DOI: 10.6977/IJoSI.202312_7(8).0002
G. J. Naidu etc./Int. J. Systematic Innovation, 7(8),12-23(2023)

Deep attention network with sentence-level classification-based sentiment analysis in Telugu considering linguistic feature

G. Janardana Naidu[1]*, M. Seshashayee[2]

[1,2]Dept. of Computer Science, GIS, GITAM (Deemed to be University)
Visakhapatnam, Gandhi Nagar, Rushikonda, Andhra Pradesh- 530045, India.
[1]jana.766@gmail.com; [2]mseshashayee@gmail.com

(Received 3 November 2023; Final version received 15 September 2023; Accepted 30 October 2023)

Abstract

Sentiment analysis in conversations has gained increasing attention for the growing number of applications like human-robot interactions. Inaccurate emotion identification in existing sentiment analysis methods is due to a lack of concentration on explicit and implicit factors in sentiment detection. Hence a novel deep attention expression analysis technique has been introduced in which a deep attention network with parallel level fuzzy classifier identifies the nature of the words using sequential word N-gram functions by incorporating distributed semantics thereby the implicit and explicit nature of the sentence is identified and classified. Moreover, negations and tone in the sentence create a perplexity nature of sentiment analysis. To solve this problem, linguistic feature-based classification has been presented that utilizes a POS-based tagging in the attention layer and BOW to provide word embedding. Then, lemmatization and stemming process of words the root words are identified by maximum likeliness probability, resulting in the identification of new words with linguistic features. Furthermore, Naïve Bayes classifier and ensemble clustering with lambda function have been used to identify the negations and tone of the sentence. Thus, the results provided accurate detection of the positive, negative, or neutral sentiment of the sentence with a high accuracy of 96% and a precision of 97%.

Keywords: Deep attention network, Sentiment analysis, N-gram function, Parallel level fuzzy classifier, Naïve Bayes classifier, PoS-based tagging.

Nomenclature:

c	specific class
t	the text we want to classify
P(c) & P(t)	prior probabilities of this class and text
P(t | c)	probability the text appears in a given sentence
|D|	dataset size
d_i	ith document
TN	True Negative Value
FP	False Positive Value
FN	False Negative Value

DOI: 10.6977/IJoSI.202312_7(8).0002
G. J. Naidu etc./Int. J. Systematic Innovation, 7(8),12-23(2023)

1. Introduction

In a world of digitalization, millions of people are connected through World Wide Web and social networking (Kujur, et al., 2020) which allows a new way of sharing content with other users. Social Networks, E-commerce websites, blogs, etc. are different ways, which allow users to generate and share their content, ideas, and opinions with others easily, which leads to generating a huge amount of data every day (Tran, et al., 2021). Many business organizations need to study these opinions of the users and people rely on the feedback provided by various users on the web (Urena, et al., 2019) which can significantly affect the buying behavior of the product. Hence, analyzing the opinions or sentiments of the user emerges as an essential field to study (Rehman, et al., 2019). Most Search Engines, Android/IOS applications, Social Networking Platforms, and Government Websites are nowadays available in Indian Languages, and this generates a huge amount of data over the Internet which allows a researcher to explore the research field in Natural language processing for Indian languages (Shelke, et al., 2020).

Telugu is a Dravidian language that is indigenous to India. It is one of the few languages with official status in more than one Indian state, alongside Hindi, English, and Bengali. Within the states of Telangana and Andhra Pradesh, as well as the city of Yanam in Puducherry, it is the first language. The Government of India has designated one out of every six languages as a classical language of India. Telugu is the most widely spoken Dravidian dialect and ranks third in terms of native speakers in India (74 million, according to the 2001 census) (Badugu, et al., 2020). Hence it becomes necessary to analyze people's opinions, feelings, evaluations, appraisals, attitudes, and emotions regarding entities such as products, services, organizations, individuals, situations, events, themes, and their attributes (Kumar, et al., 2021) in Telugu language as well.

Natural language processing (NLP) is a field of computer science and artificial intelligence that studies how human and computer languages interact (Juhn, et al., 2020). Sentiment Analysis is the subfield of NLP that deals with the extraction of sentiment from a source of data. It tries to understand the data in human language and categorize it into positive, negative, and neutral sentiments. Sentimental analysis of Code-Mix social media data allows us to deduce the underlying sentiment of a word or sentence, which has a wide range of real-world applications (Joshi, et al., 2019). Sentiment analysis has been used in various applications like a product, event, or movie feedback. The accurate forecasting of sentiment analysis in the Indian movie based on the users' opinions in the microblog data helps these industries to earn more profit (Minaee, et al., 2019). Many existing methods have been applied to sentiment analysis and the standard methods like SVM, and random forest provide considerable performance (Saad, et al., 2019). Moreover, the NLP completed nicely in marketing evaluation, competitive analysis, and locating unsuccessful gossip for threat control in massive records surroundings. Sentiment analysis in NLP (Hasan, et al., 2019) is a complicated undertaking that distributes unstructured textual content and classifies it as both a wonderful, terrible, or impartial sentiment to explain the opinions, feelings, and attitudes present in a text or a fixed of textual content (Zhang, et al., 2019). Many researchers find numerous system studying models till the specification of sentiment price is a query mark (Yang, et al., 2019) because of one-of-a-kind unstructured dataset with distinctive languages (Chiranjeevi, et al., 2019) with numerous challenges in identifying tone, polarity, and negations. Hence need arises in the implementation of a technique to carry out sentiment analysis in the Telugu language.

The main contributions of this paper are as follows:

- The nature of the words has been identified using deep Attention Expression Analysis Technique with a parallel level fuzzy classifier.
- Linguistic Feature-based Classification with a POS-based tagging in the attention layer utilizing BOW with word embedding which utilizes Naïve Bayes classifier to identify the negations in the sentence and ensemble clustering to determine the tone of the sentence.

These techniques thus help in categorizing the nature of sentences and words with effective identification of tone, negations, and new words in the sentimental analysis of Telugu language.

The content of the paper is organized as follows: Section 2 presents the literature survey; the novel solutions are presented in Section 3; the implementation results and its comparison and conclusion are in Section 4 and Section 5 respectively.

2. Literature survey

Jonnalagadda et al (Jonnalagadda, et al., 2019) employed Telugu SentiWordNet and a Rule-Based Approach for Telugu sentiment analysis. SentiWordNet was used to obtain the sentiment, and the results were confirmed using ACTSA (Annotated Corpus for Telugu Sentiment Analysis), an annotated corpus data set. However, to enhance accuracy, the work must be examined using Bi-grams and Tri-grams.

In this paper, Garapati et al (Garapati, et al., 2019) employed a Rule-Based Approach to create SentiPhraseNet. SentiPhraseNet was used to obtain sentiment, and the results were confirmed using ACTSA, an annotated corpus data set. Furthermore, the approach's weaknesses can be mitigated by increasing the number of rules and dynamism.

Priya et al (Priya, et al., 2020) suggested a novel framework that was created specifically for sentiment analysis of text data in the Telugu language. To gather and analyze sentiment in tweeter data in the Telugu language, the suggested framework was merged with the word embedding model Word2Vec, a language translator, and deep learning methodologies such as Recurrent Neural Network and Naive Bayes algorithms. In terms of accuracy, precision, and specificity, the results are promising. Furthermore, new algorithms can be added to the suggested framework in the future, and the dataset size can be extended.

Kumar et al (Kumar, et al., 2019) used a Bi-directional Recurrent Neural Network (BRNN) to improve sentiment analysis performance in regional languages. The BRNN technique has the benefit of representing high and low resource texts in a shared space, and the sentiment is assessed using a similarity metric. Based on Twitter data, the suggested method is evaluated and compared to existing methods such as Random Forest and Support Vector Machine (SVM). The proposed method's future work will focus on employing appropriate text representation and optimization techniques to improve sentiment analysis performance.

Surya Chandra et al (Suryachandra, et al.) used a descriptive analysis approach to classify Telugu Amazon reviews in their study. The research is divided into three stages: pre-processing, classification, and semantic analysis. Sentiment analysis in Natural Language Processing (NLP) is a difficult task that deals with unstructured textual content and categorizes it as either positive, negative, or neutral. Sentiment analysis is a subset of text mining that attempts to explain the thoughts, sentiments, and attitudes expressed in a text or set of textual content. Cleaning the received data, performing missing value treatment, and isolating the essential data from the reviews are all part of the pre-processing stage. Using the suggested Adabooster classifier, semantic analysis is also performed to determine the users' sentiment ratings and the compound polarity of each review. The analysis of images will be part of future work.

V.K. Singh et al (2013) in their work developed aspect-level sentiment analysis of movie reviews, a new type of domain-specific feature-based algorithm. They used SentiWordNet based scheme with two different linguistic feature selections comprising of adjectives, adverbs, and verbs and n-gram feature extraction. Also, for each movie reviewed, they calculated document-level sentiment and compared the findings to those obtained using the Alchemy API. Moreover, practically most of the reviews have a mixture of positive and negative sentiments about different aspects which leads to false sentiment detection. It should be concentrated on future work.

From the survey, to enhance accuracy, the work must be examined using Bi-grams and Tri-grams, (Garapati, et al., 2019) the approach's weaknesses can be mitigated by increasing the number of rules and dynamism, (Priya, et al., 2020) new algorithms can be added to the suggested framework in the future, and the dataset size can be extended, for (Kumar, et al., 2019) employing appropriate text representation and optimization techniques to improve sentiment analysis performance and for [Suryachandra, Palli, et al.,]the analysis of images is required as a part of future work. (Singh, et al., 2013) A mixture of positive and negative sentiments should be concentrated in the future. Hence to overcome the above-mentioned issues a novel methodology has to be implemented.

3. Deep attention network with sentence level classification-based sentiment analysis in Telugu language

Sentiment analysis has tremendous importance in assisting decision-making in a variety of applications particularly in the extraction of sentiments from text. However, detecting implicit and explicit aspects is a critical factor in determining the sentence's contextual character, which has been overlooked in earlier studies. As a result, a novel Deep Attention Expression

Analysis Technique is presented which uses a Deep Attention Network with a parallel level fuzzy classifier to recognize the nature of the words in the sentence, and then the implicit as well as explicit nature of the sentence is identified and classified by sequential word N-gram functions Furthermore, negations and tone in the sentence make sentiment analysis more difficult. As a consequence, Linguistic Feature-based Classification has been proposed with a POS-based tagging in the attention layer using BOW with word embedding to tackle the problem. To identify the negations in the sentence, the Naive Bayes classifier is used to classify the presence of word ambiguity. The sentence's tone is also determined by connecting the words using ensemble clustering and the lambda function. Furthermore, via lemmatization and stemming to identify the inflection of root words, each word is normalized with maximum likeliness probability according to semantic aspects, resulting in the identification of new words with linguistic features. Finally, the sentiment nature of the sentence is obtained by classifying either positive, negative, or neutral depending on the polarity nature of the sentence using the deep attention network. Thus, the sentiment analysis of the selected sentence is efficiently identified by the proposed methodology without any perplexity considering the explicit and implicit nature, negations, and tone. The architecture of the proposed method is represented in Figure. 1.

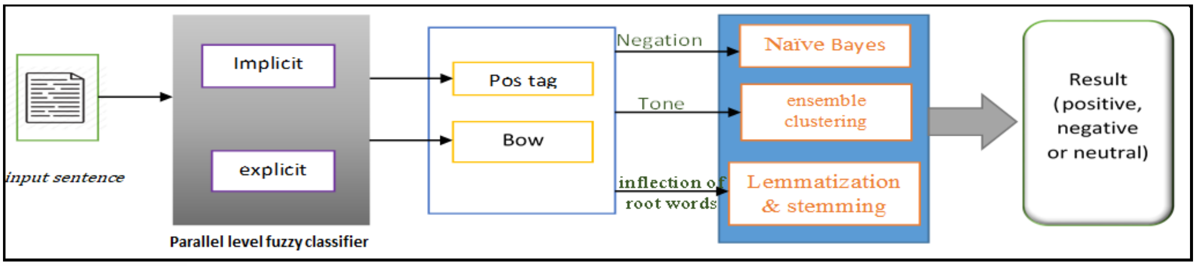

Fig. 1: Architecture of the proposed method

The input sentence is passed through a parallel level fuzzy classifier, which classifies the implicit and explicit nature of the sentence. Then to identify the tone of the sentence, the sentence is tagged with POS tagging. From the tagged sentence, the negation is identified by Naïve Bayes and tone classification is performed by ensemble clustering. Furthermore, root words were found by stemming and lemmatization. Then, the sentence is classified as either positive, negative, or neutral.

3.1 Deep attention expression analysis technique

In this technique, the implicit and explicit nature of the sentence is identified using Deep attention network with parallel fuzzy classifier and N gram functions. Before feeding data to a deep attention network, sentences have to be filtered to get a proper sentence for more effective analysis. In this process, the punctuations and stop words are avoided, and also the words that are present more than once in a sentence are omitted. Then, the preprocessed sentence was given to a deep attention network. The process flow of deep attention network is shown in the figure. 2,

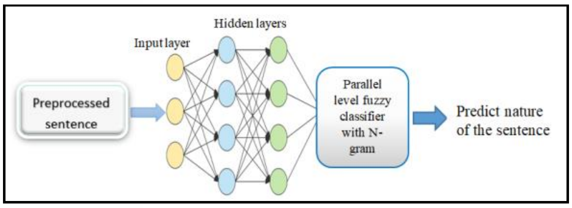

Fig. 2: Architecture of Deep Attention Network

The preprocessed sentence passed through deep attention network layers which process the sentence by ignoring some parts of the input sentence to focus on the desired words in the sentence. The deep attention layers were enabled to dynamically highlight relevant features of the input data, which is typically a sequence of textual elements. It is applied directly to the raw input or its higher-level representation. The attention layer computes a weight distribution on the input sequence by assigning higher values to more relevant elements. Then a parallel level fuzzy classifier was used to identify the nature of the sentence by using the Fuzzy Unordered Rule Induction algorithm (FURIA). Where the Fuzzy classifier groups the words into a fuzzy set with a membership function specified by truth value.

Algorithm 1: Implicit and explicit nature classification

Input: Collection of sentences $\{S1, S2, S3,...,\}$
Output: nature assigned to sentences
initialize nature
for all sentences **do**
stanford_tagger = SPOS(sentences$_i$) /∗ Applying Stanford Part-Of-Speech Tagger on each sentence ∗/
if NN in Stanford _tagger **then**
nature ← NN
end if
end for
initialize nature_groups
for all nature **do**
WordNet sets = WNSS(nature$_i$) /∗ Applying WordNet synonym set on each nature ∗/
if TRUE in WordNet_sets **then**
nature_groups ← nature$_i$
end if
end for
frequent_nature = frequency_measure (nature, nature_groups) /∗ Filtering the frequent nature ∗/
fuzzy_rules = FURIA (sentences, frequent _nature) /* Building Fuzzy rules */
initialize nature_assigned_sentences
for all sentences **do**
nature_identification = FURIA(sentences$_i$)/∗ Applying Fuzzy rules on each sentence ∗/
if TRUE in nature_identification **then**
nature_assigned_sentences ← nature_identification
end if
end for
return nature_assigned_sentences

Then the sequential word N-gram functions for sentence-level classification are used for the identification of the nature of the sentence.

A sequence of n objects in a text document, which comprises words, numbers, symbols, and punctuation are fed to N-gram models. The most frequent n sizes

DOI: 10.6977/IJoSI.202312_7(8).0002
G. J. Naidu etc./Int. J. Systematic Innovation, 7(8),12-23(2023)

are 2 (bigrams), 3 (trigrams), and 4(four grams). Considering an n-gram where the units are characters and text with t characters, where $t \in N$. There are $t - n + 1$ strings, where each string requires n units of space. Thus, the total space required for n-gram is $(t - n + 1) * n$ which is simplified in equation (1) as:

$$-n^2 + (t+1)n \quad (1)$$

Thus, the implicit and explicit nature combined the sentence classified by N-gram function. However, the negations and tone in the sentence are avoided in N-gram function, and the original nature of the sentence is not classified perfectly. To solve this issue, further sentence classification is done by linguistic feature-based classification method.

3.2 Linguistic feature-based classification

In this Linguistic Feature-based Classification method, PoS (part of speech) tagging has been done in the attention layer with BoW (bag of words) word embedding which assigns a special label to each token (word) in a text corpus to denote the part of speech as well as other grammatical categories like tense, number (plural/singular), and so on. Then if any negations are present in the input sentence which is identified by the Naïve Bayes classifier, the tone was identified by using ensemble clustering with a lambda function that evaluates an expression for a given sentence. Linguistic features are identified by utilizing maximum likeliness probability according to the semantic features by lemmatization and stemming considering the inflection of root words. During the stemming process, it removes the last few characters of the word, and lemmatization process makes the dictionary form a word, then analyzes the word's nature.

The process flow of the Linguistic Feature-based Classification method is given in Figure 3,

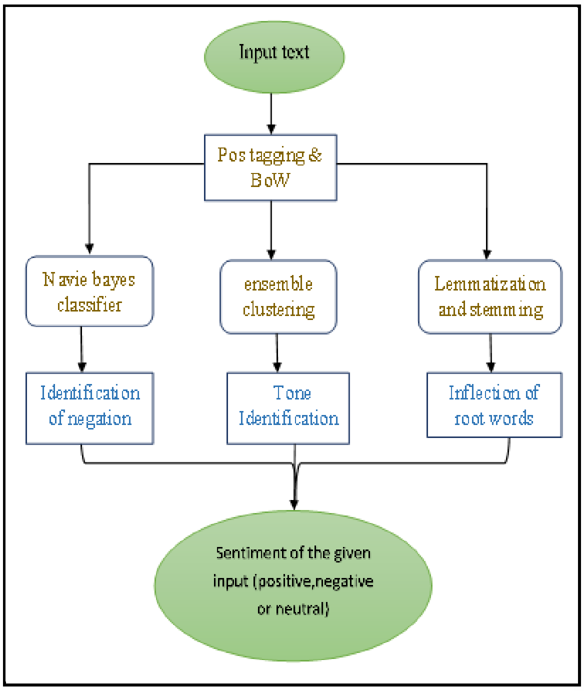

Fig. 3. Linguistic Feature-based Classification

Step 1:

Every word in a phrase is in correspondence with a particular part of speech, depending on the definition of the word and its context cataloged by part of speech (PoS) tagging which utilized BoW (bag of words) with word embedding. The PoS tagger assigns a tag to each word from BoW like JJ, JJS, VB, VBS, RB, NN, NNS, DT, etc., as shown in Table 1.

Table 1. PoS tags of the proposed system

PoS id	PoS name	PoS abbreviation
1	Noun	NN
2	Adjective	JJ
3	Verb	VB
4	Adverb	RB
5	Nouns	NNB
6	Adjectives	JJS
7	Verbs	VBS

From the tagged words naive Bayes classifier is used to determine the sequence of words in the sentence that is affected by negation words such as లేదు, కాదు which is shown in Figure 4.

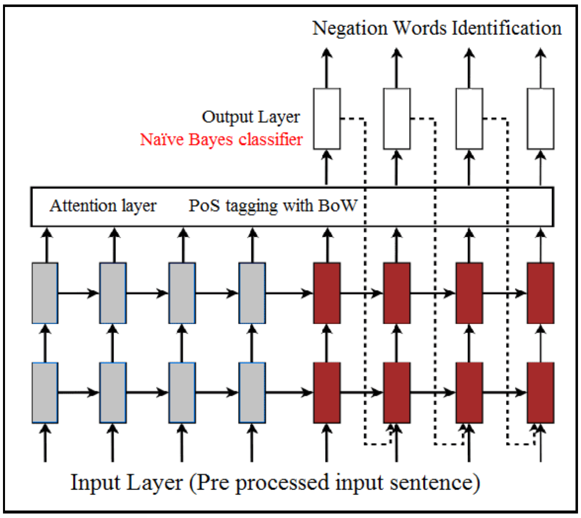

Fig. 4. Architecture of proposed negation identification

In multinomial Naive Bayes to find the probabilities of classes assigned in texts, joint probabilities of words and classes have been used. The Naive Bayes classifier assumes that the existence of one feature in a class is unrelated to the presence of any other characteristic in that class. Even though these characteristics are interdependent, they all contribute to the fact to be identified by Bayes Theorem, for a word t and a class c [22],

$$P(c|t) = \frac{P(c)P(t|c)}{P(t)} \quad (2)$$

In equation (2),

c is a specific class

t is a text we want to classify

$P(c)$ & $P(t)$ are the prior probabilities of this class and text. And

$P(t \mid c)$ is the probability the text appears in a given sentence.

In our method probability $P(c|t)$ which classifies the probability of word t occurs in class c to identify word negation in the sentence.

Step 2:

To identify the tone of the sentence correlating the words with ensemble clustering. In the group, similar words are obtained with a lambda function. The ensemble method correlates the word with a collection of words and it predicts the exact tone of the sentence.

In a review set R contain m number of reviews $\{r1, r2, ... rm\}$, then the tone classification process shown in fig. 5,

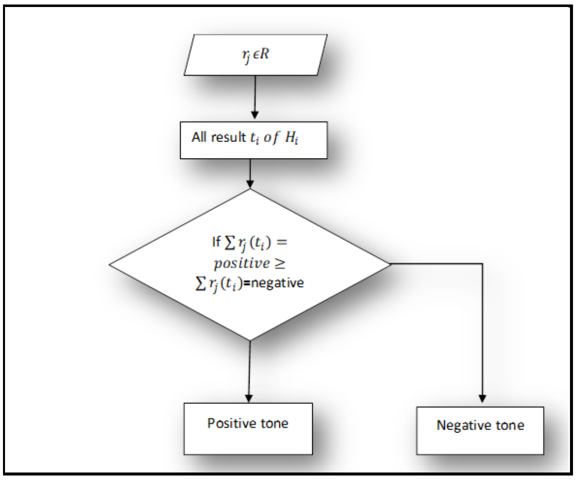

Fig. 5. Tone identification of sentence using ensemble method

Step 3:

The class prior probability is estimated as the Maximum Likelihood Estimate by following, [30]

$$P(C_k) = \frac{\sum_{i=1}^{|D|} P(C_k|d_i)}{|D|} \quad (3)$$

Where,

|D| is the dataset size

d_i is ith document

Lemmatization takes into account the context when converting a word to its meaning basic form as a Lemma. **Stemming sometimes results in stems that are not complete words so** lemmatization **was performed. For root word extraction,** lemmatization **was performed and the word was converted into a meaningful dictionary form.** Lemmatization and stemming processes have been shown in Figure. 6.

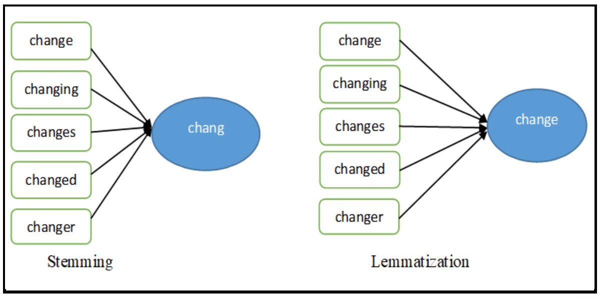

Fig. 6 Example of lemmatization and stemming

From Figure 6 in the stemming process, the words change, changing, changes, changed, and changer stemmed to chang. This word chang does not the root of the input word so that the words are lemmatized by the instance of the WordNetLemmatizer() and lemmatize() function on a single word. As a result, the

DOI: 10.6977/IJoSI.202312_7(8).0002
G. J. Naidu etc./Int. J. Systematic Innovation, 7(8),12-23(2023)

Linguistic Feature-based Classifier classified the given input sentence as positive, negative, or neutral.

Overall, the proposed deep attention network with sentence level classification based sentiment analysis considers linguistic features by utilizing the parallel level fuzzy classifier with N-gram function to find the implicit as well as explicit nature of the sentence. Also, the tone, negation, and root words were identified by naïve Bayes, ensemble clustering, Lemmatization, and stemming process by tagging the sentence with PoS tagging. Finally, the proposed method effectively analyzed the implicit and explicit nature, tone, as well as expression of the input sentence with polarity classification. The results obtained from deep attention network with sentence level classification-based sentiment analysis considering linguistic features were discussed in the next section.

4. Results and Discussion

This segment provides a detailed description of the implementation results as well as the performance of the proposed system and a comparison section to ensure that the proposed system performs valuably.

4.1 Experimental Setup

This work has been implemented in the working platform of Python with the following system specification and the simulation results are discussed below.

4.2 Dataset Description

In this work, Annotated Corpus for Telugu Sentiment Analysis (ACTSA) [23]
dataset is used to detect the sentiment analysis in social media online platforms which contains 5457 Telugu sentences. Since ACTSA is the largest resource currently accessible that compiles Telugu sentences from many sources and then goes through pre-processing and hand annotation by Telugu speakers, it is chosen as the research dataset. In the proposed system dataset was split into a training set and test set, the proposed system used the training set to train the proposed model in which 80% of the dataset was assigned to the training set while 20% of the dataset was assigned a test set. The below diagram shows the training data set,

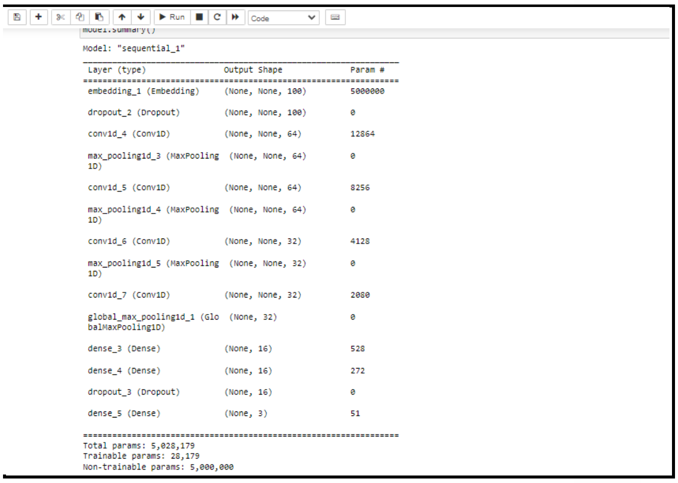

Fig. 7. Simulation view of a training data set

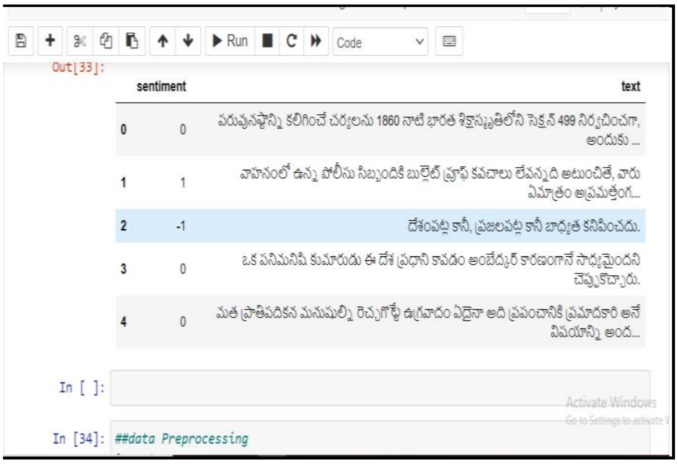

Fig. 8. Result of the proposed method

Figure. 8 After the preprocessing, the text messages are classified into three different types positive, negative, and neutral. In a deep attention network, a parallel-level fuzzy classifier divides the implicit and explicit nature of the sentence with a sequential N-gram function, which gives the result of the implicit and explicit nature of the sentence. Then linguistic features are analyzed by Pos tagging correlating with BoW. From the tagged set of data, negation words are identified by using the naïve Bayes classifier also the tone of the sentence is classified by the ensemble clustering with lambda function. Then to find the inflection of root words stemming the sentence lemmatization is done. Then extract data from the layer's deep attention network and linguistic feature-based classification the proposed dataset assigned the labels as, positive that is 1, negative that is -1, and neutral that is 0. The proposed system considers negation word, tone, and inflection of root words so that the accuracy of the proposed method increases when compared to other existing systems.

4.3 Performance metrics of the proposed system

4.3.1 Accuracy

The accuracy of the input data is calculated using,

$$\text{Accuracy} = \left[\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}\right] \quad (4)$$

TP- True Positive Value
TN- True Negative Value
FP- False Positive Value
FN- False Negative Value

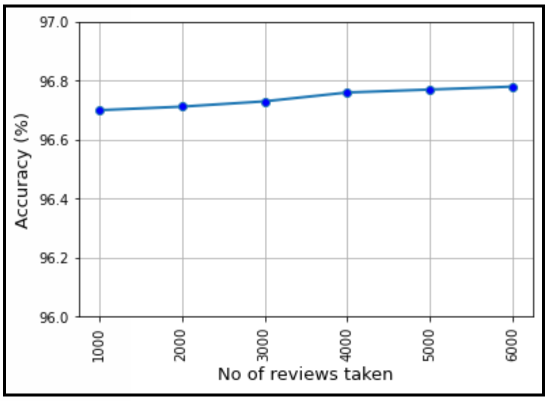

Fig. 9. Accuracy of the proposed System

The above-mentioned graph clearly explains the accuracy of the proposed system. From the graph, as the number of reviews is increased from 1000 to 6000 the proposed system's accuracy is also increased. The proposed system by using a deep attention network in which the perplexity is reduced by using the lambda function thereby increasing the accuracy of the proposed system from 96.70 to 96.79 with an increase in reviews.

4.3.2 Recall

The recall of the input data is calculated using,

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (5)$$

TP- True Positive Value
TN- True Negative Value

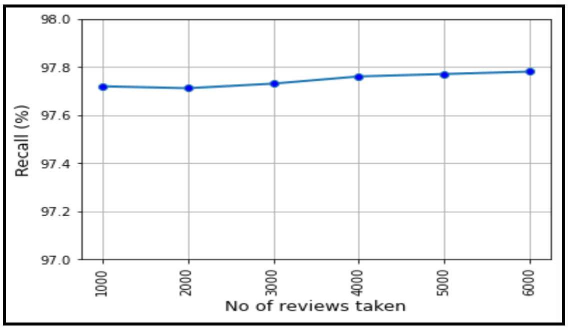

Fig. 10. Recall of the proposed System

Figure. 10 concludes that the number of reviews increases recall of the proposed system also increases. When the number of reviews taken is 1000 then recall is 97.70% and recall increases corresponding with the number of reviews increase. At the point of 6000 reviews recall is the maximum value of 97.79%. Due to the normalization of the word according to the semantic features by Lemmatization and stemming recall is increased in our proposed method.

4.3.3 Precision

The precision of the input data is calculated using,

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

TP- True Positive Value
FP- False Positive Value

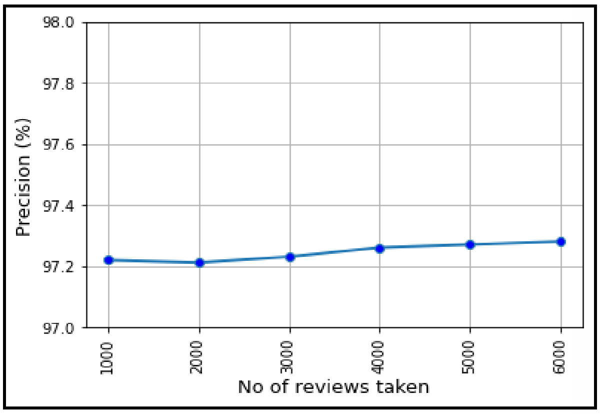

Fig. 11. The precision of the proposed System

The precision of the proposed system is shown in Figure 11. It shows the precision of the proposed system is increased when the number of reviews taken increases. When the number of reviews is 1000, 2000, 3000, 4000, 5000, or 6000 then the corresponding precision values are 97.21%, 97.21%, 97.23%, 97.25%, 97.27%, 97.30% respectively. consideration of negation and tone of the sentence by using the naïve Bayes classifier increases the precision value of our proposed method.

4.3.4 F1-Score

The F1-Score of the input data is calculated using,

$$F1=2*\frac{Precision*Recall}{Precision+Recall} \quad (7)$$

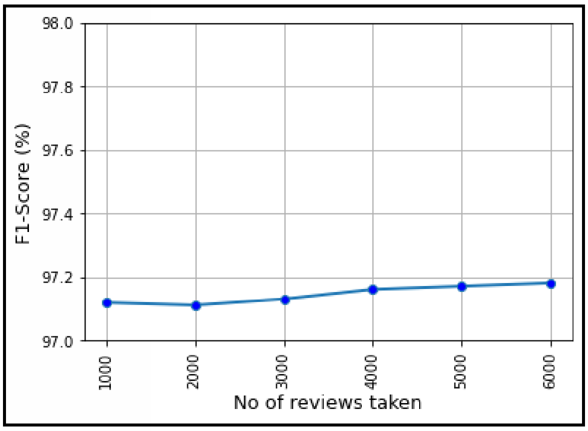

Fig. 12. F1 score of the proposed System

From Figure 12 when we increase the number of reviews from 1000 to 6000 also F1-score increases from 97.1% to 97.19%. From the graph, the number of reviews is increased as well and the proposed system f1 score is also increased in which the deep attention network incorporates a parallel level fuzzy classifier with sequential N-gram function. The overall performance of our system is given in Fig. 13,

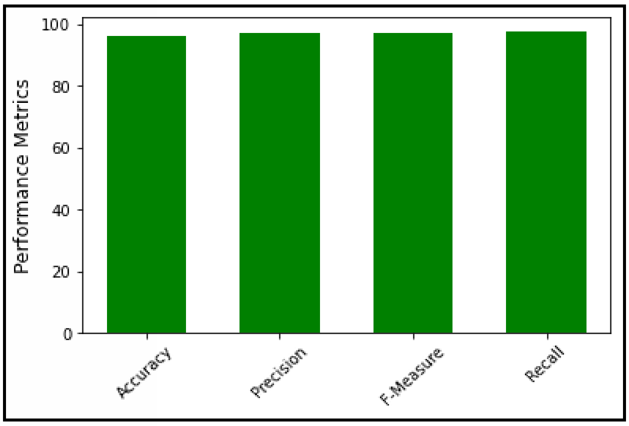

Fig. 13. Overall performance of the proposed System

4.4 Performance comparison of the proposed method:

This section describes the various performances of the proposed method compared with the results of previous methodologies like MNB+LSTM, naïve Bayes, SVM, Random forest, stacking, Ada-boost.

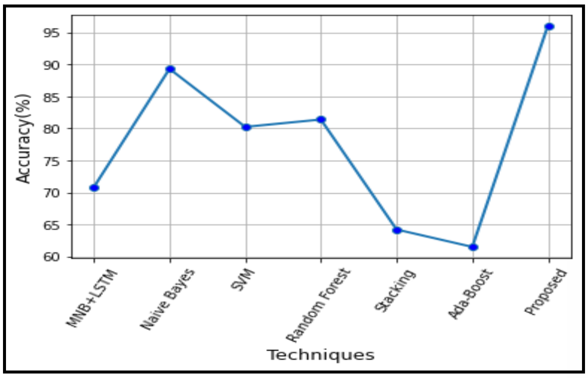

Fig. 14. Accuracy comparison of the proposed System

Figure. 14 shows the accuracy of our proposed system is 96% which is higher than the other methods that are taken for comparison. The accuracy of MNB+LSTM [27], naïve Bayes [25], SVM [28], Random forest [24], stacking [29], Ada-boost [26] are 71%, 89%, 80%, 82%, 64%, 61% respectively. This clearly shows that our proposed system was performed better than other methods.

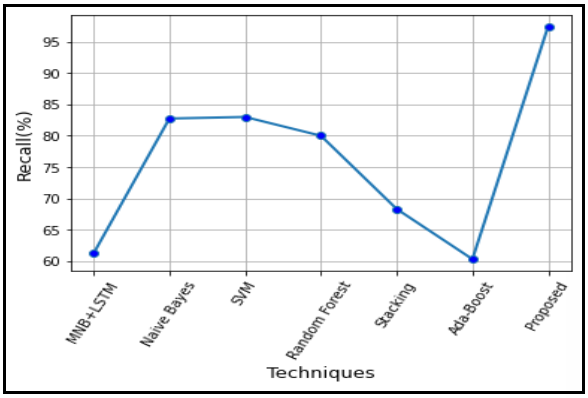

Fig. 15. Recall comparison of the proposed System

When recall is compared to the existing systems such MNB+LSTM, naïve Bayes, SVM, Random forest, stacking, Ada-boost the proposed system recall is higher than the existing systems as shown in fig. 15. The proposed recall is 98% highest when compared with the recall of other existing techniques. Recall percentages of existing systems like MNB+LSTM, naïve Bayes, SVM, Random forest, stacking, and Ada-boost are 61%, 83%, 83%, 80%, 67%, and 60%. Hence the proposed system has the highest value of 98% than other methods.

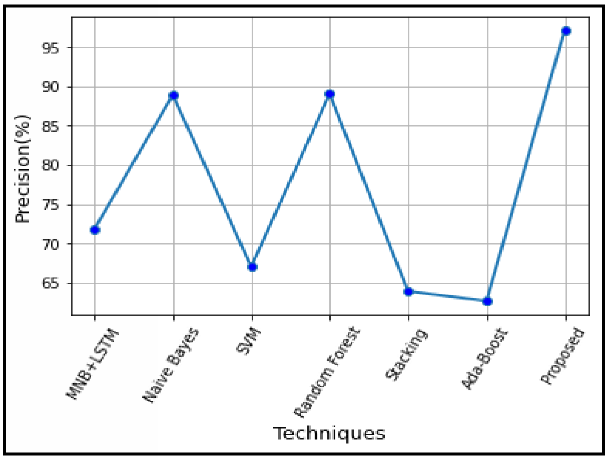

Fig. 16. Precision comparison of the proposed System

Precision comparison of the proposed system is shown in Figure 16. The precision value of MNB+LSTM is 72%, The precision value of naïve Bayes is 88%, The precision value of SVM is 63%, The precision value of Random forest is 88%, The precision value of stacking is 64%, The precision value of Ada-boost is 62%. When it comes to the proposed system precision value reaches a maximum of 97%. This shows that the proposed method performed better compared to other methods.

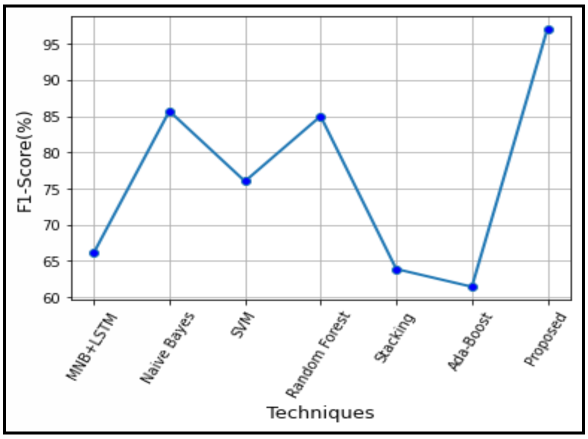

Fig. 17. F1-score comparison of the proposed System

The above figure. 17 shows the F1-score comparison of the proposed system. When a proposed method has a value of 97%, other comparison methods have the lowest value than the proposed system. Corresponding values for MNB+LSTM, naïve Bayes, SVM, Random forest, stacking, and Ada-boost methods are as follows 66%, 86%, 76%, 85%, 64%, 61%. These values are lower than the proposed value of 97%. This proves that the proposed system has better performance.

Overall, the proposed sentiment analysis technic outperforms existing methods like MNB+LSTM, naïve Bayes, SVM, Random forest, stacking, Ada-boost. The proposed method achieved an accuracy of 96% using deep attention analysis technique with linguistic feature-based classification by considering tone, negation, and root words.

5. Conclusion

Finding annotated Telugu datasets for NLP applications such as POS tagging and sentiment analysis is difficult. Hence in this work classification of sentiment was analyzed with a deep attention network which determines the implicit and explicit future of the sentence by parallel-level fuzzy classifier and N-gram function. Also, linguistic feature-based classification used to identify the negation, tone, and root words is identified and classified by using naïve bays classifier, ensemble clustering, lemmatization, and stemming. Finally, the extracted value from the attention network and linguistic classification proposed system classified as positive, negative, or neutral with an accuracy of 96% with a system precision of 97%. The recall of the approach is increased by 98%, with a 97% F1-score, and also depicted enhanced performances comparatively with other existing approaches.

References

Badugu, & Srinivasu. (2020). Telugu Movie Review Sentiment Analysis Using Natural Language Processing Approach. *Data Engineering and Communication Technology*. Springer, Singapore, 685-695.

Chiranjeevi, P., Teja Santosh, D., & Vishnuvardhan, B. (2019). Survey on Sentiment Analysis Methods for Reputation Evaluation. *Cognitive Informatics and Soft Computing. Springer, Singapore,* 53-66.

Garapati, Abhinav, et al. (2019). SentiPhraseNet: An extended SentiWordNet approach for Telugu sentiment analysis. *International Journal of Advance Research, Ideas and Innovations in Technology*.

Hasan, Mahmudul, Ishrak Islam, & Azharul Hasan, K. M. (2019). Sentiment Analysis Using Out of Core Learning. *2019 International Conference on*

Electrical, Computer and Communication Engineering (ECCE). IEEE.

Jonnalagadda, Pravarsha, et al. (2019) A rule-based sentiment analysis in Telugu. *International Journal of Advance Research, Ideas and Innovations in Technology*.

Joshi, & Aditya. (2019). *Sentiment Analysis and Opinion Mining from Noisy Social Media Content*. Diss. PhD Thesis, International Institute of Information Technology Hyderabad.

Juhn, Young, & Hongfang Liu. (2020). Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy and Clinical Immunology,* 145(2), 463-469.

Kujur, Fedric, & Saumya Singh. (2020). Visual communication and consumer-brand relationship on social networking sites-uses & gratifications theory perspective. *Journal of theoretical and applied electronic commerce research,* 15(1), 30-47.

Kumar, R. G., & Shriram., R. (2019). Sentiment analysis using bi-directional recurrent neural network for Telugu movies. *Int J Innov Technol Explor Eng,* 9(2), 241-245.

Kumar, Satheesh, R., et al. (2021). Exploration of sentiment analysis and legitimate artistry for opinion mining. *Multimedia Tools and Applications,* 1-16.

Minaee, Shervin, Elham Azimi, & AmirAli Abdolrashidi. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. *arXiv preprint arXiv:1904.04206*.

Priya, Balakrishna, G., & Usha Rani, M. (2020). A Framework for Sentiment Analysis of Telugu Tweets. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(6).

Priya, Balakrishna, G., & Usha Rani, M. (2020). A Framework for Sentiment Analysis of Telugu Tweets. *International Journal of Engineering and Advanced Technology (IJEAT),* 9(6).

Rehman, Anwar Ur, et al. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications,* 78(18), 26597-26613.

Saad, Shihab Elbagir, & Jing Yang. (2019). Twitter sentiment analysis based on ordinal regression. *IEEE Access,* 7, 163677-163685.

Shelke, Mahesh B., & Sachin Deshmukh, N. (2020). Recent Advances in Sentiment Analysis of Indian Languages. *International Journal of Future Generation Communication and Networking,* 13(4), 1656-1675.

Singh, Vivek Kumar, et al. (2013). Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. *2013 International mutli-conference on automation, computing, communication, control and compressed sensing (imac4s)*. IEEE.

Suryachandra, Palli, & Venkata Subba Reddy, P. Classification of The Sentiment Value of Natural Language Processing in Telugu Data Using Adabooster Classifier.

Suryachandra, Palli, & Venkata Subba Reddy, P. Classification Of The Sentiment Value Of Natural Language Processing In Telugu Data Using Adabooster Classifier.

Tran, Huyen Thi Thanh, et al. (2021). Social Media Insights During the COVID-19 Pandemic: Infodemiology Study Using Big Data. *JMIR Medical Informatics*, 9(7), e27116.

Urena, Raquel, et al. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Information Sciences,* 478, 461-475.

Yang, Chao, et al. (2019). Aspect-based sentiment analysis with alternating coattention networks. *Information Processing & Management,* 56(3), 463-478.

Zhang, Yuebing, et al. (2019). A cost-sensitive three-way combination technique for ensemble learning in sentiment classification**.** *International Journal of Approximate Reasoning,* 105, 85-97.