Emotion Classification on Social Media Comments Using Categorical Feature Extraction Along With the Bidirectional Encoder-based Recurrent Neural Network Classification

S. SARANYA, G. USHA
Department of Computing Technology,
SRMIST, Tamil Nadu,
INDIA

Abstract: - All across the world, social media is one of the most widely used platforms for information exchange. Comments on relevant information might be made in response to a video or any other piece of information. A remark may include an emotion that may be recognized by an automated recognition system. On Facebook, Twitter, and YouTube comments, we performed studies to determine their emotional categorization. A set of comments is gathered and manually classified using six fundamental emotion labels (happy, sad, angry, surprised, disgust, and fear) and one neutral label, with each emotion label representing a different emotion category. A prominent approach in natural language processing (NLP), deep learning has been used in a wide range of categorization applications. This procedure begins by preprocessing the input data with normalization, followed by categorizing characteristics in feature extraction utilizing the Linguistic and word count analysis (LIWC). Finally, for the categorization stage, the classify features might be supplied. Finally, for categorizing emotions, the Bidirectional Encoder based recurrent neural network classification approach is used. The studies have been carried out with the use of typical social media data that has been acquired from the kaggle data repository. The findings show that the suggested model outperforms all other existing mechanisms in terms of overall performance.

Key-Words: - Social media, emotion, Linguistic and word count analysis, Bidirectional Encoder based recurrent neural network.

Received: July 25, 2023. Revised: October 22, 2023. Accepted: November 23, 2023. Published: December 31, 2023.

1 Introduction

Because of the widespread usage of the Internet, the variety of social networks, the simplicity with which views and ideas can be expressed, the availability of brands, and the speed with which people can engage, people spend the majority of their time online, as shown in Figure 1. Meaningful and nonsensical data are generated as a result of this activity on social media. Work in this subject is possible because of the rapid development of technology and the rise in data, as well as its speed and costs. Text mining, natural language processing, and other artificial intelligence techniques are all employed in the field of emotional analysis to mine texts for information about people's thoughts, emotions, and attitudes. An excellent source of emotional analysis is social networking reviews. Sentiment Analysis is a term used to describe opinion mining or emotional intelligence. Analyzing unstructured and disorganized content from various social media and internet sources, such as Twitter, WhatsApp, Youtube, and Facebook conversations, is known as sentimental analysis.

	Number of Active Users	Predominant Age Groups
Twitter	350 million	18-49
TikTok	689 million	16–24
YouTube	More than 2 billion	15–25
Facebook	Nearly 3 billion	25–34
Instagram	Over 1 billion	18–34
Tumblr	More than 475 million	18–24

Fig. 1: Usage of the Social Networking

In most instances, the dialogues would be informal; the attitudes and feeling of the individuals who are taking part in the argument would be mirrored in these discussions. This provides the door for further consideration of the behavioral patterns of the persons who are taking part in the debate. It is necessary to construct rules-based automated systems that analyze data in accordance with a set of established rules to do opinion mining. Mechanical systems are also created to do opinion mining utilizing some of the ideas of machine learning. These mechanical systems are also used to perform opinion mining. There are various scenarios in which deep learning algorithms may be used to create a sentimental analysis model, all of which are discussed here. The sentimental analysis also be a useful tool for identifying the emotions expressed in the textual content. Sentimental analysis has shown to be a trustworthy source for providing insightful thoughts on a wide range of market-rolling products and breakthroughs. At the same time, sentiment analysis plays an important part in evaluating a person's opinions on certain films or other items, among other things, depending on user input made on social media platforms. Global networking may have taken the world by surprise, and the world's aspects may have been diminished as a result of this phenomenon. Using the results of this study, we suggest a novel framework Categorical LIWC and BEBRNN for the classification of people's emotions when they receive positive social media feedback. To narrow down the search space, a LIWC-based category feature selection filter has been implemented. The BEBRNN was employed as a classifier to distinguish between the different feelings. The main objective of this paper was,

- To develop a framework that is independent of the learning algorithm

- Develop a computerized framework that will precisely classify the emotions within a limited period

The remaining section of the paper can be organized as follows, Section 1 shows an overview of opinion mining impact and the paper's goal contributions; Section 2 shows an analysis of other existing technologies; Section 3 illustrates the problem statement, Section 4 shows the implementation of a novel methodology for the people emotion classification over people emotion classification; and Section 5 shows the methodology's effectiveness based on its findings. An overall summary depicted in section 6 may be used to provide a conclusion to our implemented document.

2 Related Works

With social networking sites systems such as Twitter, Facebook, Instagram, YouTube, and WhatsApp sweeping the communication world, it has become critical that the records residing all over these social networking sites systems share information relevant to the point of view, mood, and also conviction of people regarding any kind of item, suggestion, or even plans. Before this, several studies were carried out to examine the contents of social media and to undertake a perspective investigation of the material included in social media accounts. The author of [1], describes in their paper a unique deep learning system for classifying multiple emotions on Twitter. A major focus of [2], is on classifying tweets into extremist and non-extremist subcategories to create a framework for content analysis linked to terrorism. The researchers construct a tweet classification system based on user-generated social media postings on Twitter, which employs deep learning-based sentiment analysis methods to discriminate between tweets that are extremist and non-extremist. Plutchik's wheel emotion detection rule-based patterns are developed, learned, and employed in previously encountered texts in [3]. [4], used a unique multilayer bidirectional long short-term memory (BiLSTM) developed on top of pre-trained word embedding vectors. In [5], Experiments were done utilising data from Taiwan's largest internet forum, Militarylife PTT. To conduct our research, we built a social media sentiment analysis infrastructure that included a military-specific sentiment vocabulary and tested the efficacy of numerous deep learning models with various parameter calibrations. Social media messages from Bangladeshi citizens regarding the coronavirus were examined in [6]. Similar studies can be found in [7], [8]. Three classes have dealt with their emotions. moods of analysis, depression, and rage. The data collection was done in Bangla. Various deep-learning algorithms have been used in their work. In their research, they offer a semi-monitored sentiment analysis that combines the lexical method with machine learning. Eight-gram sensations are generated by employing a random sample of product reviews and numerical rating scales with five decimal places. T-test annotations provided by robots are statistically equal to those of human annotators in terms of precision and precision. Also in [9], a technique for sentiment classification based on positive and negative phrases in the text of documents is presented. To more accurately classify microblogs as sentimental, we'll need to use flexible sentimental lexicons. A genetic algorithm is used to address optimization and classification problems in this situation. Begin introducing the new community decision-making process. [10]. There should be free-form text and comparisons to various possibilities for procedures. This is a paradigm for social networking. Analysis of the free text and expert opinions on possible replacements was done using sentiment analysis. [11], [12]. The proportion of emotional data in each evaluation has been raised, although the sample sizes have remained the same. The researcher also proposed the Lexicon-Integrated 2-chain

Sensorial Model (LSTM-CNN), which combines LSTM and CNN simultaneously. It's time to reflect on the Twitter Task Sentiment Analysis's fourth year of existence. [13], analyzing Twitter sentiment is a complex process. Talk about it in depth. This is the tenth and final SemEval-2015 work item. It's the most established mission in the last three years, with 40 personnel on board. shows MLSTM as an alternative to typical hardware design approaches (Memristor depends on Long Short Term Memory). Internal structure modeling may be simplified by transferring parameters between LSTM cells. Each LSTM cell unit in the circuit has a crossbar comprised of a single memristor. Piecewise function models using equal hardware activate the sigmoid and tanh activation functions in the device under test. [14], BLSTM has offered a feasible design that includes various channel characteristics and self-attention mechanisms (SAMF-BiLSTM). When it comes to transmitting one's feelings, this strategy uses the most up-to-date instruments and strategies. Rather than relying on lexicons that must be assembled by hand, the proposed method automates the polarity of emotions and goals. According to SAMF-BiLSTM, SAMF BiLSTM-D was indicated for paper applications in the research. In [15], [16] the author Suggested analysing and mining internet customer reviews for emotion. Amazon.com reviews and other user-generated content are the major focus of this study. This causes the views to be instantly deleted. The data in this study will be analyzed using a logistic regression model. The limitation does not operate as well in this task. [17], biomedical data hypertext mining is being examined. To extract biological data from hypertext pages, this study will make use of text-mining algorithms based on "biomedical ontology" (e.g., online information mining). The conventional method of medical language analysis, the met thesaurus, is the topic of this study. The document is constrained by its lack of scientific relevance. To develop generative adversarial networks, hierarchical attention processes are employed in this research to identify the theorized and denied components in biological literature (GAN). Classifiers are used in [18], [19], [20], [21], [22], to classify positive and negative attitudes. To do sentiment analysis, they make use of the hard dataset. Use the Lexicon of Positive and Negative Polarity from the NCCA-ALSTM to classify sensations.

3 Problem Statement

Use an emotion-related grammatical category to describe the speaker's approach toward a declaration of certainty and speculative scope to find out which part of the phrase is unsure. An uncertain segment of a phrase is designated by the emotion-related grammatical category associated with a particular word (for instance, "seems," "possible"). an ineffective effort to determine the extent of a negative sign word's linguistic influence (e.g., "failed"). A few approaches have been tried in the past, but none of them seem to have reached the pinnacle of success. To effectively identify emotions, an effective method must be developed.

4 Proposed Work

The LIWC and Bidirectional encoder-based recurrent neural network model is utilized in this study to enhance emotion classification using a contemporary system, changing the conventional word form into a numerical type and categorizing emotions into positive or negative polarity. The general depiction of the suggested technique is shown in Figure 2.

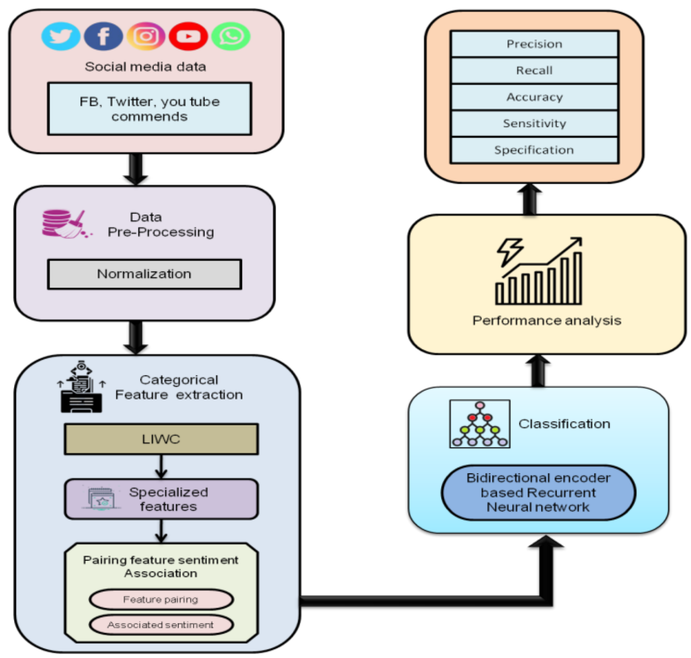

Fig. 2: Pictorial representation of the suggested methodology

a.Dataset

The dataset obtained from the kaggle belongs to FB, Twitter, and you tube

- https://www.kaggle.com/seungguini/bts-youtube-comments
- https://www.kaggle.com/mortena/facebook-comments-sentiment-analysis
- https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis

b.Preprocessing

It is necessary to do text normalization, stemming, part of speech tagging, and stop word deletion during the pre-processing step to prepare the text for the main processing stage. It is necessary to process the chosen comments to improve the accuracy of the categorization. The accuracy and performance of the system will be improved as a result of the pre-processing portions. The initial stage in this preprocessing procedure is tokenization. The input is broken down into little word bits known as tokens, which eliminates the characteristics that are of low importance. Then stop eliminating words from the matrix, since this will help to reduce the matrix's size and increase the degree of discrimination across texts. The stemming of words is then used to consider a unit for all of the words that are connected to a single root. As a result, it has little impact on the system's performance. Stop words contain grammatical terms like conjunctions and prepositions, as well as other phrases. There is no meaning or contextual information in these words. In the case of studies that are primarily concerned with word frequency, the high frequency of stop words that occur in documents produces inaccuracies in the findings of the processing. Two goals serve as the basis for the use of syntactic tagging in words. In this case, stemming is performed based on the syntactic category of the words. It may be expressed in the following way:

$$\overrightarrow{S_0} = \frac{\sum_{i=1}^{M} \overline{S_i}}{M} \tag{1}$$

Where represents the process of stemming, M is the word polarity.

The equation can be rewritten as follows,

$$S = \left(\frac{\sum_{i=1}^{M} (\overrightarrow{a_i} - \overrightarrow{a_0})^2}{M}\right)^{\frac{1}{2}} \tag{2}$$

After that, the syntactic tags may be deleted, and the data can be used to discover the most often occurring syntactic nouns, verbs, and adjectives that surpass the stated threshold,

$$D = \left(\frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (\overrightarrow{a_i} - \overrightarrow{a_0})^2}{M(M-1)}\right)^{\frac{1}{2}} \tag{3}$$

Where D represents the tags, a represents the syntax and stylistic error

Finally, the stem of the word or text is determined via the stemming procedure, which entails translating a single stem into numerous stems. The TF-IDF model incorporates information about the more important and less significant phrases, while Bag of Words is only a collection of vectors expressing the number of words in the text.

- Term Frequency: illustrates a frequency calculation.
- Inverse Document Frequency: represents the uneven records.

$$\frac{\partial s}{\partial x}\left(\int_0^s ps(x)dx\right) = \partial s(s)(x^{-1})(s))D/ds \tag{4}$$

Information from the Sentiment Lexicon is used to integrate Sentiment's texts. In our method, emotional words take priority over verbal ones. Everything in the lexicon of emotions has a positive or negative value that determines the strength and emotional range of the experience. Every value in the normalized shape set [-1, +1] may be mapped to several positive scales and values very near to 1, but Negation is explicitly forbidden. Because phrases with a clear negative or positive polarity provide more sentiment data than neural concepts, they should be given more priority intuitively. To ascertain the emotional effect of a particular word. Convergence is improved quickly using the suggested strategy,

$$X_{-1}^{+1} = \left(\frac{1}{t}\right)^{\left|\frac{High\, polarity\, f(t) - f_{i(t)}}{less\, polarity\, f(t) - High\, polarity\, f(t)}\right|} \tag{5}$$

Finally, the processed dataset was obtained.

c. Feature extraction

Social media vocabulary words are used in this study's categorical feature extraction methodology. The suggested Feature Extraction method was used to extract the target categorical features from reviews and then filter out any irrelevant categorical features that were found. Sentences were also taken from comments and then adjusted to account for any previous shifts that could have changed their polarity. There is just one word used in this method of category feature extraction: feature. To analyze a text's emotional, cognitive, and structural components, the team behind LIWC created a text analysis program. Each word in the text is searched for and matched against a term in the lexicon by LIWC. In the dictionary, some terms represent the word's linguistic, psychological, and social aspects, such as pronouns, pleasant feelings, and social processes. LIWC increases the percentage value of a word category. After categorizing every word in the document, the findings will be shown as a table with the percentages of each category's words. The procedures employed to extract categorical features based on LIWC for sentiment analysis were as follows.

1. The first step is to familiarize themselves with all the words in the text.
2. Compile a list of all the words that appear in the dictionary under each of the class labels (positive and negative)
3. Use the equation to get the feature class ratio.

$$C_j^X = [C_1^X, C_2^X, C_3^X, \ldots.. C_k^X] \tag{6}$$

The equation can be rewritten as follows,

$$C_j^{nX} = \left[C_1^{kX}, C_2^{kX}, C_3^{kX}, \ldots.. C_k^{kX}\right]\aleph_{k,\xi j,\beta} \tag{7}$$

Where C represents the features

Here
$\aleph_{k,\xi j,\beta} = \aleph_{\xi j,\beta} + \alpha \times \left(\aleph_{v,\xi j} - \aleph_{\xi j,\beta}\right) \times \Re$

As a result of these reviews being processed and the categorical feature characteristics they extracted, the review knowledge base has been expanded.

$$Au_{i,o} = \begin{cases} 1 + \log i, o, pj_{i,o} > 0 \\ 0, pj_{i,o} \leq 0 \end{cases} \tag{8}$$

Algorithm:1 Categorical Feature Extraction

Input: Processed Reviews R, youtube video comments-review.
Output: Extracted features
Insert variables
//Extracting feature based on categories
Core Concepts= [Extract (Review[i]-youtube comments)] //
Core Concept (information, entertainment, script)
Basic facts = [Extract (Review[i]-youtube comments)]
Beach view
Video name= [Extract (Review[i]-youtube comments)] //
Video name (Beach view)
//Extracting categorical feature
Comment name= Identify [feature]
Grammar analysis = Identify [emotions]
Feature matrix generation $\left(v = -\frac{1}{2}\left[\left(\sum_{j=1}^{k} \alpha_j w_j H(s_j, s^-)\right) + \left(\sum_{j=1}^{k} \alpha_j w_j H(s_j, s^+)\right)\right]\right)$
Feature score analysis $\left[Positive\ Score(m) = \sum_{l=1}^{a} TermScore(V_l)\right]$
K= (Count feature)
Do for m=i,...n
If(Core concepts[m] in uppercase(BEACH,SAND)
End if
End for
End

Then, the matrix F was generated from the review, and it was merged with the statistical matrix

$$v = -\frac{1}{2}\left[\left(\sum_{j=1}^{k} \alpha_j w_j H(s_j, s^-)\right) + \left(\sum_{j=1}^{k} \alpha_j w_j H(s_j, s^+)\right)\right] \tag{9}$$

For the last step, feature scaling (i.e., each column) and instance scaling were used to ensure that the matrix was consistent.

$$Feature\ Score = \frac{\sum_{k=1}^{d} polarity\ (r)/r}{\sum_{k=1}^{d} 1/r} \tag{10}$$

Where n denotes the number of features

Finally, the sentimental feature score was evaluated.

$$Positive\ Score(m) = \sum_{l=1}^{a} TermScore(V_l) \tag{11}$$

Whenever a negative word comes in front of a phrase, the following word's emotion score is simply reversed. Following the example below, the sentiment value of the target comments may be calculated:

$$Negative\ Score(s) = PosScore(m) + NegScore(sv) \tag{12}$$

Where s denotes a comment with m positive and sv negative phrases, the positivity and negativity of the associated comment p are indicated by PosScore and NegScore.

An association technique, shown in Figure 3, was used to link the emotive categorical traits with their related feelings, and then their sentiments were paired.

$$\varepsilon_j = \sum_{p=1}^{k} \alpha_p O_{jp*}\, \delta\alpha_j \tag{13}$$

Where,
$\delta\alpha_j = min\{\max[\gamma(1-\varepsilon_j), -\alpha_j], C - \alpha_j\}$

Finally, the features can be extracted and it can be associated with their corresponding sentiments.

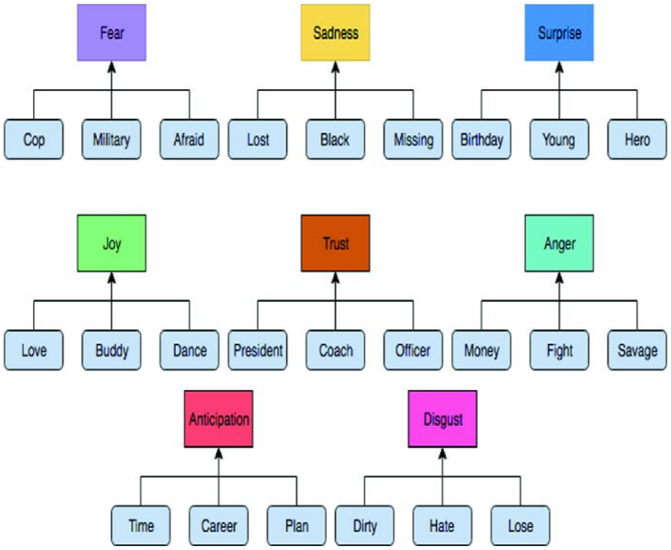

Fig. 3: Emotion association

<u>Algorithm:2 Categorical Feature Sentiment Association</u>

Input: Extracted categorical features
Output: Sentiments association
Lexicons for the extracted feature
Sentence → Analysis (Comments)
//Form Sentiment pairs
For count → (Category)
For m-1,...n
Category → (List of sentiments)
Matching Sentiments → (word Matching)(match/hot)
If
Match (Associate)
Else
Discard word
Count (Associate pairs)
End if
End

---

d. Classification

The last step in the classification process is to utilize a recurrent neural network based on a Bidirectional Encoder to categorize emotions. Most of this level involves classifying emotions as either good or negative using an RNN. For example, this RNN enables you to determine the difference between an independent variable and a shared one with ease. The proposed approach first interprets and redistributes the data before calculating the emotions during categorization using its class probability. The cycle may be started by activating a neuron with an equation.

$$Neurons(n,m) = layer(f_p(n,m)n \tag{14}$$

Where,
$f_p(n_i, m_j) = \tanh[n_i; m_j].W_{att})$

The equation can be written in the form of vectorized form as shown in the equation ,

$$f_p(V,X) = [o^1; o^2 \ldots., o^n].W_{kh} \tag{15}$$

The quadratic set in which the training set can be merged is shown in the equation ,

$$X = \frac{1}{2}\left\|Y - a^l\right\|^2 = \frac{1}{2}\sum_j (y_j - a_j{}^l)\hat{}2 \tag{16}$$

The gradient output is given by equations 17 and 18,

$$\mathcal{L}_{lpr}(\theta) =- \sum_{i=1}^{C} \hat{y}^c \log(y^c) + \mathcal{L}_{lpr} + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{17}$$

Where,
$\mathcal{L}_{lpr} =- D_{nL}(u(n)||p_\theta)$

Where y is the emotional feature

Emotion polarity can be represented by using the following equation,

$$d(x) = h(x-1) = o(x+1) \tag{18}$$

Assume,
$d(x)$ denotes Centre polarity
$h(x)$ denotes upper focus word limit
$o(x)$ denotes a lower focus word limit.

The polarity characteristics can be calculated by using the formula,

$$y(x) = \sum_{k=o(x)}^{h(x)} V_l(k)|X_b(k)| \ , \ l = 1,2, \ldots., N \tag{19}$$

The equation can be rewritten as follows,

$$V_l(y(x)) = \begin{cases} \frac{k-o(l)}{c(l)-o(l)} & , o(l) \le k \le d(l) \\ \frac{h(l)-k}{h(l)-c(l)} & , c(l) \le k \le h(l) \end{cases} \tag{20}$$

Finally, the following formula may be used to classify the data.

classify opinions$(V_l(y(x))) =$
$k_j^l R_j^l$ (21)

The CNN classification is concluded in equation 22

$$F(\text{Polarity}) = k_i^l R_i^l - m\left(k_i^l R_i^l\right)^2 \qquad (22)$$

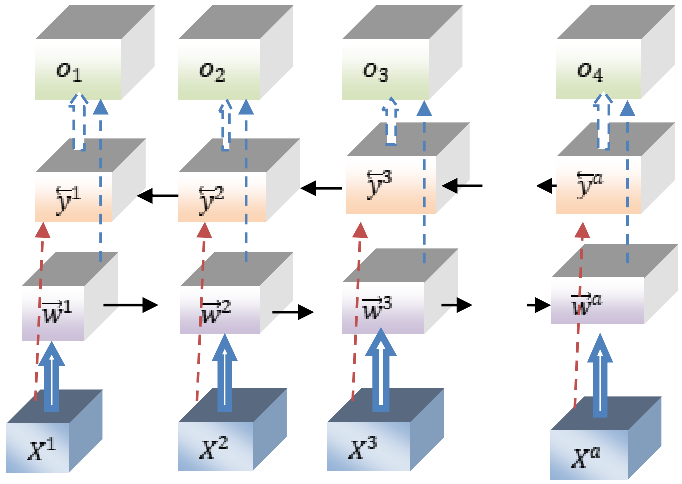

Fig. 4: BEBRNN layers

Finally, the comments can be classified depending upon the emotions.

Algorithm: 3 Opinion Classification

Input: Associate categorical features
Output: Classified opinion
Train fea= 70% of input
Test fea=30% of input
Number of Associated Categorical Features
//you tube comments review (Query box { type of emotion})
for Evaluate
Number of positive sentiments
$\updownarrow$
Number of Negative sentiments
$\updownarrow$
Neutral Sentiments
Identify group polarity;
//Insert likelihood Value to the matrix
End for
End
end
end

---

5 Performance Analysis

This section illustrates the outcome of sentiment analysis. The method was performed using the resource for emotional analysis which distinguishes positive and negative opinions. The whole experimentation was carried out in a Matlab environment.

Table 1. Classified output

Sample comments from social media	Emotions
"the anticipation of when the power is going to go out! I NEED TO STUDY WHAT IS HAPPENING STOP SANDY"	"Anger"
" Oh God #Its amazing!	"Positive"
"Sandy just made landfall on the great State of New Jersey & NYC. Hang tight, you guys".	"Anger"
"Sandy has denied me my jog. I'm crying as much as it's raining right now..."	"Anger"
"Shed in the backyard was knocked over #see you in the next video"	Other
"Lovely, there are fallen tree branches in my swimming pool. Eh, It could be worse... #413Sandy #MASandy #Sandy"	"Positive"
"So my childhood the town is being destroyed. That's cool. Stupid nature"	"Anger"
"So much food in my house because my moms stocking up for Sandy. I'm cool with it"	"Anger"
"Hurricane Sandy might not kill me but this boredom sure will"	"Anger"
"This movie was scary stuff"	"Fear"
"Hurricane Sandy is powerful af!!! This wind is NO joke!!!"	Other
"Power back on. Not sure how much longer that will last. Damn you #sandy - get up off my #raw!"	"Anger"
"I'm like really scared.... stuff like this doesn't happen in Ohio! #Sandy #Manhattan"	Fear
"NZ's embassy in Washington is closed as the city hunkers down ahead of #Sandy"	Other
"11 killed in #Cuba, #Sandy toll reaches 51 in #Haiti"	Other

Here a small set of keywords that were likely to indicate emotional content belonging to any of the emotional classes positive, fear, or anger[a]. The list of identified keywords looks as follows:
"anger: anger, angry, bitch, fuck, furious, hate, mad"
"fear: afraid, fear, scared"
"Positive: :), :-), =), :D, :-D, =D, glad, happy, positive, relieved"

A feature extraction model based on LIWC may be used to match the words that are associated with this set of keywords Lists of synonyms for the terms were added automatically. The words that emerged from this process were then culled because they were deemed to be inadequate descriptors of human emotion. As shown in Table 1, the proposed classifier can discriminate between the various emotions represented by the "positive," "anger," and "fear" tags.

To prove the efficiency of the suggested methodology it can be compared With the existing mechanisms, [23]. The proposed technique's behavior is tested using metrics such as accuracy, sensitivity, F1 score, AUC, and specificity. It is necessary to consider the following four ideas while doing this evaluation: False positives and false negatives must be distinguished from real positives and real negatives. Data values that are detected as positive by the algorithm are referred to be "TP." TN refers to data values that the system appropriately recognizes as negative. Some FPs are recognised as positive, but they aren't the exact numbers involved. That which is designated as negative but does not include the exact numbers is referred to as a "negative integer."

We also contrast the suggested technique with a few of the existing techniques concerning these parameters.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Negative + False\ Positive + True\ Negative} \tag{23}$$

$$Specificity = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{24}$$

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{25}$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{26}$$

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{27}$$

$$FPR = \frac{False\ Postive}{False\ Positive + True\ Negative} \tag{28}$$

$$Precision = \frac{TP}{TP + FN} \tag{29}$$

$$Recall = \frac{TP}{TP + FP} \tag{30}$$

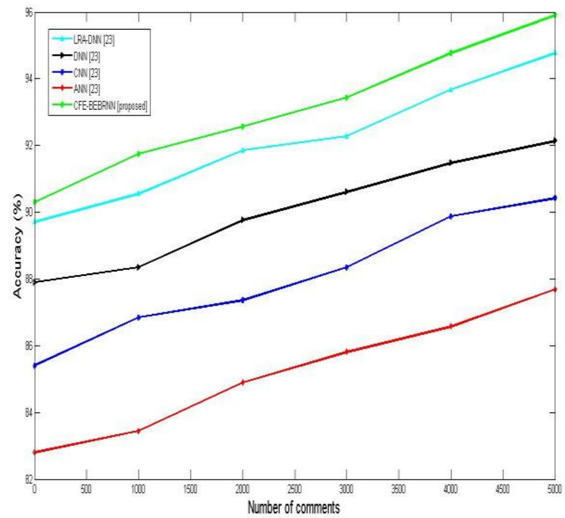

Fig. 5: Number of Comments Vs. Accuracy

Figure 5 shows the proposed classification method, showing a maximum accuracy yield of 96%, which is better than other existing methodologies.

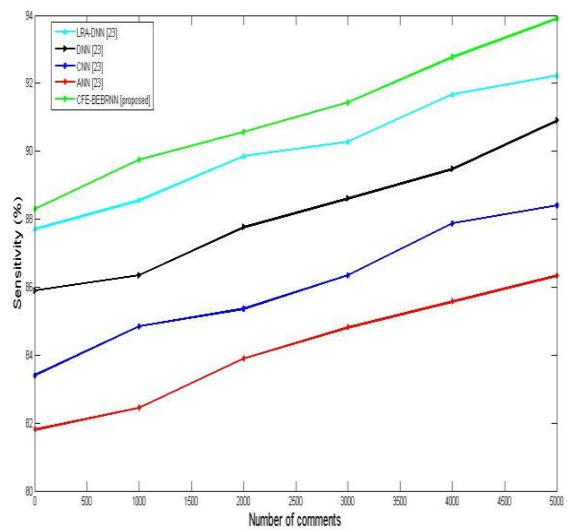

Fig. 6: Number of Comments Vs. Sensitivity

Figure 6 shows the proposed classification method, showing a maximum sensitivity yield of 94%, which is better than other existing methodologies.

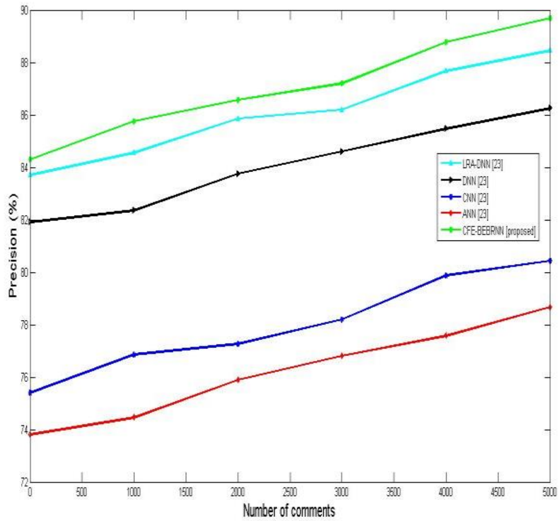

Fig. 7: Number of Comments Vs. Precision

Figure 7 shows the proposed classification method, showing a maximum precision yield of 89%, which is better than other existing methodologies.

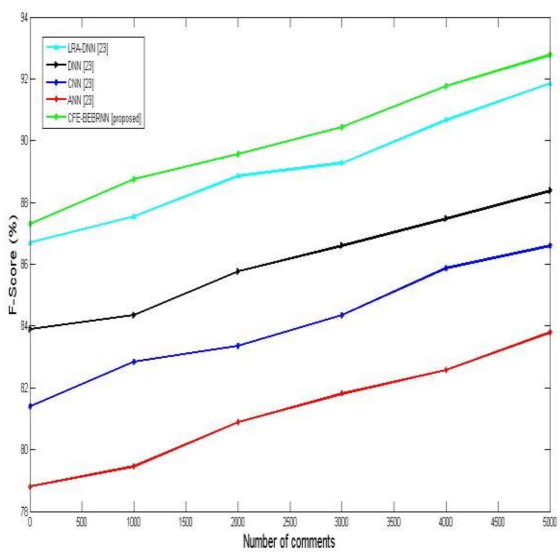

Fig. 8: Number of Comments Vs. F score

Figure 8 shows an F1 rating for the solution. The results demonstrate that, according to the suggested strategy, the coefficient of F1 score was 93 percent.

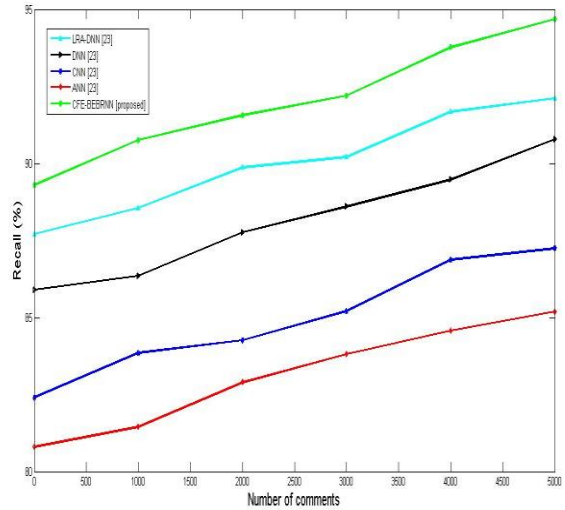

Fig. 9: Number of Comments Vs. Recall

Figure 9 shows the proposed classification method, showing a maximum recall yield of 95%, which is better than other existing methodologies.

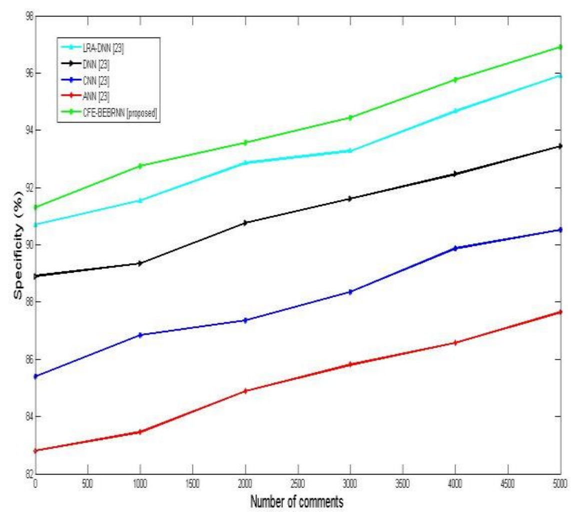

Fig. 10: Number of Comments Vs. Specificity

Figure 10 depicts the suggested method, which has a 96.5 percent specificity rate when compared to the current system.

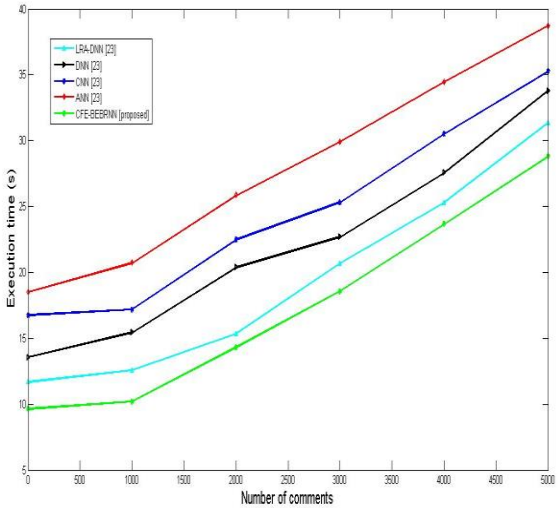

Fig. 11: Number of Comments Vs. Execution time

As of from Figure 11 the suggested methodology excecutes the classification process with in 27 seconds which is much lower than other existing mechanisms. From the result obtained it was revealed that the suggested methodology outperforms well than other existing mechanisms over emotion classification.

5 Conclusion

Researchers from a variety of fields, such as psychology, neuroscience, social science, and computer science, are working to better understand how to recognize emotions in text. identifying emotions and providing actionable suggestions has a broad range of uses, such as: enhancing the teaching models and learning outcomes from student assessments, understanding customer satisfaction surveys, and how it may assist improve the company. in this study, we employ bebrnn's categorical feature extraction to categorize the emotion class for our social media data automatically. there is a 96 percent accuracy rate with the proposed approach, which is quite high when compared to previous methods. in the future, we hope to expand our work on student evaluations of teaching by identifying student emotions toward active learning models and elements that help increase student happiness and performance via practical pattern discovery.

References:

[1] M. Jabreel and A. Moreno, "A deep learning-based approach for multi-label emotion classification in tweets," Applied Sciences, vol. 9, p. 1123, 2019.

[2] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and I. Awan, "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques," Human-centric Computing and Information Sciences, vol. 9, pp. 1-23, 2019.

[3] E. Tromp and M. Pechenizkiy, "Pattern-based emotion classification on social media," in Advances in social media analysis, ed: Springer, 2015, pp. 1-20.

[4] E. A. H. Khalil, E. M. El Houby, and H. K. Mohamed, "Deep learning for emotion analysis in Arabic tweets," Journal of Big Data, vol. 8, pp. 1-15, 2021.

[5] L.-C. Chen, C.-M. Lee, and M.-Y. Chen, "Exploration of social media for sentiment analysis using deep learning," Soft Computing, vol. 24, pp. 8187-8197, 2020.

[6] M. S. A. Pran, M. R. Bhuiyan, S. A. Hossain, and S. Abujar, "Analysis Of Bangladeshi People's Emotion During Covid-19 In Social Media Using Deep Learning," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6.

[7] F. H. Khan, U. Qamar, and S. Bashir, "A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet," Knowledge and Information Systems, vol. 51, pp. 851-872, 2017.

[8] A. Dey, M. Jenamani, and J. J. Thakkar, "Senti-N-Gram: An n-gram lexicon for sentiment analysis," Expert Systems with Applications, vol. 103, pp. 92-105, 2018.

[9] H. Keshavarz and M. S. Abadeh, "ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs," Knowledge-Based Systems, vol. 122, pp. 1-16, 2017.

[10] J. A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma, "Carrying out consensual group decision making processes under social networks using sentiment analysis over comparative expressions," Knowledge-Based Systems, vol. 165, pp. 335-345, 2019.

[11] W. Li, L. Zhu, Y. Shi, K. Guo, and Y. Zheng, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN-LSTM family models," Applied Soft Computing, p. 106435, 2020.

[12] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "SemEval-2016 task 4:

Sentiment analysis in Twitter," arXiv preprint arXiv:1912.01973, 2019.

[13] B. T. Hung, "Integrating Sentiment Analysis in Recommender Systems," in Reliability and Statistical Computing, ed: Springer, 2020, pp. 127-137.

[14] S. Wen, H. Wei, Y. Yang, Z. Guo, Z. Zeng, T. Huang, et al., "Memristive LSTM network for sentiment analysis," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019.

[15] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," Artificial Intelligence Review, pp. 1-51, 2019.

[16] H. Jiang, C. Kwong, G. O. Kremer, and W.-Y. Park, "Dynamic modeling of customer preferences for product design using DENFIS and opinion mining," Advanced Engineering Informatics, vol. 42, p. 100969, 2019.

[17] N. A. Swidan, S. K. Guirguis, O. G. Abood, and A. S. Hameed, "Text Document Clustering using Hashing Deep Learning Method."

[18] F. T. Al-Khawaldeh, "Hierarchical Attention Generative Adversarial Networks for Biomedical Texts Uncertainty Detection," International Journal of Advanced Studies in Computers, Science and Engineering, vol. 8, pp. 1-12, 2019.

[19] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," Expert Systems with Applications, vol. 72, pp. 221-230, 2017.

[20] N. Saleena, "An ensemble classification system for twitter sentiment analysis," Procedia Computer Science, vol. 132, pp. 937-946, 2018.

[21] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," Decision support systems, vol. 57, pp. 77-93, 2014.

[22] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on Twitter aspect-based sentiment analysis," Applied Intelligence, vol. 48, pp. 1218-1232, 2018.

[23] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, "An efficient way of text-based emotion analysis from social media using LRA-DNN," Neuroscience Informatics, p. 100048, 2022.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
All authors contributed equally to this work.

Compliance with Ethical Standards

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding is involved in this work.

Conflict of Interest
Authors declare that they have no conflict of interest.

Data availability statement
Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study

Declaration
Human and Animal Rights:
No violation of Human and Animal Rights is involved.

Ethical approval
All authors ensure ethical approval.

Consent to Participate
All authors have the consent to participate.

Consent to Publish
All authors have the consent to publish the paper in this journal.