International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

Emotional Tendency Analysis of Twitter Data Streams

Gunasundari Ranganathan[1], Clara Barathi Priyadharshini Ganesan[2], Balakumar Chellamuthu[3]
[1]Professor, [2,3]Assistant Professor ,
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India

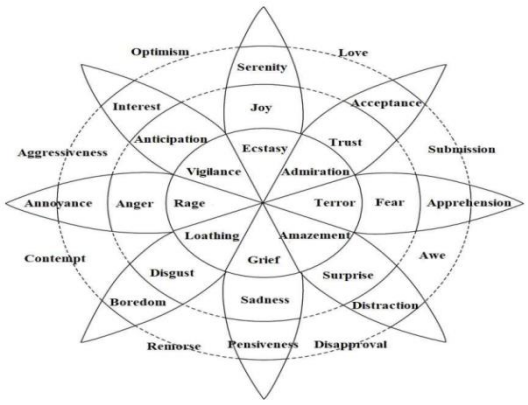
Abstract—The web now seems to be an alive and dynamic arena in which billions of people across the globe connect, share, publish, and engage in a broad range of everyday activities. Using social media, individuals may connect and communicate with each other at any time and from any location. More than 500 million individuals across the globe post their thoughts and opinions on the internet every day. There is a huge amount of information created from a variety of social media platforms in a variety of formats and languages throughout the globe. Individuals define emotions as powerful feelings directed toward something or someone as a result of internal or external events that have a personal meaning. Emotional recognition in text has several applications in human-computer interface and natural language processing (NLP). Emotion classification has previously been studied using bag-of words classifiers or deep learning methods on static Twitter data. For real-time textual emotion identification, the proposed model combines a mix of keyword-based and learning-based models, as well as a real-time Emotional Tendency Analysis

Keywords: Natural language processing, Emotional tendency analysis

I INTRODUCTION

In today's internet, which has become a primary platform for individuals to communicate their views and sentiments? Millions of individuals use social media to share, debate, publish, and comment on the world's events, news, and activities. For example, "I'm really saddened by the chemical assaults in Syria! :(" or "I had a wonderful meal at Copper Chimney!" There was a gasp of terror throughout the room. It is possible to analyze how individuals respond to diverse circumstances and events by capturing their feelings in text, particularly those uploaded or distributed on social media. People's sentiments and views about a company's goods may be tracked by business analysts using this data. As a general rule, the majority of Sentiment Analysis is done today fails to explain the actual sentiments of consumers and the strength of their response, which is a major issue. Using the proposed emotional analysis method, the researcher has been able to get a deeper understanding of their markets, which has resulted in increased profits.

Darwin carried out the first study of emotions via the analysis of facial expressions and body language of humans .and animals[5]. Ekman P (1992) defined six basic emotion categories: anger, disgust, fear, joy, sadness, and surprise [1]. This model is one of the most popular ones for building emotion models based on collective information from previous studies. Explained in fig 1 sample list of emotions. Based on that, a Profile-of-mood-states (POMS) was developed which is basically an instrument that defines a six-dimensional emotional-state representation; each dimension representing one of the six basic emotions. This instrument can be extended for the detection of more emotional categories.



This approach defines a twelve-dimensional POMS that makes use of a combination of basic as well as supplementary emotion categories such as anger, trust, anticipation, disgust, fear, joy, depression, surprise, fatigue, vigour, tension and confusion. Each dimension represents one of the formerly mentioned twelve emotion categories. This approach has included the supplementary emotions of depression, fatigue, tension and confusion as these are crucial emotions for the assessment of the mental state of any individual.

II REVIEW OF LITERATURE

Rodrguez et al. (2016) developed a fuzzy logic technique to follow the suggestion on Twitter. This technique approaches suggestions as if it were a link prediction issue. It employs three types of likeness between two users: similarity of tweets, followed ids, and followee tweets. These

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

commonalities are calculated during the extraction of user profiles. Tran et al. (2018) introduced the hash tag recommendation approach, which greatly enhances the performance of hash tag recommendation systems based on research of tweet content, user attributes, and presently popular Twitter hash tags.

Corcoglioniti et al. (2018) proposed a recommendation system that uses features predicted from information using machine learning algorithms, ranging from basic SM traits to particular domain-relevant user profile characteristics. Jiménez-Bravo et al. (2019) have created a recommendation system that presents other users with expert profiles. Experts must be able to deliver fascinating facts to consumers. The expert recommended to a user will be chosen based on the information they post and if it is of interest to the user.

Zappavigna (2014) investigated ambient affiliation to determine how a user of the micro blogging service Twitter executes relational identities while enacting discourse fellowships. Faulkner et al. (2014) introduced a novel method to document-level stance classification that employs two feature sets to capture the linguistic characteristics of the stance-taking language. They also created a corpus of annotated student writings for stance at the essay level to examine the usefulness of features based on linguistic research involving argumentative language. That corpus serves as a more typical sample of argumentative language than the chaotic online discussion material often employed in attitude categorization research.

Sobhani et al. (2015) developed a new paradigm for labeling arguments. News comments were grouped based on the subjects retrieved by NMF. Following that, the top keywords for each cluster were used to name the clusters. Hercig et al. (2015) created a technique for determining a viewpoint in online debates. A fresh database of Czech news comments was created, and a maximum entropy classifier, a support vector machine classifier, and a convolutional neural network were tested. Cao et al. (2017) used numerical ratings and textual information from many mobile platforms to build an app recommendation matrix factorization-based latent factor model. Rating matrices from these various platforms have been factorized, and the textual material has been included in the topic models.

III EMOTIONAL TENDENCY ANALYSIS (ETA)

Recognizing user emotions in social media postings is a difficult challenge when analysing posts from various platforms on the web. Text is one of the most prevalent ways people convey their feelings, especially on social media. Emotional recognition in text has several applications in human-computer interface (HCI) and natural language processing (NLP), since it is essential in human contact (NLP). Emotion classification has previously been studied using bag-of-words classifiers or deep learning on static Twitter data. For real-time textual emotion identification, this model combines a mix of keyword-based and learning-based models, as well as a real-time ETA. Natural Language Processing (NLP) approaches such as PoS tagging and topic modelling, together with random forest classification, are used to extract textual features. According to the results, this new suggested model outperforms the classic Unison model by an average of 99.39%.

Emotion Recognition Approaches

Emotion recognition is mostly done on two levels: low-level/coarse-grained level analysis or high-level/fine-grained level analysis. The low-level analysis encompasses a binary classification of the text into positive or negative which is also called as Sentiment Analysis. The high-level analysis is basically a higher form of sentiment analysis which is nothing but a further classification into crisp emotional categories. Basically, there are four different approaches to emotion recognition: machine learning based, keyword-based, deep learning based and hybrid/combination approach as explained in Table.3.1.

Table 3.1: Emotion Recognition approaches

Approach	Characteristics
Keyword based	• Traditional and Easy • Emotions recognition via predefined emotion words (keywords) which are detected via some rules and vocabularies.
Machine Learning based	• Builds classification model using ML algorithms • Training with large emotion data and emotion recognition of new incoming data.
Deep Learning based	• Emotions recognition via deep learning models, no explicit feature extraction required and needs large amount of training data
Hybrid/Combination	• Uses a combination of the above approaches

Emotion recognition is undoubtedly an important research aspect in the field of text mining. Timely detection of the stress- state or suicidal tendency of a person on the basis of emotion recognition is one of the examples why the research is an indispensable necessity. Also, the public mood of a current topic can be inferred effectively using real-time data such as electoral tweets, trending topics, etc. Also, the analysis on real-time data is crucial as it proves quite beneficial for deployment in real-time applications.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

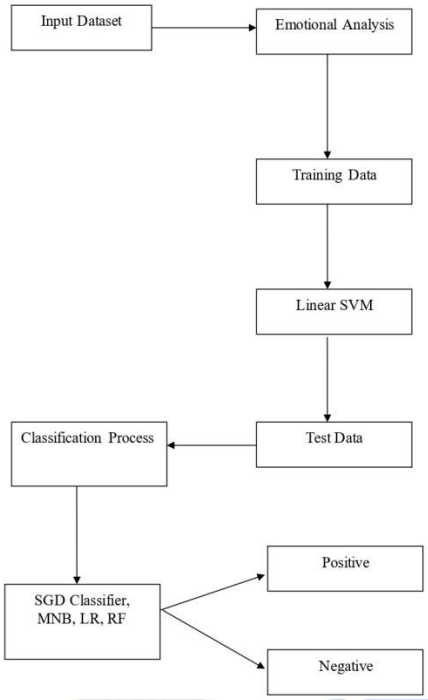


Figure 3.1: ETA Architecture

Emotional Tendency Analysis

The ETA Model performs emotional analysis of the customer reviews. The ETA model performs sentiment emotion mining based on LDA topic model by explicitly capturing the sentiments of the specific entities. The model focuses on sentiment analysis depending on topics and word emotion lexicon consisting of eight emotion labels as stated by plutchik's theory of emotions. The theory of emotions is based on the eight emotion labels. As illustrated in table 5.1, the initial step involves in the collection of reviews from social media followed by pre-processing of the reviews. An emotion corpus has been created based on the PoS tagging feature extracted from the processed dataset. Sentiment emotion classification has been performed that results in output of sentiment distribution of each review into the labels of emotion. The emotion labels defined under positive are "surprise", "anticipation", "joy", "trust", and emotion labels defining negative sense are "fear", "angry", "sadness" and "disgust". Finally, LDA topic model was applied on the emotive words that results in the topic distribution over sentiment reflected emotion words. And the word distribution over topics clearly indicates the way people express their emotion towards certain entity.

The first level of the process generates a sentiment label for the words from the sentiment distribution specific to the document. The second level generates a latent topic from the sentiment specific distribution and lists the topics with their word distribution. Plutchik's eight-dimension sentiment lexicon has been used to determine the emotional tendencies of words in social media to increase the accuracy of the sentimental classification.

Twitter is an extraordinary social network for text analysis, sentiment evaluation, and social-web evaluation. Various kinds of software languages deliver various techniques to acquire information programmatically from Twitter. Towards extracting tweets from Twitter, a Twitter application is created by the registered user to communicate with the Twitter API; Twitter developer site for data extraction. It is claimed that the relevant application will provide authentication by enabling the required consumer key and secret, access token, and access token secret of the system. The key information can be included in the program and get attached in extracting the data from Twitter.

Pre processing for Data Cleaning

In order to begin pre-processing data, one must first obtain the needed data and then clean it to extract the desired characteristics. This method uses the Twitter API to get big publicly accessible Twitter datasets.

Importance of Pre processing Techniques

Expressing the difficulty involved in grasping public opinions about specific events due to the associated gross size and diversity features of social-media data, the necessity of an automatic and real-time sentiment analysis system is pointed out. Relevantly, obtaining online sentiments and opinions from the social-media data for sentiment analyses, the underlying text classification tasks are regarded as challenging as a result of unstructured social media text data (tweets) being noisy and infested with inappropriate and redundant information.

Table 3.2: Tweets creations

Date	Tweet created date
Favorite Count	Number of times the tweet has been favorited
favorited	Whether the tweet is favorited or not
Tweetsid	The unique identifier of the tweet
IsRetweet	Whether the tweet is retweeted or not
latitude	Geographical coordinate latitude of tweet location
longitude	Geographical coordinate longitude of tweet location
RetweetCount	Number of times the tweet has been retweeted
screenname	The user's name that is displayed on the tweet
text	The actual UTF-8 text of the Tweet

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

Sentiment Analysis

It is a process of classifying a given text as expressing a positive, negative or neutral sentiment. For that purpose, this method is using the SentiWordNet dictionary in which each word is given a score according to positive or negative information displayed. Using that scores, a tweet is probabilistically classified as positive, negative or neutral. Thus, this method achieves low/coarse-grained level analysis using above approach.

Feature Extraction Using NLP

In this step, repetitive tweets' content, that occur mostly due to retweets by other users to a particular user's tweet, are removed using n-gram generation by using tweet-id. For extracting textual features to train this classification model, this method used two NLP techniques viz. PoS-Tagging and Topic Modelling. Part-Of-Speech tagging is also called as grammatical tagging in which the words of a particular text are tagged with the corresponding Part-of-Speech. This process is very helpful for identifying the most useful features which in turn are useful for identifying the emotions.

Classifying Emotion Labels

Parts of speech tagging is a technique of partitioning the sentences into words and mapping an appropriate tag such as noun, verbs, adverbs, adjectives etc. Each word is on the basis of the PoS tagging feature.

Table 3.3 Emotions Scores For Different Pos Tag Features

PoS tags	Nouns	Adjectives	Adverbs	All words
Anger	0.04	0.2	0.3	0.26
Anticipation	0.02	0.45	0.2	0.1
Joy	0.03	0.3	0.3	0.1
Trust	0.3	0.35	0.15	0.2
Fear	0.015	0.4	0.2	0.17
Surprise	0.015	0.1	0.015	0.015
Sadness	0.1	0.35	0.25	0.2
Disgust	0.015	0.1	0.015	0.017

The emotion lexicon provides emotion score for the individual words on the emotion of "anger", "anticipation", "joy", "trust", "fear", "surprise", "sadness" and "disgust". Based on the National Research Council (NRC) word emotion lexicon, the emotion score for the sample reviews in the experimental dataset on PoS tagging feature is tabulated in Table 3.3. The outputs of the parts of speech include nouns, adjectives, adverbs as well as the combination of all words. The emotion ratings for the input text file is shown along with the scores captured by nouns, adjectives, adverbs and combination of it. The emotion scores of nouns, adjectives, adverb and its combination for each of the eight emotion labels. The pictorial representation clearly states that adjectives and adverbs majorly contribute to the emotion of the customer.

Table 3.4 Positive Sentiment Based Emotion Topics

Rooms	Variety	Stay	Comfort
worth	flavourful	awe	influence
accept	crispy	wonder	rule
certain	delicious	delight	loved
charge	fresh	never	amazing
believe	tasty	hope	capture
outlook	buffets	expect	sudden
charge	foretaste	duty	worthy

The positive sentiment based emotion topics are tabulated in Table 3.4. The emotion based positive topics are "Rooms", "variety", "stay" and "comfort". A closer observation of the words denotes the emotional words of the positive emotion labels "joy", "Trust", "surprise" and "anticipation". Negative sentiment based emotion topics. are listed in Table 3.5.

Table 3.5 Negative sentiment based emotion topics

Food	Cost	Location	Service
burnup	worst	bad	disappoint
dry	annoy	worse	feel
overcooked	terrible	never	horror
brunt	upset	hassle	terror
salty	afraid	old	panic
sucks	die	neveragain	doubt
old	overpriced	regret	worry

The emotion based negative topics are "Food", "cost", "location" and "service". A deep look at the word list indicate that the emotional words belong to the negative emotion labels "anger", "fear", "sadness" and "disgust". The analysis of the words under the positive and negative sentiment based emotion topics shows that ETA discovers coherent topics with high quality. Particularly the more coherent emotion. based topics are captured and also the results are highly interpretable.

Table 3.6: Adjective Emotion Value Vector

Word	Happiness	Anger	Sadness	Fear	Disgust
absurd	1.28	1.95	1.34	1.32	1.64
accidental	1.06	2.82	3.61	3.69	1.88
addictive	1.17	2.57	3.19	2.77	3.04

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

aggressive	1.75	3.23	1.89	2.73	1.95
alcoholic	1.41	304	3.64	3.11	2.71
alone	1.27	1.84	3.65	3.30	1.56
angry	1.17	4.47	2.83	2.83	2.09
athletic	4.05	1.91	1.77	1.56	1.36
bastard	1.19	3.73	2.57	1.85	2.59
beastly	1.24	2.06	1.65	3.14	1.97
beautiful	4.51	1.22	1.32	1.31	1.18
blissful	4.6	1.11	1.08	1.14	1.05
blind	1.18	2.37	3.44	3.07	1.35
blond	2.88	1.36	1.25	1.18	1.40
broken	1.14	2.89	3.23	2.41	1.70
brutal	1.16	3.65	2.99	3.28	2.86
capable	3.42	1.31	1.35	1.56	1.26
champion	4.4	1.25	1.28	1.22	1.21
cold	1.47	2.21	2.23	1.88	1.48
confused	1.17	2.32	2.28	2.10	1.43
controlling	1.31	2.82	1.99	2.19	2.11
grateful	4.06	1.13	1.18	1.01	1.06
happy	4.68	1.08	1.17	1.11	1.10
hard	1.65	2.22	1.75	2.21	1.40

Table 3.7 : Verb & Adverb Strength [1]

Verb	Strength	Adverb	Strength
Love	1	complete	+1
adore	0.9	most	0.9
like	0.8	totally	0.8
enjoy	0.7	extremely	0.7
smile	0.6	too	0.6
impress	0.5	very	0.4
attract	0.4	pretty	0.3
excite	0.3	more	0.2
relax	0.2	much	0.1
reject	-0.2	any	-0.2
disgust	-0.3	quite	-0.3
suffer	-0.4	little	-0.4
dislike	-0.7	less	-0.6
detest	-0.8	not	-0.8
suck	-0.9	never	-0.9
hate	-1	hardly	-1

Emotion Classification

In this step, two classification models are built and trained with emotion-word features obtained from [6]. The first is the Unison model which is basically the traditional emotion classification model based on Bag-of-Words model. POMS is used for its implementation. The second model is this proposed Random Forest (RF) model. It is built by training it with emotion-labelled tweets. The random forest algorithm is used for solving Quadratic Programming (QP) problems and hence it is chosen to build a multiclass classification model for emotion detection.

Table 3.8: Emotion values

Tweet	Happiness	Anger	Sadness	Fear	Disgust
She is not very beautiful	0.098	0.756	0.736	0.738	0.764
My heart taken by u; broken by u and now its shattered #bitch#heartbreaker	0.587	0.800	0.930	0.361	0.556
Each day is a bless #life. Being happy is not a destination, it's a journey. #life	0.879	0.213	0.223	0.213	0.215
Just watched Trisha's abduction; scared#24	0.223	0.595	0.645	0.820	0.530
Alone in d world. #simpleplan #lost	0.420	0.344	0.548	0.506	0.310

Algorithm 3.1 ETA

Input: Dataset contains emotional details
Output: Emotional states Y;
Step 1: Initialize the emotional states Y;
Step 2: output←Y;
Step 3: max←€
Step 4: repeat
Step 5: For each dataset training using Linear SVM
Step 6: Y←q(Y′, Y)
Step 7: End for
Step 8: Test Data and Classification using SGD Classifier, MNB, LR, RF
Step 9: If (Emotion Y′|Y) // positive or negative or neutral
Step 10: Y ←output
Step 11 End if

IV RESULTS AND DISCUSSIONS

The Twitter API is used to obtain real-time tweets for analysis by developers via the Twitter4J library. The retrieved tweets are preprocessed in suitable format. The SentiWordNet dictionary is used for sentiment analysis of the tweets. NLP algorithms are then used for extracting informative features. The Unison model is the previously developed model for emotion recognition based on POMS [6]. Proposed model for emotion recognition is developed as explained in section 5.3. The classification is performed using both the classifiers. The labels of real-time test data are assigned according to the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

probability of occurrence of emotion words predefined in which later helps in performance evaluation of the classifiers. Their performances are expressed in terms of performance measures. The metrics of classification performance evaluation used are accuracy, precision, recall and f1-score.

$$Accuracy = \frac{\text{Correctly Predicted Observation}}{\text{Total number of Observation}} \text{----------(eq-5.1)}$$

$$Precision = \frac{\text{Correctly Predicted Positive Observation}}{\text{Total Predicted Positive Observation}} \text{----------(eq 5.2)}$$

$$Recall = \frac{\text{Correctly Predicted Positive Observation}}{\text{Total Positive Observation}} \text{----------(eq 5.3)}$$

$$f1 - score = \frac{2}{1/\text{Precision} + 1/\text{Recall}} \text{----------(eq 5.4)}$$

The classifier's accuracy is defined as the ratio of properly predicted data samples to the total number of data samples, as shown in Equation (5.1).
According to Eq.(5.2), precision is the percentage of projected positive cases that are really positive.

As indicated in Eq. 1, the recall measures the proportion of accurately anticipated positive observations to the total number of positive observations (5.3).

For the sake of completeness, Eq. (5.4) defines f1-score as the harmonic mean of accuracy and recall. In order to assess the performance of the two classification models, all of these metrics are employed. Table.3 shows the specifics of these measurements. The average of 10 executions is used since the data is real-time. For the testing set, ETA used seed words to extract tweets, and then used the first approach to filter and categorise the emotional tweets extracted from these seed terms. In all, there are 900 tweets in this collection, with around 150 tweets each Emotion-Category to guarantee consistency. Aside from these two points, the testing set is assured to include tweets that are unique from those found in training.

IMPLEMENTATION SETUP

This study utilized the Python programming language and the spider environment. Python is an object-oriented, high-level programming language with interpreted dynamic semantics. Python's straightforward, easy-to-learn syntax places a premium on readability, which lowers the cost of program maintenance. Python allows modules and packages, promoting the modularity and reuse of code in programs.
Python version used: 3.7
IDE: Anaconda (Spider)
Datasets: Kaggle.com

DATASETS

Twitter Datasets are collected from the kaggle.com website and Twitter API also. Opinion classification enables the polarity score of the Tweet from the Kaggle website.

Tweets	Positive Words Count(A)	Negative Words Count(B)	Score =A-B	Opinion
He is Suresh Tanna, A small businessman.has faced great slowdown after note ban & He has faith in GST	2 (Great, Faith)	1 (Slowdown)	1	Positive
6 years of ADMK Govt, capped by a disastrous GST implementation have seriously damaged TN's Industrial Growth.	0	2 (disastrous, damaged)	-2	Negative
Gujarat elections: Saurashtra traders miffed after GST nixes 'kutcha' transactions.	0	0	0	Neutral

Table4.1SampleTweetswithScoreandOpinion

SUMMARIZATION OF DOCUMENTS

The summarizer of the opinion analyzer counts and clusters the documents into three groupspositive,negative,andneutral.Table4.2 shows the total number of summarizations of day-wise tweets grouped into positive, negativeor neutral.

Table4.2 Summarization of Day-Wise Results

Creation Date	Negative	Neutral	Positive	Total
12/10/2021	167	410	130	707
13/10/2021	184	532	201	917
14/10/2021	88	275	108	471

EXPERIMENTAL RESULTS

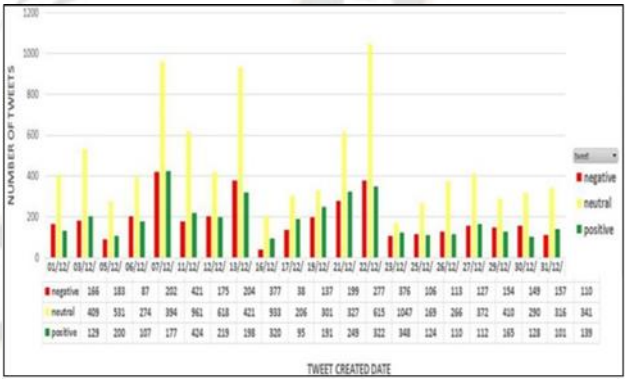


Figure 4.2 Results of Twitter Opinion Analyzer

In Figure 4.2, the opinion analyzer is shown, with green representing positive values, red representing negative values, and yellow representing neutral values. The graph below depicts the overall analysis of the results. The statistics suggest that 24% of consumers expressed a favorable impression, while most respondents (54%) left indifferent comments, and 22% of customers tweeted negatively.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

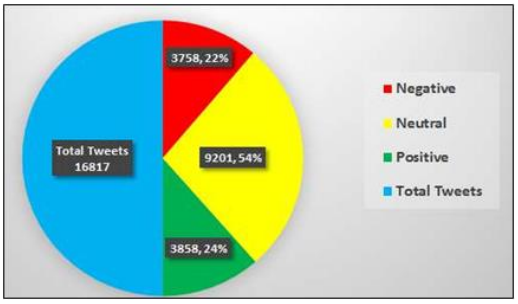


Figure 4.3 Overall Result Analysis.

This method is very easy to implement & finds many sentiments word with their orientation quickly and quickly.

CLASSIFICATIONACCURACY

The experiment is conducted by implementing the working DL as a classification algorithm. Classification accuracy by using D-E-MSB and CNN as feature selection methods is compared.

Precision Comparison

The precision value obtained for negative opinions and positive opinions in D-E-MSB and CNN is graphically represented in Figure4.4.

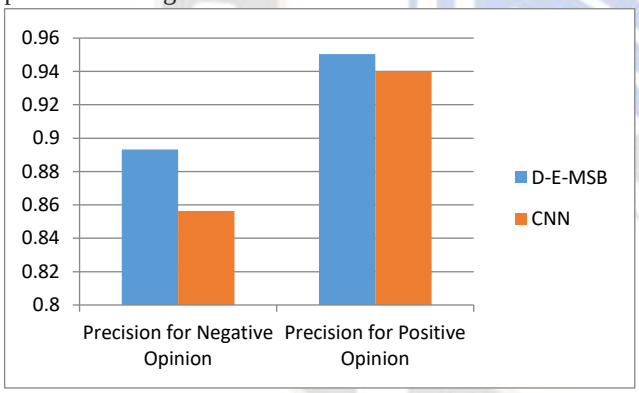


Figure 4.4: Precision Comparison chart

The value clearly shows that precision is high when D-E-MSB is used as a feature selection technique than CNN in a wrapper-based classifier model where CNN is used as a classification algorithm. Comparing the precision values of positive and negative opinions, the precision value of a positive opinion is higher than that in both algorithms. On average, D-E-MSB provides a 1%improvement in precision value for positive opinion and a 3.7% improvement for negative opinion over CNN.

Recall Comparison

The Recall value obtained for negative opinions and positive opinions in D-E-MSB and CNN is graphically represented in Figure 4.5.

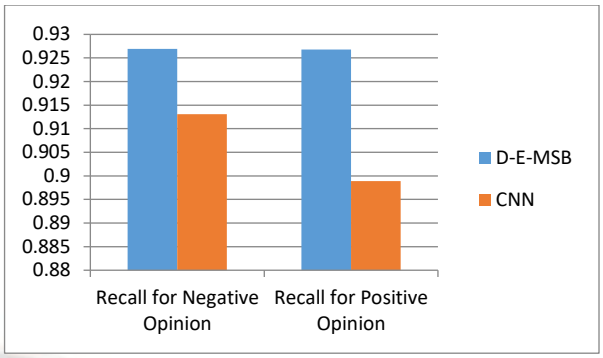


Figure 4.5 Recall Comparison chart

The value clearly shows that precision is high when D-E-MSB is used as a feature selection technique than CNN in the ML-based wrapper model for classification. Comparing the Recall values of positive and negative opinions, the Recall value of a positive opinion is higher than the negative opinion in both algorithms. On average, D-E-MSB provides a 2.8 % improvement in Recall value for positive opinion and a 1.3 % improvement in Recall value for negative opinion over CNN.

F-Measure Comparison

The F-Measure value obtained for negative opinions and positive opinions in D-E-MSB and CNN is graphically represented in Figure6.6.

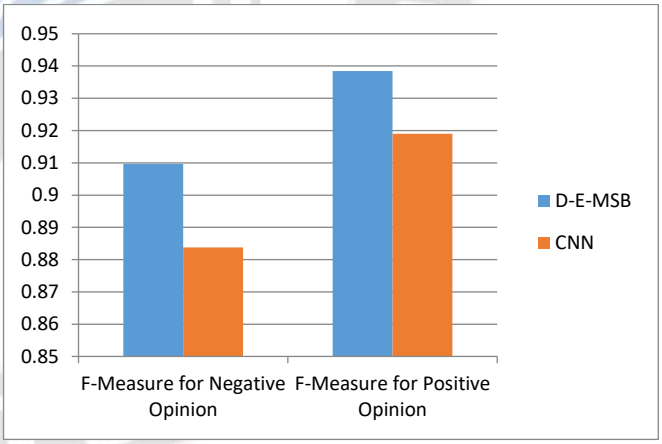


Figure 4.6 F-Measure Comparison chart

The value clearly shows that precision is higher when D-E-MSB is used as a feature selection technique than CNN in a wrapper-based classifier model where Misused as a classification algorithm. Comparing the F-Measure values of positive and negative opinion, the F-Measure value of a positive opinion is high compared to the negative opinion in both the algorithms. On over age, D-E-MSB provides a 2 % improvement in F-Measure value for positive opinion and 2.6%improvement in F-Measure value for negative opinion over CNN.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

Classification Accuracy Comparison

The classification accuracy and overall performance obtained for the given opinions by D-E-MSB and CNN are graphically represented in Figure 4.7.

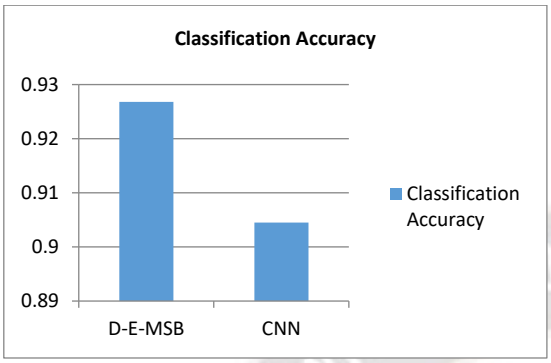


Figure 4.7 Classification accuracy

D-E-MSB produces a 2.2%improvement in overall accuracy when compared with CNN. Since ML has adaptive moments between the available features, it considerably improves over DL in accurately classifying the given opinions.

V CONCLUSION AND FUTURE SCOPE

CONCLUSION

Sentiment analysis is the process of tracking public reviews about a particular topic, product or services. This process involves collecting the available information, extracting the features, selecting the needed features and finally making the classification to arrive at the opinions in Twitter. With the rapid growth of ecommerce sites, public forums and social media on the Web, individuals and organizations are increasingly using public opinions available in these media for making their decision. In addition the information available form a king the decision is also increased. Sentiment analysis process involves feature extraction, feature selection and finally sentiment classification. The feature selection step in sentiment analysis has high impact in determining the accuracy of sentiment classification.

FUTURE SCOPE

The future scope includes combining existing algorithms to obtain performance for large amounts of datasets, and the ratio of spam to non-spam tweets should be considered for improving performance metrics. Also, different extraction techniques can be combined to obtain a new feature set, which would be helpful in improving the classification performance.

The problem can be extended as a big data optimization problem where the attempts can be made to implement distributed computing frameworks as a platform to implement these proposed algorithms. It has been observed from literature that inmost of the works in distributed computing frame work uses traditional algorithms where these emetic algorithms in such frame work have no contribution in handling the huge volume of data. The dimension of the solutions pace has to be reduced.

Acknowledgement

Authors of this paper acknowledge the financial support provided by the Karpagam Academy of Higher Education under the Seed Money Project Scheme (2022-2023).

REFERENCES

[1] Aggarwal, CC, Zhai, C (Eds.), “Mining Text Data”, Springer-Verlag New York, (2012)

[2] Aggarwal, CC, “Opinion Mining and Sentiment Analysis”, In: Machine Learning for Text.Springer, Cham, (2018).

[3] Alex Marin, Bin Zhang and Mari Ostendorf “Detecting Forum Authority Claims in OnlineDiscussions” workshop on language in social media (LSM 2011), page 39-47(23 june2011).

[4] Alom, Z., Carminati, B., & Ferrari, E,”Detecting Spam Accounts on Twitter”. 2018, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). doi:10.1109/asonam.2018.8508495.

[5] Angela Fahrni& Manfred Klenner, “Old Wine or Warm Beer: Target-Specific Sentiment Analysis of Adjectives”, Institute of Computational Linguistics, University of Zurich, Switzerland,(2010).

[6] Anubha Sharma &NirupmaTivari, “A Survey of Association Rule Mining Using GeneticAlgorithm”, International Journal of Computer Applications & InformationTechnology, vol. 1, no. 2, pp. 5-11,(2012).

[7] Apoorv Agarwal, BoyiXie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau“Sentiment Analysis of Twitter Data” workshop on language in social media (LSM 2011), page 30-38(23 june2011).

[8] Bara, I.-A., Fung, C. J., & Dinh, T,”Enhancing Twitter spam accounts discovery using cross- account pattern mining”. 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). doi:10.1109/inm.2015.7140327.

[9] Bilai, S&Abdelouahab, M, “Evolutionary algorithm and modularity for detecting communities in networks”, Elsevier, vol. 473, pp. 89-96,(2017).

[10] Cao, D, He, X, Nie, L, Wei, X, Hu, X, Wu, S & Chua, TS, “Cross- platform app recommendation by jointly modeling ratings and texts”, ACM Transactions on Information Systems, vol. 35, no. 4, pp. 1-27,(2017).

[11] Chao, Y, Robert, H &Guofei, G, “Empirical evaluation and new design for fighting evolving twitter spammers”, IEEE Transactions on Information Forensics and Security, vol. 8,no. 8, pp. 1280-1293,(2013).

[12] Chen, J, Nairn, R, Nelson, L, Bernstein, M & Chi, E,”Short and tweet: Experiments on recommending content from

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

information streams", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp.1185-1194,(2010).

[13] Chu, Z., Gianvecchio, S., Wang, H., &Jajodia, S., " Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" IEEE Transactions on Dependable and Secure Computing, 9(6), 811–824. doi:10.1109/tdsc.2012.75.

[14] Corcoglioniti, F, Nechaev, Y, Giuliano, C, Zanoli, R, "Twitter user recommendation for gaining followers", In: Ghidini C., Magnini B., Passerini A., Traverso P. (eds) AI*IA 2018 – Advances in Artificial Intelligence. AI*IA 2018. Lecture Notes in Computer Science, vol. 11298. Springer, Cham,(2018).

[15] Dangkesee, T., &Puntheeranurak, S. "Adaptive Classification for Spam Detection on Twitter with Specific Data". 2017 21st International Computer Science and Engineering Conference (ICSEC). doi:10.1109/icsec.2017.8443779.

[16] Dong Nguyen and Carolyn P. Rose "Language use as a reflection of socialization in online communities" workshop on language in social media (LSM 2011), page 76-85(23 june2011).

[17] Eshraqi, N., Jalali, M., &Moattar, M. H., "Detecting spam tweets in Twitter using a data stream clustering algorithm". 2015 International Congress on Technology, Communication and Knowledge (ICTCK). doi:10.1109/ictck.2015.7582694.

[18] EvandroCunha, GabrielMagno, GiovanniComarela, VirgilioAlmeida, MarcosAndr´eGonc¸ alves and FabricioBenevenut,"Analyzing the Dynamic Evolution of Hashtags on Twitter: a Language-Based Approach" workshop on language in social media (LSM 2011), page 58-65(23 june2011).

[19] Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3-4), 169–200. doi:10.1080/02699939208411068

[20] Faulkner, A, "Automated classification of stance in student essays: An approach using stance target information and the Wikipedia link- based measure", In: Proceedings of the Twenty-Seventh International Flairs Conference, 174-179,(2014).

[21] Feng, Y., Li, J., Jiao, L., & Wu, X., "BotFlowMon: Learning-based, Content-Agnostic Identification of Social Bot Traffic Flows". 2019 IEEE Conference on Communications and Network Security (CNS). doi:10.1109/cns.2019.8802706.

[22] Fulin Wu, Shifei Ding, Huajuan Huang&Zhibin Zhu, "Mixed Kernel Twin Support VectorMachines Based on the Shuffled Frog Leaping Algorithm", Journal of Computers, vol. 9, no. 4, pp. 947-955,(2014).

[23] Giatsoglou, M, Vozalis, MG, Diamantaras, K, Vakali, A, Sarigiannidis, G &Chatzisavvas, KC, "Sentiment analysis leveraging emotions and word embeddings", Expert System with Applications, Elsevier, vol. 69, pp. 214-224,(2017).

[24] Guang Qiu, Bing, Liu, Jaijun Biu & Chun Chen, "Opinion word Expansion and Target Extraction through Double Propagation", Association of Linguistics, vol. 37, no. 1, pp. 9-27,(2010).

[25] Guidi, B., &Michienzi, A,"Users and Bots behaviour analysis in Blockchain Social Media".2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS). doi:10.1109/snams52053.2020.93365.

[26] Gupta, A., & Kaushal, R, "Improving spam detection in Online Social Networks". 2015 International Conference on Cognitive Computing and Information Processing(CCIP). doi:10.1109/ccip.2015.7100738.

[27] Gupta, H., Jamal, M. S., Madisetty, S., &Desarkar, M. S, "A framework for real-time spam detection in Twitter". 2018 10th International Conference on Communication Systems & Networks (COMSNETS). doi:10.1109/comsnets.2018.8328222.

[28] Haiying, S & Ze, L, "Leveraging social networks for effective spam filtering", IEEETransactions on Computers, vol. 63, no. 11, pp. 2743-2759,(2014).

[29] Tuama, B. A. . (2023). Bigdata Based Disaster Monitoring of Satellite Image Processing Using Progressive Image Classification Algorithm . International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 70–77. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2573.

[30] Han, J, Kamber, M & Pei, J, "Data mining: Concepts and techniques", 3rd ed. Boston: Morgan Kaufmann. J Han, J Pei, M Kamber,(2012).

[31] Hercig, T, Krejzl, P, Hourová, B, Steinberger, J &Lenc, L, "Detecting stance in czech news commentaries", In: Proceedings of the 17th ITAT: CEUR Workshop, pp. 176 – 180,(2015).

[32] Heredia, B., Prusa, J. D., &Khoshgoftaar, T. M,"The Impact of Malicious Accounts on Political Tweet Sentiment". 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC). doi:10.1109/cic.2018.00035.

[33] Huang, Z & Cardenas AF, "Extracting Hot Events From New Feeds Visualization and Insights", in Distributed Multimedia Systems Conference, pp. 10-12,(2009).

[34] Huberman, BA, Romero, DM & Wu, F, "Social networks that matter: Twitter under the microscope", available at SSRN 1313405,(2008).

[35] Ilias, L., &Roussaki, I., "Detecting malicious activity in Twitter using deep learning techniques", Applied Soft Computing, 107, 107360. doi:10.1016/j.asoc.2021.107360.

[36] Ismaila, I & Ali, S, "Improved email spam detection model with negative selection algorithm and particle swarm optimization", Elsevier journal of Applied Soft Computing, vol. 22, pp. 11-27,(2014).

[37] Ismaila, I, Ali, S &Sigeru, O, "Hybrid email spam detection model with negative selection algorithm and differential evolution", Elsevier journal of Engineering Applications of Artificial Intelligence, vol. 28, pp. 97-118,(2014).

[38] Jain, VK, Kumar, S & Fernandes, SL, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics", Journal of Computational Science, vol. 21, pp. 316-326,(2017).

[39] Jennifer Golbeck, J, "Introduction to social media investigation: A hands-on approach", Elsevier,(2015).

[40] Jiménez-Bravo, DM, De Paz, JF &Villarrubia, G, "Twitter's experts recommendation system based on user content. In:

Rodríguez S. et al. (eds) Distributed Computing and Artificial Intelligence", Special Sessions, 15th International Conference DCAI 2018. Advances in Intelligent Systems and Computing, vol. 801. Springer, Cham,(2019).

[41] Kamble, S., &Sangve, S. M, "Real Time Detection of Drifted Twitter Spam Based on Statistical Features". 2018 International Conference on Information , Communication, Engineering and Technology (ICICET). doi:10.1109/icicet.2018.8533767.

[42] Mr. Bhushan Bandre, Ms. Rashmi Khalatkar. (2015). Impact of Data Mining Technique in Education Institutions. International Journal of New Practices in Management and Engineering, 4(02), 01 - 07. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/35.

[43] Kantepe, M., &Ganiz, M. C., " Pre-processing framework for Twitter bot detection", 2017 International Conference on Computer Science and Engineering (UBMK). doi:10.1109/ubmk.2017.8093483.

[44] Latah, M., "Detection of Malicious Social Bots: A Survey and a Refined Taxonomy. Expert Systems with Applications", 113383. doi:10.1016/j.eswa.2020.113383,(2020).

[45] Li, Bryan, Dimitrios Dimitriadis& Andreas Stolcke, "Acoustic and Lexical Sentiment Analysis for Customer Service Calls", ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE,(2019).

[46] Lingam, G., Rout, R. R., &Somayajulu, D, "Detection of Social Botnet using a Trust Modelbased on Spam Content in Twitter Network". 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS). doi:10.1109/iciinfs.2018.8721318.

[47] Lingam, G., Rout, R. R., &Somayajulu, D., "Deep Q-Learning and Particle Swarm Optimization for Bot Detection in Online Social Networks". 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). doi:10.1109/icccnt45670.2019.8944.

[48] Madisetty, S., &Desarkar, M. S., "A Neural Network-Based Ensemble Approach for Spam Detection in Twitter". IEEE Transactions on Computational Social Systems, 1–12. doi:10.1109/tcss.2018.2878852.

[49] Neha Upadhyay, Angad Singh &NorlelaSamsudin, "A Survey on Twitter for Sentimental Analysis using Machine Learning Methods", vol.6, no. 5, pp. 4890-4893,(2016).

[50] Okamoto, M & Kikuchi, M, "Discovering volatile events in your neighborhood: Local-area topic extraction from blog entries", in Asia information Retrieval Symposium, Springer, Berlin, Heidelberg, p. 181-192.

[51] Wartena, C&Brussee, R 2008, 'Topic Detection by Clustering Keywords', Database and Expert Systems Application, DEXA'08.19th International Workshop on IEEE, pp. 54-58,(2009).

[52] Pio Sajin, R. Ashwini, B. Baron Sam, "Distributed and Improvised Up-GrowthApproachfor Utility Based Mining" Research India Publications, (2015).

[53] Praveen Kumar Rajendran, A. Asbern, K. Manoj Kumar, M. Rajesh, R.Abhilash. "Implementation and analysis of Map Reduce on biomedical bigdata. "Indian Journal of Science and Technology 9.31,(2016).

[54] Rahman, M. A., Zaman, N., Asyhari, A. T., Sadat, S. M. N., Pillai, P., &Arshah, R. A., "SPY-BOT: Machine learning-enabled post filtering for Social Network-Integrated Industrial Internet of Things", Ad Hoc Networks, 121, 102588. doi:10.1016/j.adhoc.2021.102588.

[55] Rajiv Bajpaiet, "Aspect-Sentiment Embeddings for Company Profiling and Employee Opinion Mining Computation and language",(2019).

[56] RamnathBalasubramanyan, William W. Cohen, Doug Pierce and David P. Redlawsk "What pushes their buttons? Predicting comment polarity from the content of political blog posts" workshop on language in social media (LSM 2011), page 12-19(23 june2011).

[57] Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani and Joseph King "How can you say such things?!?: Recognizing Disagreement in Informal Political Argument" workshop on language in social media (LSM 2011), page 2-11(23 june2011).

[58] Rodríguez, FM, Torres, LM & Garza, SE, "Followee recommendation in Twitterusing fuzzy link prediction", Expert Systems, vol. 33, no. 4, pp. 349-361,(2016).

[59] Maria Gonzalez, Machine Learning for Anomaly Detection in Network Security , Machine Learning Applications Conference Proceedings, Vol 1 2021.

[60] Rohan, A, Julie, R &Johanvan, D, "Detecting targeted malicious email", IEEE Transactions on Computer and Security, vol.10, no. 3, pp. 64-71,(2012).

[61] Sahoo, S. R., & Gupta, B. B., "Hybrid approach for detection of malicious profiles in twitter.Computers & Electrical Engineering", 76, 65–81. doi:10.1016/j.compeleceng.2019.03.

[62] Samaneh Moghaddam & Fred Popowich, "Opinion Polarity Identification through Adjectives", Computation and Language, Cornell University, (2010).

[63] Sedhai, S., & Sun, A., "Semi-Supervised Spam Detection in Twitter Stream. IEEE Transactions on Computational Social Systems", 5(1), 169–175. doi:10.1109/tcss.2017.2773581.

[64] Gyawali, M. Y. P. ., Angurala, D. M. ., & Bala, D. M. . (2020). Cloud Blockchain Based Data Sharing by Secure Key Cryptographic Techniques with Internet of Things. Research Journal of Computer Systems and Engineering, 1(2), 07:12. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/5.

[65] Shetty, G., Nair, A., Vishwanath, P., &Stuti, A, "Sentiment Analysis and Classification on Twitter Spam Account Dataset", 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA). doi:10.1109/accthpa49271.2020.921.

[66] Shi, P., Zhang, Z., & Choo, K.-K. R., "Detecting Malicious Social Bots based on Clickstream Sequences". IEEE Access, 1–1. doi:10.1109/access.2019.2901864.

[67] Shoukry, A &Rafea, A, "Preprocessing Egyptian Dialect Tweets for Sentiment Mining", the Fourth Workshop on Computational Approaches to Arabic Script-based Languages. AMTA2012, San Diego, CA USA,(2012).

[68] Sinha, P., Maini, O., Malik, G., & Kaushal, R, "Ecosystem of spamming on Twitter: Analysis of spam reporters and spam

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8077
***Article Received:** 23 June 2023 **Revised:** 20 August 2023 **Accepted:** 01 September 2023*

reportees". 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.2016.7732293.

[69] Sobhani, P, Inkpen, D &Matwin, S, "From argumentation mining to distance classification", In: Proceedings of the Workshop on Argumentation Mining, pp. 67–77,(2015).

[70] Stephan Gouws, Donald Metzler, Congxing Cai and Eduard Hovy "Contextual Bearing on Linguistic Variation in Social Media" workshop on language in social media (LSM 2011), page 20-29(23 june2011).

[71] Szde Yu, "Covert communication by means of email spam: A challenge for digital investigation", Elsevier journal on Digital Investigation, vol. 13, pp. 72-79,(2015).

[72] Tingmin Wu, Shigang Liu, Jun Zhang and Yang Xiang, "Twitter Spam Detection based on Deep Learning",ACSW 17, January 31-February 03, 2017, Geelong, Australia.

[73] Tran, VC, Hwang, D, Nguyen, NT, "Hashtag recommendation approach based on content and user characteristics", Cybernetics and Systems, vol. 49, no. 5, pp. 368-383,(2018).

[74] Treeratpituk, P & Callan, J, "Automatically labelling hierarchical clusters", in Proceedings of the 2006 international conference on Digital government research, Digital Government Society of North America, pp. 167-176.

[75] Velavan, P &Subashini, S, "Correlation Based Feature Selection with Irrelevant Feature Removal", International Journal on Computer Science and Mobile Computing, vol. 3, no. 4, pp. 862-867,(2014).

[76] Walaa Medhat, Ahmed Hassan, B & Hoda Korashy, "Sentimental Analysis Algorithms and Applications: A Survey", Ain Shams Engineering Journal, vol. 5, pp. 1093-1113,(2014).

[77] Wang, X., Kang, Q., An, J., & Zhou, M. , "Drifted Twitter Spam Classification using Multiscale Detection Test on K-L Divergence". IEEE Access, 1–1. doi:10.1109/access.2019.2932018.

[78] Wei, F., & Nguyen, U. T., " Twitter Bot Detection Using Bidirectional Long Short-Term Memory Neural Networks and Word Embeddings". 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). doi:10.1109/tps-isa48467.2019.00021.

[79] Wu, B., Liu, L., Yang, Y., Zheng, K., & Wang, X, "Using Improved Conditional Generative Adversarial Networks to Detect Social Bots on Twitter". IEEE Access, 8, 36664–36680. doi:10.1109/access.2020.2975630.

[80] Wu, Y., Fang, Y., Shang, S., Jin, J., Wei, L., & Wang, H, "A novel framework for detecting social bots with deep neural networks and active learning". Knowledge-Based Systems, 211, 106525. doi:10.1016/j.knosys.2020.106525.

[81] Xianghan, Z, Zhipeng, Z, Zheyi, C, Yuanlong, Y &Chunming, R, "Detecting spammers on social networks", Elsevier Journal on Neurocomputing, vol. 159, pp. 27-34,(2015).

[82] Xu, Y, Quan, G, Xu, Z & Wang, Y, "Research on Text Hierarchical Topic Identification Algorithm Based on the Dynamic Diverse Thresholds Clustering", Conference on Asian Language Processing, pp. 206-210,(2009).

[83] Zappavigna, M, "Enacting identity in microblogging through ambient affiliation", Discourse &Communication,vol. 8, pp. 209–228,(2014).

[84] Zhang, Lei, Shuai Wang & Bing Liu, "Deep learning for sentiment analysis: A survey",WileyInterdisciplinary Reviews: Data Mining and Knowledge Discovery vol. 8.4, pp.