Defence Science Journal, Vol. 74, No. 5, September 2024, pp. 671-681, DOI : 10.14429/dsj.74.19555

Deep Learning for Unearthing Emotions in Twitter: A Hybrid Emotional Recognition Model

Aman Raj[#,*], Vivek Kumar[$] and Divya Chaudhary[!]

[#]*Department of Computer Science & Engineering, Netaji Subhas University of Technology, Delhi - 110 078, India*

[$]*DRDO-Defence Scientific Information & Documentation Centre, Delhi - 110 054, India*

[!]*University of Delhi, New Delhi - 110 021, India*

[*]*E-mail: raj.academic8@gmail.com*

ABSTRACT

With the intensification of new classes of media such as Twitter, the Internet has become a primary route for individual and interpersonal messaging. Many individuals share their thoughts regarding news-related topics on Twitter, an established SNS network built on people's relationships. It offers us with a Source of data from which we can dig people's thoughts, which is useful for product reviews and community monitoring. A Hybrid Emotional Recognition Model (HERM) is proposed in this research. Hashtags are recognized as the tag for emotional cataloging based on gathered posts from Twitter. Meanwhile, emoji and the N-grams are dug and used to classify the gathered topic comments into four distinct sentiment groups using the distorted emotional models. Machine learning approaches are applied of categorizing the emotional information set, yielding an 92 % accuracy result. Furthermore, entities underlying emotions might be obtained using the deep learning model SENNA.

Keywords: Entity identification; Point of view evaluation; Emotions categorization; Selecting attributes

1. INTRODUCTION

At present, social media platforms are highly advanced and can usually be accessed from almost any location. Several initiatives in this area merit further investigation. Twitter users may log in using the website, mobile devices, or other clients. persons may submit up to 140-character message to Twitter whenever they want to communicate information, and followers who follow these persons may repost or distribute these communications. Tweets are 140-character communications that reflect users' ideas, thoughts, and emotions. Twitter has over 259.4 million daily users worldwide across the world at the close of November 2022. Twitter users post approximately 500 million tweets every day, consuming over 22 TB of space for storage, and the number is continually growing.

As a result of these factors, Twitter is among the top 8 most visited websites and twitter is able to attracts an immense amount of natural language processing professionals to do explore in this area[1]. Twitter data mining and examination may be applied in a variety of sectors, including epidemic estimate, people movement, public estimation monitoring, etc.

To be more precise, product remarks in tweets are worthwhile mining. Merchants may get consumer feedback in actual time and then alter their personal items to be more competitive in the marketplace; purchasers can learn about other people's experiences These remarks play a vital role in helping people decide on whether or not to buy a product. Tweets have a considerable effect on network public sentiment transmission due to qualities such as their ability to be created in real-time and transmitted far and swiftly[2].

It is vital for the administration to understand and analyze the people's views on some burning social problems to keep the people from being misled by offenders who may damage the country and civilisation. As a result, recognizing emotions and things in Twitter statistics has a high orientation charge. The significance of this study lies in its innovative approach to harnessing social media, particularly Twitter, as a rich source of data for sentiment analysis. With the proliferation of platforms like Twitter, it has become a primary channel for individual expression and interaction. This research introduces the Hybrid Emotional Recognition Model (HERM), a pioneering method that leverages hashtags, emojis, and the N-grams for emotional categorisation of Twitter posts.

By employing machine learning techniques, we can classify these posts into distinct sentiment groups, achieving an impressive accuracy rate of 92 %. This breakthrough not only enhances our understanding of how emotions are expressed and perceived on social media, but also holds immense potential for applications in product reviews and community monitoring. Furthermore, the integration of deep learning with the SENNA model allows us to delve deeper into the underlying emotions of entities. This novel approach promises to contribute significantly to the field of sentiment analysis and open new avenues for research in emotional recognition. In essence, this study not only advances the methodology of sentiment analysis but also presents a practical framework for

Received : 18 October 2023, Revised : 13 March 2024
Accepted : 22 March 2024, Online published : 19 September 2024

extracting valuable emotional insights from the vast reservoir of data available on Twitter. The implications of this research extend far beyond the academic realm, offering valuable tools for businesses, marketers, and analysts seeking to understand and engage with online communities more effectively.

2. RELATED WORKS

Twitter has emerged as a prominent platform for global communication within the realm of social media, solidifying its position as the foremost microblogging platform worldwide. Tweets encapsulate a diverse array of patterns and emotional nuances, and the extraction of these elements holds immense potential for applications in public opinion monitoring, marketing strategies, and rumor control. The evaluation of text sentiments is typically categorized into three distinct groups: positive, neutral, and negative. The overarching objective is to provide individuals with a genuine insight into the collective voice and sentiment of society.

The landscape of Twitter sentiment analysis has evolved considerably since its inception, with notable contributions shaping the field. Pioneering the domain, Way, *et al.* laid the foundation in 2005 by manually annotating over 20,000 tweets with emotion tags, creating a pivotal emotion polarity dataset. This seminal work incorporated machine learning techniques such as Support Vector Machine (SVM), Naive Bayes, and Conditional Random Fields (CRF) to formulate an effective emotion classifier. Building on this, Braig, *et al.* harnessed Twitter's Application Programming Interface (API) to amass a substantial repository of emotional icons, delving into their nuanced impact on mood categorization. These seminal studies collectively form the cornerstone of advancements in Twitter sentiment analysis[13-16].

In a groundbreaking approach, Go[4], *et al.* devised tri machine learning classifications based on Maximum Entropy, Naive Bayes, and SVM, integrating emotion icons as key features. This resulted in an impressive emotional inclination discrimination accuracy of over 80 %. This research has found applications in various commercial sectors, spanning online shopping, film reviews, and online automotive dealership communications. Fei Hong Chao, for instance, delved into analyzing review content directed towards Yahoo English Athletics, using text analysis to unveil investor sentiment towards the stock market.

Way, *et al.* employed the Ling Pipe tool for emotion classification, augmenting classifier accuracy through manual curation of the training set and discerning emotional tendencies within actual manuscripts. The field of sentiment-based text analysis is expanding, with concurrent growth in related research areas. R. Pavitra[5] introduced a sentiment-topic analysis model that operates in a semi-supervised manner, establishing a sentimentality lexicon encompassing positive and negative lexemes for Bigram sentiment division.

Wang and Xiaohui[6-7] delved into sentiment analysis for group proceedings and disease surveillance, further enriching the data resource by incorporating hypermedia for sentiment analysis studies[8]. Named-entity recognition has gained substantial traction in the domain of natural language processing, owing to advancements in computer technology for information retrieval and search engines. Coh[9], *et al.* leveraged SVM for the automatic identification of names and organizations, yielding promising outcomes.

Tan proposed a method for identifying place names using context-based rules generated by transferring erroneous cues, achieving 97 % accuracy in data tests. Wang[10], *et al.* amassed extensive data from vast volumes of real-world text, evaluating the reliability of dynamically-constructed words and word structures for each entity. Through the amalgamation of multiple rules, named entities can be the to automatically identified.

Turkish researchers[11] investigated named-entity identification through their local Twitter, presenting a novel termed-entity-annotated tweet dataset & assessing various tweet-specific dialectal occurrences. It was explored in a similar theme[12], as did Azerousal[13], *et al.*, who holds a patent focusing on named-entity recognition research.

In the pursuit of fine-grained sentiment analysis for tweets through data mining, our approach delves into a comprehensive exploration of impactful techniques. This includes adept Twitter data collection, meticulous tweet pre-processing, and judicious knowledge extraction. Our methodology is anchored in an emotional dictionary tailored for tweets, employing nuanced sentiment analysis through the weighted meanings of emotional words and a sophisticated multi-feature fusion approach. Given the overwhelming volume and rapid generation of tweets, manual monitoring of trending events and public sentiment analysis becomes impractical. To address this challenge effectively, our research leverages cutting-edge machine learning and deep learning methodologies, as documented in[17-19] marking substantial progress in advancing text processing capabilities. In the realm of sentiment analysis, we aim to devise a circumplexes sentiment model. This model will employ hashtags for classification, and adeptly incorporate the N-gram and emoji attributes. Through the utilization of the advanced SENNA model, we can effectively discern and categorize emotions into four distinct sentiment types as predefined.

2.1 Objectives

The objective is to analyze the emotions of a user based on their tweets. To define this problem formally, we establish the following definitions:

- Explanation 1 (Tweet words w): Since every single term in a tweet can be connected to a user's emotion, we aggregate all the words in the tweet and represents it using a two-tuple as w = {t, a}, where it is the text form of w, and a is the occurrence of w in the tweet.
- Explanation 2 (Sentiment dictionary D): For every sentiment, we create a vocab that represents it distinctly, called a sentiment vocab. Because the definition influences the total assessment of sentiment, various sentiment dictionaries may include the same terms. We use a two-tuple to express each emotion vocabulary as *Di={d, v,}* where, d symbolises each word in the dictionary and v is the sentimentality's center vector. The closer an individual's vector visualization is to the center vector,

the more probable this feeling will be expressed. Every term in the dictionary may alternatively be expressed as a two-tuple, $d = \{t, c,\}$ where, it is the world's text form and c represent the word's relevance to the feelings[20-21].

3. METHODOLOGY

3.1 Gathering of Data

The solution you described has three basic techniques for acquiring Twitter data: acquisition of API, site evaluation acquisition, and APP capturing collection. A word table of terms is used in the training set data, and a tweet collection system utilizing a search API is being built of categorizing four different moods. The system is made up of three major modules:

First is the keyword vocabulary module, which constructs information vocabularies and sends desires to the API based on the arguments in the vocab. Second is pre-progression unit, which preprocesses the data in real-time, including extracting useful data and processing time. Lastly, the database storage module, which uses MongoDB, a NoSQL record, based on the provided data type, creates an arrangement of four separate emotive word dictionaries.

3.2 Text Feature Extraction

Currently, the primary focus of research in text feature extraction is on developing effective algorithms for feature selection and text representation models.

The fundamental building block of a text feature is referred to as a "text feature", which must possess certain characteristics, including the ability to differentiate between features present in targets and non-targeted texts, the capacity to prompt and identify text contents, ease of implementation in classification between features, and avoiding the excessive number of features.

The basic goal of the extraction of features is to minimize the dimension of vectors of features while preserving the essential substance of the texts, hence reducing the number of phrases that must be processed. This reduces computing complexity while increasing the speed and effectiveness of the processing of text. Text feature extraction approaches include the phrase recurrence technique, the analysis of principal components way, The TF-IDF process, The N-gram technique, & emoji technique. Although the frequency of a word method may exclude some useful information-containing words due to their relatively low rate, and the TF-IDF method might not determine the weight of words according to their particular locations in the text, the N-gram technique can filter textual data based on a given thresholds value. Moreover, emojis are widely used on Twitter to express emotions. Given the nature of Twitter, we will use the N-gram and emoji methods to mine features from the text in this scheme. The N-gram model is used to continuously identify a large amount of data words, which can be letters, words, syllables, etc., to achieve the programmed cataloging of emotions.

3.2.1 Revised N-Gram Model

In mathematical terms, the N-gram technique may be thought of as every term in a text, like an algorithm win, and its rise in likelihood has only a tenuous relationship through the periods it appears. This determines the independence of the winner. We may use the probabilistic conditional probabilities to forecast the likelihood of the win appearing.

If W is a piece of arrangement with N elements, then W=& the sequence's occurrence chance is P(W):

$$P(W) = P()P(\ |\ )P(\ |) \ldots P(|\ldots) \tag{1}$$

When dealing with a huge quantity of data, the Markov supposition can be applied, which states that the chance of a word appearing is only related to the chance of the word that came before it. This simplifies the problem. Therefore, the uni-gram model can be modified to the bi-gram model.

$$P(W) \text{ material } P()P(\ |\ )P(\ |) \ldots P(|) \tag{2}$$

Similarly, we may generate trigrams, which are solely tied to the chance of two arguments appearing forward-facing it.

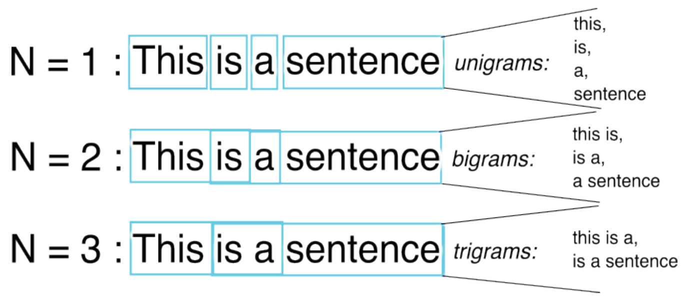
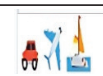
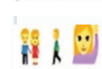
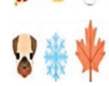
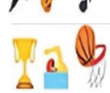


Figure 1. Definition of the N-gram.

Table 1. Twitter's emoji categorisation table

Category	Emoji	Names
Traveling/ commuting		car, airplane, sailboat
Events		party popper, jack-o-lantern, graduation cap
Places		school, European, castle, home + garden
Other activities		artist palette, books, television
Feelings		smiling face with heart eyes, unamused face, crying face
People		man and woman holdsing hands, person walking, person raising one hand
Eating & drinking		pizza, doughnut, hot beverage
Nature & animals		dog, snowflake, maple leaf
Music		microphone, guitar, musical notes
Sport		trophy, swimmer, basketball and hoop

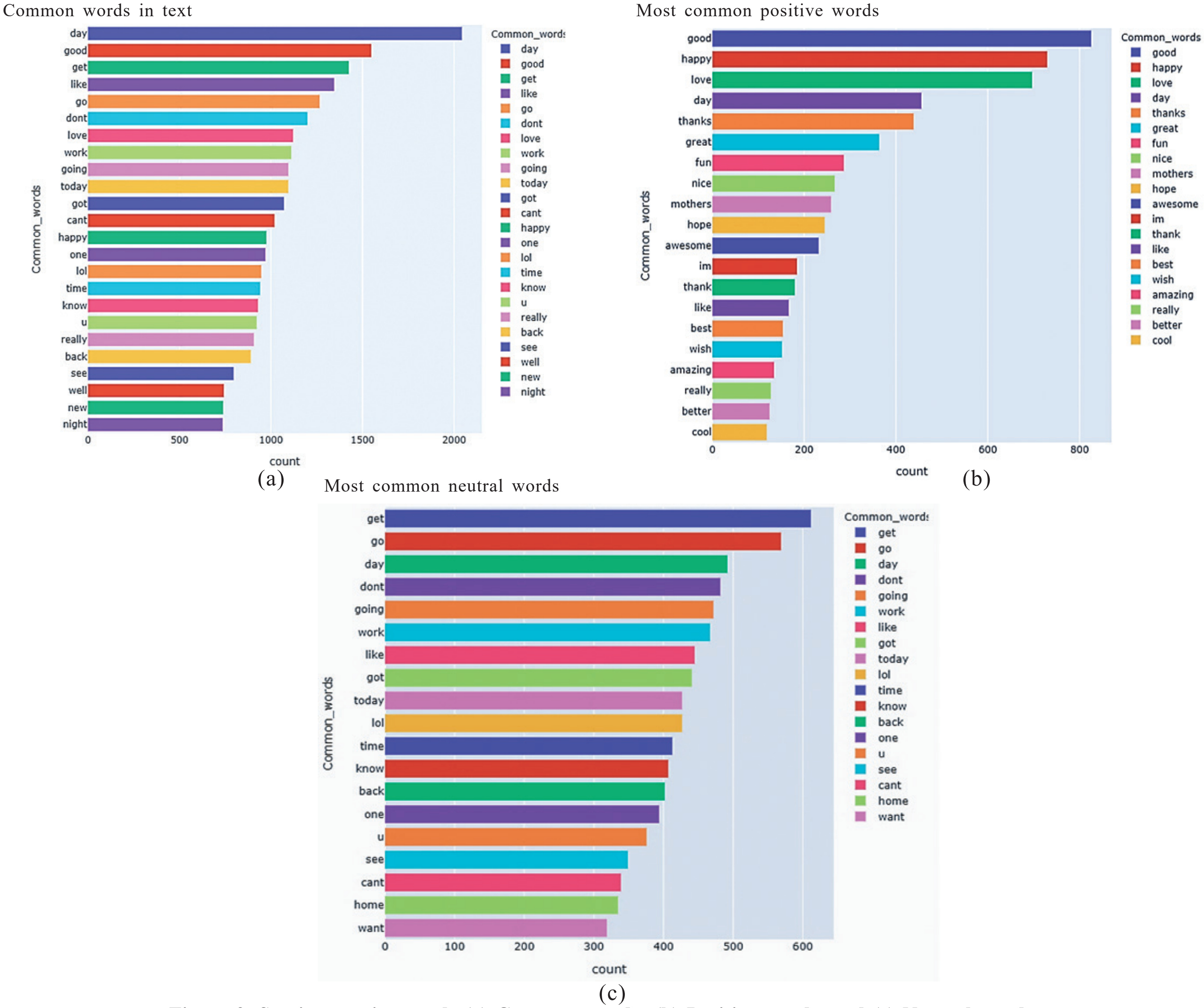


Figure 2. Sentiment wise words (a) Common words; (b) Positive words, and (c) Neutral words.

P(W) material P()P(|)P(|)P(|) ...P(|)   (3)

3.2.2 *N-gram Model*

It can also be enhanced by inserting padding elements at the starting point of each bi-gram or tri-gram, such as typical blanks or whitespace. This raises the number of grammes and enhances the prediction models' accuracy. We employed this upgraded version of the N-gram technique in our study and discovered that it boosted the number of characteristics and enhanced the accuracy of predictions. However, for short-word tweets, the tri-gram algorithm can only detect a few properties. Nevertheless, adding the padding characteristic still increases the feature quantity significantly. This is illustrated in Figure 1. Table 1 contains the Twitter emoji categorization table.

3.3 Analysis of Sentimental entity

This unit discusses the applications of sentimental techniques. for analyzing emotions and detecting emotions in Twitter text. The majority of sentimental models are currently inspired by psychological theories and emotive ideas from social networks. The emotional model and keywords in this work are the circumplexes and 4 coordinates pole edge. depending on the required theme and variety of sentiment. Figure 2 illustrates the most commonly used words based on their sentiment. To upholds the freedom of the 4 types of feeling and decrease overlap, sentimental words between axes are removed, resulting in the acquisition of four kinds of expressive tags in the vocabulary. These tags serve as the training dataset used for data training. The vocab slab is presented in Table 4.

3.4 Text Classifier Training

In this process, When the identifier, text, and arguments are passed into the adaptation procedure, the tags that accompany them are processed for model fitting.

Data mining relies heavily on classification. Utilizing a classifier based on ML to teach a unit allows the machinetrained classifiers the to automatically categorise the content of fresh data, freeing up human resources for artificial categorisation. As a result, it is critical to specify which type of classifier should be used of categorising data within the arena of data mining.

Of categorising data, four distinct classifiers are available: Nave Bayesian, logistic regression, ML, support vector machines, and the KNN method are some of the techniques used.

Nave Bayesian is a generative model that can handle various categorizations by estimating the probability of categorizing. Logistic regression is simple to apply and requires little storage space. The best available classifier is SVM, which may be used directly without modification. SVM works well in classifying choices on data beyond the training set and has a lower rate of errors. The K-neighbor technique is useful not just for categorization but also for regression analysis. Table 2. contains Precision of classifiers and attributes and Table 3 contains the rate of recall between the classifier and attributes.

In comparison to KNN algorithms, Naïve Bayes and, SVM offers several benefits such as high speed of classification, accurate performance with varying training data sizes, and universality. We chose the SVM classifier for our investigation because of its non-linear mapping approach of translating the sample space to a feature with high dimensions space. This aids in the conversion of non-linearly separable issues in the first sample frame to linearly distinct issues within the feature space. We count the performance of our classifier based on accuracy and memory rates. Assume…

Table 2. Precision of classifiers and attributes

Characteristics	Naïve bayes	Logistic regression	SVM	KNN
Uni-gram	86.3	85.2	92	88.1
Uni-gram, emoji	88.2	84.4	91.1	88.5
Uni-gram, punctuation	86.6	85.6	90.9	88.1
All feature	86.9	85.9	91.8	91.1

Table 3. Rate of recall between classifiers

Characteristics	Naïve bayes	Logistic regression	SVM	KNN
Uni-gram	86.3	85.5	91.3	88.3
Uni-gram, emoji	88.2	84.9	91.7	88.5
Uni-gram, punctuation	86.5	85.3	88.6	88.9
All feature	87	86	90	90

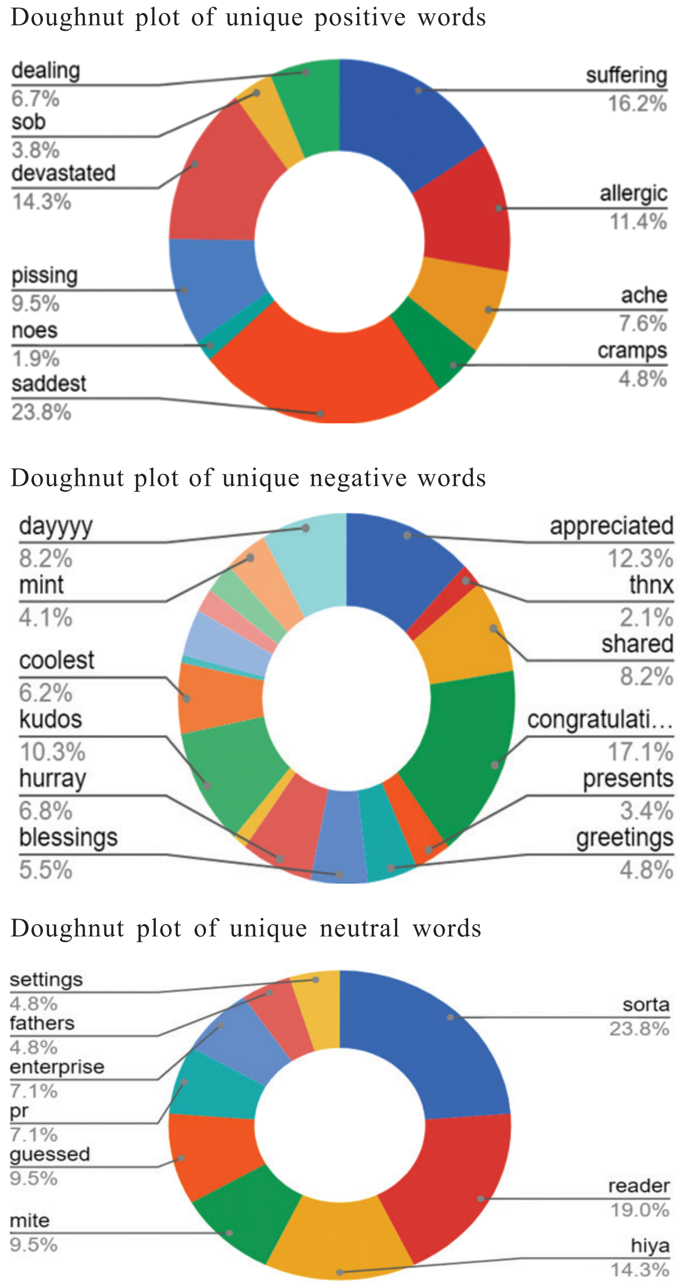


Figure 3. Doughnut plots of different unique sentiment.

Table 4. Four types of sentimental dictionaries

Sentiment category	Positive words	Negative words
Happy ACTIVE	#amazing, #overjoyed, #happy, #feelhappy, #proud, #wonderful, #enjoy, #delighted, #sohappy, #enthusiastic, #excellent, #excited, #joyful, #superhappy, #veryhappy	#fearful, #distressed, #distress, #annoying, #disturbed, #angry, #nervous, #irritated, #tense, #mad, #worried, #annoyed, #bothered, #stressed, #stressful, #furious
Happy Inactive	#resting, #quiet, #serene, #contentment, #calm, #sleepy, #calming, #convinced, #restful, #satisfied, #placid, #relaxed, #gloomy, #nothappy, #contented, #asleep, #fatigued, #sleepyhead	#hapless, #sorrow, #suicidal, #downhearted, #feelsad, #miserable, #unhappy, #dispirited, #disappointed, #depressed, #depression, #ifeelsad, #sosad, #verysad, #hopeless, #depress, #sad, #supersad

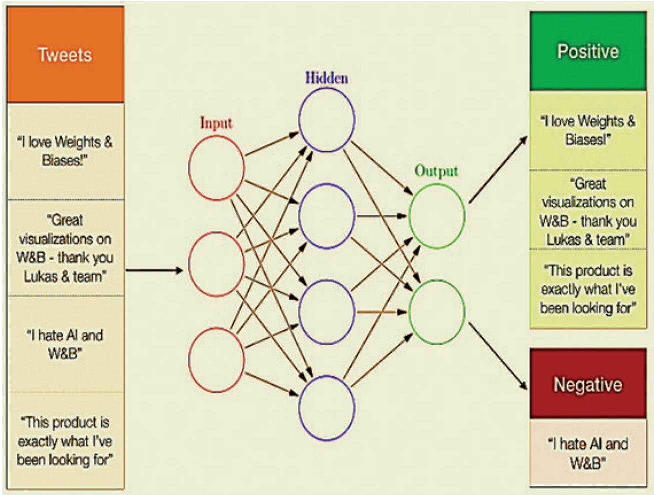


Figure 4. Training model of deep learning.

M_{right_i} = represents the number of tweets which is mistakenly divided into class C_i

M_{wrong_i} = represents the number of tweets which is mistakenly divided into class C_i

M_{all_i} = represents the number of tweets which is included in the class Ci actually Then:

$$\text{Precise rate}(i) = \frac{M_{right_i}}{M_{right_i} + M_{wrong_i}} \times 100\%$$

$$\text{Recall rate}(i) = \frac{M_{right_i}}{M_{all_i}} \times 100\% \quad (5)$$

And the average precise rate and average recall rate can be calculated according to the formulas:

$$\text{Average precise} = \frac{\sum_{i=1}^{m} \text{Precision rate}_i}{m} \quad (6)$$

$$\text{Average recall} = \frac{\sum_{i=1}^{m} \text{Recall rate}_i}{m} \quad (7)$$

Algorithm SVM: The SVM based on the analysis of Twitter user's sentiment.

Input:

- D: sentimental classification dictionary
- T: all target tweet
- S: sentimental class table comprised in scheme

Output: the sentiment E of marktweet t

1. For every T's each word w do
2. For each α1 not fulfilled KKT conditions
3. If min = distance (α1,α2)
4. Checks $\sum_{i-1}^{m} \alpha i.\text{label} = 0$
5. Compute α1, α2
6. Checked label*(W^TX+ b)≥1.0
7. Compute b1, b2, b3
8. Return α matrixs & b matrixs

3.5 The Detection of Entity

Named-Entity Recognition (NER), similarly known as entity recognition, is the process of identifying specific entities with distinct meanings in a given set of text, including people's names, locations, organisations, and proper nouns. There are four primary methods for NER:

- Statistical-based recognition: Statistical models that include decision tree models, conditional random fields models, support vector machine (SVM) models, maximum entropy models, and hidden Markov models, are used in this technique.
- Rule-based recognition: This technique is primarily dependent on two types of data: restrictive clauses & named-entity terms
- Combined method: This method uses a combination of rules and statistics to recognise entities. The system first uses statistical methods to identify entities and then applies rules to correct and filter the results.
- Machine learning-based recognition: This method utilizes machine learning algorithms to recognize entities.
- In English, using SVM-based methods for entity recognition achieves an accuracy of over 99 % for recognising places and people's names.

Deep learning is a new subfield of machine learning that simulates the functions of the human mind. It originated from deep belief networks, which were developed from Boltzmann machines presented in a paper by Hinton. Deep learning is based on the usage of an algorithm S with N levels (S1, S2, SN) that receives an input I and creates an outcome O as I= > S1= > S2= > S3= > Sn= > O. This system learns aspects that help in decision-making on its own. The outputs of a low level can be utilized as contribution to the advanced level by modifying its parameters in every level of the framework, resulting in a logical arrangement of input information[22-23].

The training model for deep learning is shown in Fig. 4. Natural language is human communication that needs to be converted into computer-readable symbols for computational identification. Words are incorporated in deep learning to portray them as highly dimensional vectors with 200 to 500 dimensions. Bengio presented this word-embedding approach over a decade ago.

Each word is allocated geographic coordinates in space with multiple dimensions when deep learning is used to train word vectors. Figure 3 depicts a donut plot of several emotion words.

Each word is assigned a random vector at the start of the word vector retraining procedure. Deep learning, for example, may be utilised to forecast the veracity of a five-word statement, such as "Benjamin enjoys basketball." By changing any of the terms in this statement, for example, "the" to "theory," "Benjamin enjoys playing theory basketball." becomes grammatically incorrect. Deep learning models can predict the truthfulness of changed five-word phrases. This is achieved by using mathematical equations and symbols to represents the process of learning and prediction. SENNA not solitary developed a technique for implanting words but additionally addressed the NLP.

The neural network language framework is used to process tasks (POS, Chunking, NER, and SRL). Each word in SENNA may be accessed straight from the lookup table.

SENNA utilises HLBL to generate word vectors that capture different semantics and grammatical usage. These word vectors are then collective in various ways to represents the words. To classify emotions, the SENNA model processes these representations and can classify them into one of the four previously described emotions.

4. EXPERIMENT AND ANALYSIS

4.1 Data Acquisitions

The paper's dataset was obtained from Twitter's open platform using the search API between March 1st and May 1st, 2015. A keyword dictionary consisting of four different sentiment categories was used to collect training data. Due to the large volume of tweets, it was resource-intensive to manually classify them, so the paper utilized hashtags as the automatic tags for classification using machine learning methods. Hashtags are used in tweets as a level of tag to designate a certain topic, and the article assumes that tweets labelled with a given emotion category correspond to that category, enabling machine categorisation.

4.2 Data Preprocessing

Tweets are a unique type of data that differs from traditional news or media data in that they often contain a lot of errors and noise. To prepare the data for classifications, certain preprocessing steps are necessary, including:

4.2.1 Removing Usernames and URLs

Some usernames contain sentimentality words and are even included in the sentiment dictionaries, but usernames themselves have slight use in sentiment analysis. To avoid intrusion with the trial results, all labels with the "@" symbol are replaced with the term "USERID". Additionally, tweets often comprise URLs, which are not useful for text dispensation, and are therefore swapped with the term "URL".

4.2.2 Spells Errors

Tweets frequently contain terms with spelling errors, such as "asd" (sad), "Yayyyyyy!!", "Awesoomeeee" or purposeful misspelling for emphasis. To avoid interference, the module uses a database of frequent misspelling detected in tweets to fix misspelt terms.

4.2.3 Abbreviations

Tweets frequently use abbreviations, such as "good4you" for "good for you," which can pose a challenge for sentiment analysis. The module corrects these abbreviations to their full word equivalents.

4.2.4 Conflict Between Sentiment Categories

Some tweets contain hashtags that belong to two different sentiment categories, making them conflicting tweets. To avoid interference with classification precision, the module removes these tweets from the dataset.

4.3 Text Feature Extracting

The module employs the N-gram text feather during pipeline creation and enhances English signs, sensitive symbols, and an array of arguments as fluffs. Tweets with the same emojis category, for example, are defined as belonging to the similar category. (Table 6).

Because emojis is preserved as Unicode, it is the primary of categorising and mark the Unicode appeal. (Table 5).When an emoji is seen, every tweet is scanned and a key-value combination is added to a list. (Table 7). Wordnet creates an association of all synonyms & may substitute the same term for all terms in the same word network. As a consequence, the replacement is used in the sentiment classification module to reduce feature dimensionality while enhancing classification accuracy.

Table 5. Twitter emotion categorisation label table

Emotion type	Label
Happy-active	1
Happy-inactive	2
Unhappy-active	3
Unhappy-inactive	4

Table 6. Twitter sentiment categorisation sample table

Tweets	Type
I really love this place	Happy-active
He is a bad person	Unhappy-active
He surprises me on my Birthday	Happy-active
Please Pray for him!	Unhappy-inactive

Table 7. Twitter's labeling example

Tweets	Label
I really love this place	1
He is a bad person	3
He surprises me on my Birthday	1
Please Pray for him!	4

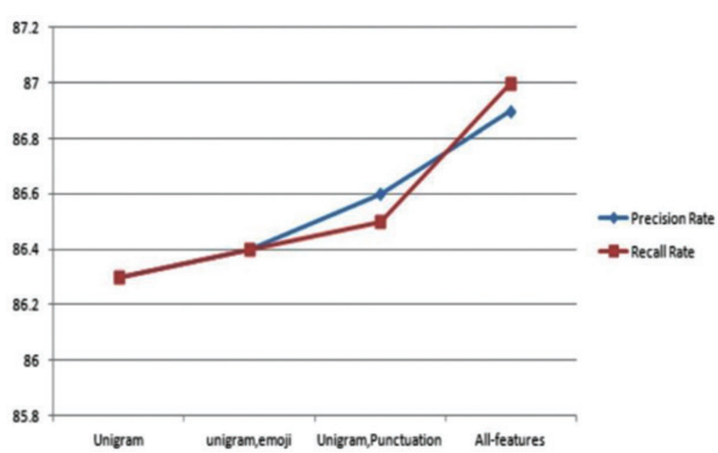


Figure 5. Nave bayes precision and recall rates.

4.4 Result Analysis and Classifier Training

The graphs show that the accuracy and recall rates of Nave Bayes change when different methods of feature extraction are used. Previously the accuracy rate was around 86.5 % (86.386.9 %), while the rate of recall was approximately 86.5 % (86.387 %).. (Fig. 5). It shows that by combining unigrams, emoji, and punctuation, the accuracy rate may reach 92 %. (Fig. 9).

The graph show that the accuracy and recollection rates of Logistic regression change when dissimilar approaches of feature extraction are used. The accuracy rate was around 85 % (84.285.9 %), but the rate of recall is around 85 % (84.986 %).

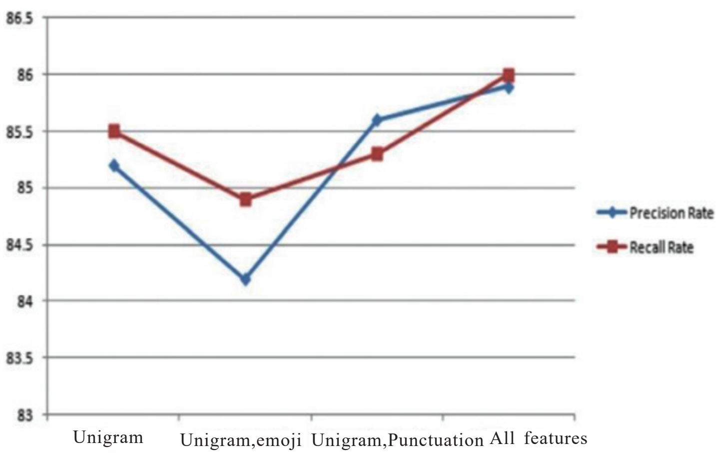


Figure 6. logistic regression.

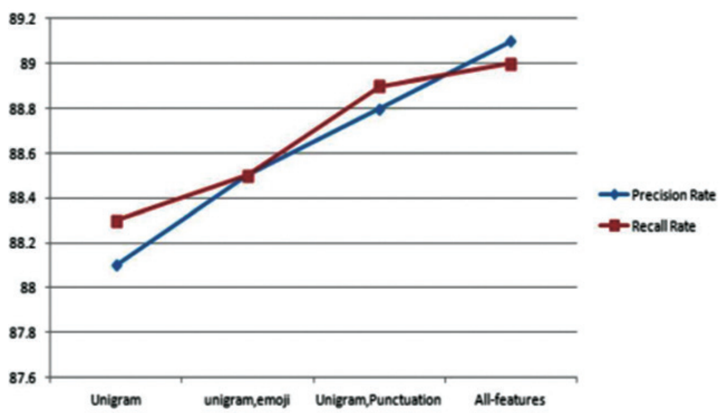


Figure 7. Precision rate & recall rate of KNN.

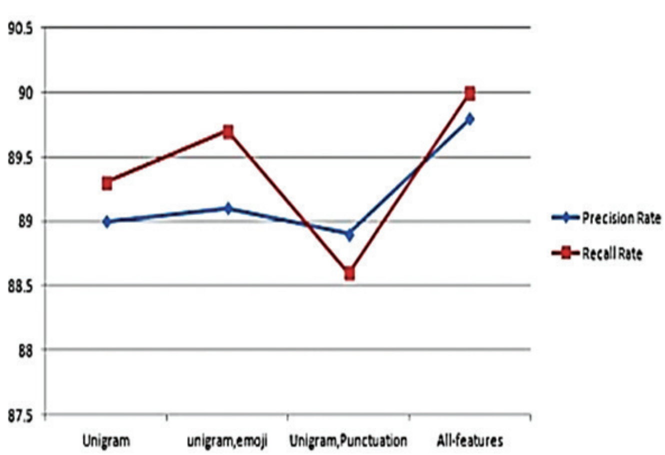


Figure 8. Precision &recall rate of SVM.

It shows that by combining unigrams, emoji, and punctuation, the accuracy rate might reach up to 85.9 %. (Fig. 6).

The graph shows that the accuracy and recall rates of KNN change when different methods of feature extraction are used. The accuracy rate was around 88.5 % (88.189.1 %), while the rate of recall was approximately 88.5 % (88.39 %). It shows that by combining unigrams, emojis, and punctuation, the precision rate may reach 89.1 %. (Fig. 7). Also, the Fig. 9 and Fig. 10 show Precision of classifiers & attributes and Rate of recall between classifiers.

The graph shows that the accuracy rate & remember rate of SVM change when different methods of feature extraction are used.The accuracy rate was around 92 % (91.98 %), while the recall ratio is approximately 92 % (91.70 %).

It shows that by combining unigrams, emoji, and punctuation, the accuracy rate may reach 92.8 %. (Fig. 9 and 10). When the data in this research is compared, it is understandable that the accuracy of classification may reach 92.8 % by employing unigrams, emoji, and language as attributes and SVM emotive classification methods. SVM is the most effective sentiment classification approach for our experiment.

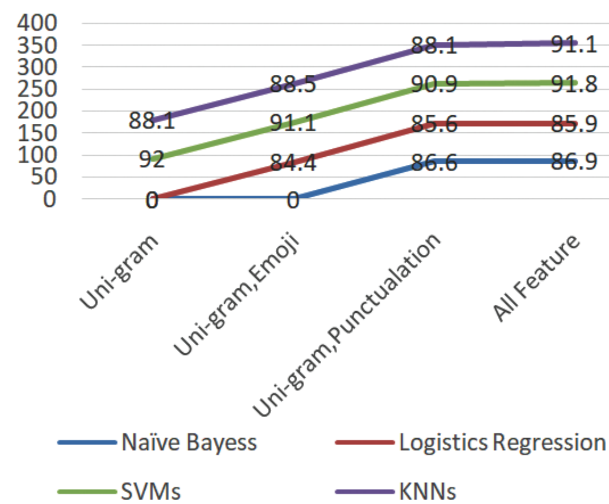


Figure 9. Precision of classifiers and attributes.

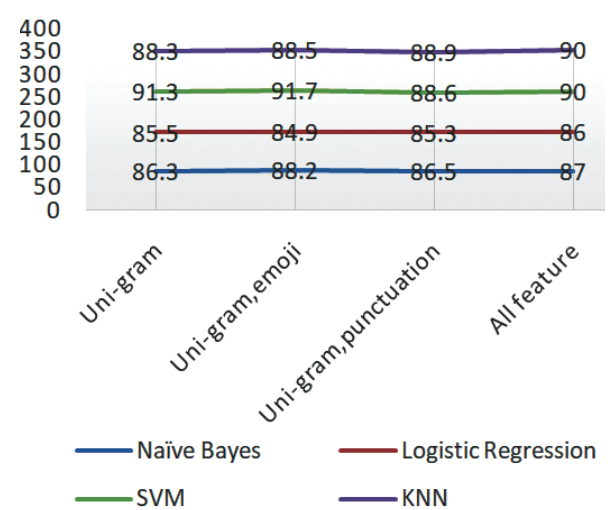


Figure 10. Rate of recalls between classifiers.

4.5 Results and Analysis

After emotional classifier training, the automatic classifier was built to handle 5000 fresh values. Furthermore, named-entity identification was conducted for each data type. The SENNA deep learning toolkit is utilized in this section for entity extraction for every kind of information, and all of these phrases are organized at the identical time. The top 10 results are shown in Table 8. The visual graphic displays for each form of emotional entity are as follows: (Fig. 11 and 12).

Figure 11. Positive words clouds.

Figure 12. Negative word cloud.

Table 8. Quartet of sentiments and the distilled essence

Happy active	Happy inactive	Unhappy active	Unhappy inactive
Excited	Serene	Anxious	Lonely
Joyful	Netflix	Stressed	Depressed
Thrilled	Instagram	Frustrated	Grieving
Energetic	Relaxed	Angry	Miserable
Enthusiastic	Peaceful	Irritated	Desperate
Vibrant	Cozy	Tense	Isolated
NDSU	Blissful	Nervous	Heartbroken
Cheerful	Leisurely	Worried	Hopeless
Ecstatic	Yoga	Overwhelmed	Disappointed
Inspired	Calm	Agitated	Lonely

We retrieved emotive articles from 5000 fresh values using SENNA. These organizations are the explanations for why consumer exhibit certain feelings. The outcome demonstrates that Netflix can make its consumers feel.

Simultaneously, when we recite the original piece, we noted that international users frequently devote their free time on Netflix. Most individuals are ecstatic to have landed in Canada, which is a breathtakingly gorgeous country. And NSUT students always exhibit their activities throughout test or graduation periods.

5. CONCLUSION

This project intends to improve on current Tweet sentiment investigation technologies, such as classic boldness categorization for positive-negative and median-type sentiment, as well as multivariate emotion model classification. Simultaneously, we change the machine learning approach for emotion categorization and the deep learning method for entity recognition. The following are the project's innovative features:

- Most current research solely analyses the cataloging of positive and negative emotion's view, and nobody has established the fundamental components behind emotional study globally. Though, emotional composition research of current topics and people opinion monitoring offer a wide variety of submissions in our daily lives.
- Previous research used pos tags to apply deep learning algorithms to named-entity recognition. However, when data is with a lot of noise, it can be hard to get a meaningful level using this technique. Entity acknowledgement theory's identification accurateness in deep learning, on the opposite pointer, can surpass 95 %.

This study's technique may be used to track the general public on social media. We noticed in this investigation that the size of the data cast-off in the research had a considerable influence on the accuracy of the final outcomes. So, in order to progress the classifier's accuracy, it must be extensible for every tweet data. Furthermore, in the upcoming, we will leverage emotive mining tools to mine users' feelings from tweets.

REFERENCES

1. Wang, P.; Luo, Y.; Chen, Z.; He; L. & Zhang, Z. Orientation analysis for Chinese news based on word embedding and syntax rules. *IEEE Access*, 2019, **7**, 159888-159898. doi: 10.1109/ACCESS.2019.2950900.
2. Wang, Y.; Guo, J.; Yuan, C. & Li, B. Sentiment analysis of Twitter data. *Appl. Sci.*, 2022, **12**, 11775. doi: 10.3390/app122211775
3. Braig, N.; Benz, A.; Voth, S.; Breitenbach; J. & Buettner, R. Machine learning techniques for sentiment analysis of COVID-19-related Twitter data. *IEEE Access*, 2023, **11**, 14778-14803. doi:10.1109/ACCESS.2023.3242234.
4. Go, A.; Bhayani, R. & Huang, L. Twitter sentiment classification using distant supervision. Cs224n Project Report, 2009, 1–12. https://www-cs-faculty.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf
5. Pavitra, R. & Kalaivaani, P.C.D. Weakly supervised sentiment analysis using joint sentiment topic detection

with Bigrams. *In* IEEE 2015 2nd International Conference on Electronics and Communication Systems (ICECS), 2015, 889–893.
doi: 10.1109/ECS.2015.7125042

6. Wang, Z.; Hu, K.; Chen, L.; Du, H.; Wang, S.; Wang, Z. & Cui, X. Exploiting feature selection algorithm on group events data based on news data. *In* 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI), Beijing, China, 2015, pp. 62-65.
doi: 10.1109/IIKI.2015.20.
7. Xiaohui, C.; Nanhai, Y. Zhibo, W.; Cheng, H.; Weiping, Z.; Hanjie, L.; Yujie, J. & Cheng, L. Chinese social media analysis for disease surveillance. *Pers. Ubiquit. Comput.,* 2015, **19**, 1125–1132.
doi: 10.1109/IIKI.2014.11
8. Zhibo, W. Chongyi, Z.; Jiawen, S. Ying, Y. Weiping, Z. & Xiaohui, C. Key technology research on user identity resolution across multi-social media. CCBD2015, 2015.
doi: 10.1109/CCBD.2015.28
9. Coh, CL.; Asahara, M. & Matsumoto, Y. Chinese unknown word identification using character-based tagging and chunking. Proceedings of the 41st Annual Meeting of the association of Computational Linguistics, Sapporo, 2003, pp. 197–200. https://aclanthology.org/P03-2039.pdf
(Accessed on 02 January 2024)
10. Wang, Yuan-Yuan & Zhong-Shi, He. A new approach to Chinese person names recognition based on part-of-speech detecting. Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Shanghai, China, 2004, **5**, pp. 2969-2972.
doi: 10.1109/ICMLC.2004.1378541
11. Küçük, D.; Jacquet, G. & Steinberger, R. Named entity recognition on Turkish tweets. *In* Language Resources and Evaluation Conference., 2014.
https://aclanthology.org/L14-1328/
(Accessed on 02 January 2024)
12. Panandikar, M. & Raghavan, A. Fine-tuning a named entity recognition model using data augmentation and oracle-based learning. *In* 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gautam Buddha Nagar, India, 2023, pp. 945-949.
doi: 10.1109/UPCON59197.2023.10434751.
13. Azeroual, O.; Nacheva, R.; Nikiforova, A.; Störl, U. & Fraisse; A. Predictive analytics intelligent decision-making framework and testing it through sentiment analysis on Twitter data. *In* Proceedings of the 24th International Conference on Computer Systems and Technologies (CompSysTech '23). Association for Computing Machinery, New York, NY, USA, 2023, PP. 42–53.
doi: 10.1145/3606305.3606309
14. Plaza-Del-Arco, F.M; Molina-González, M.D.; Ureña-López, L.A. & Martín-Valdivia, M.T. A multi-task learning approach to hate speech detection leveraging sentiment analysis. *IEEE Access*, 2021, **9**, 112478-112489.
15. Harnain, K. & Gupta, M.K. An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. *Multimedia Tools and Appl.*, 2022, **17**(81**)**, 23649-23685.
doi: 10.1007/s11042-022-12648-y
16. Bandhakavi, A.; Wiratunga, N.; Deepak, P. & Massie, S. Generating a word-emotion lexicon from# emotional tweets. Proceedings of the third joint conference on lexical and computational semantics (* SEM 2014), 2014, pp 12-21). https://aclanthology.org/S14-1002.pdf
(Accessed on 05 January 2024)
17. Singh, L.; Gupta, P. Katarya, R. & Jayvant, P. Twitter data in emotional analysis: A study. *In* Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2020, pp. 1301-1305.
doi: 10.1109/I-SMAC49090.2020.9243326
18. Tahayna, B. & Ayyasamy, R.K. Applying English idiomatic expressions to classify deep sentiments in COVID-19 Tweets. *Comput. Syst. Sci. & Engin.*, 2023, **47**(1).
doi: 10.32604/csse.2023.036648
19. Yadav, J.; Yadav, A.; Misra, M.; Rana, N.P. & Zhou, J. Role of social media in technology adoption for sustainable agriculture practices: Evidence from twitter analytics. *Commun. Assoc. Inf. Syst.*, 2023, **52**(1), 35.
doi: 10.17705/1CAIS.05240
20. Shrivastava, V.D. Exploring sentiment in tweets: An ordinal regression analysis. Congress on Control, Robotics, and Mechatronics 2023 Mar 25 (pp. 331-343). Singapore: Springer nature singapore. https://link.springer.com/chapter/10.1007/978-981-99-5180-2_27
(Accessed on 05 January 2024)
21. Anusha, K. & Vasumathi, D. A comparative survey on multi-language sentiment analysis methods. *In* Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1172-1178.
doi: 10.1109/ICAISS58487.2023.10250624.
22. Thakur, N. Investigating and analysing self-reporting of long COVID on Twitter: Findings from sentiment analysis. *Appl. Syst. Innovation*., 2023, **6**(5), 92.
doi: 10.3390/asi6050092
23. Doğan, B.; Balcioglu, Y.S. & Elçi, M. Multidimensional sentiment analysis method on social media data: Comparison of emotions during and after the COVID-19 pandemic. *Kybernetes*, 2024.
doi: 10.1108/K-09-2023-1808

CONTRIBUTORS

Mr Aman Raj is currently pursuing a Bachelor's degree in Computer Science and Engineering at Netaji Subhas University of Technology, New Delhi, India.
In the current study: He did coding, implementing intricate algorithms, and critically analysing their outcomes.

Mr Vivek Kumar obtained his MBA from Faculty of Management Studies at the University of Delhi. He is working as a Scientist "E" at DRDO-DESIDOC, Delhi. His areas of interest include:

Requirements analysis, design and development of cryptosystems, threat modelling, information security management, and testing. His expertise also extends to project management, holding certification in this domain. He is further recognised as an ISMS Lead Auditor, reflecting his proficiency in Information Security Management Systems.

Contribution in the current study: His guidance significantly influenced the direction of the research, providing crucial insights and periodic supervision.

Dr Divya Chaudhary obtained PhD in Computer Engineering from University of Delhi and working as an Assistant Professor at Northeastern University in Seattle, Washington, United States. Her areas of interest include: Cloud computing, deep learning, and machine learning.

Contribution in the current study: Her guidance and mentorship were instrumental in shaping the research paper.