

Examination of Twitter Data to Assess the Presence of False Reporting in a Food Review

M. Tech. Scholar Sakshi Patidar, HoD & Prof. Kamlesh Patidar

Department of Computer Science and Engineering
Jawaharlal Institute of Technology,
Borawan Dist. Khargone, MP, India

Abstract- In today's internet age, micro blogging sites, personal blogs, and reviews spread ideas and opinions. The reviews cover a variety of topics, including products, companies, businesses, individuals, forums, corporations, brands, movies, and more. Text mining requires sentiment analysis. Public tweets were categorized as positive, negative, or neutral. Twitter's API will collect work data, and keywords will determine tweet sentiment and paper assessments. Next, the percentage of positive tweets will determine tweet polarity. Negative. Then, a supervised model was used to analyze more data sets. Machine learning is applied. NB, ME, RF, and SVM are machine learning classifiers. This work uses SVM classifiers for data set testing, training, and tweet sentiment polarity. Classifiers test and train data sets. Analyze the parameters to prove the completed output has the best classifier performance. RF, DTs, and SVM classifiers and more tweets will improve feature evaluation accuracy. Future research using the same method might include more features, improving prediction accuracy.

Keywords- Sentiment analysis, twitter data, machine learning techniques, opinion mining, social media, Support vector machine (SVM), Naïve Bayes(NB), Maximum Entropy, Decision Tree (Dts), Random Forest(RF).

I. INTRODUCTION

In recent year people in these days depending on micro blogging sites like facebook instagram, tumblr, twitter youtube millions people share the posts, live news and express their opinion about different subjects such as a political affair, product review, educational, women issue and general topics, extracting knowledge from the twitter data [1]. Sentiment analysis is the mining of opinion and analysis of twitter data and that describe as a positive negative and neutral category which explore data from various social media platforms [2]. The aim of this analysis in research determining the subjectivity opinion. Result of this analysis based on this sentiment analysis and review of tweets or classified opinions which are based on the data size and document type [4]. Twitter application is an excellent medium for creation of tweets presentations [5].

Twitter analysis is a popular topic for research. Such analysis is useful because it's gathering by crawler data which are used for collect to data from twitter and classified public opinion by analyzing of vast social media data [6]. The aims of this study that analyze the level of sentiment from the social media [7]

In this sentiment analysis we are using twitter API for extracting data then cleaning the data and after these processes fed data into three classified tweets on the basis of sentiment (new data) [8]. This Analysis helps to understand the way of thinking about any research topic brands, products etc [9].

Through the advertisement campaign can see how people are reacting from this campaign in personal marketing. There is a way to analyze sentiment related to them. [10]. Use of the same campaign can

be seen as reacting for Political parties and can be analyzed.

There are several reasons for sentiment analysis where we can choose twitter data as given below.

- On twitter more than 500 million numbers of tweets on daily bases and that is a vast level of data for sentiment analysis.
- On Twitter there is number of all age groups people, with a high percentage of business executives' people being present from many countries on social media.
- 50 million or more people download from many browser twitter applications.

In this study we have used of supervised learning Classifiers to analyze the sentiment of the people for this analysis. Such as Support Vector machine learning classifiers (SVM), Decision tree (DTs) and Random forest (RF). In this result we will compare all classifiers based on accuracy which gives the best result. Finally for this research we also used machine learning techniques.[29]

In our work we introduced of score vector of tweets and our external features with n-gram of features and show that impact of SVM classifiers on for improve our classification performance level.

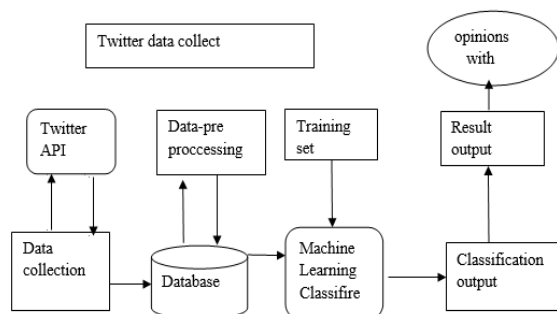


Fig 1. Working Architecture of Twitter classification.

II. LITERATURE REVIEW

Santhosh Kumar and others, IEEE Through the use of machine learning strategies, the fields of data mining and analysis of opinions on Twitter data are able to investigate items originating from a variety of social networks. This kind of communication, in which a person sends information to be read by other users and then posts it, is known as a "tweet." The comments on the posts made by other users are their method for disseminating information and

expressing opinions.[30] These are some of the best places for advertising other tweets and opinions. This model will extract the data from Twitter from the social media, delete any unnecessary data, and then classify the remaining data into one of three categories: good, negative, or neutral. For the purpose of this article, we will acquire hotel data from Twitter for the purpose of doing this analysis using machine learning algorithms. We will then discover the outcome utilizing a variety of metrics in order to get an accurate output result. Analysis of tweets based on opinions, employing a variety of classifiers, with the goal of obtaining reliable results. Both supervised and unsupervised machine learning methods have been included into this model [9].

L. L. Bo Pang et all In this research, we have used a variety of deep learning approaches for the purpose of conducting a sentiment analysis using data from Twitter. Deep learning is a technology that can simultaneously answer a broad variety of issues or carry out complicated tasks, in addition to gaining popularity among academics. When it comes to feature extraction, deep learning algorithms on their own can create the high order features that are necessary to make accurate predictions about an object. [31] This contributes to the development of respect for the thing in question. Deep learning's features make it possible to process massive amounts of data, both organized and unstructured. In general, employ two different kinds of neural networks: convolutional neural networks (CNN) for image processing and recurrent neural networks (RNN) for natural language processing. Both of these forms of neural networks are useful. We are able to make use of a variety of embedded systems, such as a word2Vec or a global vector (GloVe). A number of different combinations are tried out in an effort to find the model with the highest possible score value, and the results are compared [10].

Twitter is a platform where people exchange their knowledge and views as tweets, and it is one of the excellent sources for sentiment Analysis. V. Lakshmi et all, in this analysis, we gather data with opinion in this modern age from micro blogging sites such as twitter and facebook. People's opinions may be grouped into one of three categories: positive, negative, or neutral. [32] This helps in performing analysis on the various types of opinion, which is one of the necessary steps for performing sentiment analysis. Data mining and text mining, both forms of

natural language processing, are used in order to remove undesired information gleaned from social networking sites. The classification of tweets using machine learning methods is the focus of this study. The goals of this research are to enhance classification results for sentiment analysis while also boosting the effectiveness and reliability of proposed methodologies. Obtaining the maximum possible accuracy from the classifier may be accomplished by using decision trees, hybrid trees, and ad boosted trees. The suggested model is composed of two stages of preprocessing and one level of classification. The accuracy of classifiers and their f-measures may be improved with the use of hybrid models [11].

Levy, M et al People throughout the globe now communicate and share information with one another via social networking sites such as Facebook and Twitter. Twitter is a social networking website that allows users to communicate with members of various online groups. Where users may submit postings, sometimes called as tweets, and read other users' messages user-updated opinion about things like daily news, brand, and other locations. [33] The goal of this model is to first collect the actual data from the Twitter account, and then to do sentiment analysis on that data. In order to do this analysis for this model, we are making use of supervised machine learning methods. For the purpose of this study of the paper, we are carrying out sentiment analysis by extracting the data straight from Twitter API, followed by a cleaning procedure and the finding of the data on Twitter.

After the data have been brought in, the training of the data will be done on multiple different models. Based on the results of the sentiment analysis, each tweet was placed into one of three categories: good, negative, or neutral. Where the data are collected on two topics, McDonald's and KFC, the former of which is more well-known. For these models, numerous machine learning methods are used, and the results are discovered via the usage of cross validation, f-score, maximum entropy, and other testing metrics, amongst others. In further work, we will be able to do sentiment analysis for a variety of domains, such as the detection of rumours on Twitter data pertaining to the spread of illnesses [12].

W. Medhat, A. Hussan, and all of the others the current age is known as the modern era, and it is

based on the internet. In this era, individuals communicate their opinions and ideas via various forms of social media, including micro blogging sites, personal blogs, reviews, and so on.

Text mining includes a number of different processes, including one called sentiment analysis. The opinions of other individuals were analyzed and then classified into positive, negative, and neutral tweets. The findings of this study are provided in the form of people's evaluations and the sentiments of tweets, which are then recognized with the assistance of searching for a keyword. Finally, an assessment of the positive or negative polarity of tweets is made. [34] Using Naive Bayes classifiers (NBC) can be used to both test the data and features of words as well as evaluate the polarity of sentiment of each tweet. Additionally, it can be used to compare the performance of different machine learning classifiers, such as Random forest, Naive Bayes, and support vector machine, using evaluation parameters such as accuracy and precision. By using Classifiers such as RF, NBC, and SVM, an increase in both the estimated accuracy and the three features of the number of tweets may be achieved. It is possible that in future work some other characteristics that are employed for enhancing the accuracy of prediction may be introduced [13].

Sidharth, darsini et al People all around the world share their thoughts and information using various social media platforms on a daily basis. The Twitter programme is one of the most popular platforms for sharing opinions, reviews, postings, and special subject concerns. It is also one of the platforms that have the most users. The primary objective of this study was to undertake a sentiment analysis of people's opinions and examine the societal challenges facing women using the model that was provided.

This is a very pressing issue in many nations, and it affects every woman. Using a twitter scraper to capture data from Twitter, a dataset can be constructed in python programming; this dataset can then be cleaned, and noise can be removed from the dataset. Python programming tools like text blob are used for classifying each tweet using the technology. [35] These tools are among those that are used. The Text Blob will classify each tweet according to whether it is positive, negative, or neutral depending on the polarity of the feeling. #Women and the

#MeToo movement. There are two different sets of data. By using a variety of machine learning techniques, one's hypotheses may be evaluated using the model. After obtaining the results, compare the effectiveness of each model with the data that was evaluated using the different testing settings. In order to get a greater level of accuracy with both hashtags (#Women's and #Metoo), support vector machines are used. The material is shared using the hashtag #women, which is more popular than #Metoo. In further research, we will apply the same technique to data from additional Twitter accounts in order to improve the effectiveness of other classifiers used for sentiment analysis [14].

The goal of Harpreet Kaur and her colleagues in this study is to determine the polarity of the word and categories tweets based on whether they are positive or negative. [36] In this particular piece of research, we make use of lexicon-based classifiers in conjunction with machine learning. Support vector machine (SVM), Logistic Regression (LR), Multinomial Naive Bayes (MNB), Recurrent Neural network (RNN), and Recurrent Neural network (RNN). Both types of preexisting data sets, "Sentiment140" from Stanford University, which is comprised of 1.6 million tweets and "original" data from 'Crowd flower' data, which is comprised of every library entry numbering 13870, have been utilized in this study. Both data sets are categorized according to sentiment.

A number of different classifiers are applied to both sets of data, and the results that are achieved are then compared with one another. Apply this model to determine people's feelings in order to anticipate fresh data. Data will be taught and properly categorized based on standard dictionaries by using text machine learning models [15].

B. O'Connor et al. have found that during the last several years, sentiment analysis has been more popular on Twitter. Within the scope of our investigation of tweets, we focused on ordinal regression. The purpose of this study is to carry out sentiment analysis on the data obtained from Twitter by using machine learning methods and ordinal regression. In this method, a feature extraction technique is used to the preprocessing of tweets, which results in the creation of an effective feature. [37] You may utilize it in a few different classes for scoring features and balancing out the game. This

study makes use of supervised learning classifiers such as multinomial logistic regression (MLR), support vector machines (SVM), decision trees (DsT), and random forests (RF). In this investigation, we used the NLTK corpora resources and the Twitter data set that was provided freely accessible for the creation of this system using various machine learning techniques. Discovering that detection of ordinal regression provides the highest degree of precision for experimental work. Nevertheless, you will get the greatest results by utilizing the Decision tree rather than the other tech

niques. The accuracy of the Decision tree is rather excellent, coming in at 91.81 percent. In the work that will be done in the future, the utilization of bigrams and trigram together with other deep learning methods will be used in enhanced ways [16].

Seyedali Mirjalili et al Analysis of sentiment is a sort of text mining that determines if a piece of written material has a good, negative, or neutral attitude. This is an example of opinion mining or polarity analysis using material. The utilization of a variety of different individuals who are accessible on social media platforms may help increase the quality of this study. Because restricting the amount of characters in tweets makes it easier to do analysis, a lot of users choose to limit their tweets to 280 characters when posting them on Twitter.

There are around 550 million new tweets made every single day on Twitter. [38] The purpose of doing sentiment analysis from Twitter data is to determine the overall attitude of society. In this research, we make use of a number of different machine learning approaches, such as the support vector machine classification method, the maximum entropy classification method, and the Naive Bayes Classification method. Using this strategy, we will be able to determine the high accuracy and precision of the data obtained from Twitter. The maximum entropy approach achieves the same level of success as utilizing the baseline learning method, but it is far more successful than the baseline method. The tweets were analyzed in this research and categorized according to whether they were good, negative, or neutral. Work to be done in the future to make the performance measure better [17].

B. Pang, L. Lee In the context of this research, sentiment analysis can be performed using specific

opinions that are found in social media via online micro blogging sites such as Twitter and Facebook. This analysis is then utilized to take into consideration a factor in the field of research that is associated with products, movie reviews, and the stock market. For the purpose of study, movie data was obtained through Twitter's application programming interface (API), and machine learning techniques were used to predict the emotion of data associated with reviews of movies. Using unigram and bigram features, classifiers such as Support vector machine (SVM), Naive bayes (NB), and Maximum entropy classifier categories data. [39]

After doing preprocessing, cleaning, and training the data using a classifier, the data were then analyzed. Use 15,000 tweets for the training set, and use just 2,000 tweets for the testing set. It outperforms all other classifiers using SVM, and its accuracy is 84% when combined with the selection feature. In addition, Maximum Entropy, Naive Bayes, and Bigram are included. Because Maximum Entropy performs better than Naive Bayes, and Support Vector Machines (SVM) provide superior results when compared to another classifier [18].

B. Pang, L. Lee at all Twitter is now the most significant social media network, and it may be used for a variety of purposes. Use Twitter in a variety of situations, such as to get an opinion on the most recent product, to read a review of a movie, to estimate an event, and to identify someone's data along with their weightage and concept, and so on. This analysis may be used either in support of users or against a political party, and it can also be used to illustrate a divided view. [40] In this research, we investigate the possibility of predicting the opinions of users based on the identities of several political parties. In this research, either SVM or Naive Bayes algorithms were used to examine people's sentiments about Congress and BJP, which are two separate political parties.

A sentiment analysis of 140 characters was also constructed with the use of tweets. For the purpose of improving the precision of a greater number of data sets. Using support vector machines (SVM) to analyze the data set of tweets from #congress and #BJP, with an accuracy of 77.33 and 75.48 percent respectively. For two different reasons, we have been using domain-based tweets. The first purpose is to improve the quality of tweets and make predictions

about the labels in the data set. The second is used, with respect to the data set, in order to increase the accuracy of the greatest data collection, such as sentiment 140. The usage of a hash tag for one political party but not another is considered to be informal slang and is not sanctioned or permitted [19].

Darwish and all of them The vast majority of individuals disseminate their knowledge, thoughts, and perspectives on many topics over the internet via social media platforms like Twitter. In the course of this research, the amount of sentiment information was increased. Through the use of the research technique, we acquire insight into how individuals feel and how they react in a variety of circumstances. In this research, we compare the performance of the machine learning algorithm with that of the deep learning algorithm. In order to classify people's feelings, we made use of a combination of traditional computer algorithms and neural networks. a piece of work consisting of a data collection including over one million tweets gathered from five different domains. In this particular system, 75 percent of the dataset is used for training, while 25 percent is used for testing.

The hybrid learning strategy, which made use of both unsupervised and supervised learning methods, produced the most accurate results with an efficiency rate of 83.7%. Convolution neural network, Naive bayes, decision tree, and recurrent neural network are some of the several classification algorithms. Using hybrid models allowed for the maximum level of accuracy to be achieved, which was 83.1 percent specificity and 83.3 percent sensitivity. The data set. In further study, sentiment will be paired with either feelings or text for analysis [20].

The study conducted by T. Joachims and his colleagues on Twitter focuses on the analysis of tweets in the context of the phrase "research for the twitter sentiment categorization." In the models that have been presented for the purpose of extracting various characteristics from N-grams of Twitter data using the machine learning domain. [41] In addition to this, we used the various weighting strategies in order to better comprehend the influence on the classifier's accuracy. For the purpose of enhancing the performance of an SVM classifier by utilizing a score vector composed of tweets as a source of external knowledge. Within this experiment, we have

included a total of four distinct n-gram features, each of which uses a unique method of weighting. The Unigram function displays the highest level of performance with regard to the correctness of the output. The hybrid model utilizes an external feature and compares itself to the superior performance of the SVM classifier in terms of accuracy. In order to increase our performance in the task that we will be doing in the future, we need to plan with more relevant external information [21].

Surnar, Avinash et al Twitter platforms are a manner of expressing ideas, and social networking sites serve as platforms on which users may express their thoughts and share them with millions of other users on a daily basis. Therefore, the primary emphasis of this review study is on. Text mining and natural language processing-based sentiment analysis In this research, a machine learning algorithm is employed to execute extraction sentiment from the sentiment of tweets, and a variety of approaches were also applied for this analysis of study using tweets. In the survey study, we went over a variety of strategies, methodologies, and methodologies. It is essential to have an understanding of the relevance of Extrication tweets, as well as their structure and twitter.

Word Net's use of these techniques has been demonstrated in published research to result in an increase in accuracy. Examples are Support Vector Machines, Naive Bayes, and Maximum Entropy [22].

O. Almatrafi et al When machine learning and natural language processing are combined, sentiment analysis becomes a reality. Using a variety of content forms, such as audits, news pieces, and articles, you may categories people's thoughts as either favourable, negative, or neutral. It is quite challenging to get results from tweets while attempting prediction in the Indian language. Using the archiver provided by Twitter, you may retrieve tweets written in the Hindi language.

In the election that took place in India in 2016, we gathered tweets over the course of time that mentioned five national political parties in order to conduct text mining on 42,235 tweets. We also employed supervised and unsupervised methods in this study. The SVM, Naive Bayes, and Dictionary-based classification algorithms were used to construct the positive, negative, and neutral categories of the tested data, respectively. The results showed that the SVM strategy was the most

successful for the BJP (78.4 percent), followed by the Naive Bayes approach (62.1 percent), and the dictionary approach for the Indian National Congress. The BJP came in first with 60 out of 126 seats, but came in second as a political party with 26 out of 126. Through the use of surveys, we have come to the conclusion that social media platforms like Twitter and Facebook will continue to expand. The accuracy of SVM algorithms is much higher than that of Naive bayes algorithms (62.1% vs 78.4%), as shown in the preceding text. [23].

Agarwal, Apoorv et al In most of the public locations in Indian cities, women and girls have been subjected to a great deal of violence and harassment nowadays. This is especially true in public places. The primary topic of this piece of writing. The primary function of social media is to increase the level of protection afforded to women in urban areas of India via the use of social media platforms and apps such as Facebook and Twitter. The majority of the significant duties and threats to women's safety that surround them are discussed in this article, as well as how these threats emerged for the goal of protecting women in India. We also focus on twitter data, or tweets, which typically consist of images, written messages, written messages and text.

These tweets are related to the topic of women's safety in Indian cities and can be read as a message in the youth culture of India and used to strike action through educated people. When people are on public transportation for work, they may use Twitter handles with hashtag messaging to communicate with one another and share how they feel about your perspective. These messages are worldwide dispersed via Twitter and how I can sense them in my own body, while being surrounded by others. In addition to that, we made use of machine learning methods, namely the SPF algorithm and the linear algebra factor model approach [24], both of which assist in classifying data into meaningful categories.

F. A. Pozzi et al Analyzing the public's feelings and thoughts on a product or service, whether they come from politicians or celebrities, is what we mean when we talk about doing sentiment analysis. In this investigation, a python-based programming language for Twitter analysis was used to carry out research on the contenders for president of the Republic of Indonesia in 2019. In this work, we employ Naive Bayes techniques for collecting data

from Python libraries. We also make use of text processing for testing and text categorization of data, and we finally categorize the collected information according to different categories or degrees of sentiment.

The conclusion that can be drawn from this study is that the value of 'JoKowi-ma' ruf Amin paire is negative 45.5 percent of the time, while positive emotion accounts for 54.55 percent of the time after which they integrated the data that had been tested as well as the training data that had been utilized for each presidential contender. And get an accuracy of 80.90 percent. For the purpose of this research, a comparison was made between several accuracy values, such as the accuracy value of Naive Bayes, which was 75.58 percent, the accuracy value of SVM, which was 63.69 percent, and the accuracy value of k-NN, which was 73.34 percent. Classify the data using SVM and K-NN, dividing it into positive and negative categories. In conclusion, in the result from our experiment, then compare between methods with better accuracy levels, such as the naive bayes method, which has an accuracy level of 80.90 percent, and another method, k-NN, which has an accuracy level of 75 percent, and SVM, which has an accuracy level of 63.99 percent. In our subsequent work, we may make plans to do an analysis of how satisfied the general public is [25].

K. Arun, et al The majority of individuals in today's society get their information, including news, business updates, communication, scores, and so on, from various social media and network platforms, such as Facebook, Snapchat, Twitter, and so on. Due to the fact that Twitter is a micro blogging website, it has become the most popular channel for attracting the attention of the media. Tweets are short pieces of text that are posted on Twitter. These tweets are used to distribute information across social media platforms, and users have the ability to get direct comments to their tweets from the general public.

Using the cleaning process, remove any unwanted data using the Python language code to retrieve tweets based on specific schemes and the creation of offensive words. After the algorithm has been applied to the data, a negative word is provided as training. Following the cleaning of the data, machine learning classifiers such as naive bayes were used. Specifically, random forest and naive bayes both yield improved accuracy in sentiment prediction,

incorporating a few parameters for boosting the accuracy of the prediction. Determine the accuracy of the prediction rate for people's views by employing the numerous models about the public [26].

A Pak and P.paroubek et al The number of people who use the internet today is constantly growing, and data is being produced on a wide variety of sites. When this occurs, it becomes vital to know the emotion of the people and to assess data provided by other military or government institutions. Through the assistance of organizations, individuals may take control of their actions and select which steps should be taken to accomplish them. In this day and age of micro blogging sites, where millions of users communicate with one another and share their millions of views, opinions, and various day-to-day life issues on social platforms such as Twitter, Facebook, and Tumblr, data can be efficiently extracted from these sites for use in marketing or social media.

In order to do sentiment analysis, we have used a variety of machine learning methods. Using a variety of machine learning classifiers, polarity-based sentiment analysis was performed, and the results were used to categories tweets as having either a positive or negative sentiment. It is possible to further deploy these classification models in order to categories tweets and any subject that is discussed on Twitter. When compared to the machine learning algorithm, the use of various deep learning methods, such as RNN, CNN, and LSTM, results in higher accuracy. After training on multiple data sets across a variety of domains, the data were tested and found to have a high percentage of correct answers. In work that will be done in the future utilizing this model, it will be possible to create a percentage accuracy that is as good as human accuracy [27].

S. Y. Yoo, J. I. Song et al The largest number of people use Twitter compared to any other social media network. Today, we are able to see that the number of people using Twitter has reached 330 million people all over the world. Twitter users often utilise the platform to express their opinions in the form of tweets. Identifying the viewpoints or categories expressed in tweets. As either the book itself or the topic of this investigation of feelings. Techniques from natural language processing (NLP) should be used at the beginning of the text. The elimination of stop words is one of the methods that

can be utilized in the analysis of sentiment. The development of sentiment analysis is going to be the primary emphasis of this study, which will make use of lexicons and multiplication polarity. As a consequence, lexicon-based approaches provide less accurate results due to the unfinished nature of a lot of their aims. In further work for these sentiment analyses, the polarity of the opinion will be identified. Because interpretations of lexical methods, such as uni-grams, B-grams, and tagger posts, are evaluated according to rules of sentiment [28].

Table 1. Literature review in details.

Reference No.	Author name, publication, and Year	Title of Paper	Methodology	Data set Description	Result	Future Work
[12]	Levy, M at all., IEEE, 2019	Sentiment analysis	Maximum entropy, naïve bayes, SVM	7000 tweets	78% 45% 60%	Improve accuracy and increase data set.
[11]	V.Lakshami, K.Harika at all IEEE. 2018	Sentiment analysis	Decision tree Ado boosted D-tree, SVM	1,600,000 1000 posi& neg tweets	84% 67% 82%	Improve accuracy of classifier.
[10]	C.N dos santos at all.IEEE.2019	Sentiment Analysis	Word embedded model, CNN	300.000 tweets	0.65 % accuracy	Improve accuracy performance of different data set.
[9]	Levy, M at all., IEEE, 2016.	Sentiment analysis	SVM	7000twittes	67% accuracy	Improve accuracy and increase data set.
[17]	Seyedal mirjalili at all IEEE 2020	Sentiment analysis	NB, SVM, Maximum entropy	Tweets review like Hotel .film product using twitter API.	86% 74.6% 82.6%	Improve the performance measure .
[16]	B.o'connor, R . bala subramanyan at all IEEE	Sentiment analysis	SoftMax, SVR, DTs, RF	10,000 twitter post, 7000 training 3000 test	67.2% 81.95% 91.8% 83.2%	Improve our approaches using bigrams ,trigrams
[15]	Harpreet kaur at all IEEE. 2020	Text classification	SVM, RNN, LSTM	Stand for d university, crowdfower, 13871 samples library, 8493 negative, 3142 neutral, 2236 positive	77.98% 82%	Improve accuracy ration using RNN .
[14]	Siddharth,Darshini at all IEEE 2020	Sentiment analysis	NB, SVM, RF, LR	#Metoo50429 #Women50430	85.08% 95.62% 93.07% 93.72%	Improve accuracy of model and same use for product review etc.
[13]	W.Medhat,A.Hassan at all IEEE 2019	Sentiment analysis	Naïve bayes Algorithm	tweets	70%	improve the accuracy of prediction.

[21]	T.Joachims at all IEEE . 2018	[20]	A.Darwish and K.L.lakhtaria at all IEEE 2018	[19]	Nikhil lekkysetty at all IEEE 2019	[18]	Bopping L.Lee at all IEEE 2016
Sentiment classification	Sentiment analysis	Sentiment analysis	Sentiment analysis	Sentiment analysis	Sentiment analysis	Twitter analysis	
SVM classifiers, (unigram, bigram, trigram), Weighting schemes, Tf, Tf-idf, Binary, OPAL, VCU-TSA, Proposed approach	Random forest, Navie bayes DTs, RNN-LSTM, NN, CNN	Random forest, Navie bayes DTs, RNN-LSTM, NN, CNN	Random forest, Navie bayes DTs, RNN-LSTM, NN, CNN	SVM, Naïve bayes	SVM, Naïve bayes	SVM, Naïve bayes Maximum entropy	
Tweets classification Sem Eval2016 data set	1,048,000 tweets from GitHub and 140 characters for each tweet.	1,048,000 tweets from GitHub and 140 characters for each tweet.	1,048,000 tweets from GitHub and 140 characters for each tweet.	Political tweets on congress and BJP Sentiment 140 dataset.	Political tweets on congress and BJP Sentiment 140 dataset.	17000 tweets from teitter API (140 word) limit 15000 training 2000test	
79.1% 80% 81%	73.8% 77.5% 72.5% 82.3%79.5% 79.6%	73.8% 77.5% 72.5% 82.3%79.5% 79.6%	73.8% 77.5% 72.5% 82.3%79.5% 79.6%	77.33% 67.77%	77.33% 67.77%	84% 70% 74%	
Improve n – gram based classifiers for improve performance .	Including classification techniques for Arabic language.	Including classification techniques for Arabic language.	Including classification techniques for Arabic language.	Improve accuracy of largest data set sentiment 140	Improve accuracy of largest data set sentiment 140	Automatic sentiment for multilingual message .	

[26]	[25]	[24]	[23]	[22]
K.Arun ,A. Srinagesh at all IEEE 2019	F. A. Pozzi at all IEEE 2018	Agarwal ,Apoorv at all IEEE 2019	O .Almatrafi at all IEEE 2016	Surnar ,Avnish at all IEEE .2018
Sentiment analysis	Sentiment analysis	Sentiment analysis	Sentiment analysis	Sentiment analysis
Naïve bayes, Namely, Random forest	Naïve bayes algorithm, SVM ,text mining, KNN	SPC algorithm, SVM	Naïve bayes, SVM, Dictionary based approach	Naïve bayes, SVM, Maximum entropy
Twitter data #gst #demonetization #my clean india	Data tweets from twitter token	Tweets from twitter API #Me Too India #MeToo,#Refussed, #Rape	Political tweets #BJP #Congress #BSP #AAP #NCP	Twitter data customer review, posts data set.
72.4% 80.1% 84.78%	80.9% 75.5% 63.99%		62.1% 78.4% 34%	89% 90% 90%
Implementation of algorithms for sentiment analysis.	We plan to analyze sentiment of public satisfaction for social media A facebook,instagram.	Using machine learning techniques can be analysis huge amount twitter data	Can be predictions in using other machine learning algorithm. such Regression etc.	Improve accuracy using hybrid approach.

[27]	A.pak and P. paroubek at all IEEE 2020	Sentiment analysis	Naïve bayes, MNB,SVC LSTM ,RNN, CNN	Twitter API Kaggle	97% 95%	Improve accuracy and performance u using deep learning algo.
[28]	S. Y. Yoo at all IEEE 2019	Sentiment analysis	Linear SVM Naïve bayes NLP, Lexicon based	Twitter API 8000 data	89% 79%	Determine of f sentiment analysis polarity.

III. CONCLUSION

In this study work of our research opinion-based sentiment analysis of tweets from twitter in the term of machine learning techniques Such as SVM, Naïve Bayes, Maximum entropy, DTs, Random Forest. Our proposed work will extract data from twitter API on twitter and. Main focus of our work enhances the accuracy of the machine learning classifiers of classification of tweets. We have both supervised and unsupervised machine learning used in proposed methodology.

The twitter analysis of twitter data is possible by various aspects/parameters of data sets to mine the sentiment. For this result we conduct our experiment using machine learning algorithms. The proposed model involves a supervised machine learning algorithm. After extracted of data into fed datausing supervised model for training and testing purpose then classification of tweets is done using Support Vector Machine **SVM** Classifier for highest accuracy. Show that Result of our proposed work better in the term of SVM classifiers.

For future work use of the same methodology can be sentiment analysis in various fields like prediction of stock market and to analyze people's opinions

regarding corona virus issues that are current topics globally for health issues.

REFERENCES

- [1] El Rahman, S. A., AIOtaibi, F. A., & AIShehri, W. A. (2019). Sentiment Analysis of Twitter Data. 2019 International Conference on Computer and Information Sciences (ICCIS).
- [2] Sahar A. El_Rahman, Feddah Alhumaidi AIOtaibi and Wejdan Abdullah AIShehri, Sentiment Analysis of Twitter Data, 978-1-5386-8125-1/19/\$31.00 ©2019 IEEE
- [3] Rekha V, Raksha R, Pradnya Patil, Swaras N and Rajat GL, Sentiment Analysis on Indian Government Schemes Using Twitter data, 978-1-5386-9319-3/19/\$31.00 ©2019 IEEE
- [4] Sonia Saini, Ritu Punhani, Ruchika Bathla and Vinod Kumar Shukla, 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University, Sentiment Analysis on Twitter Data using R Nann Hwan Khun and Hninn Aye Thant, Visualization of Twitter Sentiment during the Period of US Banned Huawei,
- [5] Sani Kaniş and Dionysis Goularas, 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data.
- [6] Lei Wang, Jianwei Niu, and Shui Yu, SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis. JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2018
- [7] Alaa S. Al Shammari Real-time Twitter Sentiment Analysis using 3-way classifier, 978-1-5386-4110-1/18/\$31.00 ©2018 IEEE
- [8] Santhosh Kumar K L and Jayanti Desai, Jharna Majumdar, "Opinion Mining and Sentiment Analysis on Online Customer Review" In 2016 IEEE International Intelligence and computer Research.
- [9] L. L. Bo Pang, «Opinion Mining and Sentiment Analysis Bo», Found. Trends® Inf. Retr., vol. 1, no. 2, pp. 91–231, 2008.
- [10] V. Lakshmi, K. Harika, H. Bavishya, Ch. Sri Harsha, "SENTIMENT ANALYSIS OF TWITTER DATA," vol.04, February 2017. Link- "https://www.irje.t.net/archives/V4/i3/IRJET-V4I3581.pdf"

- [11] Levy, M (2016) .Playing with twitter data .[Blog] R-bloggers. Available at: [http://www.r-bloggers.com/playing-with-twitter data/\[Access\]](http://www.r-bloggers.com/playing-with-twitter-data/).
- [12] W. Medhat, A hussan, and H. Korashy. "Sentiment analysis algorithm and application: A survey," *Ain shams Engineering journal*, vol. 5 no.4, pp.1093-1113, 2014.
- [13] Sidharth, darsini, and sujithra, " Sentiment Analysis on youtube & Twitter Data Using Machine." *Int. j. Res.Appl. Sci. Eng.Technol.*, vol. 8 no.5pp. 755-758,2020.
- [14] Harpreet Kaur, Veenu Mangat, Nidhi, "A survey of sentiment analysis techniques", February 2017, 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).
- [15] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series." *Icwsn*, vol. 11, no. 122-129, pp. 1-2, 2010.
- [16] Evolutionary Machine Learning Techniques: Algorithms and Applications (Algorithms for Intelligent Systems) by Seyedali Mirjalili, Hossam Faris, Ibrahim Aljarah.
- [17] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79- 86, 2002.
- [18] Nikhil Lakkisetty, Phani Deep, Balamurugan J, "Social media and its impacts on politics", *International Journal of Advance Research, Ideas and Innovations in Technology*, Volume 4, 2018.
- [19] A. Darwish and K. I. Lakhtaria, "The Impact of the New Web 2.0 Technologies in Communication, Development, and Revolutions of Societies," *Journal of Advances in Information Technology*, vol. 2, no. 4, pp. 204-216, Nov. 2011.
- [20] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features." In *European conference on machine learning*, pp. 137-142. Springer, Berlin, Heidelberg, 1998.
- [21] Surnar, Avinash, and Sunil Sonawane. "Review for Twitter Sentiment Analysis Using Various Methods."
- [22] O. Almatrafi, S. Parack, and B. Chavan, "Application of Location-Based Sentiment Analysis Using Twitter for Identifying Trends Towards Indian General Elections 2014," *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, Article No. 41, Jan. 2015.
- [23] Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics*, 2009.
- [24] F. A. Pozzi, E. Fersini, E. Messina and B. Liu, in *Sentiment Analysis in Social Network*, United States, Todd Green, 2017, p. 228.
- [25] K. Arun, A. Srinagesh, M. Ramesh, "Twitter Sentiment Analysis on Demonetization tweets in India Using R language", *International Journal of Computer Engineering in Research Trends (IJCERT)*, Volume 4, N o.6, pp. 252-258, June-2017.
- [26] A Pak and P.paroubek, "twitter as corpus for sentiment analysis and opinion mining," *LREc*. Vol.10. No.2010, 2010.
- [27] S. Y. Yoo, J. I. Song, and O. R. Jeong, "Social media contents-based sentiment analysis and prediction system," *Expert Syst. Appl.*, vol. 105, pp. 102-111, 2018.
- [28] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, "Stock market predication using a linear regression," in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212716.
- [29] V. Prakaulya, R. Sharma, U. Singh, and R. Itare, "Railway passenger forecasting using time series decomposition model," in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212725.
- [30] R. Yadav, A. Choorasiya, U. Singh, P. Khare, and P. Pahade, "A Recommendation System for E Commerce Base on Client Profile," in *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018*, 2018, doi: 10.1109/ICOEI.2018.8553930.
- [31] V. S. Tomar, N. Gupta, and U. Singh, "Expressions recognition based on human face," in *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 2019, doi: 10.1109/ICCMC.2019.8819714
- [32] R. Verma, P. Choure, and U. Singh, "Neural networks through stock market data prediction,"

- in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212717
- [33] P. Kewat, R. Sharma, U. Singh, and R. Itare, "Support vector machines through financial time series forecasting," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212859.
- [34] A. Sharma, D. Bhuriya, and U. Singh, "Survey of stock market prediction using machine learning approach," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212715.
- [35] S. Sable, A. Porwal, and U. Singh, "Stock price prediction using genetic algorithms and evolution strategies," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212724.
- [36] A. Roshan, A. Vyas, and U. Singh, "Credit Card Fraud Detection Using Choice Tree Technology," in Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, 2018, doi: 10.1109/ICECA.2018.8474734.
- [37] H. Soni, A. Vyas, and U. Singh, "Identify Rare Disease Patients from Electronic Health Records through Machine Learning Approach," in Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, 2018, doi: 10.1109/ICIRCA.2018.8597203.
- [38] A. Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.
- [39] A. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.
- [40] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8
- [41] Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.
- [42] R. Baghel, P. Pahadiya and U. Singh, "Human Face Mask Identification using Deep Learning with OpenCV Techniques," 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1051-1057, doi: 10.1109/ICCES54183.2022.9835884.
- [43] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha and U. Singh, "Cancer Prediction Using Random Forest and Deep Learning Techniques," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 227-231, doi: 10.1109/CSNT54456.2022.9787608.
- [44] Singh, Upendra, Gupta, Puja, and Shukla, Mukul. 'Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM'. 1 Jan. 2022: 6505 – 6520.
- [45] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8
- [46] Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology using synchronized clock zone scheme. J Super comput (2022). <https://doi.org/10.1007/s11227-022-05012>.
- [47] Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.
- [48] Jindal, R, Kumar, N, Patidar, S. IoT streamed data handling model using delta encoding. Int J Commun Syst. 2022; 35(13):e5243. doi:10.1002/dac.5243
- [49] S. Patidar, A. Jain and A. Gupta, "Comparative Analysis of Machine Learning Algorithms for

- Heart Disease Predictions," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022, pp. 1340-1344, doi: 10.1109/ICICCS53718.2022.9788408.
- [50] B. Jain, S. Patidar and D. Sudershan, "Heterogeneous Software Defect Prediction using Generative Models," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 367-372, doi: 10.1109/CSNT54456.2022.9787607.
- [51] Singh, Shani Pratap and Shukla, Jayesh and Sharma, Shaili and Daga, khushhal and Bhalavi, Brahman Singh and Singh, Upendra, Face Mask Detection using Multi-Stage CNN Architecture (July 10, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems - ICICNIS 2021, Available at SSRN: <https://ssrn.com/abstract=3884022> or <http://dx.doi.org/10.2139/ssrn.3884022>
- [52] Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.
- [53] Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.
- [54] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8
- [55] Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology using synchronized clock zone scheme. J Supercomput (2022). <https://doi.org/10.1007/s11227-022-05012-2>
- [56] Singh, Upendra, Gupta, Puja, and Shukla, Mukul. 'Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM'. 1 Jan. 2022: 6505 – 6520.
- [57] U. Singh and L. S. Songare, "Analysis and Detection of Monkeypox using the GoogLeNet Model," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1000-1008, doi: 10.1109/ICACRS55517.2022.10029125.