IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 5 (May. 2019), ||S (IX) || PP 38-43

A Survey: Sentiment Analysis Using NLTK and Challenges

Sayali A. Salkade1,Viral Patel2, Abhijit Pandey
[1]*Department of Computer Science Engg, Jhulelal Institute of Technology, RTMNU, Nagpur, India*
[2]*Department of Computer Science Engg, Jhulelal Institute of Technology, RTMNU, Nagpur, India*
[3]*Department of Computer Science Engg, Jhulelal Institute of Technology, RTMNU, Nagpur, India*

Abstract: *Millions of users share and extract their opinion about different products,events,different organizations or persons on social networking sites.Sentimental analysis deals with study of opinion and thoughts of human as well as different attitudes of human towards an object.The main objective of paper is to provide different challenges faced and brief description about sentimental analysis.One of the applications is analysis of tweets on twitter, another is review of different products on website.In this paper we will discuss about sentimental analysis on twitter,news blogs,product reviews etc.The results of review using NLTK software have also been shown in paper.*

Keywords: *Sentimental analysis,naïve bayes classification,NLP,social media,python*.

I. Introduction

Thoughts or emotions about particular topic is expressed by feelings called sentiments(ie:good or bad,positive or negative).A study of thoughts and opinions of humans is given by sentimental analysis.[5]

Different products are tracked and determine whether they are viewed in positive or negative manner using web.NLP is study of human and computer interaction.There are different levels on which sentimental analysis can be classified such as polarity(negative or positive).Now a days there is incremental growth in users on various social media websites such as twitter,facebook,product review sites etc.Availibility of large users made sentimental analysis one of interesting and important topics.Sentimental analysis can be described as follows –

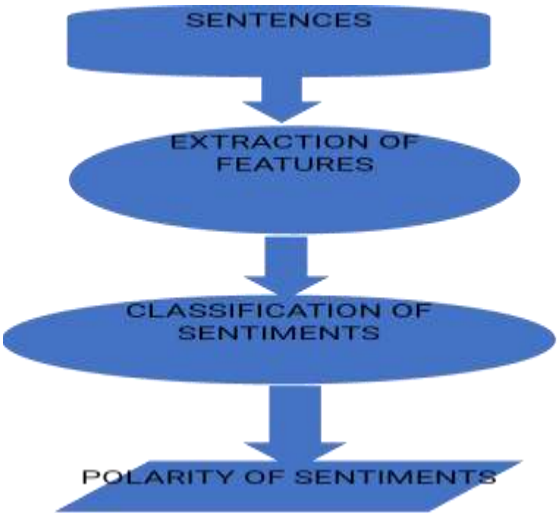

Fig.1- Sentimental analysis process

II. Literature Survey

Mostafa Karamibekr and Ali A. Ghorbaniss in 2013,did focussed on topic related opinion mining where only topic related data is considered[1].

Turney gave an approach in which he used “bag of words” ,he neglected individual words and only group of words were considered in which Altavista was used for review,only adjectives and adverbs were selected from that[2].

Fang and Bi Chen in used Machine learning and Lexicon lookup Lexicon lookup uses lexicon of negative and positive words.[3]

III. Extraction of Features in Sentimental Analysis-

1.First step is extracting features for analysis of sentences.
It includes all adjectives relevant to human thoughts or opinions.
2.Frequencies and terms present-Frequency count denotes value of features.
3. Phrases used for opinion-Different phrases and words such as good or bad, low or high, positive or negative

IV. Applications of Sentimental Analysis

A. Tweeter analysis-[2]
Tweeter data is collected using tweeter api.

1.Collection of Tweets-

Collection of tweets about particular area of interest is done. Model efficiency is done by dividing dataset into training and test data set.

2.Tweets Preprocessing-

a.Removal of duplicate tweets .
b.Converting upper case to lower case. Two same words should not be considered as different words.
c.Removal of stop word-Stop words(still ,or) should be removed using weka by matching them against dictionary of words that are available.
d.Correction of spelling
e.Removal of features in twitter-Urls and usernames which are not important from future use should be removed.
f.Special character and digits removal -Removal of special character and digits will help to remove semantics which don't convey meaning

B. Product Review Analysis-[3]

Review is done by customers in form of ratings like thumbs up,down,star rating and textual,emoji
1.It involves creation of website and receiving feedback
(rating-star and textual-in form of sentences)
2.Preprocessing and NLP
3.Labelling of features.

C. Scientific Journal Reviews-[4]

Score that is provided by reviewers is inconsistent with what is given in review. Strong critics negative score is provided and non-strict critics-positive score is considered. Consistency should be provided between them.

V. Algorithm

Naïve Bayes Algorithm-

It states that presence of any particular feature in class is not related to presence of any other feature. It is used to calculate posterior probabilities.

Steps of performing naïve Bayes algorithm-
1.Dataset is converted into frequency table
2.Find likelihood from probabilities like overcast probability is0.66
3.Calculate posterior probability for each class. Class with highest posterior probability prediction output.
It predicts different class probability on different classes.
Formula-

$$P1(C|X) = \frac{P1(X|C)P1(C)}{P1(X)}$$

P1(C|X)-posterior probability
P1(C)-prior probability
P(X)-prior predictor probability
P(X|C)-which is the probability of given class

VI. Tool Used For Semantic Analysis-

NLTK package-

It is platform for building python programs to work with data in human language. NLTK is open source free software. It is wonderful library for natural language processing and teaching. NLTK is available for windows, Linux and mac OS. It involves main features like removal of stop words, stemming, tagging, tokenization, character count and word count .[7]

1.Install python-
sudo apt-get install python3
To download and install nltk
3. import nltk
nltk.download()

VII. Challenges

1.Fake opinion-Readers or customers are misguided by providing fake reviews.
2.It is difficult to build domain specific opinion mining.
3.Interrogative sentence may not be positive or negative but keywords may be positive or negative.
4.Reader and author may understand points in different ways.
5.Some competitors or politicians may use spam review for their publicity or increasing value of product.
6.Many reviews have both positive and negative comments.Since some people might not give opinions in similar way.[3]
7.Sometimes due to comments of some people bad reviews of products are given.Comments given by user may be negative or positive in various situations.[3]

VIII. Results

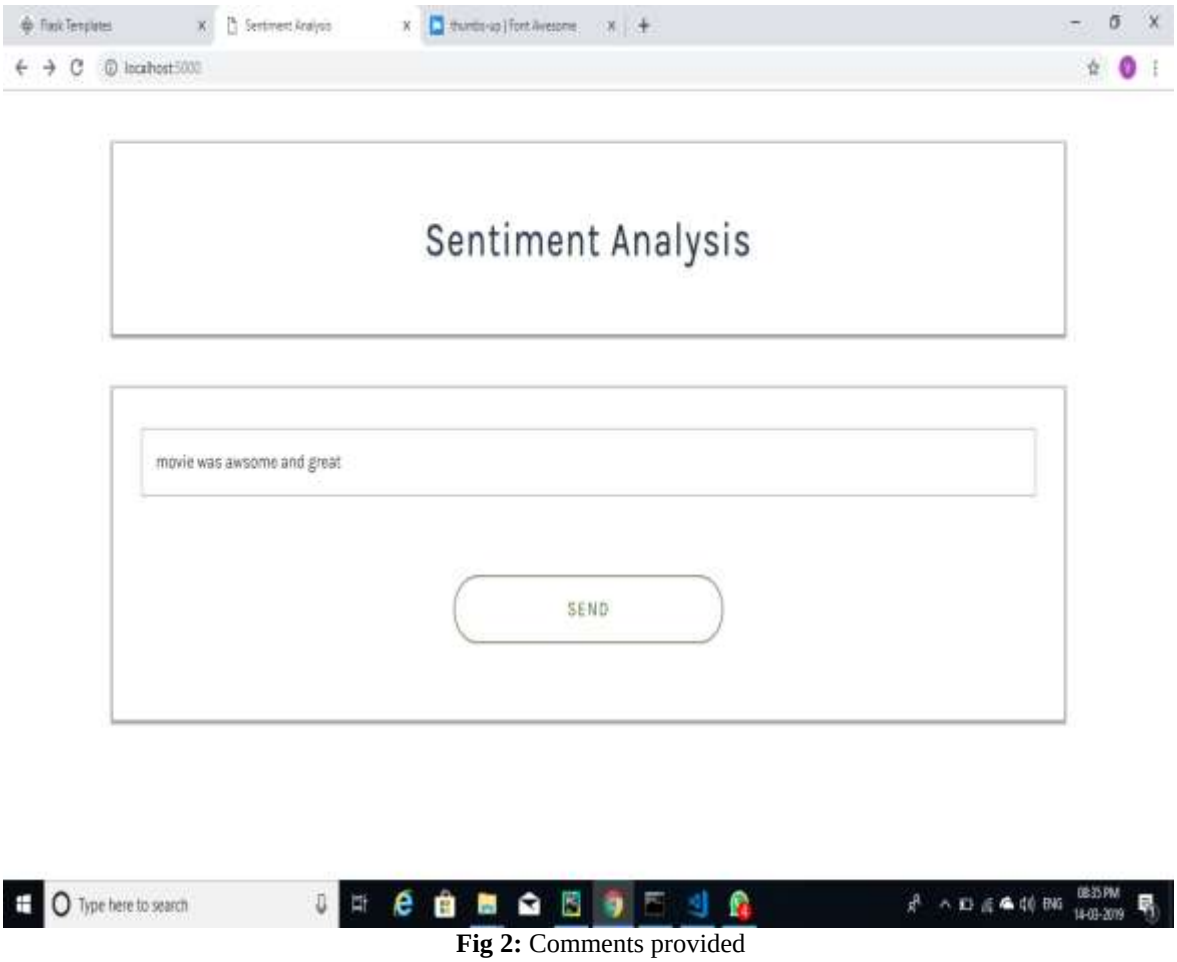

Fig 2: Comments provided

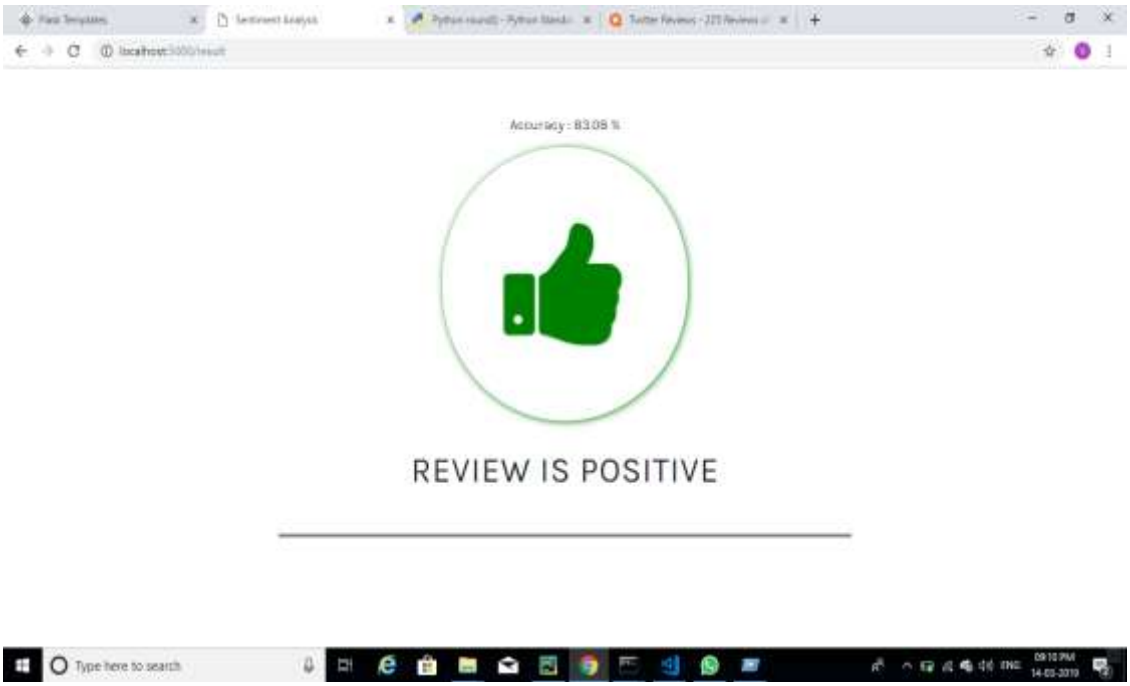

Fig.3 Positive comment

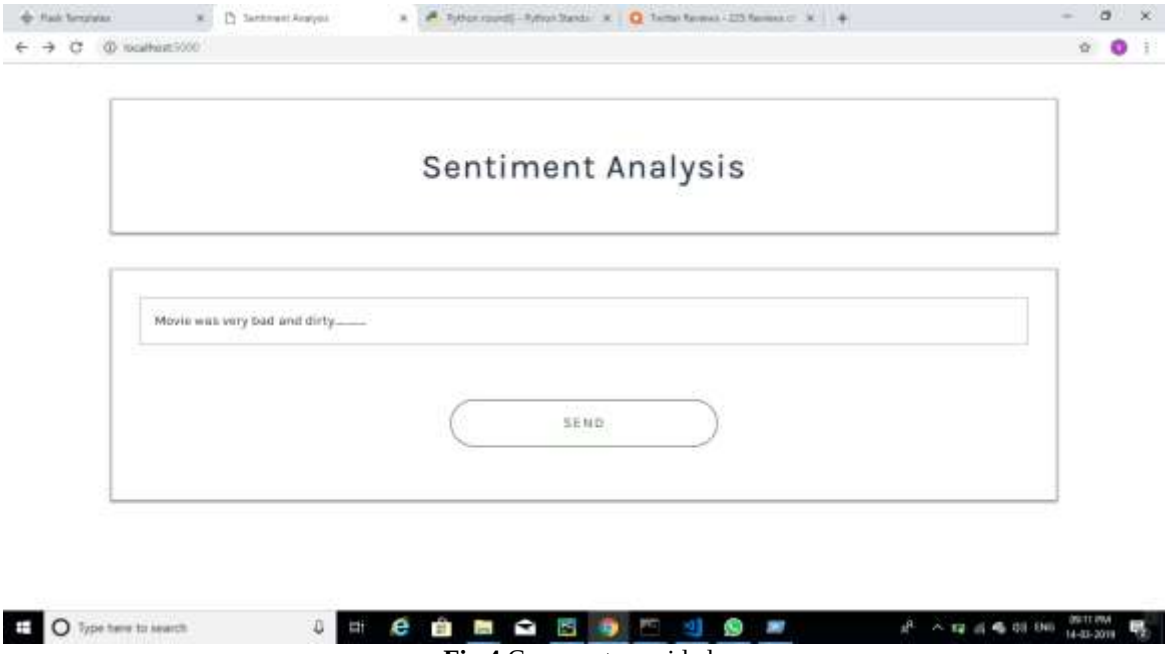

Fig.4 Comment provided.

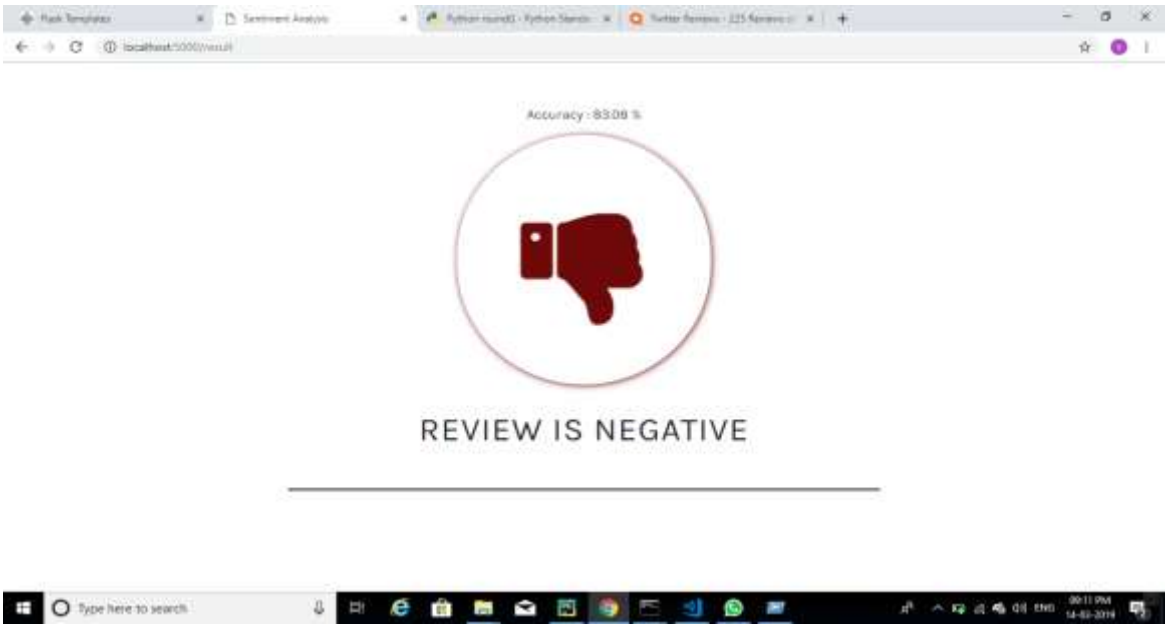

Fig.5 Negative comment

IX. Conclusion

Most popular field of research is sentimental analysis.Main aim of sentimental analysis is finding out users interest and their opinion about certain topics.Much research has been done in this field but there are many issues related to data that is not structured.Accuracy is not upto mark in dictionary approach as compared to supervised learning but it provides lesser processing time.This paper will help many researchers to explore information about sentimental analysis.[1]

References

[1]. Upma Kumari, Dinesh Soni, Dr. Arvind K Sharma," A Cognitive Study of Sentiment Analysis Techniques and Tools: A Survey", IJCST Vol. 8, Issue 1, Jan - March 2017.
[2]. Dr.Balasaravanan.K, Bharathi Bhaskaran R, Prabhakaran R,Saravanan S ,Vinoth M,"Twitter Sentiment Analysis", International Journal of Pure and Applied Mathematics, Volume 119 No. 10 2018.
[3]. Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal," Sentiment Analysis On Online Product Review", International Research Journal of Engineering and Technology, Volume: 04 Issue: 04, Apr -2017.
[4]. BrianKeith, ExequielFuentes, ClaudioMeneses," A Hybrid Approach for Sentiment Analysis Applied to Paper Reviews", In Proceedings of ACM SIGKDD Conference, Halifax, Nova Scotia, Canada, August 2017
[5]. Chandni, Nav Chandra, Sarishty Gupta, Renuka Pahade," Sentiment Analysis and its Challenges, International Journal of Engineering Research & Technology (IJERT),Vol. 4 Issue 03, March-2015
[6]. https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/
[7]. https://www.guru99.com/nltk-tutorial.html