iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5133

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 V, May 2018- Available at www.ijraset.com

Improving Probability of winning Elections

Sharvari Mehta[1], Rohini Kshirsagar[2], Shama Marathe[3], Tejas Mawande[4], Pankaja Alappanavar[5]
[1, 2, 3, 4, 5]*Department of InformationTechnology SAE. Kondhwa SavitriBai Phule Pune University*

Abstract: Our country is a democratic country. Main aim of democracy is to devolve political power to citizens. That means political leaders are decided by opinion of people. So to know opinion of people prior to election, opinion polls are conducted. Opinion polls represent opinion of people by conducting a series of questions. Opinion polls are used throughout course of Election campaigns by candidates, media and general public. Major problem with offline opinion polls conducted by media is that the questions are asked to a representative representing a large group of people so it affects accuracy of prediction. We propose a system that will collect opinion of people individually where we ask a series of questions and collect answers of people. Sentiment Analysis would then be performed on comments given by people and system gives output in form of score. The candidate with highest score has more probability of winning the Election. We use a dictionary based approach for Sentimental Analysis for better accuracy.

Keywords: Sentiment analysis; data mining; verb oriented approach, machine learning, dictionary-based approach.

I. INTRODUCTION

In democratic countries political leaders are elected by people among themselves by Elections. Future of country is directly dependent on its political leaders. Various political organizations and media conduct opinion polls prior to election to predict result of an upcoming Election. Opinion poll is an assessment of public opinion by questioning a representatives of large groups of people. Modern opinion polls are conducted through telephone surveys. Quality polls use random sampling to decide who is called. It is called random digit dialling [1].Another method is registration based sampling. It predicts politician's probability of winning election based on already available data. Polls result seemed inconsistent and inaccurate[16]. For Example, Most surprising failure of opinion polling till date was the prediction that Thomas Dewey will defeat Harry S. Truman in United States presidential Election in 1948. Major polling organizations indicated victory for Dewey. Another failure of opinion polling system was in 2016 presidential Election. It was predicted that Donald Trump would lose 2016 US presidential Election to former US Secretary of State Hillary Clinton; however Donald Trump was elected 45thPresident of United States [16]. In United Kingdom most polls failed to predict conservative Election victories of 1970 and 1992, and labor's victory in1974 [16]. One of the reasons for inaccurate prediction is that they are conducted off line that means the real problem is not with pollster's mind but is with pollster's methodologies. Primary cause was unrepresentative samples. In an off line polling when a representative of group is asked, the general opinion of people is taken into consideration. As individual opinion is neglected, it largely affects accuracy of prediction. Another problem is nonresponse bias. This occurs when certain kinds of people systematically do not respond to surveys despite of equal opportunity outreach to all parts of electorate. Also many of those who polled were simply not honest or were not able to express their opinion correctly. Another possibility is the way pollsters identify likely voters. Because no one knows in advance who is actually going to vote. In this project we introduce a system for online opinion poll. In which user will create an account register himself and then express his opinion individually. In this way we overcome inaccuracy caused only considering opinion of representative of group of people. Also since we ask a series of questions on particular candidate we get direct feedback from people for him. In this system we will collect comments in form of answers to questions about political candidate and perform sentiment analysis on them to know which candidate has more possibility of winning an upcoming Election. Such a system can be used by political organizations, media and people. Sentiment analysis has many problems and aspects to work on. It is known as many other names such as opinion mining, sentiment classification and sentiment extraction. Basic task of sentiment analysis is to classify polarity of sentence. In this paper we consider negation handling, Elimination of stop words, Intensifiers, Exclamation mark, Capitalization of words, opinion verbs and other frequently used opinion terms for enhanced and comparatively more accurate Sentiment analysis. Sentiment analysis can be applied in many domains such as shopping, entertainment, politics, education, marketing and research. This paper focuses on sentiment analysis in social domain.

The rest of paper is organized as follows: Section 2 explains sentiment analysis and machine learning and section 3 mentions some previous work. Section 4 introduces our verb oriented approach for performing sentiment analysis on comments received from

people in opinion polls. Section 5 expresses analysis and predicted accuracy of method proposed in this paper. In Section 6 we conclude with expected results of and future scope for sentiment analysis using this method.

II. SENTIMENT ANALYSIS

Sentiment is a view or an opinion for something or someone [14]. It can be positive, negative or neutral. Sentiment analysis is defined as "Process of computationally identifying and categorizing opinions expressed to determine attitude of writer towards particular thing, person, etc "[5]. There are two types of sentiment classification: binary classification and multi-class classification. In binary classification sentiments are classified into two classes: positive or negative. In multi-class classification sentiments are classified into five classes: strongly positive, positive, neutral, negative or strongly negative. Sentiment analysis is carried out on different levels:

A. Word level

in word level sentiment analysis sentiment of each word is individually computed using lexical resource like SentiWordNet [2].

B. Document Level

dentify if document (E.g. product reviews, blogs, etc) express opinions and opinions are positive, neutral, negative.

C. Sentence Level

Identifies if a single sentence has some opinion and if yes it is positive, negative, neutral.

D. Attribute Level

Extracts the attributes of product (E.g. For mobile: battery, camera, zoom size, etc) that are subject of opinion and opinion expressed for particular attribute. It expresses possibility that a negative review does not mean author dislikes all attributes of topic.

E. Input for Sentiment Analysis

Organization's internal data: Customer's feedback from email, call center, letters, etc. News and Reports: Opinions in news articles and commentaries. Word-of-mouth on web: Comments, postings, reviews, etc on social networking sites, e-commerce websites, etc.

F. Output of Sentiment Analysis:

Percent. Pie chart, bar graph and Sentiment: positive, negative, neutral. From a sentiment perspective sentence can be objective or subjective. While objective sentence contains one or more facts about product, topic or issue, a subjective contains expressed opinion(s) about a product, feature, topic or an issue. A list of opinion terms is used to distinguish between subjective and objective sentences. From technical perspective two main approaches for sentiment analysis are Machine Learning approach and Lexicon based approach [3]. In machine learning approach we use various machine learning algorithms to perform sentiment analysis and in lexicon based approach we use dictionaries for determining sentiment of sentence. In machine learning, there is a term called 'Classification'. Classification is identifying which object falls under which category. Machine learning algorithms like nearest neighbor, naïve bayes, k-means, etc can be used to classify sentiment expressed in sentence into positive or negative category. In machine learning approach, we require both training and testing datasets to perform sentiment analysis. An algorithm is trained by training dataset and then it is performed on testing dataset for required output. Ratio of training dataset to testing dataset is 80:20%.

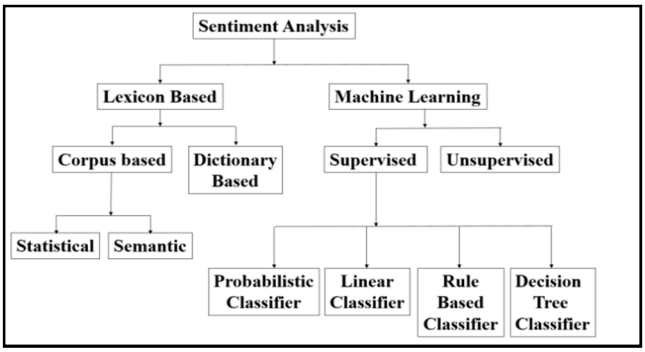


Figure 1: Methods of Sentiment Analysis

Accuracy of sentiment analysis varies according to dataset used and method of sentiment analysis used. For example probabilistic classifiers are used to predict if the players will play or not and if the weather will be sunny or rainy. The result of sentiment analysis in various formats in various domains: positive/negative, like/dislike, support/ against (politics) [4], excellent/boring (film), favorable/unfavorable, buy/don't buy, bullish/bearish, optimistic/pessimistic.

In this paper main focus is on opinion of people about politicians. It is as important to politicians and media as customer reviews for manufacturers.

III. LITERATURE REVIEW

Until now most of research is done based on sentiment analysis is in domains such as movies and products. Various methods used are Support Vector Machines, Naïve Bayes, Maximum Entropy and verb oriented approaches also [5]. It was reported that when both Support Vector Machine and the Naïve Bayes techniques were performed they had the best and worst performances respectively [8].

While dealing with an enormous amount of text data, maintaining model's performance and accuracy becomes a challenge. The performance of sentiment analysis method significantly depends on the type of words used in corpus and types of features used for analysis [9]. Various techniques used for improving Sentiment Analysis were using only domain specific features in corpus and using exhaustive stop word list and also by using Noise-Free corpus. Noise-Free corpus refers to unimportant entities of text such as links, numerical values, urls, punctuation marks etc. Normalized corpus is what we get after performing Lemmatization. Lemmatization is normalizing to its root form. For example: 'Playing' and 'Player' both are normalized to 'Play'. Research has also been done on various aspects of sentiment analysis like negation handling [6] and stop words Elimination [7].

The Twitter is popular and availiable as a service and source of data has increased the interest in sentiment analysis. Previous research reflected on the challenges like contextualizing effects and linguistic complexities are threat for the accuracy of sentiment classification of tweets.

Rezvaneh Rezapour tested the results of adding annotated, and based on corpus hashtags to a sentiment lexicon; found out that hashtags in combination with negation detection increase prediction accuracy by almost 8%. It was used as advanced model to identify and rank the candidates of political parties like Republican party and Democratic Party in the 2016 primary election of New York [10]. Khin Zezawar Aung proposed a system for education domain. In education system, students' feedback is important to measure the quality of teaching. Author analyzed Students' feedback using lexicon based approach to identify the students' positive or negative attitude. Author proposed a system to analyze the students' text feedback automatically using lexicon based approach to predict the level of teaching performance. A database of English sentiment words is created as a lexical source to get the polarity of words. By analyzing the sentiment information including intensifier words extracting from students' feedback, it was possible to determine opinion result of teachers, describing the level of positive or negative opinions [11]. Lu Zhang's research on Online Public Opinion targets at collecting, analyzing, summarizing and monitoring massive public opinions on the Internet in real time. Meanwhile, OPOS often have the ability to identify the key or sudden events, and thus notify related people immediately for rapid responses to these events. As part of this end eavor, author introduces the architecture and techniques of an OPOS that has been used by several large enterprises. This designed OPOS generally contains data layer, computation layer and application layer from bottom to up. Experimental results on real-world data validate the effectiveness of algorithms fixed in the system. The system is demonstrated in a ship-building company is provided to justify the value of the Online Public Opinion System for real enterprises [12].

IV. PROPOSED SYSTEM

Most of research has been done for products and services using adverbs, adjectives and nouns as features to classify the document. In this paper we propose to perform series of techniques on sentence and calculate its sentiment score at the end. This system will accept opinions of people prior to election about candidates participating in election and give output in form of sentiment score. The candidate with highest positive score has more probability of winning the Election.

Following techniques will be performed on Sentence accepted from user. Sequence of the techniques performed may change depending upon runtime problems. Also the final score of sentence will only be calculated after all the techniques are performed on the sentence. The total score for each candidate will be calculated by adding score of all voters.

A. Tokenization

Tokenization is the process of breaking a stream of text into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens further becomes input for further processes like text mining or parsing. Tokenization is useful both in linguistics (where it is a form of text segmentation), and in computer science, where it forms part of lexical analysis.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Potato	Suffered	from	Potato Blight

Figure 1: Tokenized output

B. Stemming

Words are often used with different variations in text depending on their grammar. Normalizing words to their root forms is called stemming.

1) INPUT

a) Playing
b) Player
c) Plays
d) Players
e) Played

2) OUTPUT

Play It saves memory since we need to store only root words.

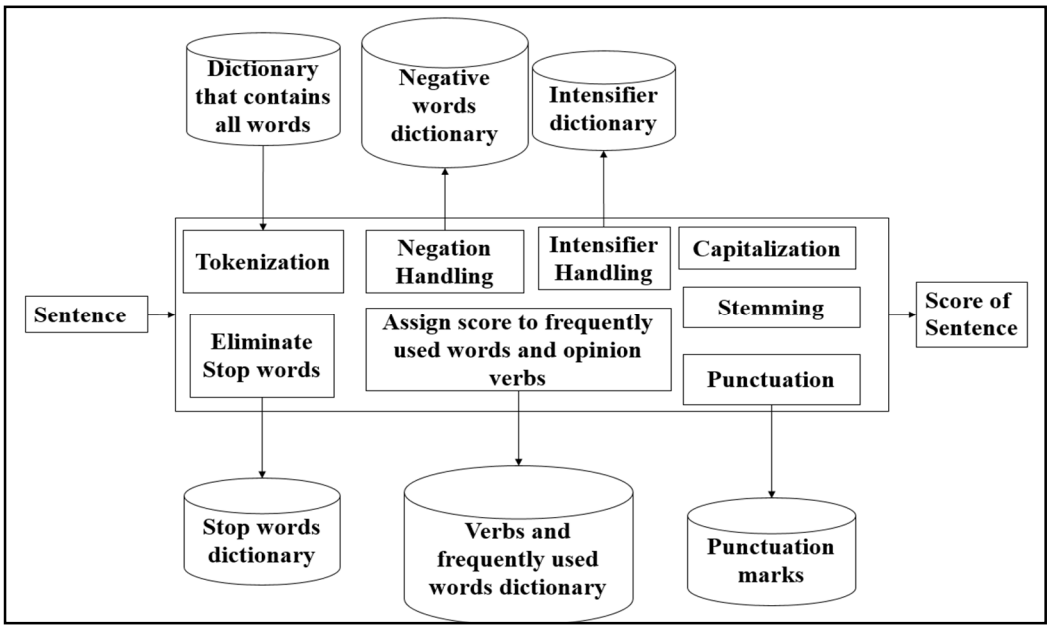


Figure 2: Proposed System

C. Eliminating Stopwords

Stopwords are defined as the most commonly used words in a corpus. (Ex. if, the, an, etc). These words are used to define structure of sentence but are of no use in defining sentiment. Eliminating them improves performance of model.

Procedure for eliminating Stop words: [7]

First step involved in the Eliminate stop Words algorithm is breaking the sentence into tokens and store this tokens into the array. Each stop word from the array is compared with the database to assure if the word is present in the database or not. This task is performed through sequential search Technique. If the word from the array matches with database, then the word is removed from the array and this process continues till all stop words are compared with the database. The Main task of this procedure to assure that the original sentence remains devoid of the stop words.

D. Punctuation Handling

For sentiment analysis we classify punctuations into two types. First type of punctuations do not express any sentiments like full stop, comma, semicolon, etc. Second type of punctuation is exclamation. It intensifies the sentiment of word before it.

Procedure for Exclamation [5]

In this procedure we check if exclamation is used in comment and if the exclamation is present in comment then we increase intensity of comment accordingly.

E. Intensifier Handling

Intensifiers are words that increase or decrease intensity of word before which they are used. For example in the phrase 'difficult' and 'really difficult' there is change in intensity. Here we would perform intensifier handling which would increases accuracy of sentiment analysis.

Procedure for Intensifier Handling [9]

In this procedure we check if intensifiers are used in comment and if the intensifiers are present in comment then we increase intensity of comment accordingly.

F. Negation Handling

Negative words affect the sentiment of sentence significantly hence, it is very important to perform negation handling on sentence. It also increases the accuracy of sentiment analysis. However, it gives inaccurate results when sarcasm is expressed.

Procedure for Negation Handling [6]

Negation Handling Algorithm also Tokenize words and store those words into the array. A single Negation word is read from the negation list .The Negation word is compared to the database of the Negation Handling Procedure and this is done by using Sequential Search Technique. If the number of negation words are odd in the sentence then the score assigned to the sentence is multiplied by -1 whereas if sentence consists of even number of Negation word then score is multiplied by +1.

G. Capitalization

Capitalization increases intensity of sentiment expressed in sentence. Hence, it is important that we consider capital words for performing sentiment analysis.

Procedure for Capitalization:

Extract each character from sentence. If the word is in upper case set flag= 1 else set flag = 0 Convert to lower case. Search in database assign score. If the word is positive then add 1 to score of word else subtract 1 from score of word.

H. Opinion verbs and frequently used terms

Verbs are action words which may or may not express opinion. English sentence cannot be constructed without verbs. So we use an opinion verb dictionary which contains opinion verbs (Ex. Protest, support, accept, reject, etc) along with their sentiment score for calculating score of sentence. Similarly, we use dictionary of frequently used opinion terms which are not verbs (Ex. Good, against, wise, etc) and assign score to words in sentence. Opinion terms are important because not all opinions can be expressed in form of verb, some opinions are expressed in forms of adjectives, adverbs, nouns, etc.

Procedure for verbs and opinion words [13]

Verb and opinion words algorithm have various sequential steps involved in it.Precisely, the very first step is accepting a particular sentence from user and then splitting each and every word into tokens which is further used for sentiment analysis. The next step involves assigning Parts Of Speech (POS) tags to the tokens which means identifying nouns, verbs, Adjectives, etc of the sentence. Verb Oriented algorithm later checks for the Verb Tags in the sentence and assigns score to it by referring the database .If several other than verb tags are present then those are also checked in the database for assigning score. For those terms missing from dictionaries we use Trie data structure here to search a word in dictionary. Trie is an efficient information re**Trie**val data structure. Using Trie, search limit can be bought to optimal limit [15]. We only use dictionary containing all words in English (ex. Oxford Dictionary) if the word is not available in our other dictionaries. If we find a word in dictionary we assign it score, else we eliminate the word.

V. EXPECTED RESULT

Goal of this project is to perform sentiment analysis on opinion of people using the verb oriented approach and to see how effective it is in social domain. A. Kennedy used adjective terms for analyzing sentiment but did not consider negative words. The author worked in Product and movie domain achieving the accuracy of 68.6% and 66.7% respectively. Similarly M. Dadvar performed sentiment analysis based on frequency terms and used adverbs and adjective of determining opinion of sentence. The author used negation words (like no, not, rather, hardly, etc) to perform negation handling. The accuracy of this method for movie review

analysis is 70.00%. In the proposed method a dictionary based approach (Dictionaries with words and their sentiment values) is used and the part of speech focused on is verb. The negation handling, stop word elimination, intensifier handling, Exclamation mark, punctuation marks, opinion verb Handling and frequently used opinion word handling, etc are also implemented to increase the accuracy of sentiment analysis. Predicted accuracy of this method is 79.16%, which is comparatively higher than other sentiment analysis methods.

VI. CONCLUSION AND FUTURE WORK

Current research on sentiment analysis shows that there is growing need to develop approaches to cope up with variety of commonly generated text. This paper presents sentiment Analysis by using verb oriented approach and is based on fact that every English sentence compulsorily contains a verb. Thus it can be implemented in feedback systems or elections by changing Features. It can also be implemented in IT companies to hire people or to select people for important responsibilities. This project also increases accuracy of result predicted using opinion polls. It also increases use of technology among common people. Politics in one of important factors for any country if the level of literacy and technology is increased in politics it directly increases development rate of the country.

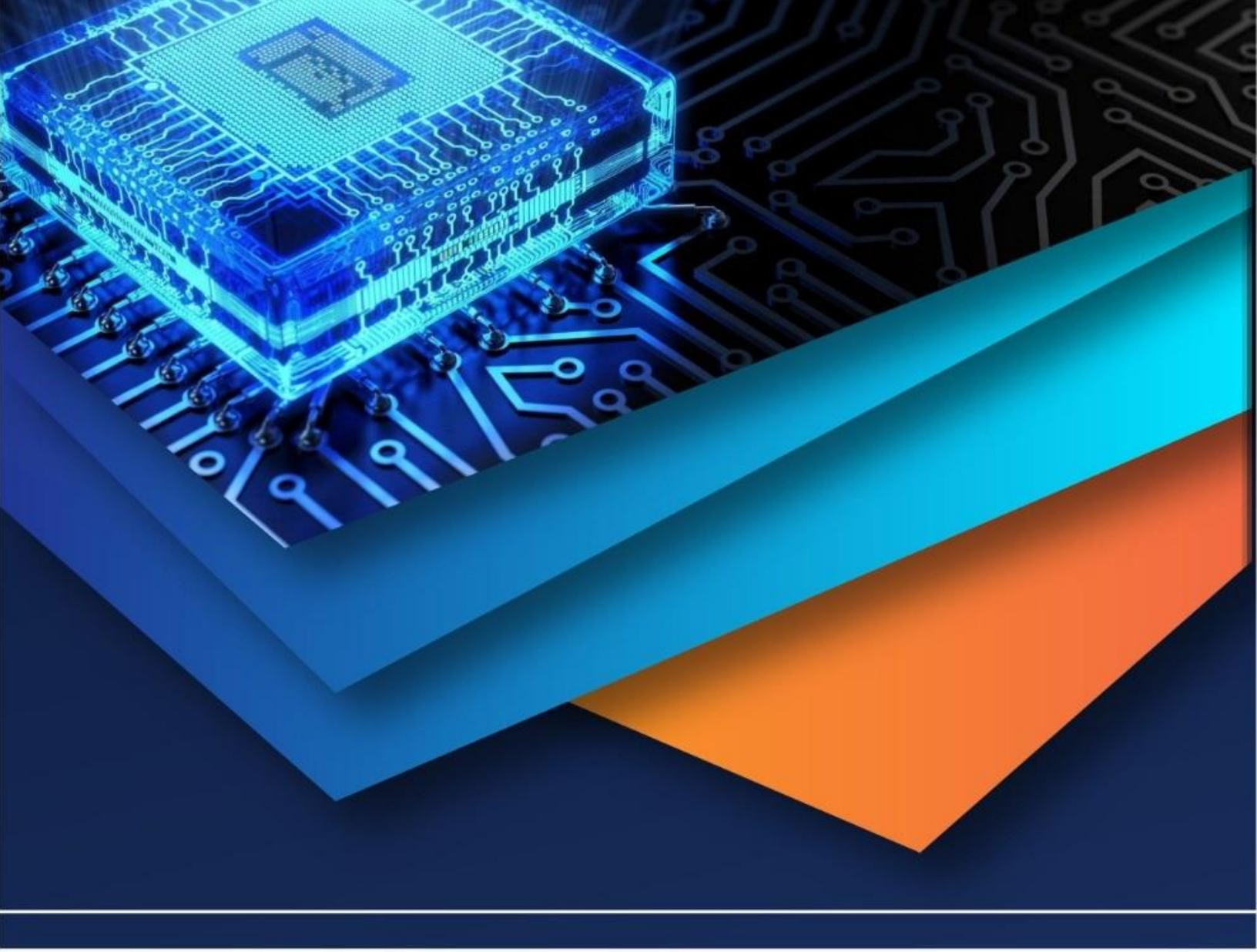
In future this project can be modified to detect word sense ambiguition and sarcasm completely. It can also implement autocorrect and autocomplete. Also this project can be modified for handicapped or uneducated people. The input can be accepted in various languages.

VII. ACKNOWLEDGEMENT

We hereby take this opportunity to record our sincere thanks and heartily gratitude to our guide Prof. Pankaja Alappanavar for her useful guidance and making available to us her intimate knowledge, experience and all necessary facilities which were indispensible in the completion of this project report. We express our special thanks and heartily gratitude to respective staff members of Department of Information Technology of Sinhgad Academy of Engineering, Pune for their valuable time, support, suggestions and persuasion. We also express our sincere thanks to the institute for providing us required facilities, Internet access and important books.

REFERENCES

[1] https://en.m.wikipedia.org/wiki/Random_digit_dialing
[2] sentiwordnet.isti.cnr.it
[3] Kishori K. Pawar, Pukhrah P. Shrishrimal, R. R. Deshmukh Twitter Sentiment Analysis: A Review IJSER in IJSER Volume-6, issue-4, April 2015.
[4] V. P. H. Binali and W. Chen. A state of art opinion mining and its application domains. In IEEE International conference on Industrial Technology. February 2009.
[5] Mostafa Karamibekr, Ali A. Ghorbani. Verb oriented Sentiment Classification. In IEEE International Conferences on Web Intelligence and Intelligent Agent Technology.2012.
[6] Amna Asmi, Tanko Ishaya. Negation Identification and calculation in Sentiment Analysis. In IMMM The second International Conference on Advances in Information Mining and Management. 2012.
[7] www.ijcaonline.org/archieves/volumeiso/number2/raulji-2016-ijca-911462.pdf.
[8] B. Pang, S. Vaithyanathan and L. Lee. Sentiment Classification using Machine Learning Techniques. In EMNLP02 Proceedings of the 40th Annual Meeting, Association for Computational Linguistics (ACL), 2002.
[9] https://www.analyticsvidhya.com/blog/2015/10/6-practices-to-enhance-performance-text-classification-model/.
[10] Lufan Wang, Rezvaneh Rezapour, Jana Diesner, Omid Abdar,.Identifying the Overlap between Election Result and Candidates' Ranking. In 11th International Conference on Semantic Computing, IEEE. 2017.
[11] Khin Zezawar Aung, Nyein Myo. Sentiment Analysis of Students' Comment Using Lexicon Based Approach IEEEICIS 2017, May 24-26, 2017, Wuhan, China. Pages 149-154.
[12] Lu Zhang, Xiaopeng Wang, Zhan Bu, Jie Caoand Zhiang Wu. Online Public Opinion System: Design and Applications. In International Conference on Advanced Cloud and Big Data. 2016.
[13] Shailendra Kumar Singh and Sanchita Paul. Sentiment classification of social issues using contextual Valence shiftersIn International Journal of Engineering and Technology August 2015.
[14] https://www.google.co.in/searchclient=msandroidomlge&ei=4qZJWsnxFof4vATi7rO4CA&q=sentiment&oq=se&gs_I=mobile-gws-serp
[15] https://www.geeksforgeeks.org/trie-insert-and-search/
[16] https://en.m.wikipedia.org/wiki/Opinion_poll
[17] Extraction of Agricultural Elements Using Unsupervised Learning by Nitika Sinha, Aakash Rathod, Pranay Gupta, Pradnya Lanke, Pankaja Alappanavar.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY