# Sentiment analysis of student feedback in Telugu language

**E.SUNEETHA**[1]
Assistant Professor, Department of Computer Science and engineering

**P.V.V.SANDEEP**[2]
University college of Engineering (A) Jawaharlal Nehru technological university, Kakinada-3, East Godavari district, Andhra Pradesh, India

**Abstract**

In recent times, Sentiment Analysis in low resourced languages and regional languages has become emerging areas in Natural Language Processing. Researchers have shown greater interest towards analyzing sentiment in Indian languages such as Hindi, Telugu, Tamil, Bengali, Malayalam, etc. In best of our knowledge, microscopic work has been reported till date towards Indian languages due to lack of annotated data set. In this Project, we are going to implement a two-phase sentiment analysis for Student Feedback using Telugu SentiWordNet which consists of 25000 annotated lexicons created based on dictionary approach. Initially, it identifies subjectivity classification where sentences are classified as subjective or objective. Objective sentences are treated as neutral sentiment as they don’t carry any sentiment value. Next, Sentiment Classification has been done where the subjective sentences are further classified into positive and negative sentences.

**Keyword*s:*** Natural Language Processing; Sentiment analysis; SentiWordNet; Lexicon; Subjectivity; Objective.

## 1. Introduction

Users can share their opinions and sentiments on a variety of topics in new interactive forms where users are not only passive information receives.. Because of the importance of this information in several areas (political, commercial or individual), it would be interesting to treat opinions automatically. The term “sentiment analysis” is used to refer to the automatic processing of opinions, sentiments and subjectivity in texts. This field is known as the opinion mining or sentiment analysis.

In Natural language processing (NLP), sentiment analysis is a technique that deals with analyzing the emotions, sentiments, opinions of an individual towards a product, movies, events, news or organizations, etc. The Primary task of sentiment analysis is to identify the polarity of a text in a given document. The polarity may be either positive, negative or neutral.

Sentiment analysis can be applied to text in three categories namely, sentence level, document level, and aspect level. Sentence level analysis focuses on identifying sentence-wise polarity value in a given document. Document level analysis determines the polarity value based on consideration of the whole document. In aspect level analysis, it identifies the polarity of every aspect (word-wise) in a given text.

Telugu is the second most popular language in India after Hindi. According to ethnologue list of most-spoken languages worldwide, Telugu ranks fifteenth in the list, and a total of 85 million Telugu native speakers exist across the world.

SentiWordNet is a lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. According to esuli and sebastiani “SentiWordNet is the result of the automatic annotation of all the synsets of WordNet towards the notions of positivity, negativity, and neutrality”. Each synset is associated with three numerical scores Pos(S), Neg(S), and Obj(S) which indicate “positive”, “negative”, and “objective” i.e., neutral respectively.

There exist several sentiment analyzers for the English language but, in the context of Indian languages, little work has been done. The primary reason behind is the lack of the available resources in Indian languages.

In this paper , we proposed a document-level sentiment analysis and sentence level analysis for Telugu student feedback. The document level sentiment analysis level is to determine the overall opinion of the document sentiment analysis at document level assumes that each document expresses opinions on a single entity. And the sentence-level sentiment analyzer level is a two-step sentiment analysis. In subjectivity analysis and sentiment analysis. In subjectivity analysis, we classify the subjective and objective sentences from the given corpus. Further, we analyze the sentiment of subjective sentences either negative or positive. The objective sentences are treated as neutral sentences as it doesn’t carry any sentiment value for the sentence. Therefore, in the first phase, the system classify the sentences as either subjective (positive, negative) or objective (neutral). In the second phase, the system classify the subjective sentences as either positive or negative.

The rest of the paper is organized as follows: Section 2 we present some basic definitions helpful in understanding the main idea of this work. Section 3 explains the proposed model for sentiment analysis. Programs and algorithms discussed in Section 4. Section 5 explains the Equations Section 6, the results of the experiments are analysed. A summary and conclusions are presented in Section 7.

## 2. Sentiment analysis – basic definitions

Sentiment Analysis is the process of ‘computationally’ determining whether a piece of writing is positive, negative or neutral. It’s also known as **opinion mining**, deriving the opinion or attitude of a speaker. Sentiment analysis involves analyzing behavior, interests and disgusts, thoughts, views, and attitude of a person from the content generated from the web. It is an emerging sub-discipline at the structure of natural language processing and information retrieval, which has numerous commercial and academic applications, which deals with hidden information of the text.

### 2.1. *LIWC Dictionary*

Linguistic inquiry and word count software tool to explore various text corpora for hidden insights from linguistic patterns.

### 2.2. *Strip & Split*

Strip and split both are the methods of the string class. Both served different-different purposes. The strip method is used to strip the particular substring from ends of the given string whereas split method is used to split the string based on some delimiter.

/n – New line

The strip function eliminates whitespaces from both ends of the string.

The split function splits the string into smaller strings with a delimiter (that is literal space).

### 2.3. *Tkinter:*

The tkinter (Toolkit intergace) package is the standard python interface to the Tk Gui toolkit. Most of the time, tkinter is all you really need, but a number of additional modules are available as well. The Tk interface is located in a binary module named _tkinter. This module contains the low-level interface to Tk, and should never be used directly by application programmers. It is usually a shared library, but might in some cases be statically linked with the Python interpreter.

While working with GUI one may need  to open files and read data from it or may require to write data in that particular file.

From tkinter.filedialog import askopen filename:

One can achieve this with the help of open() function but one may not be able to select any required file unless provides a path to that particular file in code.

### 2.4. *Tkinter standard dialog boxes :*

There are many common programming tasks that can be performed using pre-defined Gui dialog boxes.

Messages –message box

Single value data entry.

If you want to ask the user for a single data value, either a string, integer or floating point value, you can use a simple dialog object.

### 2.5. *File Chooser:*

A common task is to select the names of folders and files on a storage device. This can be accomplished using filedialog object. These dialog boxes always return you a “fully qualified file name” that includes a full path to the file. Also note that if a user is allowed to select multiple files, the return value is a tuple that contains all of the selected files. If a user cancels the dialog box, the returned value is an empty string.

## 3. Proposed scheme

In this section, we proposed an sentiment analyzer for Telugu student feedback sentences. A model is shown in Figure 1. It Starts with data collection and annotation. Further, using Telugu SentiWordNet, it classifies the sentiment of each sentence in news corpus. Finally, it compares the classification result with the automatic annotated result for error analysis.

### 3.1. *Text Data gathering:*

In this paper, data has been gathered through internet or students in Telugu Language with Student feedback of a faculty member in a sentence form or in a document form to analyze the data. The Student feedback is a data collection of negative, positive, and neutral words. This dataset is provided to the four annotators who have proficiency in the telugu language, and belong to states of Andhra Pradesh and telangana to annotate the sentiment of sentences in the dataset. This Manually annotated data is uses as the baseline for comparison with system result.

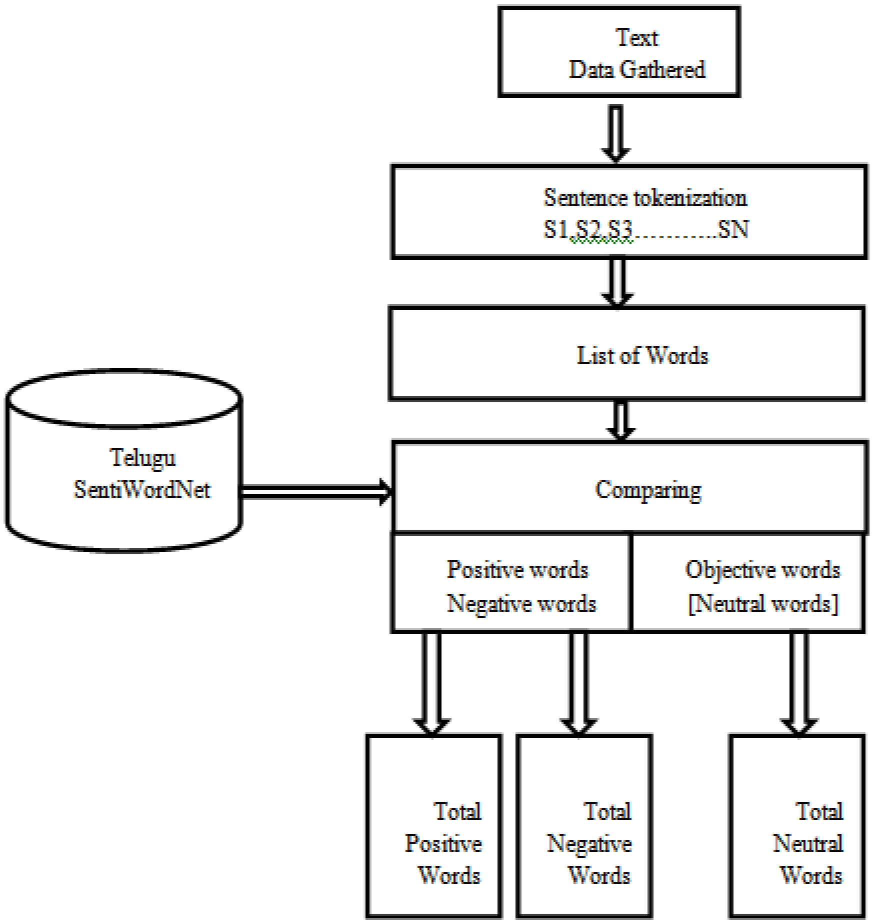


Fig 1. Model for Sentiment analysis

### 3.2. *SentiWordNet for Sentiment Analysis*

SentiWordNet is a sentiment lexicon that associates the sentiment information to each word synset. We can represent SentiWordNet as Wordnet + sentiment information. In this paper, we have used Telugu SentiWordNet consists of four files which contain negative, positive, neutral and ambiguous words respectively. The words in each file are categorized into five parts-of-speech tags namely, adjective (a), noun (n), adverb (r), verb (v) and unknown (u). We have used neutral words file for the subjectivity classification, negative and positive words file for the sentiment classification. The list of words in the Telugu SentiWordNet and their categorization is shown in  Table 1.

Table 1: Telugu SentiWordNet Data categorization

| | *Negative* | *Positive* | *Neutral* | *Ambiguous* |
|---|---|---|---|---|
| Adjective | 1116 | 659 | 86 | 515 |
| Noun | 1066 | 544 | 124 | 320 |
| Verb | 833 | 363 | 60 | 156 |
| Adverb | 102 | 90 | 11 | 6 |

| | | | | |
|---|---|---|---|---|
| Unknown | 959 | 480 | 78 | 96 |

### 3.2.1 Subjectivity Classificaton.

The subjectivity classification which takes the corpus of telugu news sentences as the input and outputs the subjective news sentences (SNS) file. It has performed by comparing each word in the sentence with the SentiWordNet neutral keywords file. If the word is present, the sentences are treated as objective sentences and discards in this level as they don’t carry any sentiment value and the remaining are treated as subjective sentences and stores in SNS file.

### 3.2.2 Sentiment Classificaton.

The sentiment classification which takes the corpus of subjective news sentences (SNS) as the input and outputs the sentiment of a sentence. It has performed by comparing each word in the sentence with the SentiWordNet Positive keywords file (poskf) and negative keywords file (negkf). If the word is present in poskf, the sentiment of that sentence is considered as positive, and if the word is present in negkf, the sentiment of that sentence is considered as negative. Otherwise, the sentence is simply discarded as any word of that sentence is not matched with any of the keywords in negkf and poskf.

## 4. Programs and Algorithms:

**Algorithm 1: Test**

```
positive_list<- []
negative_list<- []
neutral_list<- []
pos<- 0
neg<- 0
neu<- 0
with open("Telugu_SentiWordNet/TE_NEG.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
for line in file:
   ENDFOR
with open("Telugu_SentiWordNet/TE_NEU.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
for line in file:
   ENDFOR
with open("Telugu_SentiWordNet/TE_POS.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
for line in file:
   ENDFOR
count<- 1
with open("Reviews_Dataset/book_neg.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
for line in file:
pos<- 0
neg<- 0
neu<- 0
      IF len(arr) > 0 AND (line.startswith('___________________________') = False):
for word in arr:
         ENDFOR
pos<- pos + neu
```

```
        IF pos = 0 AND neg = 0:
            OUTPUT str(count)+" neutral"
        ELSEIF pos> 0 AND neg> 0:
            ENDIF
        ELSEIF pos> 0 AND neg = 0:
        ELSEIF pos = 0 AND neg> 0:
neg<- neg/len(arr)
            OUTPUT str(count)+" negative "+str(neg)
        ENDIF
    count<- count + 1
```

**Algorithm 2 : TeluguSentimentDetection**

```
import tkinter
importnumpy as np
importmatplotlib.pyplot as plt
main<- tkinter.Tk()
main.title("Telugu Sentiment Detection Using SentiWordNet") #designing main screen
main.geometry("1300x1200")
global filename
positive_list<- []
negative_list<- []
neutral_list<- []
global variable declaration pos,neg,neu,fscore,pos_count,neg_count,count
FUNCTION uploadSentiwordnet():
text.delete('1.0', END)
text1.delete('1.0', END)
text2.delete('1.0', END)
text.insert(END,"Sentiment Negative Words List\n\n");
text1.insert(END,"Sentiment Neutral Words List\n\n");
text2.insert(END,"Sentiment Positive Words List\n\n");
with open("Telugu_SentiWordNet/TE_NEG.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
    ENDFOR
with open("Telugu_SentiWordNet/TE_NEU.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
    ENDFOR
with open("Telugu_SentiWordNet/TE_POS.txt", "r",encoding="utf-8") as file:  #reading LIWC dictinary
ENDFUNCTION
    ENDFOR
FUNCTION uploadSentences(): #function to upload tweeter profile
global filename
filename<- filedialog.askopenfilename(initialdir="Reviews_Sentences")
pathlabel.config(text=filename)
ENDFUNCTION

FUNCTION sentimentAnalysis():
      IF len(arr) > 0 AND (line.startswith('____________________________') = False):
count<- count + 1
for word in arr:
            IF word in neutral_list:
```

```
neu<- neu + 1
        ENDIF
        IF word in negative_list:
neg<- neg + 1
        ENDIF
        IF word in positive_list:
pos<- pos + 1
        ENDIF
      ENDFOR
pos<- pos + neu
      IF pos = 0 AND neg = 0:
      ENDIF
count<- count + 1
    ENDIF
   ENDFOR
text3.insert(END,"\n\nAccuracy : "+str(((fscore*100)/count)*20)+"\n\n\n\n\n")
ENDFUNCTION

FUNCTION graph():
height<- [count,pos_count,neg_count]
bars<- ('Total Sentences', 'Positive Sentences','Negative Sentences')
y_pos<- np.arange(len(bars))
plt.bar(y_pos, height)
plt.xticks(y_pos, bars)
plt.show()
ENDFUNCTION

font<- ('times', 16, 'bold')
title<- Label(main, text='Sentiment Analysis Using Telugu SentiWordNet', justify=LEFT)
ENDIF
graphbutton.config(font=font1)
font1 <- ('times', 12, 'bold')
text=Text(main,height=20,width=20)
text3.config(font=font1)
main.config(bg='brown')
   main.mainloop()
```

## 5. Equations

In Algorithm 2, there is a high chance that some words in the sentence are matched with the negative keywords file, and some words in the same sentence are matched with positive keywords. In that scenario, it is hard to decide the sentiment of the sentence. To resolve this issue, we are keeping count variable to identify this kind of sentences. If the count is greater than one, the sentence is matched in both the lists poskf and negkf. So, we are adopting sentiment score to identify the actual sentiment of a sentence. To find the sentiment score of the sentence, calculate the number of positive words (PWS) and negative words (NWS) in the same sentence. Then, Calculate the positive ratio and negative ratio and Total sentiment score of the sentence using the equations, 1,2 and 3 respectively.

Formula for ratio

*Positive Ratio =*
*Total no positive words/total no of words in sentence* (1)

*Negative Ratio =*
*Total no negative words/total no of words in sentence.* (2)

*Sentiment Score = Positive ratio – Negative Ratio* (3)

There are three statistical parameters namely, precision, recall and F- score are also evaluated to test the performance of the experimented work using the equations 4, 5, 6 respectively. The results are shown in terms of statistical parameters for subjectivity classification and sentiment classification in Table 2.

$$Precision = \frac{Tp}{Tp + Fp} \tag{4}$$

$$Recall = \frac{Tp}{Tp + Fn} \tag{5}$$

$$F - Score \ = \frac{2 * Precision * Recall}{Precision + \ Recall} \tag{6}$$

Where,
T$p$ = true positive, F$p$ = false positive, F$n$ = false negative

TABLE 2: Results in terms of Accuracy, Precision, Recall, F – score

| | Accuracy | Precision | Recall | F – Score |
|---|---|---|---|---|
| Subj-Class | 80% | 0.83 | 0.72 | 0.81 |
| Senti – Class | 83.47% | 0.71 | 0.64 | 0.70 |

Where,
Subj – class = Subjectivity classification,
Senti – class = Sentiment Classification.
This result entirely depends on the quality of SentiWordNet. The Obtained accuracy can be improved by improving the Telugu SentiWordNet.

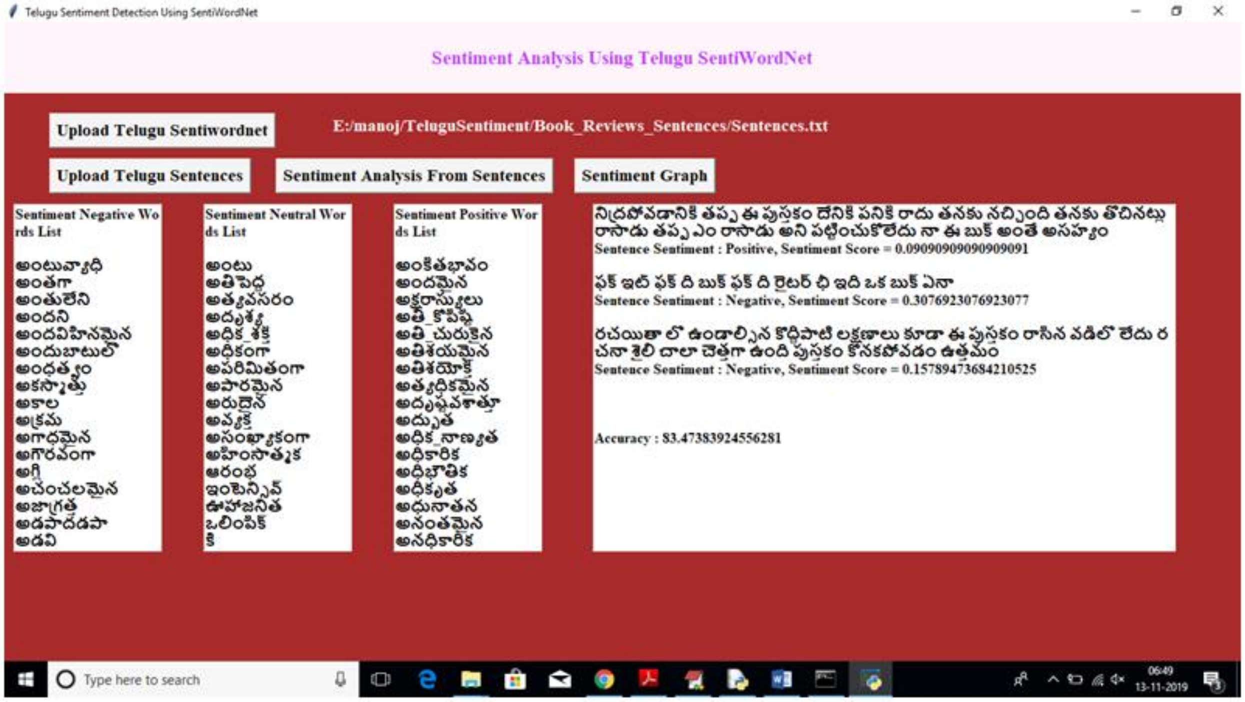
## 6. Results of the experiments

We data gathered Student Feedback sentences from students in Telugu language to detect sentiment and downloaded SentiWordNet List also.

Student Feedback sentences store inside 'Reviews – Sentences' folder and SentiWordNet list saved inside 'Telugu – SentiWordNet' folder.

Initially run the project and upload Telugu SentiWordNet to load sentiment words list such as Neutral, Positive, Negative to application from SentiWordNet database. All three list read from SentiWordNet and display inside different text areas.

For Sentence tokenization upload sentences text file to detect sentiment from each sentence. In Fourth text area we can see each sentence and below that sentence we can see whether sentences contains positive or negative sentence and beside we can see score also at end of text area all sentences and its sentiment will displayed.


Telugu Sentiment Detection Using SentiWordNet
Sentiment Analysis Using Telugu SentiWordNet
Upload Telugu Sentiwordnet
Upload Telugu Sentences
Sentiment Analysis From Sentences
Sentiment Graph
Type here to search
06:49
13-11-2019

E:/manoj/TeluguSentiment/Book_Reviews_Sentences/Sentences.txt

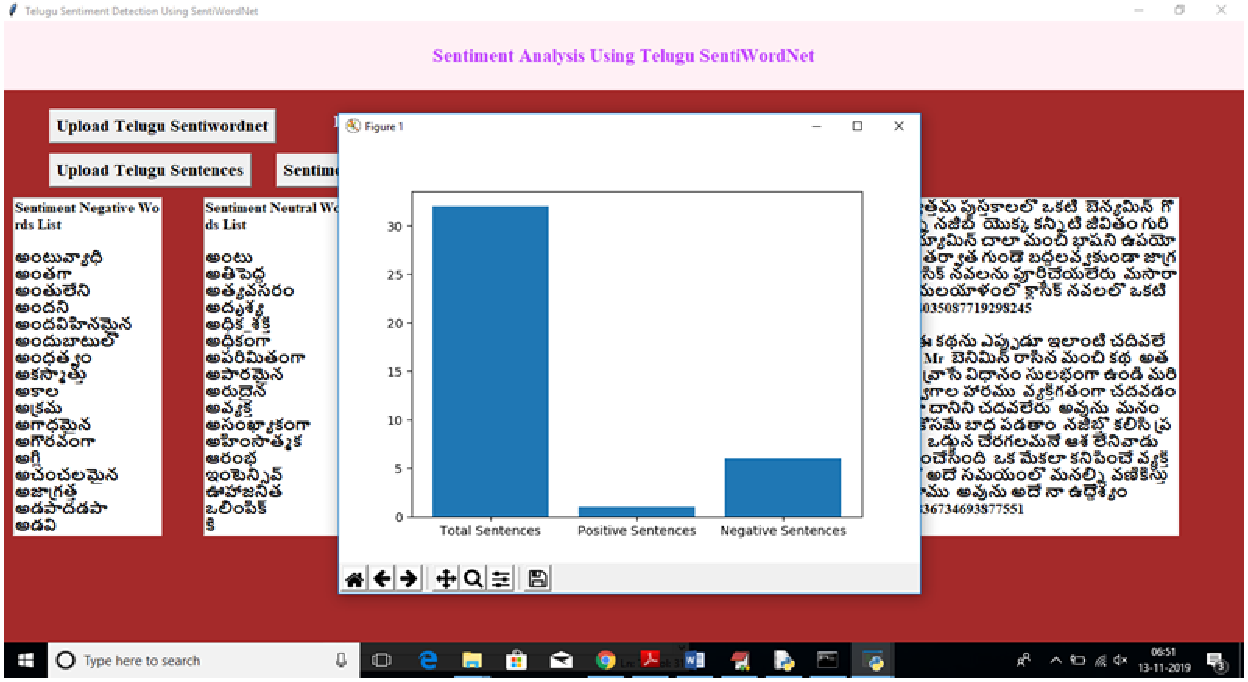
```
Sentiment Negative Wo
rds List

అంటువ్యాధి
అంతగా
అంతులేని
అందని
అందవిహీనమైన
అందుబాటులో
అంధత్వం
అకస్మాత్తు
అకాల
అక్రమ
అగాధమైన
అగౌరవంగా
అగ్ని
అచంచలమైన
అజాగ్రత్త
అడపాదడపా
అడవి
```

```
Sentiment Neutral Wor
ds List

అంటు
అతిపెద్ద
అత్యవసరం
అదృశ్య
అధిక_శక్తి
అధికంగా
అపరిమితంగా
అపారమైన
అరుదైన
అవ్యక్త
అసంఖ్యాకంగా
అహింసాత్మక
ఆరంభ
ఇంటెన్సివ్
ఊహాజనిత
ఒలింపిక్
కి
```

```
Sentiment Positive Wor
ds List

అంకితభావం
అందమైన
అక్షరాస్యులు
అతి_కోపిష్ఠ
అతి_చురుకైన
అతిశయమైన
అతిశయోక్తి
అత్యధికమైన
అదృష్టవశాత్తూ
అద్భుత
అధిక_నాణ్యత
అధికారిక
అధిభౌతిక
అధీకృత
అధునాతన
అనంతమైన
అనధికారిక
```

```
నిద్రపోవడానికి తప్ప ఈ పుస్తకం దేనికి పనికి రాదు తనకు నచ్చింది తనకు తోచినట్టు
రాసాడు తప్ప ఎం రాసాడు అని పట్టించుకోలేదు నా ఈ బుక్ అంతే అసహ్యం
Sentence Sentiment : Positive, Sentiment Score = 0.09090909090909091

ఫక్ ఇట్ ఫక్ ది బుక్ ఫక్ ది రైటర్ ఛీ ఇది ఒక బుక్ ఏనా
Sentence Sentiment : Negative, Sentiment Score = 0.3076923076923077

రచయితా లో ఉండాల్సిన కొద్దిపాటి లక్షణాలు కూడా ఈ పుస్తకం రాసిన వడిలో లేదు ర
చనా శైలి దాలా చెత్తగా ఉంది పుస్తకం కొనకపోవడం ఉత్తమం
Sentence Sentiment : Negative, Sentiment Score = 0.15789473684210525



Accuracy : 83.47383924556281
```

We got accuracy value also, Now for sentiment graph to see total sentences, positive and negative sentences count in graph will displayed. The graph
x – axis represents sentence type as total, positive and negative in Y – axis represents count.



Sentiment Negative Words List

అంటువ్యాధి
అంతగా
అంతులేని
అందని
అందవిహీనమైన
అందుబాటులో
అంధత్వం
అకస్మాత్తు
అకాల
అక్రమ
అగాధమైన
అగౌరవంగా
అగ్గి
అచంచలమైన
అజాగ్రత
అడపాదడపా
అడవి

Sentiment Neutral Words List

అంటు
అతిపెద్ద
అత్యవసరం
అదృశ్య
అధిక_శక్తి
అధికంగా
అపరిమితంగా
అపారమైన
అరుదైన
అవ్యక్త
అసంఖ్యాకంగా
అహింసాత్మక
ఆరంభ
ఇంటెన్సివ్
ఊహాజనిత
ఒలింపిక్
కి

ుత్తమ పుస్తకాలలో ఒకటి బెన్యమిన్ గ
్ నజీబ్ యొక్క కన్నీటి జీవితం గురి
మ్యామిన్ చాలా మంది భాషని ఉపయో
తర్వాత గుండె బద్దలవ్వకుండా జాగ్ర
సిక్ నవలను పూరిచేయలేరు మసాలా
మలయాళంలో క్లాసిక్ నవలలో ఒకటి
4035087719298245

ఈ కథను ఎప్పుడూ ఇలాంటి చదివలే
Mr బెనిమిన్ రాసిన మంచి కథ అత
్రాసే విధానం సులభంగా ఉండి మరి
్యగాల హారము వ్యక్తిగతంగా చదవడం
దానిని చదవలేరు అప్పను మనం
కోసమే బాధ పడతాం నజీబ్ కలిసి ప్ర
ఒడ్డున చేరగలమనే ఆశ లేనివాడు
ంచేసింది ఒక మేకలా కనిపించే వ్యక్తి
అదే సమయంలో మనల్ని వణికిస్తు
ము అప్పను అదే నా ఉద్దేశ్యం
36734693877551

## 7. Summary

In Telugu Languages, it's hard to find annotated dataset to perform NLP tasks such as POS tagging, sentiment analysis, sarcasm analysis, text summarization etc. There are few annotated datasets available in this language. This paper exploits the available Telugu SentiWordNet to perform Sentiment analysis for Telugu Student feedback sentences. The proposed system for sentiment analysis has attained an accuracy of 80% for subjectivity classification and 83.47% for sentiment classification in the domain of feedback data.

In future, we need to improve the existing SentiWordNet to attains better accuracy and find an alternate way to make this SentiWordNet dynamic. It learns annotated data automatically and adds to the existing SentiWordNet.

## References

1. Machine Learning-Based Sentiment Analysis for Twitter Accounts, Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband.2018.

2. Sentiment Analysis Using Telugu SentiWordNet Reddy Naidu,1 Santosh Kumar Bharti,2 KorraSathya Babu3 and Ramesh Kumar Mohapatra4 Computer Science & Engineering, National Institute of Technology, Rourkela, Odisha-769008.2017.

3. *Lexicon Based Sentiment Analysis System for Malayalam Language*Ashna M.P PG Scholor, Ancy K Sunny,AssistantProfessor,Computer Science and Engineering,VimalJyothi Engineering College, Kannur, India.2017.

4. Multi-class emotion detection and annotation in malayalam novels, jayakrishnan r, greeshma n gopal,,santhikrishna m s,dept.of computer science,college of engineering cherthala.2018.

5. Sentiment analysis for Odia language using supervised classifier: an information retrieval in Indian language initiative,Sanjib Kumar Sahu• PriyankaBehera• D. P. Mohapatra , Rakesh Chandra.2016.

6. Prediction of Indian Election Using Sentiment Analysis on Hindi TwitterParul Sharma, Teng-Sheng Moh,Department of Computer Science,San Jose State University,San Jose, CA, USA.2016.

7. Cross-Lingual Sentiment Analysis for Indian Regional Languages, Impana P, Jagadish S Kallimani ,Department of Computer Science and Engineering,M S Ramaiah Institute of Technology, Bangalore,India.2017.

8. Sentiment analysis of movie reviews in hindi language using machine learning charunanda, mohitdua and garimananda.2018 .

9. Domain Based Sentiment Analysis in Regional Language-Kannada using Machine Learning Algorithm,Rohini V, Merin Thomas, Latha .C.A.2016.

10. A pattern-based approach for multi-class sentiment analysis in twitter,mondherbouazizi and tomoakiohtsuki, (senior member, ieee),graduate school of science and technology, keio university, yokohama 223-8522, japan.2017.