid scheme for extracting sentiment words. The recommended approach was the combination of lexicon-based and similarity-based technique.	The recommended approach could be generalized and implemented easily on other small texts based on Chinese language.
Orestes appel, et al	2017	Recommended a hybrid technique for analyzing sentiments. The recommended approach combined semantic rules, fuzzy sets, unsupervised machine learning	The recommended approach was used for computing the polarity of a text and a standard openly existing dataset based on film review.

		algorithms and a sentiment lexicon modified by the senti-wordnet for sentiment analysis.	
Biltawi, Mariam, Ghazi al-naymat, and Sara tedmori	2018	Recommended a hybrid approach for classifying the sentiments of Arabic dialect.	The recommended scheme achieved maximum accuracy rate of 96.34%.
Haripriya, A., Santoshi kumari, and C. Narendra babu	2017	Recommended hybrid process for performing real-time emotional analysis of leading events on a local basis.	The evaluation of the achieved outcomes was carried out for verifying the performance of the recommended system because of several metrics. The recommended system outperformed the other location based existing systems in terms of accuracy.
Iqbal, Farkhund, et al	2019	Recommended a hybrid system that used genetic algorithm-based feature reduction for analyzing sentiments.	The obtained results depicted that the recommended approach achieved enhanced accuracy rate of 15.4% and 40.2% over PCA and LSA approaches respectively.
Kansal, vani, and Rakesh kumar	2019	Recommended hybrid process for performing real-time emotional analysis of leading events on a local basis.	The recommended algorithm returned suitable feature sets because of positive and negative sentiments.
Korovkinas, Konstantinas, Pauliusdanėnas, and Gintautas garšva	2019	Proposed a hybrid approach for sentiment analysis.	This paper reports that better results have been achieved using our proposed strategy throughout the investigation, compared to previous results that have shown technical work.
tyagi, priyanka, and r. C. Tripathi	2019	It has been suggested that any human emotion in which emotions, attitudes, and thoughts can be transmitted.	The idea is divided between positive, negative and neutral using a supervised learning algorithm known as the k-nearest neighbor.
pamungkas, endangwahyu, and divi galihprasetyo putrid	2016	Performed lexicon-based sentiment analysis of Indonesian language. Emotional data classification is performed on three senses such as positive, negative and neutral.	The experimental results showed that the recommended method obtained a 0.68 classification accuracy.
awwad, hunaida, and adilalpkocak	2017	A hybrid stem program has been proposed to improve dictionary-focused (SA)-based emotional analysis in the Arabian language.	The achieved outcomes depicted that the recommended scheme outperformed the other approaches in terms of accuracy when these approaches were used individually.
rezapour, rezvaneh, et al	2017	Recommended and assessed an improved model for SA (sentiment analysis). In order to enhance the accurateness of SA, the recommended model integrated informative hash tags into a lexicon.	The tested outcomes depicted that the intuitive hints about a contender's fame particularly on social media could be provided by assigning ranks to contenders on the basis of sentiments expressed in tweets.
john, annet, anice john, and reshma sheik	2019	Recommended context-based sentiment analysis using hybrid lexicon. This work has included a comprehensive dictionary and background-based dictionary as well as other emotional correction rules to improve the classification of emotion-based tweets.	The tested outcomes demonstrated that the recommended approach outperformed the other existing approaches in terms of different performance metrics.
chauhan, chhaya, and smriti sehgal	2017	In this research, they have compared different existing classifiers such as SVM, ME and	The results showed that SVM is superior to naïve bayes in terms of accuracy.

		NB in order to measure depression by sentence level sentimental analysis.	
rezaei, zahra, and mehrdadjalali	2017	Hoeffding tree and Mcdiarmid tree algorithm has been used as popular tool in mining data streams and appropriate approach in order to obtain splitting attribute as data in twitter follows data stream model.	Mcdiarmid tree performance is better than hoeffding tree but the accuracy of both are same.
bouazizi, mondher, and tomoakiohtsuki	2017	Introduced another methodology for sentiment analysis that, notwithstanding the previously mentioned errands of paired and ternary classifications, goes further in the classification of writings gathered from Twitter and characterizes these writings into various sentiment classes.	In that work, they measured their magnitude in seven different categories of emotions; the proposed method is scary and can be further subdivided into several categories.
bouazizi, mondher, and tomoakiohtsuki	2018	All through this work introduced the additional parts; study the possibility of measurement, and propose a way to deal with performing it on an informational index made of tweets for 11 diverse sentiment classes	The data set is handwritten, and the results of the automated analysis are checked with human interpretation.
el-jawad, mohammed h. Abd, raniahodhod, and yasser MK omar	2018	Compared several machine learning and deep learning algorithms for performance analysis. Moreover, a novel hybrid model was recommended in this work.	The training and testing of the recommended model is done using 75% and 25% of the data part of the dataset the recommended model outperformed other existing approaches in terms of accuracy (83.7%).
rathi, megha, et al	2018	In this work, a hybrid approach was recommended for sentiment analysis. This approach combined two classification algorithms called support vector machine and decision tree together	The tested results depicted that the recommended scheme provided better classification outcomes in terms of different performance parameters.
bouazizi, mondher, and tomoakiohtsuki.	2019	They suggest a novel model to represent different emotions and show how this model helps to understand how emotions are related.	The model is used to analyze the challenges posed by the various categories and to highlight future potential improvements in the accuracy of the various categories.
astya, parmanand	2017	Introduce that emotional analysis is the process of identifying an idea or feeling expressed in a visual data, to decide the creator's demeanor towards a specific point whether it is positive, negative or impartial.	The author concludes, by using less-supervised and unsupervised learning-based models, it will be easier to reduce the absence of labeled data if a sufficient number of unlabeled data is available.
garg, pulkit, himanshugarg, and virenderranga	2017	In this paper, we read offensive tweets after being intimidated by posting them on twitter. Flow data posted on twitter is used to study things like retweet retention, number of retweets and	The author discussed the terrorist attacks which show that the worst tweets tend to survive compared to the best tweets, even though their numbers are low.

		number of preferences.	
valdivia, ana, m. Victoria luzíón, and franciscoherrera	2017	The paradoxical majority based on the combination of polarity of several sensors analyzes the use of the weight generated by the proposed order.	The author's main focus is on removing those neutral updates that are labeled in harmony with the collections.
zhao, wei, et al	2017	In order to achieve a well-maintained order using high-quality sentences and the top of the embedding layer is placed a layer on the other side in the details of the first scale used to capture the general distribution of emotions.	Examining the proposed framework has undergone separate studies that show its superiority over the foundations.
kulkarni, d. S., And S. F. Rodd	2018	A few techniques that show the release of an object based on the metaphorical mining patterns used by one research organization and avoid it with slight variations in the expression of distributional ascribes of sentiment attribute. Such strategies are not free to predict people's response successfully.	They have looked at and seen current research issues that are interesting based on an extended study of ongoing strategies. Through their research, they may think that emotional analysis is appealing to many researchers.

IV. CONCLUSION

This conclusion of this work is that the sentiment analysis makes use of NLP approach for analyzing sentiments in tweeter data. This task involves several steps of pre-processing, feature extraction and classification. In this paper, various classification algorithms are compared for analyzing sentiments. This work is based on the analysis of twitter data to study the sentiments of users. The feature extraction phase will tokenize the sentences for the further processing. The machine learning algorithm will take input of the training and testing set for the sentiment analysis. The various parameters like accuracy, precision and recall are used for the performance analysis. In future a novel hybrid method will be designed in the time ahead for the sentiment analysis to improve accuracy. N-Gram technique has been used for feature extraction. Thereafter k-nearest neighbor classifier is used for classification of the tweets into positive, negative and neutral classes.

REFERENCES

[1] Devika M D, Sunitha C, Amal Ganesh “Sentiment Analysis:A Comparative Study On Different Approaches”,Procedia Computer Science, vol.87 , pp. 44-49,2016

[2] Vishal A. Kharde, S.S. Sonawane “Sentiment Analysis of Twitter Data: A Survey of Techniques”, International Journal of Computer Applications, vol.139, pp. 11-21, 2016.

[3] Go, A., Bhayani, R., Huang, L., “Twitter sentiment classification using distant supervision”, CS224N Project Report, Stanford, vol. 46, pp. 59, 2009.

[4] Y. Kawai, T. Kumamoto, and K. Tanaka, “Fair news reader: Recommending news articles with different sentiments based on user preference,” in Proceedingsof Knowledge-Based Intelligent Information and Engineering Systems(KES), number 4692 in Lecture Notes in Computer Science, pp. 612–622, 2007.

[5] A. Kennedy and D. Inkpen, “Sentiment classification of movie reviews using contextual valence shifters,” Computational Intelligence, vol. 22, pp. 110–125, 2006.

[6] James Spencer and Gulden Uchyigit “Sentiment or: Sentiment Analysis of Twitter Data”, James Spencer and Gulden Uchyigit, vol.88, pp. 1-11, 2016

[7] DoaaMohey El-Din Mohamed Hussein “ A survey on sentiment analysis challenges”, Journal of King Saud University – Engineering Sciences ,vol. 50, pp. 1-9, 2016

[8] Shahnawaz, Parmanand Astya “Sentiment Analysis: Approaches and Open Issues” International Conference on Computing, Communication and Automation, vol. 9, pp. 1-5, 2017

[9] Anagha, M., Et al. "Fuzzy logic based hybrid approach for sentiment analysisl of Malayalam movie reviews." 2015 IEEE international conference on signal processing, informatics, communication and energy systems (SPICES). Ieee, 2015.

[10] shoukry, amira, and ahmedrafea. "A hybrid approach for sentiment classification of egyptian dialect tweets." 2015 first international conference on arabic computational linguistics (acling). Ieee, 2015.

[11] han, ping, shan li, and yunfeijia. "A topic-independent hybrid approach for sentiment analysis of chinese micro blog." 2016 IEEE 17th international conference on information reuse and integration (IRI). Ieee, 2016.

[12] appel, orestes, et al. "A hybrid approach to sentiment analysis." 2016 IEEE congress on evolutionary computation (CEC). Ieee, 2016.

[13] biltawi, mariam, ghazi al-naymat, and saratedmori. "Arabic sentiment classification: A hybrid approach." 2017 international conference on new trends in computing sciences (ICTCS). Ieee, 2017

[14] haripriya, A., Santoshi kumari, and C. Narendra babu. "Location based real-time sentiment analysis of top trending event using hybrid approach." 2018 international conference on advances in computing, communications and informatics (ICACCI). Ieee, 2018.

[15] iqbal, farkhund, et al. "A hybrid framework for sentiment analysis using genetic algorithm-based feature reduction." IEEE access 7 (2019): 14637-14652.

[16] kansal, vani, and rakeshkumar. "A hybrid approach for financial sentiment analysis using artificial intelligence and cuckoo search." 2019 5th international conference on advanced computing & communication systems (ICACCS). Ieee, 2019.

[17] korovkinas, konstantinas, pauliusdanėnas, and gintautasgaršva. "SVM and k-means hybrid method for textual data sentiment analysis." Baltic journal of modern computing 7.1 (2019): 47-60.

[18] tyagi, priyanka, and r. C. Tripathi. "A review towards the sentiment analysis techniques for the analysis of twitter data." Available at SSRN 3368718 (2019).

[19] pamungkas, endangwahyu, and divi galihprasetyoputri. "An experimental study of lexicon-based sentiment analysis on bahasaindonesia." 2016 6th international annual engineering seminar (inaes). Ieee, 2016.

[20] awwad, hunaida, and adilalpkocak. "Using hybrid-stemming approach to enhance lexicon-based sentiment analysis in arabic." 2017 international conference on new trends in computing sciences (ICTCS). Ieee, 2017.

[21] rezapour, rezvaneh, et al. "Identifying the overlap between election result and candidates' ranking based on hashtag-enhanced, lexicon-based sentiment analysis." 2017 IEEE 11th international conference on semantic computing (ICSC). Ieee, 2017.

[22] john, annet, anice john, and reshma sheik. "Context deployed sentiment analysis using hybrid lexicon." 2019 1st international conference on innovations in information and communication technology (ICIICT). Ieee, 2019.

[23] juneja, pragya, and uma ojha. "Casting online votes: to predict offline results using sentiment analysis by machine learning classifiers." 2017 8th international conference on computing, communication and networking technologies (ICCCNT). Ieee, 2017.

[24] chauhan, chhaya, and smriti sehgal. "Sentiment analysis on product reviews." 2017 international conference on computing, communication and automation (ICCCA). Ieee, 2017.

[25] hassan, anees ul, et al. "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression." 2017 international conference on information and communication technology convergence (ICTC). Ieee, 2017.

[26] rezaei, zahra, and mehrdadjalali. "Sentiment analysis on twitter using mcdiarmid tree algorithm." 2017 7th international conference on computer and knowledge engineering (ICCKE). Ieee, 2017.

[27] bouazizi, mondher, and tomoakiohtsuki. "A pattern-based approach for multi-class sentiment analysis in twitter." IEEE access 5 (2017): 20617-20639.

[28] bouazizi, mondher, and tomoakiohtsuki. "Multi-class sentiment analysis in twitter: what if classification is not the answer." IEEE access 6 (2018): 64486-64502.

[29] el-jawad, mohammed h. Abd, raniahodhod, and yasser MK omar. "Sentiment analysis of social media networks using machine learning." 2018 14th international computer engineering conference (ICENCO). Ieee, 2018.

[30] rathi, megha, et al. "Sentiment analysis of tweets using machine learning approach." 2018 eleventh international conference on contemporary computing (IC3). Ieee, 2018.

[31] bouazizi, mondher, and tomoakiohtsuki. "Multi-class sentiment analysis on twitter: classification performance and challenges." Big data mining and analytics 2.3 (2019): 181-194.

[32] astya, parmanand. "Sentiment analysis: approaches and open issues." 2017 international conference on computing, communication and automation (ICCCA). Ieee, 2017.

[33] garg, pulkit, himanshugarg, and virenderranga. "Sentiment analysis of the uri terror attack using twitter." 2017 international conference on computing, communication and automation (ICCCA). Ieee, 2017.

[34] valdivia, ana, m. Victoria luzíón, and franciscoherrera. "Neutrality in the sentiment analysis problem based on fuzzy majority." 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE). Ieee, 2017.

[35] zhao, wei, et al. "Weakly-supervised deep embedding for product review sentiment analysis." IEEE transactions on knowledge and data engineering 30.1 (2017): 185-197.

[36] kulkarni, d. S., And S. F. Rodd. "Extensive study of text-based methods for opinion mining." 2018 2nd international conference on inventive systems and control (ICISC). Ieee, 2018