iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4363

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Opinion Mining from Tweets on Goods and Services Tax (GST) using Ensemble Learning Technique and Computational Linguistics

Parminder Kaur[1], Dr. Sandeep Kautish [2]
[1, 2] *Department of Computer Science, Guru Kashi University Talwandi Sabo, Punjab (India)*

Abstract: ***In past few years, a major growth has been seen in platforms of social media. People disseminate opinions, information, behavior, and announcements via social media. They share lot regarding them and their experiences. They frequently share regarding their feelings. This provides information wealth in real-time, regarding emotional state of communities or individuals. There is explosion of interaction on social media. People use Twitter to exchange experience, opinions, and feelings. This huge amount of data can provide important information for mental health research. In this paper, extraction of emotions from the text posted on social media using data mining techniques.***
Keyword : Emotion mining, social media, data mining, tweets, ensemble learning.

I. INTRODUCTION

The field of data mining and knowledge discovery has been attracting a significant amount of research attention. An enormous amount of data has been generated every day. Data are being collected and accumulated at a dramatic pace due to the rapid growing volumes of digital data. Data mining is the process of extracting useful information, patterns or inferences from large data repositories and it is used is various business domains. It involves finding valuable information and hidden inferences in large databases [1]. There are many applications of data mining like in medical field, text data mining, sentiment analysis etc.

This paper mainly focuses on emotion mining. Emotions constitute a key factor of human intelligence, which provides indicative characteristics of human behavior, colors the way of human communication and can play an important role in human computer interaction. Emotions play an important role in successful and effective human–human communication. In fact, in many situations, emotional intelligence is more important than IQ for successful interaction. The field of emotion analysis aims at determining emotions present in text such as happy, sad, anger, surprise, love, trust and anticipation. It is also referred to as Affective Computing. Affective computing has application in many areas such as Human-Computer interaction, Depression detection, Brand perception, Social Media sentiment analysis and for Decision making. Affective computing plays a crucial role in building affective interfaces in human-computer interaction with the ultimate goal to make computer understand the emotions and attitude of human so that a computer can interact and respond effectively during the interactions.

Many researchers have put effort in analysis emotions through various sensor channels on UI such as gestures, facial expressions, voice, pitch of the sound etc. This research is the part of digital image processing. However, less effort has been put in detecting emotions from text. Nevertheless, text is an important modality for analyzing emotions because most human knowledge is transmitted via text especially with the emerging field of internet and social media sites. Emotions have cognitive bases and are shaped by several factors. They influence our decision-making, affect our social relationships and shape our daily behavior. With the rapid growth of emotion-rich textual content, such as social media text, micro blog posts, blog posts, and forum discussions, there is a growing need to develop algorithms and techniques for identifying emotions expressed in text.

Emotion	Example
Happiness	♯Happy #Enjoying #PartyyyTime
Sadness	#HeartBreak #FeelingSick
Love	#LoveMyFamily #Beautiful
Surprise	#incredibleBuilding #Wowww
Anger	#SoDisgusting #Bastard #HateIt

Table1: Twitter Emotion Analysis

It has great implications for the studies of suicide prevention, decision making in various sectors such as government, business; employee productivity, well-being of people, customer relationship management, etc. However, emotion identification is quite challenging due to the many reasons. Unlike Sentiment Analysis, it is a multi-class classification problem that usually involves at least six basic emotions. Text describing an event or situation that causes the emotion can be devoid of explicit emotion-bearing words, thus the distinction between different emotions can be very subtle, which makes it difficult to classify emotions purely by keywords.

The purpose is not to identify specific emotions but rather to tell if the text contains emotions or not [2]; in other words, if the text is subjective reflecting the writer's affect and emotional state or if it is factual and objective where the writer does not express any feelings.

II. LETRATURE REVIEW

Mitra et al. [2002] provided a survey of the available literature on data mining using soft computing. A categorization has been provided based on the different soft computing tools and their hybridizations used, the data mining function implemented, and the preference criterion selected by the model. The utility of the different soft computing methodologies is highlighted. Generally fuzzy sets are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, mixed media information and human interaction, and can provide approximate solutions faster.

Yassine et al. [2010] proposed a new framework to characterize emotional interactions in social networks, and then using these characteristics to distinguish friends from acquaintances. The goal is to extract the emotional content of texts in online social networks. The interest is in whether the text is an expression of the writer's emotions or not. For this purpose, text mining techniques are performed on comments retrieved from a social network. The framework includes a model for data collection, database schemas, data processing and data mining steps.

Jiang et al. [2017] proposed an innovative method Word Emotion Association Network (WEAN) to do emotion extraction and sentiment computing of news event. The proposed method consists of two parts: Word Emotion computation through Word Emotion Association Network and word emotion refinement. In the word emotion computation phase, microblogs with emoticons are considered to calculate the corresponding emotion present in the microblog. For refinement of the emotions derived from the first phase, they used standard sentiment thesaurus. For testing, they used Malaysia Airlines MH370 news event as dataset and computed six basic types of emotions: love, joy, anger, sad, fear and surprise.

Stojanovski et al. [2005] exploit an convolutional neural network architecture for emotion analysis in Twitter messages related to sporting events on 2014 FIFA world Cup. In this paper, seven different kinds of emotions were evaluated using hashtag labeled tweets that were collected from Twitter Streaming API. The training of the network is performed on two samples containing 1000 and 10000 tweets on which this approach achieves 50.12% and 55.77% accuracy respectively. Moreover, they have presented the analysis of this approach on three different games that have great impact on Twitter users.

Mishne et al. [2005] addressed the task of classifying blog posts on the basis of mood of the writers. They obtained a huge corpus of blog posts from one of the largest online blogging communities Livejournal. The author took the advantage of the Livejournal that allows writers to update their current mood from the 132 given categories. Yahoo API was used to get a list of 1000 web pages containing a Livejournal blog post with each kind of mood. They used variety of feature sets such as Frequency-Counts, Length related, Semantic Orientation Features and the most useful one Mood PMI-IR (Pointwise Mutual Information). For the experimental analysis, they used SVMlight (Support Vector Machine package) for classifying the mood of a blog post. Two set of experiments were performed. The first set evaluates the specific individual mood in a blog post while in the second set of experiments moods were partitioned into two mood sets as positive and negative. This was done to enhance the performance of the classification. The uniqueness of this paper is represented in the results and discussion segment which shows that the substantial increase in the training data increases the accuracy of the classification algorithm.

Yang et al. [2014] exploit a novel approach for extracting emotions. They used graphical emoticons, punctuation expressions along with a compact lexicon to label data. They provided a multi-label emotion classification algorithm (MEC) for analyzing emotions in short text of Weibo, which is a very famous online social networking site in china (just like Twitter). The approach they used is phycology independent for it worked well on different phycology theories for emotion classification. In the proposed approach they exploit K-nearest neighbor (KNN) for tweet level analysis and Naïve Bayes for Word level analysis of emotion. Moreover, their approach outperformed various state-of-art methods as discussed in experiment and results. The dataset contained tweets about Malaysia 370 missing flight and they concluded from their approach that the outbreak of Anger has a delay after breaking point of Sadness.

Catal et al. [2017] exploit a sentiment classification model based on Vote ensemble classifier utilizes from three individual classifiers: Bagging, Naïve Bayes and Support Vector Machines. Moreover, in bagging they used SVM as base classifier. The main focus of this research is to improve the performance of machine learning classifiers for sentiment classification of Turkish reviews and documents. Their experimental results show that multiple classifier system based approaches are much better for sentiment classification of Turkish documents. They performed experiments on three different domains such as book review, movie reviews and shopping reviews. The authors concluded that this approach is not restricted to just one domain and can be extended to several other domains as well.

Wang et al. [2012] exploit a technique to automatically annotate a large amount of data. They extracted large amount of tweets (2.5 million) from twitter instead of using already annotated corpus which consists of just thousands of tweets. The main focus of this research is to study the effectiveness of various feature combinations as well as the effect of the size of the training data on the emotion analysis task. To automatically annotate data they extracted the tweets using 131 relevant keywords for seven emotion categories such as joy, sadness, anger, love, fear, thankfulness and surprise. They explored variety of features such as n- grams, emotion lexicons , parts-of –speech etc. Moreover, they performed experiment on two different machine learning algorithms LIBLINEAR and multinomial Naïve Bayes. The highest accuracy of 65.57% was achieved using an enormous dataset of 2 million tweets.

Yeole et al. [2015] presented an effective technique for emotion analysis of social media text. The uniqueness of this reseach is that they have not used only direct affective words for emotion extraction, but also the indirect sentences bearing emotions are taken into account and suitable NLP techniques are applied to calculate the relevant emotion. A novel technique for emotion extraction has been presented in which they have considered direct emotion bearing words and indirect emoticons and smiley faces as well. The feature extraction phase utilizes emotion dictionaries such as SentiWordNet 3.0. Effective preprocessing was done to remove noisy data and stop words.

They used Fuzzy and rule based systems for the prediction of the emotion for a particular sentence.

Perikos et al. [2016] designed an ensemble classifier schema by combining statistical machine learning classification methods and knowledge based approach for the task of recognizing emotions in various domains such as news, headlines articles and social media posts. Moreover, the ensemble is based on the three classifiers: Naïve Bayes, Max. Entropy and knowledge based method. Furthermore, the majority voting scheme is used to combine the results of the classifiers. For training the classifiers, the corpus used is publicly available ISEAR dataset and Affective text datasets. However, for evaluation they created corpus from different sources and manual annotation was done by a human expert. They compared results with several traditional methods and the results show that ensemble classifiers are more effective than sole classifiers.

Jain et al. [2017] presented advanced framework for detection of emotions of users in Multilanguage text data using emotion theories which deals with linguistics and psychology. The emotion extraction system is developed based on multiple features groups for the better understanding of emotion lexicons. Empirical studies of three real-time events in domains like a Political election, healthcare, and sports are performed using proposed framework.

Seol et al. [2008] proposed emotion recognition system. Emotions can be expressed by various type of mediums like image, speech, facial expression, and so on. This paper focused on textual data. This hybrid system utilize two techniques, first is machine-learning method and keyword based.

Roberts et al. [2012] introduced a corpus collected from Twitter with annotated micro-blog posts (or “tweets”) annotated at the tweet-level with seven emotions: ANGER, DISGUST, FEAR, JOY, LOVE, SADNESS, and SURPRISE and analyzed how emotions are distributed in the data we annotated and compare it to the distributions in other emotion-annotated corpora. This paper used the annotated corpus to train a classifier that automatically discovers the emotions in tweets and presented an analysis of the linguistic style used for expressing emotions.

Houjeij et al. [2012] designed a system that adopts a novel approach for emotional classification from human dialogue based on text and speech context. The main objective is to boost the accuracy of speech emotional classification by accounting for the features extracted from the spoken text.

Dhawan et al. [2014] presented a new perspective for studying emotions’ expression in online social networks. The technique adopted is unsupervised; it mainly uses the k-means clustering algorithm and nearest neighbor algorithm. Experiments show high accuracy for the model in both determining subjectivity of texts and predicting emotions.

III. DISCUSSION AND FINDING

SL No.	Author / Year of Publication	Summary of work	Limitation
1.	Mitra et al. (2002)	Provided a survey of the available literature on data mining using soft computing.	Mainly focuses on the discovery algorithm and visualization techniques.
2.	Yassine et al. (2010)	Proposed a new framework to characterize emotional interactions.	Sentence structures and syntax cues are not considered.
3.	Jiang et al. (2016)	Proposed an innovative method Word Emotion Association Network (WEAN) to do emotion extraction and sentiment computing.	Emotion distance and word emotion pattern into text sentiment computation is not considered.
4.	Stojanovski et al. (2015)	Exploit an convolutional neural network architecture for emotion analysis in Twitter messages.	Don't explore the effect of various word embeddings and a deeper network architecture.
5.	Gilad Mishne et al. (2005)	Addressed the task of classifying blog posts on the basis of mood of the writers.	Success rates are relatively low, human performance on this task is not substantially better.
6.	Yang et al. (2014)	Exploit a novel approach for extracting emotions.	Only classify normal expressions into different emotions.
7.	Catal et al. (2016)	Exploit a sentiment classification model based on Vote ensemble classifier.	Does not consider emoticon features and neutral messages.
8.	Wang et al. (2012)	Exploit a technique to automatically annotate a large amount of data.	There is imbalance in the dataset.
9.	Yeole et al. (2015)	Presented an effective technique for emotion analysis of social media text.	Does not consider more advanced features for emotion classification.
10.	Perikos et al. (2015)	Designed an ensemble classifier schema by combining statistical machine learning classification methods and knowledge based approach.	Large scale evaluation is not considered for insight of system performance.
11.	Jain et al. (2017)	Presented advanced framework for detection of emotions of users in Multilanguage text data.	Data sets containing emoticons, stickers and other images with texts representing emotions are also not considered in this work.
12.	Seol et al. (2008)	Proposed emotion recognition system.	This work only focused on textual data.
13.	Roberts et al. (2012)	Introduced a corpus collected from Twitter with annotated micro-blog posts.	This work uses only certain features.
14.	Houjeij et al. (2012)	Designed a system that adopts a novel approach for emotional classification.	Appropriate features are not considered for investigation.
15.	Dhawan et al. (2014)	Presented a new perspective for studying emotions.	Training data consists of only few messages.

Table2: Techniques of Various Authors

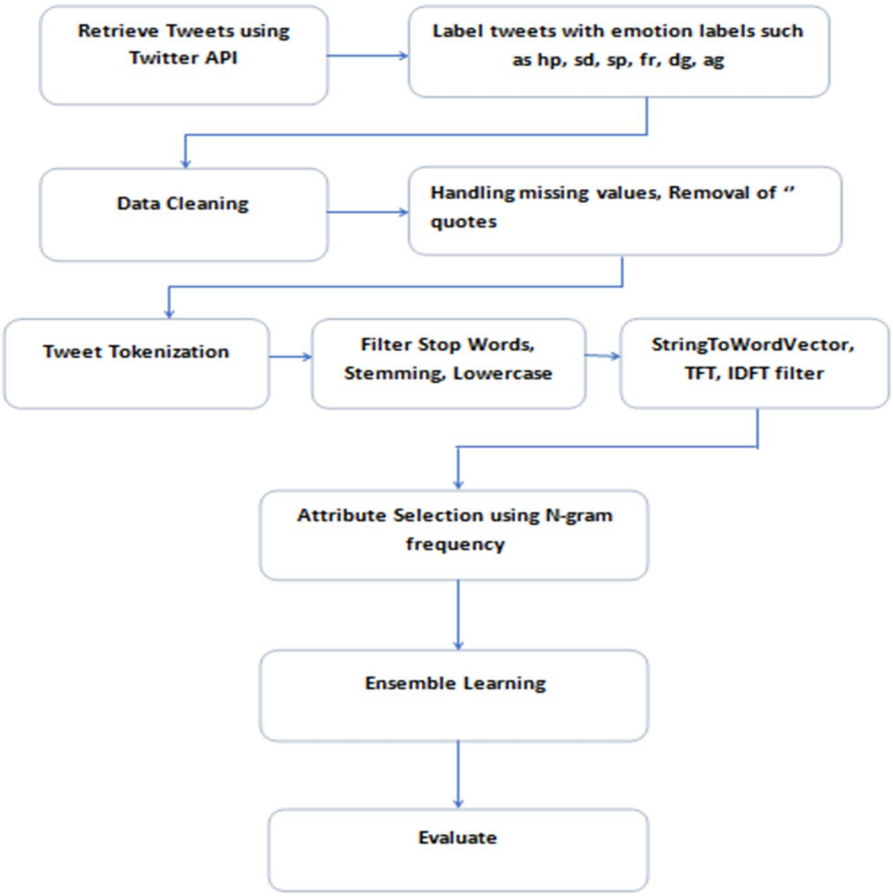


Figure 1: Flowchart of Proposed Technique

IV. CONCLUSION

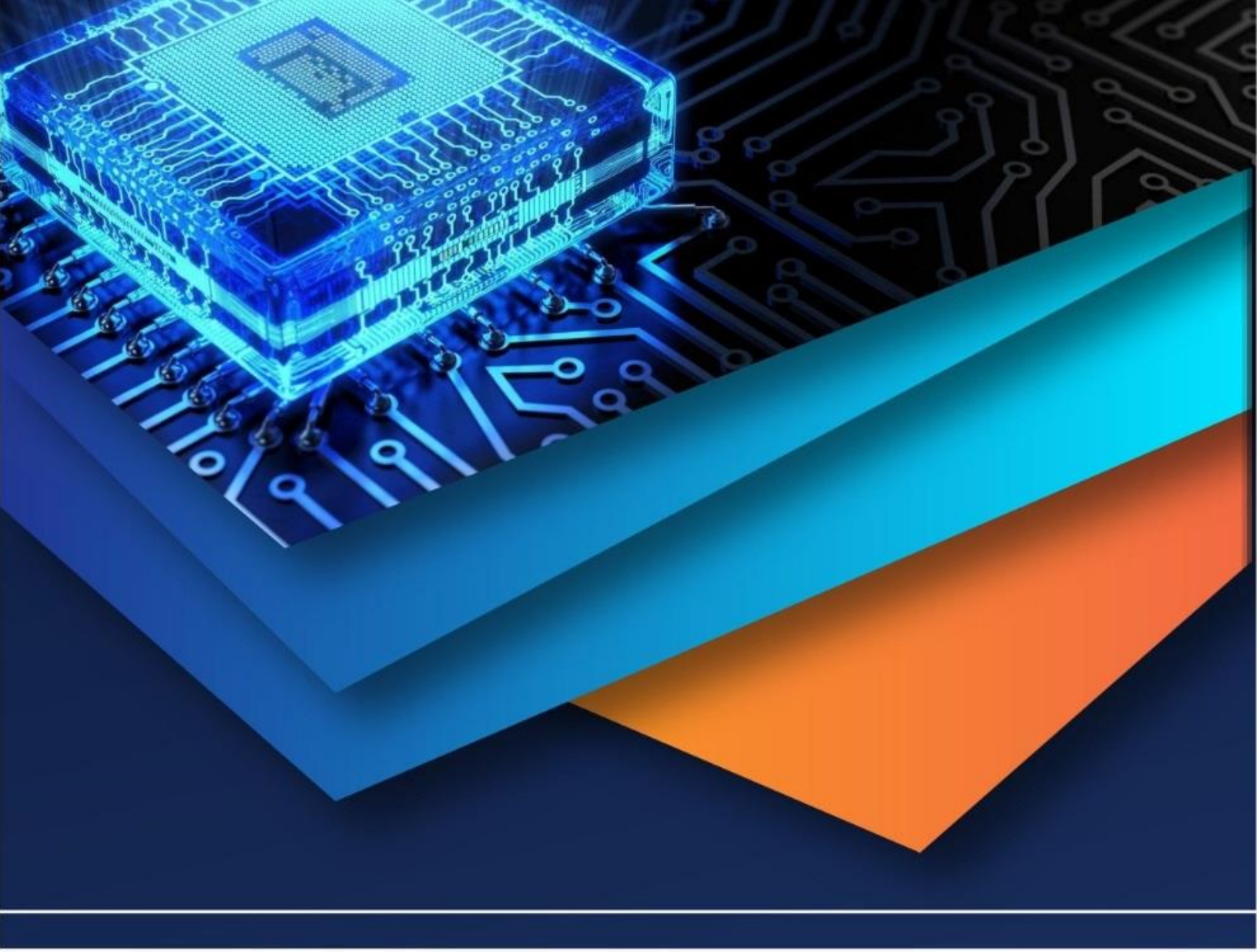
The increasing number of social media websites by Internet users has raised the interest about the opportunity to understand the relation between people's preferences and actual political behavior. We have observed that twitter is very commonly being used as a platform for deliberation by citizens of India. The social media is a powerful and reliable source of public opinion as far as a nation like India is concerned. The discussions on twitter are equivalent to traditional discussions and are capable enough to give a fair idea of emotions of general public.

Emotions have cognitive bases and are shaped by several factors. They influence our decision-making, affect our social relationships and shape our daily behavior. With the rapid growth of emotion-rich textual content, such as social media text, micro blog posts, blog posts, and forum discussions, there is a growing need to develop algorithms and techniques for identifying emotions expressed in text. It has great implications for the studies of suicide prevention, decision making in various sectors such as government, business; employee productivity, well-being of people, customer relationship management, etc. Text describing an event or situation that causes the emotion can be devoid of explicit emotion-bearing words, thus the distinction between different emotions can be very subtle, which makes it difficult to classify emotions purely by keywords. This work mainly focuses on ensemble learning based emotion extraction. The expected results would be the increase in accuracy of learning algorithm. In future, neural network with fuzzy can be used for the same purpose and also to reduce the training time of ensemble learning, features can be optimized using optimization techniques.

REFERENCES

[1] Mitra, Sushmita, Sankar K. Pal, and Pabitra Mitra. "Data mining in soft computing framework: a survey." IEEE transactions on neural networks 13, no. 1 (2002): 3-14
[2] Mohamed Yassine and Hazem Hajj, "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International Conference on Data Mining Workshops, 201
[3] Jiang, Dandan, et al. "Sentiment Computing for the News Event Based on the Social Media Big Data." IEEE Access 5 (2017): 2373-2382.

[4] Stojanovski, Dario, et al. "Emotion identification in FIFA world cup tweets using convolutional neural network." Innovations in Information Technology (IIT), 2015 11th International Conference on. IEEE, 2015.
[5] Mishne, Gilad. "Experiments with mood classification in blog posts." Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access. Vol. 19. 2005.
[6] Jun, Lan Jiang, Chongjun Wang, and Junyuan Xie. "Multi-label Emotion Classification for Tweets in Weibo: Method and Application." In Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on, pp. 424-428. IEEE, 2014.
[7] Catal, Cagatay, and Mehmet Nangir. "A sentiment classification model based on multiple classifiers." Applied Soft Computing 50 (2017): 135-141
[8] Wang, Wenbo, Lu Chen, Krishnaprasad Thirunarayan, and Amit P. Sheth. "Harnessing twitter" big data" for automatic emotion identification." In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pp. 587-592. IEEE, 2012.
[9] Yeole, Ashwini V., P. V. Chavan, and M. C. Nikose. "Opinion mining for emotions determination." In Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on, pp. 1-5. IEEE, 2015.
[10] Perikos, Isidoros, and Ioannis Hatzilygeroudis. "Recognizing emotions in text using ensemble of classifiers." Engineering Applications of Artificial Intelligence 51 (2016): 191-201
[11] Vinay Kumar Jain, Shishir Kumar, Steven Lawrence Fernandes, "Extraction of emotions from multilingual text using intelligent textprocessing and computational linguistics", Journal of Computational Science, 2017, pp. 316–326.
[12] Yong-Soo Seol, Dong-Joo Kim, Han-Woo Kim, "Emotion Recognition from Text Using Knowledge – based ANN", International Technical Conference on Circuits/Systems, 2008, pp. 1569-1572.
[13] Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, Sanda M. Harabagiu, "EmpaTweet: Annotating and Detecting Emotions on Twitter",(2012) pp. 3806-3813.
[14] Ali Houjeij, Layla Hamieh, Nader Mehdi, Hazem Hajj, "A Novel Approach for Emotion Classification based on Fusion of Text and Speech", IEEE, International Conference on Telecommunications, 2012.
[15] Sanjeev Dhawan, Kulvinder Singh, Vandana Khanchi, "A Framework for Polarity Classification and Emotion Mining from Text", International Journal Of Engineering And Computer Science, Vol. 3, Issue. 8, August, 2014, pp. 7431-7436.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY