



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

SENTIMENTAL ANALYSIS ON POLITICAL DATA

A Project Report of Capstone Project -2

Submitted by

Shubham Raj

1613101723/16SCSE101542

In partial fulfilment for the award of the degree

of

Bachelor of Technology

IN

Computer Science and Engineering

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Mr. C. Ramesh Kumar, M. Tech

Professor

APRIL / MAY- 2020



**SCHOOL OF COMPUTING AND SCIENCE AND
ENGINEERING**

BONAFIDE CERTIFICATE

Certified that this project report “SENTIMENTAL ANALYSIS ON POLITICAL DATA” is the bonafide work of “SHUBHAM RAJ (1613101723)” who carried out the project works under my supervision.

SIGNATURE OF HEAD

Dr. MUNISH SHABARWAL
PROFESSOR & DEAN
School of Computing Science &
Engineering

SIGNATURE OF SUPERVISOR

Mr. C. Ramesh Kumar
PROFESSOR & SUPERVISOR
School of Computing Science &
Engineering

ACKNOWLEDGEMENT:

This is an excellent opportunity to acknowledge and to thank, all those persons without whose support and help this project would be impossible. We might prefer to add some heartfelt words for those who were a part of this project in numerous ways.

I would prefer to because of my project guide **Mr. C. Ramesh Kumar**, for his indefatigable guidance, valuable suggestion, moral support, constant encouragement, and contribution of your time for the successful completion of project work. I'm very grateful to him, for providing all the facilities needed during the project development. At the outset, I sincerely thank all faculty members of my institution **GALGOTIAS UNIVERSITY** for his extra effort to create our session online and inspire all ideas.

I would prefer to thank all those that helped me directly or indirectly. Last but not the smallest amount, I'd prefer to acknowledge the continuing support of my friends, whose patience and encouragement during these long days and nights are paramount in making this project a reality.

THANK YOU.

DECLARATION:

I hereby declare that this submission is my very own work which, to the simplest of my knowledge and belief, it contains no material previously published or written by another person nor material which to a considerable extent has been accepted for the award of the other degree or diploma of the university or other institute of upper learning, except where due acknowledgment has been made within the text.

I inform that every data used in this report if it's taken from any site is clearly referenced under the reference section.

SIGNATURE

Shubham Raj

16SCSE101542

Date: 17-may-2020

TABLE OF CONTENTS:

TITLE	PAGE NO.
CERTIFICATE.....	2
ACKNOWLEDGEMENT.....	3
DECLARATION.....	4
ABSTRACT.....	7
LIST OF TABLES.....	8
LIST OF FIGURES.....	9
CHAPTER 1 INTRODUCTION.....	10
1.1 OVERALL DESCRIPTION	
1.2 PURPOSE	
1.3 MOTIVATION AND SCOPE	
CHAPTER 2 LITERATURE SURVEY.....	13
CHAPTER 3 PROBLEM STATEMENT.....	22
CHAPTER 4 PROPOSED MODEL.....	23
4.1 MACHINE LEARNING	
4.2 NATURAL LANGUAGE PROCESSING	

4.3 PROGRAMMING TOOLS

4.3.1 PYTHON

4.3.2 NLTK

4.4.IMPLEMENTATION OR ARCHITECTURE DIAGRAM

CHAPTER 5 OUTPUT AND RESULT.....	31
CHAPTER 6 FUTURE ENHANCEMENT/ CONCLUSION.....	34
CHAPTER 7 REFERENCES.....	35

1. ABSTRACT

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others' opinions. Second, it presents many challenging research problems, which had never been attempted before the year 2000. It is thus no surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web. In fact, the research has also spread outside of computer science to management sciences and social sciences due to its importance to business and society as a whole. In this talk, I will start with the discussion of the mainstream sentiment analysis research and then move on to describe some recent work on modeling comments, discussions, and debates, which represents another kind of analysis of sentiments and opinions. To implement an algorithm for automatic classification of text into positive, negative and neutral. Sentimental analysis to determine the attitude of the mass is positive, negative and neutral towards the subject of interest. We make Graphical representation of the sentiment in the form of pie-chart.

LIST OF TABLES:

TABLE TITLE	TABLE NO.
Table for polarity and subjectivity of few Tweets	1

LISTS OF FIGURES:

FIGURE TITLE	FIGURE NO.
Use Case Diagram of Sentimental Analysis.....	Fig 1
System Flow Diagram	Fig 2
Pie-Chart Representation	Fig 3
Bar Graph Diagram	Fig 4
Scatter Plot Representation.....	Fig 5
Screenshot of output and project.....	Fig 6

1. INTRODUCTION

1.1 OVERALL DESCRIPTION

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment. In this project, we look at one such popular microblog called Twitter and build models for classifying “tweets” into positive, negative and neutral sentiment. We build models for two classification tasks: a binary task of classifying sentiment into positive and negative classes and a 3-way task of classifying sentiment into positive, negative and neutral classes. Our experiments show that features that have to do with Twitter-specific features (emojicons, hashtags etc.) add value to the classifier but only marginally. Features that combine prior polarity of words with their parts-of-speech tags are most important for both the classification

tasks. We use manually annotated Twitter data for our experiments. One advantage of this data, over previously used data-sets, is that the tweets are collected in a streaming fashion and therefore represent a true sample of actual tweets in terms of language use and content.

1.2 PURPOSE

This project also aims to meet the following objectives:

- i. To develop Sentimental analysis model that make analysis on live twitter political data and predict the sentiments of the people of the country.
- ii. To categorized the sentiments of the people on the twitter in three categories which are (1) Positive (2) Negative and (3) Neutral.
- iii. Efficient model Which will give take live real data from twitter in form of text and then will make analysis.

1.3 MOTIVATION AND SCOPE

The goal of this model is to define whether a sentence (live twitter data) has a sentiment or not and if it does, to determine whether the emotion is positive, negative, or neutral. Then we classify that live data on the basis of emotions whether the emotion is positive, negative, or neutral and then will check how many emotions are positive, negative, or

neutral then we will implement that model and show the result in pie chart. Result will be in the terms of percentage and it will show on pie chart.

Sometimes people share their points of view without emotions. For instance, the author of the sentence “I think ruling party deserves an another chance to form government for next tenure”. However, it’s hard to understand how exactly the writer feels about ruling party. So, the sentence doesn’t express a sentiment and is neutral. Neutral sentences – the ones that lack sentiment – belong to a standalone category that should not be considered as something in-between.

2. LITERATURE SURVEY

Sentiment analysis computes the opinions or sentiments using machine learning algorithm apply on data which are present inform of text. Social media is generating an enormous amount of sentiment rich data within the sort of tweets, status updates, reviews and blog posts etc. Sentiment analysis of this user generated data is extremely useful in knowing the opinion of the gang . Twitter sentiment analysis is arduous as compared to basic sentiment analysis due to the presence of slang words and misspellings. In the twitter there limit on writing blogs which is maximum 140 character. Machine learning approach can be used for analysing sentiments from the text. Some sentiment analysis is performed by analysing the twitter posts about electronic products like cell phones, computers etc.

using Machine Learning approach. By performing sentiment analysis during a specific domain, it's possible to spot the effect of domain information in sentiment classification. They presented a new feature vector for classifying the tweets as positive, negative or neutral and extract people's opinion about products.

A. Opining Mining

Conclusion mining alludes to the expansive region of characteristic language preparing, content mining, computational phonetics, which includes the computational investigation of suppositions, sentiments and feelings communicated in content [8]. In

spite of the fact that, view or disposition dependent on feeling rather than reason is regularly conversationally alluded to as an assessment [8]. Subsequently, loaning to the equivalent for feeling mining or notion investigation. [9] said that wistful examination has numerous application including bookkeeping, law, inquire about, amusement, training, innovation, legislative issues, and showcasing or may different fields. In prior days numerous online networking have given web clients road for opening up to communicate and impart their contemplations and insights [10].

B. Twitter

Twitter is a famous ongoing microblogging administration that permits clients to share short data known as tweets which are constrained to 140 characters [2,3], [11]. Clients compose tweets to communicate their conclusion about different subjects identifying with their every day lives. Twitter is a perfect stage for the extraction of overall population sentiment on explicit issues [9,10]. An assortment of tweets is utilized as the essential corpus for assumption examination, which alludes to the utilization of sentiment mining or common language handling [1]. Twitter, with 500 million clients and million messages for each day, has immediately turned into an important resource for associations to invigilate their notoriety and brands by extricating and dissecting the assumption of the tweets by the general population about their items, administrations advertise and even about contenders [12]. [2] featured that, from the internet based life produced assessments with the mammoth development of the

internet, super volumes of supposition messages as tweets, audits, web journals or any conversation gatherings and discussions are accessible for investigation, accordingly making the internet the quickest, most containing and effectively open mechanism for estimation examination.

C. Online networking

[15] characterized an online networking as a gaggle of Internet-put together applications that make with respect to the mechanical establishments which is permitted to make and trade of client produced substance. In a conversation of Internet World Start, [16] recognized that a pattern of web clients is expanding and proceeding to invest more energy with online life by the complete time spent on cell phones and web-based social networking in the U.S. across PC expanded by 37 percent to 121 billion minutes in 2012, contrasted with 88 billion minutes in 2011. Then again, organizations utilize person to person communication locales to discover and speak with customers, business can be shown harm to profitability brought about by long range informal communication [17]. As web based life can be presented so effectively on people in general, it can hurt private data to spread out in the social world [11]. Despite what might be expected, [18] talked about that the benefits of taking an interest in online life have gone past basically social sharing to make association's notoriety and pass on in vocation openings and fiscal pay. Likewise, [15], [35] referenced that the internet based life is additionally being utilized for notice by organizations for advancements,

experts for looking, selecting, social learning on the web and electronic trade. Electronic business or E-trade alludes to the buy and offer of merchandise or administrations online which can by means of web based life, such has Twitter which is advantageous because of its 24-hours accessibility, simplicity of client care and worldwide reach [19]. Among the reasons of why business will in general utilize progressively internet based life is for getting understanding into purchaser conduct inclinations, advertise knowledge and present a chance to find out about client audit and observations.

D. Twitter Sentiment Analysis

The feeling can be found in the remarks or tweet to give valuable markers to various purposes [20]. Additionally, [12] and [36] expressed that an assessment can be classified into two gatherings, which is negative and positive words. Assessment examination is a characteristic language preparing methods to evaluate a communicated supposition or notion inside a choice of tweets [8]. There has two fundamental methodologies for extricating assessment consequently which are the vocabulary based methodology and AI based methodology [19-23]

1. Vocabulary based Approach

Vocabulary based strategies utilize predefined rundown of words where each word is related with a particular opinion [21]. The vocabulary strategies change as per the setting wherein they were made and include computing direction for an archive from

the semantic direction of writings or expressions in the reports [19]. Plus, [24] likewise expresses that a vocabulary assumption is to recognize word-conveying feeling in the corpus and afterward to foresee assessment communicated in the content. [20] has demonstrated the dictionary strategies which have an essential worldview which are:

- i. Preprocess each tweet, post by evacuate accentuation
- ii. Instate a complete extremity score (s) equivalent $0 > s = 0$
- iii. Check on the off chance that token is available in a word reference, at that point If token is sure, s will be sure (+) If token is negative, s will be negative (-)
- iv. In the wake of checking the score or extremity of tweets on the off chance that $s > \text{threshold}$ or $s > 0$, at that point positive extremity, on the off chance that $s < 0$, at that point negative extremity and in the event that $s = 0$, at that point impartial extremity

However, [21] featured one bit of leeway of inclining based strategy, is that it can adjust and make prepared models for explicit purposes and settings. Interestingly, an accessibility of marked information and consequently the low relevance of the strategy for new information which is cause naming information may be exorbitant or even restrictive for certain assignments [21].

2. AI based Approach

AI techniques frequently depend on directed grouping approaches where supposition discovery is encircled as a parallel which are certain and negative [24]. This methodology requires named information to prepare classifiers [21]. This methodology, it becomes obvious that parts of the neighborhood setting of a word

should be considered, for example, negative (for example Not delightful) and escalation (for example Extremely excellent) [19]. Be that as it may, [20] indicated an essential worldview for make a component vector is: I. Apply a grammatical form tagger to each tweet post ii. Gather all the descriptor for whole tweet posts iii. Make a well known word set made out of the top N descriptive words iv. Explore the entirety of the tweets in the trial set to make the accompanying: • Number of positive words • Number of negative words • Presence, nonattendance or recurrence of each word [19] gave some case of switch nullification, refutation just to turn around the extremity of the vocabulary: evolving lovely (+3) into not wonderful (- 3). More models: She isn't stupendous (6-5=1) however not horrendous (- 6+5=-1) either. For this situation, the refutation of an emphatically negative or positive worth mirrors a blended point of view which is effectively caught in the moved worth. In any case, [21] has referenced the confinement of AI based way to deal with be more appropriate for Twitter than the lexical based technique. Besides, [20] expressed that AI strategies can create a fixed number of the most normally happening well known words which doled out a whole number an incentive in the interest of the recurrence of the word in the Twitter.

E. Strategies of Sentiment Analysis

The semantic ideas of substances extricated from tweets can be utilized to gauge the general relationship of a gathering of elements with a given feeling extremity [12]. Extremity alludes to the most fundamental structure, which is if a book or sentence is

sure or negative [25]. In any case, assessment examination has strategies in appointing extremity, for example,

1. Characteristic Language Processing (NLP)

NLP methods depend on AI and particularly factual realizing which utilizes a general learning calculation joined with a huge example, a corpus, of information to gain proficiency with the guidelines [26]. Slant examination has been taken care of as a Natural Language Processing indicated NLP, at numerous degrees of granularity. Beginning from being an archive level characterization task [27], it has been taken care of at the sentence level [28] and all the more as of late at the expression level [13]. NLP is a field in software engineering which includes causing PCs to get significance from human language and contribution as a method for interfacing with this present reality.

2. Case-Based Reasoning (CBR)

Case-Based Reasoning (CBR) is one of the strategies accessible to actualize conclusion investigation. CBR is known by reviewing the past effectively tackled issues and utilize similar answers for take care of the current firmly related issues [29]. [25] recognized some of the advantages of utilizing CBR that CBR doesn't require a specific space model then elicitation turns into an errand of social event care chronicles and CBR framework can learn by gaining new information as cases. This and the use of database strategies make the upkeep of enormous segments of data simpler [25].

3. Fake Neural Network (ANN)

[13] referenced that Artificial Neural Network (ANN) or known as neural system is a numerical strategy that interconnects gathering of fake neurons. It will process data utilizing the associations way to deal with calculation. ANN is utilized in finding the connection among information and yield or to discover designs in data[25].

4. Bolster Vector Machine(SVM)

Bolster Vector Machine is to distinguish the notions of tweets [23]. [10] together with [37] expressed SVM can concentrate and dissect to get upto70%-81.3% of precision on the test set. Utilizing SVM prepared from these uproarious named information, they acquired 81.3% in feeling characterization exactness.

F. Application programming interface

Tweepy is a python library for get to twitter API. Tweepy API performs better than the others as far as the quality and the amount of the removed substances [14]. As time passed the Python Twitter Application Programming Interface (API) is made by gathered tweets [30]. Python computing live most recent 500 tweets from twitter and afterward get prepared on that informational index and apply order and store them into database [12]. As the Python Twitter API just included Twitter messages for the latest

six days, gathered the information should have been put away in an alternate database [14].

G. Python

Python was found by Guido Van Rossum in Netherland, 1989 which has been open in 1991[31]. Python is a programming language that is accessible and tackles a PC issue which is giving a basic method to work out an answer [31]. [32] referenced that Python can be called as a scripting language. In addition, [32] and [32] likewise upheld that really Python is an only depiction of language since it tends to be one composed and run on numerous stages. What's more, [34] referenced that Python might be a language that is incredible for composing a model since Python is a littler sum tedious and managing model gave, stand out from other programming dialects. Numerous specialists are stating that Python is productive, particularly for a luxurious undertaking, as [33] has referenced that Python is reasonable to begin up informal communities or media steaming ventures which most consistently are an online which is driving a tremendous information. [34] gave the method of reasoning that since Python can deal with and deal with the memory utilized. Other than Python makes a generator that allows an iterative procedure of things, each thing in turn and license program to snatch source information each thing in turn to go each through the full preparing chain.

3. PROBLEM STATEMENT

Following are the constraints faced when computers approach to sentimental analysis:

- The problem in sentiment analysis is classifying the polarity of a given text at document, sentence, or features/aspect level.
- whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative or neutral.
- The problem in sentimental analysis is when extracting the live twitter data because that data contains hashtag, web-link and videos link.
- Sentiment Analysis of Web Based Applications Focus on Single Tweet Only.
As the growing Internet or www(world wide web), people are using many social media like as Twitter, instagram or facebook which generates big volumes of data or opinion texts in the form of tweets or posts which is available for the sentiment analysis. This translates to a huge volume of information from a human viewpoint which make it difficult to extract a sentences, read them, analyse tweet by tweet, summarize them and organize them into an understandable format in a timely manner.
- Difficulty of Sentiment Analysis with inappropriate English

4. PROPOSED MODEL

This project has been divided into two phases. First, literature study is conducted, followed by system development. Literature study involves conducting studies on varied sentiment analysis techniques and methodology that presently in used. In phase 2, application necessities and functionalities area unit outlined before its development. Also, design and interface style of the program and the way it will move are known. In developing the Twitter Sentiment Analysis application, many tools area unit utilised, like Python 3.8, google Colab, Pycharm, Notebook.

In the proposed system, we will retrieve tweets from twitter using twitter API based on the query. The collected tweets will be subjected to pre-processing. We will then apply the supervised algorithm on the stored data. The supervised algorithm used in our system is Support Vector Machine. The results of the model i.e. the sentiment will be represented in graphical manner (pie charts/bar charts). The proposed system is more effective than the existing one.

There are primarily two types of approaches for sentiment classification of opinionated texts:

- Using a Machine learning based text classifier such as Naive Bayes
- Using Natural Language Processing

We will be using those machine learning and natural language processing for sentiment analysis of tweet.

4.1 Machine Learning

The machine learning based text classifiers are a kind of supervised machine learning paradigm, where the classifier needs to be trained on some labeled training data before it can be applied to actual classification task. The training data is usually an extracted portion of the original data hand labeled manually. After suitable training they can be used on the actual test data. The Naive Bayes is a statistical classifier whereas Support Vector Machine is a kind of vector space classifier. The statistical text classifier scheme of Naive Bayes (NB) can be adapted to be used for sentiment classification problem as it can be visualized as a 2-class text classification problem: in positive and negative classes. Support Vector machine (SVM) is a kind of vector space model based classifier which requires that the text documents should be transformed to feature vectors before they are used for classification. Usually the text documents are transformed to multidimensional vectors. The entire problem of classification is then classifying every text document represented as a vector into a particular class. It is a type of large margin classifier. Here the goal is to find a decision boundary between two classes that is maximally far from any document in the training data.

This approach needs

- A good classifier such as Naive Bayes
- A training set for each class

Class can be Positive, negative. For both the classes we need training data sets.

Naïve Bayes Classifier (NB)

The Naïve Bayes classifier is the simplest and most commonly used classifier. Naïve Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction which ignores the position of the word in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label.

$$P(\text{label} | \text{features}) = \frac{P(\text{label}) * P(\text{features} | \text{label})}{P(\text{features})}$$

$P(\text{label})$ is the prior probability of a label or the likelihood that a random feature set the label. $P(\text{features} | \text{label})$ is the prior probability that a given feature set is being classified as a label. $P(\text{features})$ is the prior probability that a given feature set is occurred. Given the Naïve assumption which states that all features are independent, the equation could be rewritten as follows:

$$P(\text{label} | \text{features}) = \frac{P(\text{label}) * P(f_1 | \text{label}) * \dots * P(f_n | \text{label})}{P(\text{features})}$$

Multinomial Naïve Bayes Classifier

Accuracy – around 75%

Algorithm :

i. Dictionary generation

Count occurrence of all word in our whole data set and make a dictionary of some most frequent words.

ii. Feature set generation

All document is represented as a feature vector over the space of dictionary words.

For each document, keep track of dictionary words along with their number of occurrence in that document.

4.2 Natural Language Processing

Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages. This approach utilizes the publicly available library of SentiWordNet, which provides a sentiment polarity values for every term occurring in the document. In this lexical resource each term t occurring in WordNet is associated to three numerical scores $obj(t)$, $pos(t)$ and $neg(t)$, describing the objective, positive and negative polarities of the term, respectively. These three scores are computed by combining the results produced by eight ternary classifiers. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing. It groups words together based on their meanings. Synet is nothing but a set of one or more Synonyms. This approach uses Semantics to understand the language. Major tasks in NLP that helps in extracting sentiment from a sentence:

- Extracting part of the sentence that reflects the sentiment
- Understanding the structure of the sentence
- Different tools which help process the textual data

Basically, Positive and Negative scores got from SentiWordNet according to its part-of-speech tag and then by counting the total positive and negative scores we determine the sentiment polarity based on which class (i.e. either positive or negative) has received the highest score.

4.3 Programming tools

4.3.1 Python

Python is a widely used high-level, general-purpose, interpreted, dynamic programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than possible in languages such as C or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale.

4.3.2 NLTK

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. *Natural Language Processing with Python* provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more.

4.4. IMPLEMENTATION OR ARCHITECTURE DIAGRAM

Use Case Diagram

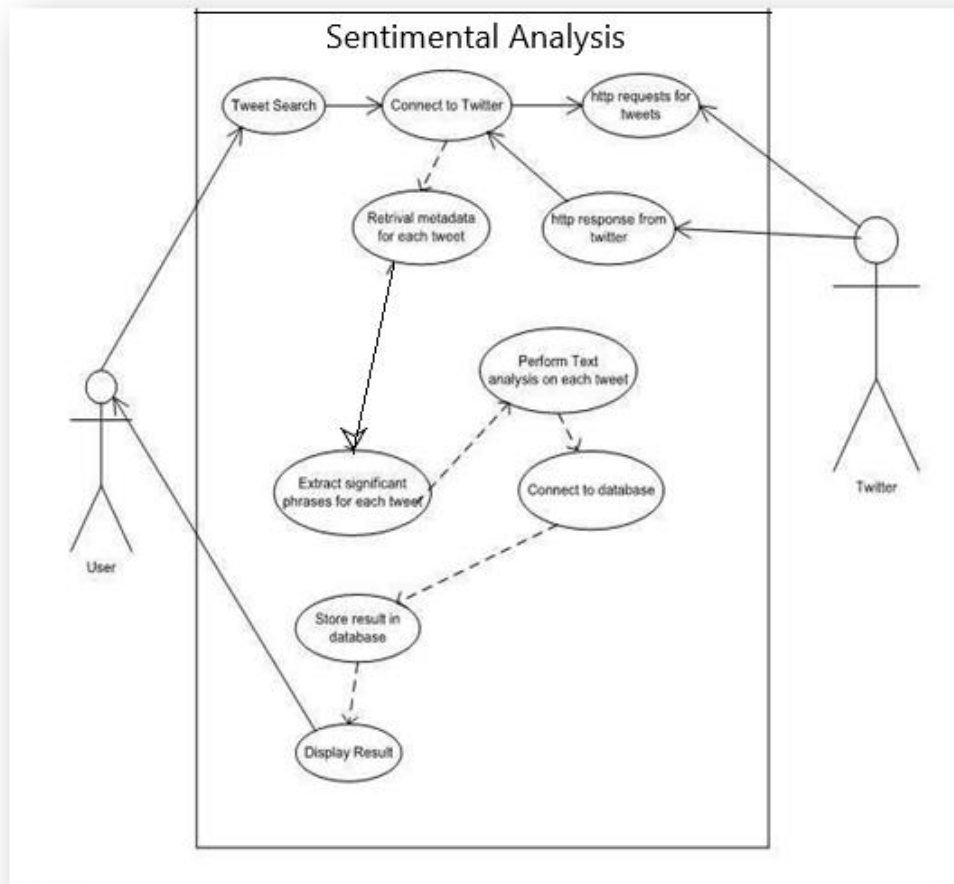


Fig 1: "Use Case Diagram of Sentimental Analysis"

System Flow Diagram

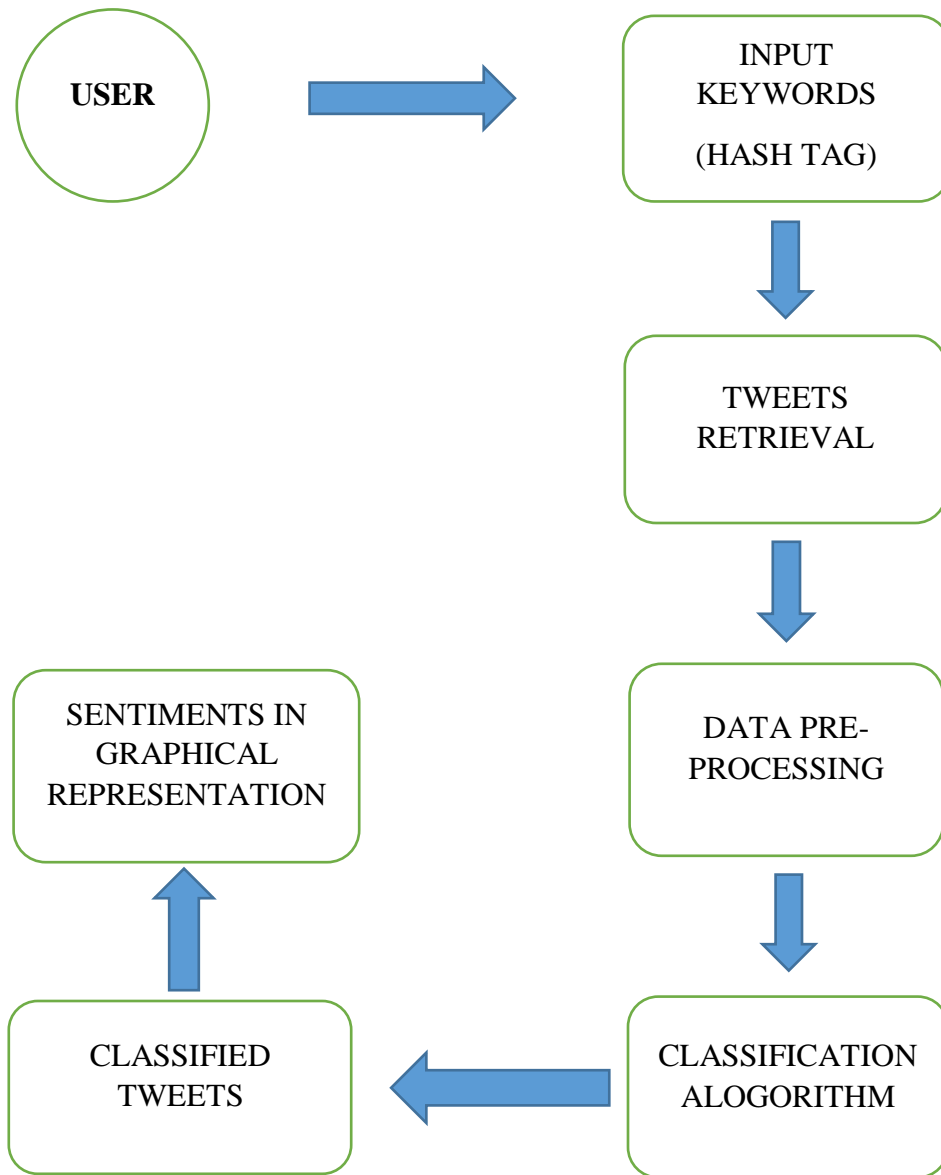


Fig 2: "System Flow Diagram"

5. OUTPUT AND RESULTS

We collected dataset containing positive and negative data from twitter using tweepy API fetching python library. Those dataset were trained data and was classified using Naïve Bayes Classifier. Before training the classifier unnecessary words, punctuations, meaning less words were cleaned to get pure data. To determine positivity, negativity and neutral of tweets we collected data using twitter API. Those data were stored in database and then retrieved back to remove those unnecessary word and punctuations for pure data. To check polarity of test tweet we train the classifier with the help of trained data. Those results were stored in database and then retrieved back using python.

After facing a number of errors, successful elimination of those error we have completed our project with continuous effort. At the end of the project output and result shown below.

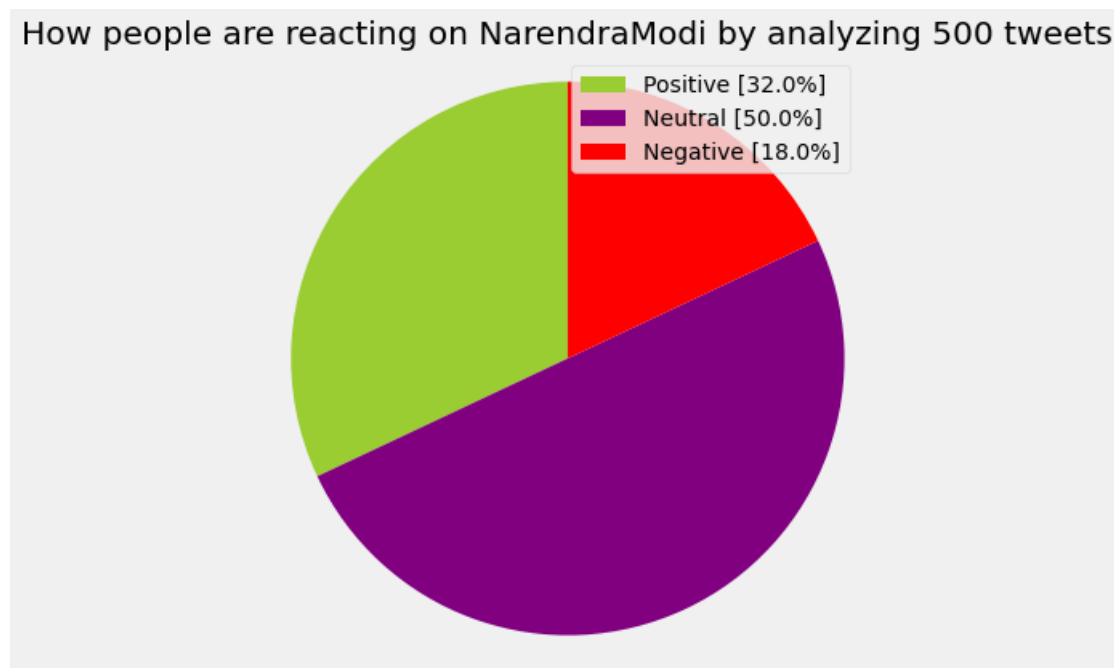


Fig 3: "Pie-Chart Representation"

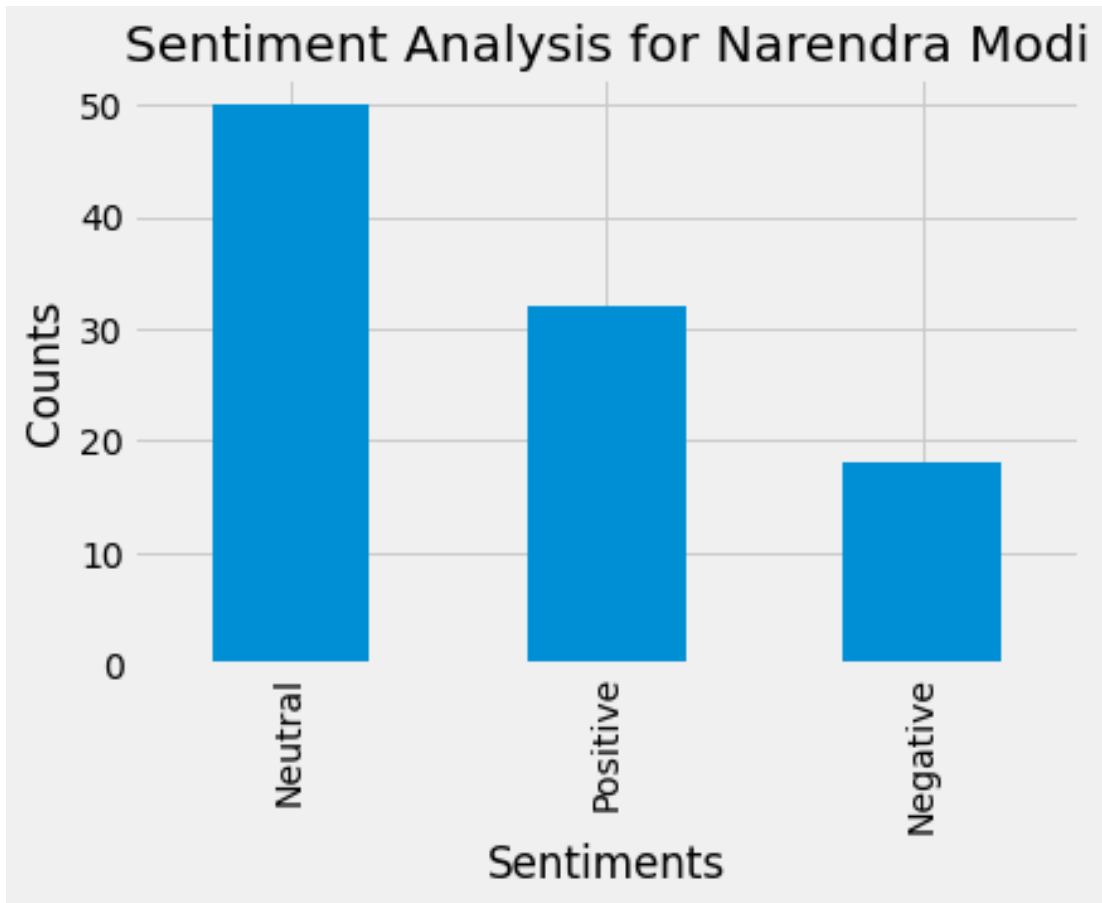


Fig 4: "Bar Graph Diagram"

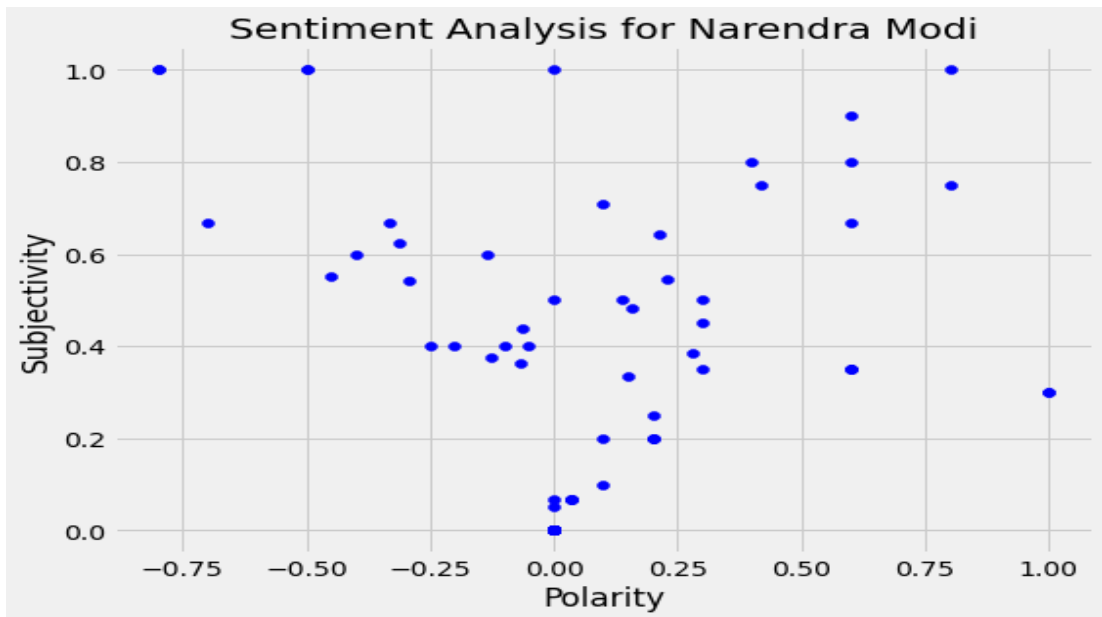


Fig 5: "Scatter Plot"

	<i>Tweets</i>	<i>Subjectivity</i>	<i>Polarity</i>
0	PM , your supporters are appalled that you ha...	1.000000	-0.800000
1	wewantmodificationlink Sir please look at...	0.000000	0.000000
2	bahl in I had been complaining about...	0.000000	0.000000
3	A small tribute by us to your resolve WeAre...	0.400000	-0.250000
4	Lekhi I will be live on NaMoAppVirtualMeet's s...	0.500000	0.136364
5	Best PM of World Shri ji\n\nModiBestPmEver	0.300000	1.000000
6	Due to sudden dismiss of g...	0.437500	-0.062500
7	\nInstead of...	0.000000	0.000000
8	Covid-19 lockdown. it is hard to understand...	0.541667	-0.291667
9	INDIA GOI Help Indians stranded in Austr...	0.000000	0.000000

Table 1 “Table for polarity and subjectivity of few Tweets”

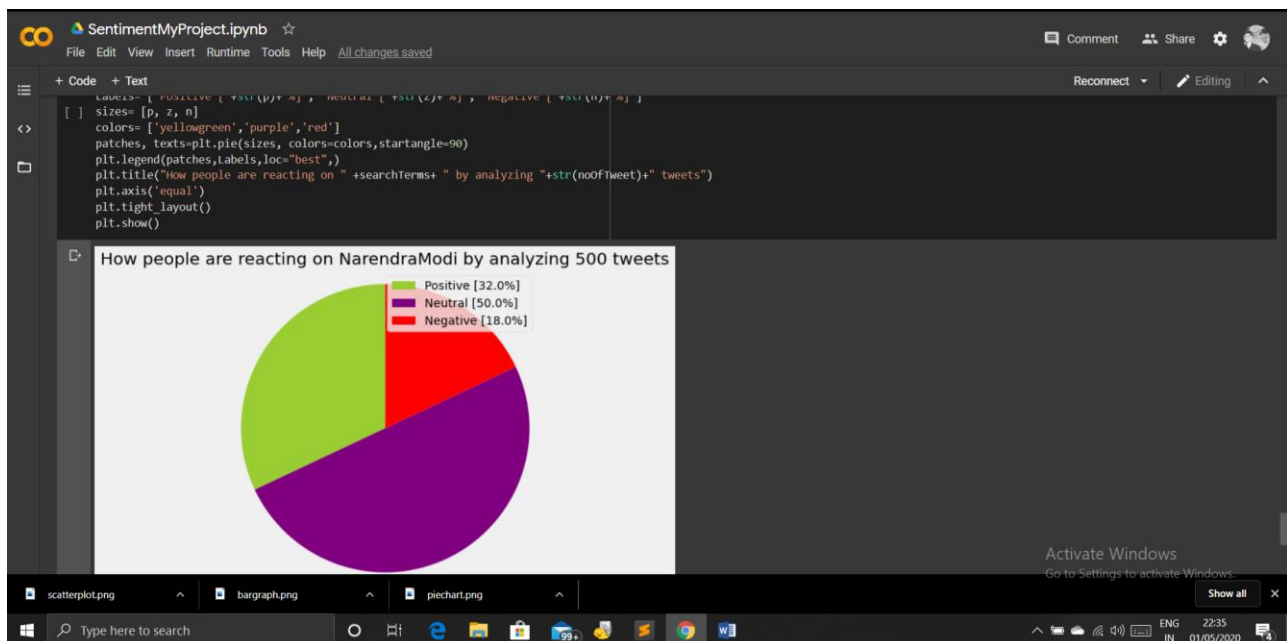


Fig 6: “Screenshot of output and project”

6. CONCLUSION / FUTURE ENHANCEMENT

We have completed our project using python as language. Although there was a problem in fetching tweets from twitter using API, through numbers of tutorial we were able to fetch it. We were able to determine the positivity, negativity and neutrality of each tweet. Based on those tweets we represented them in a diagrams like Bar graph, Pie-chart and scatter plot. All the diagrams related to outcome are shown in fig (5.1), fig (5.2), fig(5.3). This analysis is giving accurate result after analysing 500 tweets for Narendra Modi. We also shown Table (fig 5.4) which is consist of tweets and subjectivity and polarity value of 10 recent tweets.

6.1 Limitation

The system we designed is used to determine the opinion of the people based on twitter data. We somehow completed our project and was able to determine positivity, negativity and neutrality of tweet on the basis of fetched 500 tweets from twitter. It is not ready for make analysis of millions of live tweet at a time. So when we Fetch millions of tweet then that will make more accurate result.

6.2 Future Enhancement

- Analysing sentiments on emo/smiley.
- Get more accurate result after analysing millions of tweet.
- Potential improvement can be made to our data collection and analysis method.
- Future research can be done with possible improvement such as more refined data and more accurate algorithm.
- Result will be shown in Web application.
- User interface will be more attractive.

7. REFERENCES

- [1] M. Rambocas, and J. Gama, “Marketing Research: The Role of Sentiment Analysis”. The 5th SNA-KDD Workshop’11. University of Porto, 2013.
- [2] A. K. Jose, N. Bhatia, and S. Krishna, “Twitter Sentiment Analysis”. National Institute of Technology Calicut, 2010. [3] P. Lai, “Extracting Strong Sentiment Trend from Twitter”. Stanford University, 2012.
- [7] D. Boyd, S. Golder, & G. Lotan, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” System Sciences (HICSS), 2010 Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5428313
- [8] T. Carpenter, and T. Way, “Tracking Sentiment Analysis through Twitter,”. ACM computer survey. Villanova: Villanova University, 2010.
- [9] D. Osimo, and F. Mureddu, “Research Challenge on Opinion Mining and Sentiment Analysis,” Proceeding of the 12th conference of Fruct association, 2010, United Kingdom.
- [10] A. Pak, and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” Special Issue of International Journal of Computer Application, France: Universite de Paris-Sud, 2010.
- [11] S. Lohmann, M. Burch, H. Schmauder and D. Weiskopf, “Visual Analysis of Microblog Content Using Time-Varying Co-occurrence Highlighting in Tag Clouds,” Annual conference of VISVISUS. Germany: University of Stuttgart, 2012.
- [12] H. Saif, Y. He, and H. Alani, “Semantic Sentiment Analysis of Twitter,” Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media Data. United Kingdom: Knowledge Media Institute, 2011.

- [13] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment Analysis of Twitter Data," Annual International Conferences. New York: Columbia University, 2012.
- [14] J. Zhang, Y. Qu, J. Cody and Y. Wu, " A case study of Microblogging in the Enterprise: Use, Value, and Related Issues," Proceeding of the workshop on Web 2.0., 2010.
- [15] G. Kalia, "A Research Paper on Social Media: An Innovative Educational Tool", Vol.1, pp. 43-50, Chitkara University, 2013.
- [16] Internet World Stats, "Usage and Population Statistic", Retrieved 10 15, 2013 from: <http://www.internetworldstats.com/stats.htm>
- [17] A.M. Kaplan, and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," France: Paris, 2010.
- [18] Q. Tang, B. Gu, and A.B. Whinston, "Content Contribution in Social Media: The case of YouTube", 2nd conference of social media. Hawaii: Maui, 2012.
- [19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, " LexiconBased Methods for Sentiment Analysis," Association for Computational Linguistics, 2011.
- [20] M. Annett, and G. Kondrak, "A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs," Conference on web search and web data mining (WSDM). University of Alberta: Department of Computing Science, 2009.
- [21] P. Goncalves, F. Benevenuto, M. Araujo and M. Cha, "Comparing and Combining Sentiment Analysis Methods", 2013.
- [22] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", (Vol.5). International AAAI, 2011.

- [23] S. Sharma, "Application of Support Vector Machines for Damage detection in Structure," Journal of Machine Learning Research, 2008.
- [24] A.Sharma, and S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis," Association for the advancement of Artificial Intelligence, 2012.
- [25] J. Spencer and G. Uchyigit, "Sentiment or: Sentiment Analysis of Twitter Data," Second Joint Conference on Lexicon and Computational Semantics. Brighton:University of Brighton, 2008.
- [26] A. Blom and S. Thorsen, "Automatic Twitter replies with Python," International conference "Dialog 2012".
- [27] B. Pang, and L. Lee, "Opinion mining and sentiment analysis," 2nd workshop on making sense of Microposts. Ithaca: Cornell University. Vol.2(1), 2008.
- [28] M. Hu, and B. Liu, "Mining and summarizing customer reviews," 2004. [29] P. Nakov, Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, T. Wilson, Sem Eval-2013 Task2:Sentiment AnalysisinTwitter (Vol.2,pp. 312-320 ,2013.
- [29] J. Wu, J., Wang, & L. Liu, "Kernel-Based Method for Automated Walking Patterns Recognition Using Klnematics Data". 5th Workshop on Natural Language Processing. China: Xi'an Jiaotong University. 2006.
- [30] T. D. Smedt, and W. Daelemans, "Pattern for Python," Proceeding of COLING. Belgium: University of Antwerp, 2012.
- [31] A. Sweigart, "Invent your own computer games with Python. 2nd edition, 2012. Retrieved from <http://inventwithpython.com/>
- [32] C. Seberino, "Python. Faster and easier software development," Annual Conference. California: San Diego, 2012.

- [33] A.Lukaszewski, “MySQL for Python. Integrate the flexibility of Python and the power of MySQL to boost the productivity of your applications,” UK: Birmingham. Packt Publishing Ltd, 2010.
- [34] V. Nareyko, “Why python is perfect for startups,” Retrieved 01 10, 2014 from: <http://opensource.com/business/13/12/why-python-perfect-startups>
- [35] A. Hawkins, “There is more to becoming a thought leader than giving yourself the title”. Retrieved 10 18, 2013. from: <http://www.thesocialmediashow.co.uk/author/admin/>
- [36] R. Prabowo, and M. Thelwall, “Sentiment Analysis:A Combined Approach,” International World Wide Web Conference Committee (IW3C2), 2009. UnitedKingdom:Universityof Wolverhampton.
- [37] H. Saif, Y. He and H. Alani, “Alleviating Data Scarcity for Twitter Sentiment Analysis”. Association for Computational Linguistics, 2012.