

Sarcastic Sentiment Detection with Fuzzy Logic

Vijaykumar S. Bidve, Kalyani Pathak, Kruttika Bhagwat, Karishma Suryawanshi



Abstract: In the context of text classification of sentiment detection to great, instances are naturally fuzzy and therefore to get clear-cut outcome by extracting the opinion in a new innovative way i.e. In a fuzzy combination way and assign a relevant sentiment, usually either positive or negative. Due to this new approach it will be possible to have advancement for extracting the opinions of people more deeply even if instance in a sentences are manipulated more complexly.

Keywords: mixed feature rule formation algorithm, sentiment detection, twitter, machine learning, bag of words, doc2vec.

I. INTRODUCTION

Evaluation assessment is important mining of substance which perceives and removes enthusiastic information in source material and helping a business to grasp the social assessment of their picture, thing or organization while keeping an eye on the web conversations. Estimation examination is a kind of data mining that checks the propensity of people's decisions through ordinary language taking care of Natural Language Processing (NLP). Computational phonetics and substance assessment are used to remove and separate unique information from the web. Generally web based systems administration and similar sources [1] known as appraisal mining. Feeling examination is the strategy of legitimately mining substance to perceive and characterize the passionate ends imparted by the columnists e.g. end assessment fights with mockery [2]. Notion investigation is a helpful innovation that organizations can apply in online networking, client surveys, and client care. It helps check general assessment of an occasion or item [3]. Cyberhate portrays various kinds of online correspondence by disdain bundles to pull in new people, fabricating and strengthening pack character, arranging pack movement, passing on propagandistic messages and educating, prorogue counter-reactions as a segment of propagandistic campaigns, and ambush social get-togethers and individuals with contemptuous messages [4]. The reliably creating grouping of web based life content, the proportion of online detest talk is moreover extending. Feeling investigation also called assessment mining is the procedure by which content is examined to extricate conclusion and dole out an important assumption, typically positive, negative or neutral [5]. This work proposes fluffy methodology with the normal systems utilized in opinion examination.

Revised Manuscript Received on June 15, 2020.

* Correspondence Author

Vijaykumar S. Bidve*, Information Technology Department Marathwada MitraMandal's, College of Engineering, Pune

Kalyani Pathak, Information Technology Department Marathwada MitraMandal's, College of Engineering, Pune

Kruttika Bhagwat, Information Technology Department Marathwada MitraMandal's, College of Engineering, Pune

Karishma Suryawanshi, Information Technology Department Marathwada MitraMandal's, College of Engineering, Pune

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Retrieval Number: E9911069520/2020©BEIESP

DOI: 10.35940/ijeat.E9911.069520

Journal Website: www.ijeat.org

Equivocalness, or misdirection of unclearness, is a word, articulation, or verbalization which contains more than one significances. Faulty words or explanations lead to irregularity and disorder, and shape the purpose behind events of incidental humors [3].

II. MOTIVATION

With regards to content grouping, cases are normally fluffy and in this manner to get obvious result by removing the assessment in another creative in a fluffy blend way and dole out a pertinent opinion, generally either positive or negative.

In this new methodology it can be hear a progression for separating the thoughts of individuals. All the more profoundly regardless of whether case in a sentence is controlled all the more intricately.

III. REVIEW OF LITERATURE

Michael R. Berthold et al. discussed numerous fluffy guideline acceptance calculations. The vast majority of these calculations will in general scale severely with enormous elements of the element space and in a tough situation managing diverse component types or uproarious information. These guidelines can be separated from include spaces with assorted sorts of traits and handle the relating various kinds of requirements in equal. The removed guidelines rely upon singular subsets of just scarcely any properties, which is particularly valuable in high dimensional component spaces [3]. Hajime Watanabe et al. talked about fast development of informal communities and micro-blogging sites, correspondence between individuals from various social and mental foundations turned out to be more straightforward, bringing about increasingly more "digital" clashes between these individuals. Thus, loathe discourse is utilized to an ever increasing extent, to where it turned into a difficult issue attacking these open spaces. Abhor discourse alludes to the utilization of forceful, savage or hostile language, focusing on a particular gathering of individuals sharing a typical property, regardless of whether this property is their sexual orientation (i.e., sexism), their ethnic gathering or race (i.e., prejudice) or their accepts and religion, and so forth [6]. Shahin Amiriparian et al. addressed the upsides of utilizing cross area information when performing content based opinion investigation have been set up. In any case, comparative discoveries still can't seem to be seen when performing multimodal supposition examination. A potential explanation behind this is frameworks dependent on include extricated from discourse and facial highlights are vulnerable to perplexing affecting brought about by various chronicle conditions related with information gathered in various areas.

Sarcastic Sentiment Detection with Fuzzy Logic

Right now, in this investigate distinctive Bag-of-Words ideal models to help supposition recognition by giving preparing material from an extra dataset. This work investigated the impacts of cross-space preparing information when performing extremity location conclusion detection [1].

Wafa Alorainy et al. discusses automated web-based cyberhate recognition is significant for watching and getting network and territorial cultural pressure particularly in online informal organizations where posts can be quickly and generally saw and scattered. While past work has included utilizing dictionaries, sacks of words or probabilistic language parsing approaches, they regularly experience the ill effects of a comparable issue which is that cyberhate can be unobtrusive and circuitous. Subsequently relying upon the event of individual words or expressions can prompt countless bogus negatives, giving wrong portrayal of the patterns in cyberhate. It is explored the adequacy of building up a unique layer of phonetic highlights based around the utilization of 'othering' language, for example, terms and expressions that different the in-gathering (for example 'we', 'us') from the out gathering (for example 'them', 'those'), and propose activity or detachment dependent on saw representative and sensible dangers (for example 'send them', 'get out') [7].

Jianguo Jiang¹ et al. discovered recent prominent information ruptures have featured the significance of insider danger recognition explore for digital security. Oddity based insider location approaches are by and large connected with high bogus positives; subsequently, there has been expanded spotlight on including forecast of client brain science and assault inspirations. That as it may, information identifying with mental profile and character attribute of workers are trying to gather, and don't by and large satisfactorily catch assault inspirations, for example, disgruntlement [2].

Chris Jefferson et al. discussed sentiment examination means to distinguish the extremity of an archive through regular language handling, content investigation and computational semantics. In the course of the most recent decade, there has been a lot of spotlight on notion

investigation as the information accessible on-line has developed exponentially to incorporate numerous notion based reports (audits, input, articles). Numerous methodologies consider AI systems or measurable investigation; however there has been little utilization of the fluffy classifiers right now considering the vagueness of language and the reasonableness of fluffy ways to deal with manage this ambiguity [5].

RanjanaRajnish et al. talked "what others think" has consistently been a significant snippet of data for a large portion of us during the dynamic procedure. Reliance on web based surveys for this dynamic has gotten significant right now. With the developing accessibility and fame of feeling rich assets, for example, online audit destinations and individual sites, new chances and difficulties emerge as individuals presently effectively use data advancements to search out and comprehend the assessments of others [8].

U. Shree Krishna et al. told fuzzy rationale is a methodology to use for registering the rationale or the "level of truth" in light of on where it is valid or bogus (1 or 0) values are allotted to content in fluffy rationale. Fluffy rationale is nearer to in the manner our cerebrum work, to take the information and structure an incomplete truth for which further degree on for higher truth at that point if a specific limit is surpassed, for those specific outcomes, to get an engine response. Right now, technique that proposed will comprise of four fundamental stages, which are content preparing, highlight extraction, content grouping utilizing MLF calculation, lastly the assessment of the outcomes [9].

Orestes Appela et al. described, as purchasers move towards web based life stages like Twitter and Facebook to air their perspectives about an assortment of items, performing slant investigation on their reactions turns into an attractive action that can restore an abundance of data about open discernment. Not with standing, data posted in such systems is intended for human utilization rather than PCs, and the subtleties an individual can get from them are hard for a machine to decipher. Consequently most work right now constantly focused on extremity identification of the sentiment into three expansive fields of positive, negative or neutral [3].

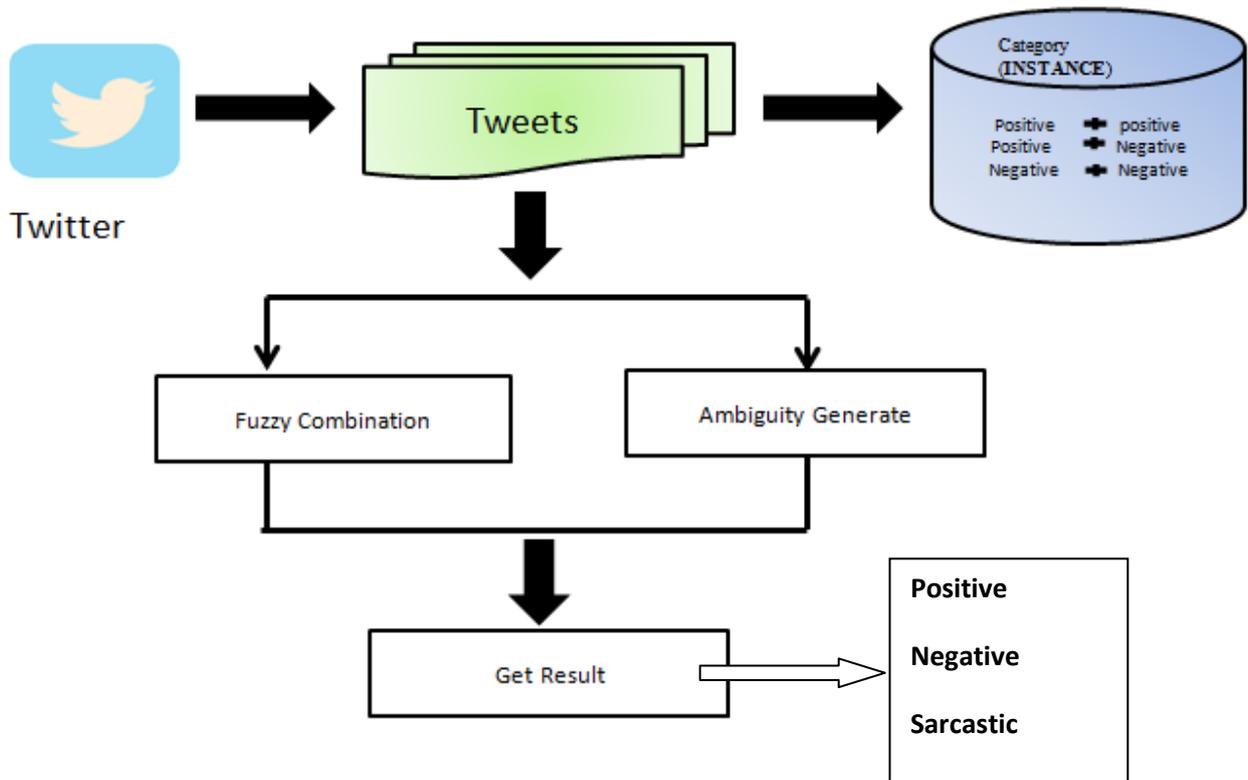


Fig. 1 System Architecture

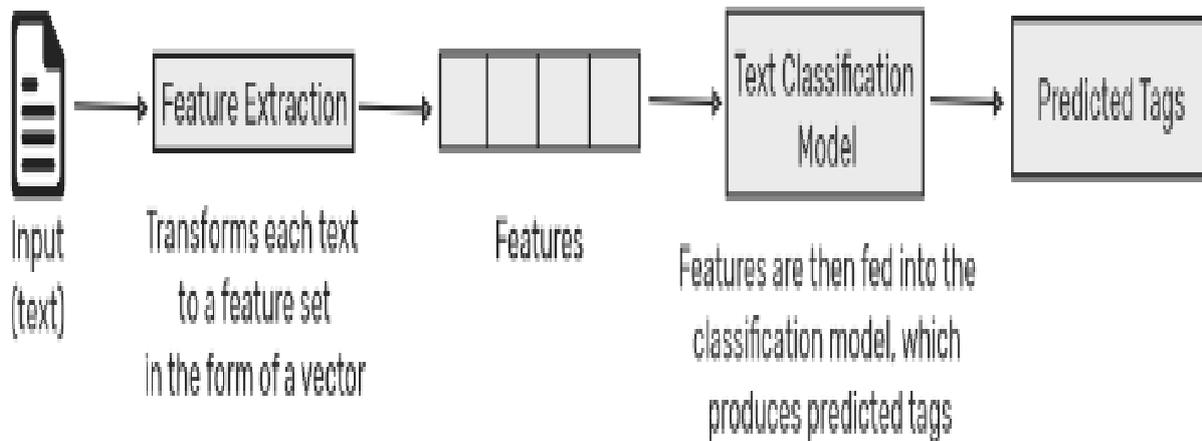


Fig. 2 Bag-of-words

IV. SYSTEM ARCHITECTURE & OVERVIEW

The system takes input from tweets in textual form. The tweets are scanned in two stages of training; first it checks the fuzzy combination and second ambiguity detection resulting in output in the terms of positive and negative sentiments. Fuzzy combination is used when the sentence is complex that is in a sarcastic way. In this work database is used for extracting the bag of words which is being stored inside each features of class to evaluate the sentence results as shown in figure 1 and figure 2. The feature means the degree of words that holds the type of sentiments for e.g., sad, happy, neutral or combination. A bag-of-words model (BOW) is a way of extracting features from text to use in modeling. The approach is very simple and flexible, and can be used in a myriad of ways for extracting features from

sentence. In this model, a text such as a sentence is represented as the bag (Multiset) of its words, disregarding grammar and even word order but keeping multiplicity. The BOW model is shown in figure 2. In word to vector, it is trained to find word vectors and then run similar queries between words. In document to vector, user tag text and get tag vectors i.e. if user only cares about tag similarities between each other. Document to vector is an unsupervised algorithm to generate vectors for sentence. The algorithm is an adaptation of word to vector which can generate vectors for words.

The vectors generated by document to vector can be used for tasks like finding similarity between sentences.

V. IMPLEMENTATION

A. Algorithm: Sentiment Detection using Sentiwordnet Dictionary

The sentiment detection algorithm use following steps,

1. polarizedTokensList ← newList()
2. while tokenizedTicket.hasNext() do
token ← tokenizedTicket.next()
3. lemma ← token.lemma
4. polarityScore ← null
5. if DomainDictionary.contains(lemma, pos) then
if SentiWordNet.contains(lemma, pos) and
6. SentiWordNet.getPolarity(lemma, pos) != 0 then
polarityScore ← SentiWordNet.getPolarity(lemma, pos)
else
7. domainDicToken ← DomainDictionary.getToken(lemma, pos) if domainDicToken.PolarityOrientation == "POSITIVE" then polarityScore ← DefaultPolarity.positive
else
8. polarityScore ← DefaultPolarity.negative
9. end if
10. end if
11. polarizedTokensList.add(token, polarityScore)
12. end if
13. end while
14. return polarizedTokensList

B. Mixed rule formation Algorithm:

Mixed fuzzy rules as used here are rules that handle different types of features. Here it is restricted up to the description of the algorithm with respect to continuous,

granulated, and nominal features but other types of features can be handled similarly as well. Each mixed rule is defined through a fuzzy region in the feature space and a class label. Fuzzy combination computes a degree of match for each rule and a corresponding input pattern. Each pattern is analyzed sub-sequentially and rules are inserted or modified accordingly.

1. For every text:

(a) Randomly choose a distribution over topics (a multinomial of length K

(b) for each word within the text:

(i) Probabilistically draw one of the K subjects from the distribution over topics obtained in (a), say topic β_j

(ii) Probabilistically draw one of the V words from β_j one for the ambiguity and other for the fuzzy combination.

C. Nearest Neighbors (KNN)

K nearest Neighbors is a simple algorithm that shops all on hand instances and classifies new instances primarily based on a similarity measure e.g., Distance functions. KNN has been used in statistical estimation and sample consciousness as a non-parametric technique. It is one of the most used getting to know algorithms.

Steps:

Stage1: Calculate the Euclidean separation between the new point and the current focuses.

Stage 2: Choose the estimation of K and select K neighbor's storage room to the new point.

Stage 3: Count the votes of all the K neighbors/Predicting Values. Load the training and test data.

a) Choose the value of K

b) For each point in test data:

Sentiment Analysis

List of Real-time Tweets for Sentiment Analysis					
ID	Name	Tweet Content	Positive Score	Negative Score	Senti Result
1	Akshay Kumar	Patience is a virtue; but in this fight against coronavirus, it has been the most powerful weapon for @MumbaiPolice. Here?s #RakhtuHausla echoing a similar sentiment, a video tribute to all our frontline warriors. https://t.co/oMiv7uF9mc	2.625	2.125	Positive
2	Narendra Modi	Birthday greetings to UP?s dynamic and industrious CM, Shri @myogiadityanath Ji. Under his leadership the state is scaling new heights of progress across all sectors. There is a marked improvement in the lives of citizens. May Almighty bless him with a long and healthy life.	4.375	3.25	Positive
3	Narendra Modi	On #WorldEnvironmentDay, we reiterate our pledge to preserve our planet?s rich biodiversity. Let us collectively do whatever possible to ensure the flora and fauna with whom we share the Earth thrive. May we leave an even better planet for the coming generations. https://t.co/nPBMthR1kr	3.125	3.0	Positive
4	Amitabh Bachchan	T 3552 - Our gratitude to them that wished us on our Anniversary .. ????? https://t.co/m3dvlMrVxE	0.625	0.5	Positive
5	Narendra Modi	Addressed #GVS2020. This is a commendable effort to bring together various stakeholders and work to eliminate COVID-19. @gavi is a symbol of global solidarity and it reminds us that by helping others we are also helping ourselves. https://t.co/eg1hk7C6IY	3.75	3.125	Positive

Fig. 3 Real time results

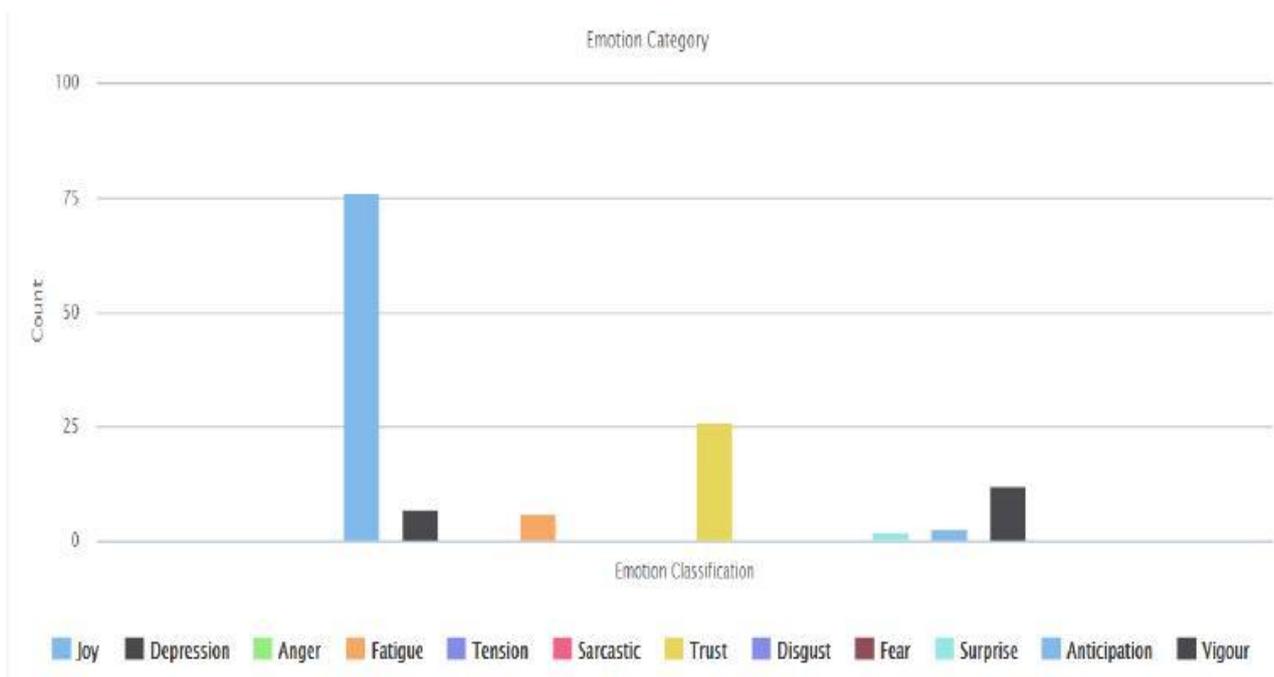


Fig. 4 Emotion recognition

i.find the Euclidean distance to all training data points
store the Euclidean distances in a list and
ii.sort it choose the first k points
Assign a class to the test point based on the majority of classes present in the chosen points.

D. Fuzzy Cluster for Vector Formation of client side Algorithm

Steps:

1. $M \leftarrow \{ \text{members in a group} \}$
2. $A \leftarrow \{ \text{sentiment, theme, emotion, stylistics, psycholinguistic} \}$

3. $M_like[m] \leftarrow \text{posts liked by member } m$
4. $Cluster[p][a] \leftarrow \text{cluster of post } p \text{ for aspect } a$
5. for $p \in M$ do
6. for $a \in A$ do
7. $x_i = 0$
8. for $p \in M_like[m]$ do
9. $c \leftarrow cluster[p][a]$
10. $m_cluster_a[x_c] += 1$
11. $cv[a] \leftarrow argmax_i(m_cluster_a[x_i])$

12. for $i := 1$ to $n(m_cluster_a)$ do
13. if $i = cv[a]$ then
14. $x_i = 1$
15. else
16. $x_i = 0$
17. $merge(m_cluster_a)$

VI. RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is web based application executed on Tomcat server. The continuous tweet posts assortment for dataset of this application utilizing Twitter API with the assistance of Twitter4j-center and Twitter4j-stream containers. A few capacities utilized in the calculation are given by rundown of containers like standfordcore-nlp container for POS labeling and so on.

Proposed work is rely upon to actualize posts suggestion framework which gathers input dataset of rundown of posts from Twitter API. All the bunching strategies like, feeling bunching, topic bunching, passionate grouping, complex grouping and psycholinguistic grouping presents on given mess free gathering condition are applied in this work. Expected result of this task is giving mess free gathering condition to twitter clients. The assistance of different angles based grouping and the area based bunching the posts with the assistance of customized warnings is used. Fig. 3 gives results of each tweet as per negative and positive score. Fig. 4 shows the categorization of various tweets as per their score.

VII. CONCLUSION

This system proposed the text classification application to scan tweets using fuzzy logic which operates on complex sentences which includes random positive plus negative words of instances and in fact it also checks the ambiguity. This new approach helps to increase the accuracy using the concept of sentiment detection and extracting the sentiments by using mixed fuzzy rule formation algorithm and to extract the keywords to differentiate for negative and positive words using the concept of bag of words.

REFERENCES

1. Nicholas Cummins,Shahin Amiriparian,Sandra Ottl, Maurice Gerczuk Maximilian Schmitt, Björn Schuller“Multimodal Bag-of-words for Cross Domains Sentiment Analysis” Journal of Computer Applications (0975 – 8887) Volume 154 – No.11, November 2016.
2. Jianguo Jiang1, Jiuming Chen, Kim-Kwang Raymond Choo, Kunying Liu1, Chao Liu1, Min Yu1, Prasant Mohapatra “Prediction and Detection of Malicious Insiders’ Motivation based on Sentiment Profile on Web Pages and Emails“, IEEE 2018.
3. Michael R. Berthold, “Mixed fuzzy rule formation”, South San Francisco, April 2002.
4. Orestes Appela, Francisco Chiclanaa, Jenny Cartera, HamidoFujitab, “A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level”, Preprint submitted to Knowledge-Based Systems May 13, 2016.
5. Chris Jefferson, Han Liu, “Fuzzy Approach for Sentiment Analysis”, IEEE School of Computing, July 2017.
6. Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki, “Hate Speech on Twitter A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection”, IEEE Access, Volume: 6, February 2018.
7. Wafa Alorainy, Pete Burnap, Han Liu, Matthew L. Williams, “The Enemy Among Us”: “Detecting Hate Speech with Threats Based

- Othering Language Embeddings”, International Journal of Scientific & Engineering Research Volume 9, Issue 3, March-2018.
8. Ranjana Rajnish, “Fuzzy Aspects in Sentiment Analysis and Opinion Mining”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 5, May 2016.
9. U. Shree Krishna, Hima Shree, K. Jayadeep, P. Lakshmi Prasanna, “Text Classification Using Fuzzy Neural Network”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7, Issue-6S4, April 2019.