|| Volume 5 || Issue 2 || February 2020 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

SENTIMENT ANALYSIS FOR LIFESAVING

Prof. Dr. P. D. Lambhate[1]
Department of Information Technology Jayawantrao Sawant College of Engineering Pune, India

Priyanka Ghate[2]
Department of Information Technology Jayawantrao Sawant College of Engineering Pune, India

Darshana Jadhav[3]
Department of Information Technology Jayawantrao Sawant College of Engineering Pune, India

Sumit Gaikwad[4]
Department of Information Technology Jayawantrao Sawant College of Engineering Pune, India

Sagar Bhirde[5]
Department of Information Technology Jayawantrao Sawant College of Engineering Pune, India

Abstract: - The Internet has become a basic requirement for everybody. There is rapid increase in social network applications; people are using these platform to voice them their opinions with regard to daily issues. People reactions toward buying a product, public services, and so on are vital. In recent years, researchers in the field of sentiment analysis have been concerned with Analyse in opinions on different topics such as movies, commercial products, and daily societal issues. The main objective of our proposed system is to perform analysis on tweets having sentiment which causes the great help for predicting the future. This paper addresses the sentiment analysis on twitter dataset. We only use tweets with negative labels. The dataset is collected which contains tweets. It has three parts: a training set, a development set, and a test set. Our main aim is to perform analysis on these tweets and conclude the tweets which are shows suicidal ideation. To accurately detect mentions of suicide ideation and Willingness to attempt suicide, our paper will explore the various Sentiment Analysis applied to the Social data. It gives an introduction to this fascinating problem and to present a framework which will perform sentiment analysis on twitter posts by associating some machine learning algorithm such as NLP, Naive Bayes, etc. Posts that describes information about the person's activities and help to find out a tendency of that particular person towards greater stress and disease development is important because by using it we can calculate depression level of user and according to it we will send motivational video links to them on their register contact information (mobile number, Gmail, etc.).

Keywords - *Sentiment Analysis, Machine Learning, Social Platform, Depression Percentage.*

I INTRODUCTION

Broadly speaking, sentiment analysis is most effective when used as a tool for business analysts, product managers, customer support directors, human resources and workforce analysts, and other stakeholders uses sentiment analysis to understand how customers and employees feel about particular subjects, and why they feel that way. A similar approach is made by us but for different applications. Many people access internet to post their views on the social platforms and also share their ideas with friends. Online platform like Twitter, Facebook , Instagram and other social platform to share their ideas. Sentiment Analysis (SA) is used as a tool for human resource, product managers and work force analysts. Sentiment Analysis is used to analyse the sentiments, emotions and opinions of the people. Sentiment analysis or opinion mining is a task that discovers the sentiments in texts. It is a system for text analysis which combines natural language processing (NLP) and machine learning techniques to assign the sentiment scores to topics, entities and categories within a sentence. The individuals, organizations, services are presented via video, text and by other means of communication. It involves different tasks like emotion analysis, sentiment mining, etc. Online social network have become means of feelings, opinion expression and also reflect bad habits of user. The messages posted by the users indicate the behavior of users. The emotions and sentiment of the users is identify, if a person is in a negative mood, message or sentence contain words with negative meaning which indicate stress ,sadness ,depression and if user is in a positive mood, this person can be emotionally and mentally stable. Machine learning algorithms are used in studies about mood monitoring system analysing messages from social networks. It is observed that users post short sentences and suffer from chronic insomnia when they are facing depression. The sentence containing certain word related to negative meaning is analysed and identifies users at a high risk attempting suicide.

II LITERATURE REVIEW

A. A Knowledge-Based Recommendation System that includes Sentiment Analysis and Deep Learning

In this paper, the system filters sentences from a On- line social network (OSN) that allows identifying potential users with depression and stress conditions. Knowledge-Based Recommendation System (KBRS) use ontology effectively and use personalized sentiment analysis. In general sentiment analysis, the recommended messages are sent to the users, in order to improve their emotional state. Sentiment analysis can be performed by three techniques.1)Machine learning 2)Hybrid technique, which combines machine learning approach 3)Lexicon-based technique using a word dictionary of textual information 3)The main aim is to send relaxing, happy, calm or motivational messages to users with psychological disturbances by using KBRS system. The research in Sociology and Psychology reports a number of mental factors related to social network mental disorders. Research indicates that young people are more addicted with OSNs and discuss potential reasons for internet addiction. It is based on sentiment analysis and ontology's, it is activated to send happy, calm and motivate messages to user. It includes mechanism to send warning messages to users, in case of depression or disturbance. In order to improve KBRS performance, user profile parameters, geographical location and sentence to identify the sentiment intensity of a message. This paper proposes a recommendation system which is knowledge based and include sentiment analysis. Recommendation System is used as a method to improve the user mood in case of negative emotional health.

B. Sentiment Analysis Using Naive Bayes Classifier

In this research paper, Naive Bayes classifier is used to classify the text that belongs to particular class and it is one of the supervised classification techniques. This classifier is used to calculate the probability of all words in document and gives the predicted result. This classification technique using machine learning gives accurate results and sentiment classification of unknown tweets by predicting the future. Naive Bayes is the probabilistic algorithm which calculates the probability of each word in the sentence and the word with highest probability is considered as output. The main objective is to perform analysis on tweets for predicting the future. The sentiment is extracted from tweet using Twitter dataset. This paper focuses on sentiment analysis on twitter dataset by using Naive Bayes classifier. The tweet is classified in terms of negative, positive and neutral. Tokenization method is used to divide the document into small parts called tokens. The concept of Bag of words is applied on tokens. Naive Bayes algorithm is explained in detail. It consists of training data set which belongs to different classes. Prior probability of classes is calculated using formula. The total number of word frequencies of both classes is calculated. Uniform distributions are to be performed to avoid zero frequency. Stop words are common words that are ignored in order to reduce the size of the dataset. Naive Bayes algorithm is applied for analysis.

C. Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data

In this paper, the suitability of Sentiment Analysis approaches for application in the big data and suggests the future work. Big data decrease the cost of computing power and storage. It includes the study of Sentiment classification techniques and application of Sentiment Analysis (SA). The sentiment classification through machine learning algorithm. The text classification method can be divided into supervised and unsupervised learning methods. The supervised method consists of labelled training documents and unsupervised learning uses large amount of unlabelled data. Support Vector Machine (SVM) is a classification method and proven to be highly effective method for traditional text categorisation than other machine learning techniques. Naive Bayes classifier is widely used algorithm that is used in document classification work. Naive Bayes is suitable for use when inputs have high dimensionality. Maximum Entropy (ME) is another machine learning classifier and makes no assumptions about the relationship between features. ME perform lexicon based method to identifying sentiment words. In order to calculate sentiment strength by manipulating the frequency of matched lexicon according to polarity.SA enables unstructured textual data to structured machine process able data. The information is extracted from social online network. The sentiment analysis of comments, number of likes and shares on posted topic.SA focuses on micro blogging because is the main source of public voice. Twitter is exploited mainly because the data is textual as compared to Facebook data. The emotion in text is explicit and implicit.SA application concentrates on both small and large scale data. Naive Bayes is evaluated for SA. The filtering and pre-processing also is to be updated for SA algorithm.SA is creating a tsunami in many organisations and expanding day by day.

III. METHODOLOGY

The method proposed in this research paper first identifies the suicidal keywords, and then we use these keywords to ex- tract tweets from Twitter using Twitter Streaming Application Programming Interface (API). After that we pre-process the data or extract features from the text document. The system consists of the following components:

- user profile and user data : Database built from the data captured from OSNs.
- **Messages :** There is a database with 360 messages, 90messages for each kind (relaxing, motivational, happy, or calm messages) to be suggested to the user. The users can previously choose one or two kinds of messages when they undergo a period of stress or depression. Depression or stress detection by machine learning: The sentences are extracted from OSN and they are filtered by machine learning to detect depression.
- **Sentiment analysis :** The sentences are filtered and scored by the sentiment metric by giving one threshold value. This range

was tested and validated. The sentiment intensity of the sentence will determine the message intensity. There are three levels of messages: extreme, intermediate and lower. The message intensity levels were determined according to the users opinions. Examples of very positive messages include intensity adverbs, such as much, very, strongly, among others. The proposed method can be divided into five segments. These segments are, identify suicidal keywords, search and Ex- tract tweets, and feature extraction, introducing Naive Bayes, model efficiency evaluation. These segments are described below:

1. Identify Suicidal Keywords

For collecting tweets from Twitter we need certain keywords that contain suicidal intent. A research that collected texts from several sources and analysed these texts to identify keywords that are usually within a suicidal note was very helpful in this regard. Then keywords were further examined and finally there was a list that happen to exist in suicidal notes. We used these keywords to extract tweets from Twitter.

2. Search and Extract Tweets

For this part we used Twitter Streaming API to extract real time tweets containing those specific keywords from Twitter. A tweet objects contains such as user id/screen name and tweet text that are necessary for us. After extracting these we need to process these data as these contain several unnecessary constituents such as special characters, emotions. We used regular expression for removing these unnecessary parts.

3. Feature Extraction

Feature extraction is transforming arbitrary data such as text and images into numerical features that can be used for Machine learning. In this we can converts a collection of text documents into matrix of token counts. This approach implements both tokenization and counting in a single class. Then we can remove words with low importance from the text. For example, “the”, “a”, “is” such type of words remain present in text documents in large number and carries less importance. Even in some cases it may affect model efficiency. Doing so we extracted features from the tweets.

4. NLP and Na¨ıve Bayes

Natural Language Processing for sentiment analysis focused on emotions is extremely useful. Natural Language Processing (NLP) is the best way to understand the language used and uncover the sentiment behind it. Even though the analysis of unstructured data allows us to manage, analyze and extract insight from billions of social media such as tweets must be able to integrate insights from NLP with structured data to get a more complete view. Matching metrics helps to ensure that the next step you choose is one that can make a difference.

Naive Bayes algorithm use to predict whether a review is negative or positive based on text alone. The objective is to train a classifier that given a text review determines if it’s positive or negative. In this test data is document tweets. There are two stages in the classification of documents. The first stage is the training of documents that are known to the category. While the second stage is the process of classification of documents that have not been known category.

Bayer’s Theorem has the following general formula:

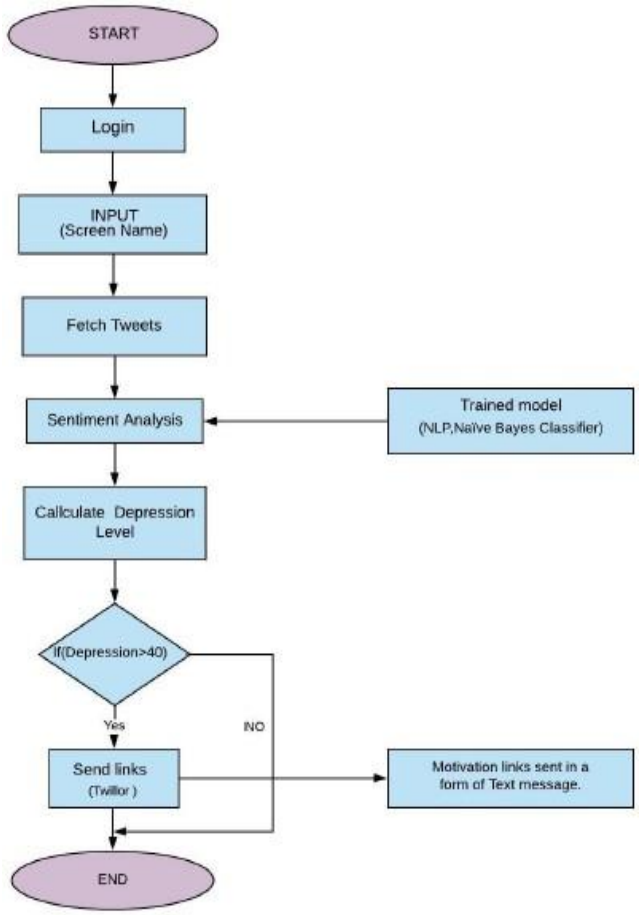

Figure 1. System flow

$$P(H \mid X) = P(X \mid H)XP(H) / P(X)$$

Where: 1. P(H l X) is the probable final probability (posterior probability) of hypothesis H occurs when given evidence E occurs. 2. P(X l H) is the probability that a proof E occurs will affect the hypothesis H. 3. P(H) is the prior probability hypothesis H occurs regardless of any evidence. 4. P(X) is the prior probability evidence of E regardless of the hypothesis or other evidence

Model Efficiency Evaluation

One way of evaluating model efficiency is the accuracy score. For this we split the dataset for training and testing. After training on the dataset we let the model predict the outcome for some cases. From there we calculate the percentage of accuracy.

Sentiment Analysis (SA) requires several basic text processing techniques. So in order to classify data first, we need to perform the following steps:

Tokenization : It is a method that divides the variety of document into small parts called tokens. These tokens may be in the form of words or numbers or punctuation marks. Ex: god is great! I won a lottery. After performing tokenization the sentence is divided into tokens as follows “god”, “is”, “great”,

"I", "won", "a", "lottery".

Stop words: These are the common words that are to be ignored which reduces the size of the dataset also the no of words (tokens). In our programming language pythonwe use a tool called natural language tool kit(NLTK) in which there is list of stop words.

Ex: I like reading, so I read. After removing stop words the sentence will be as follows Like, reading ,read.

Bag of words: This concept is applied to these tokens. A bag-of-words is a representation of text that describes the occurrence of words within a document. The occurrence of words is represented in a numerical feature. It is a way of extracting features from the text for use in modelling, such as with machine learning algorithms. The approach is very simple and flexible and can be used for extracting features from documents. But there is some complexity on two cases

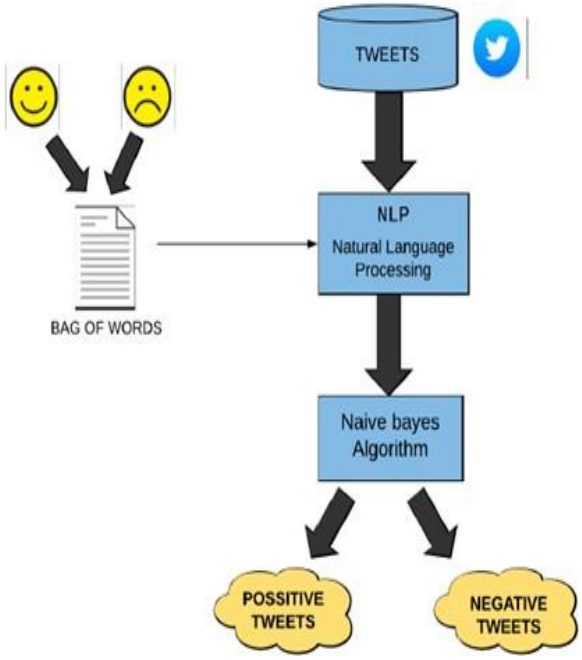

Figure 2. Classification

i.e. one is on designing the vocabulary of known words and the other is on scoring the presence of known words. Let us consider there are 2 classes i.e., positive class and negative class. Each class contains some words that is positive class contains some bag of positive words (slow, fine, good, fantastic) and negative class contains some bag of negative words (hate, terrible, heavy). We will give the input as a text and starts counting the frequency of each word in the document and this gives the result whether the text/sentence belongs to positive class or negative class.

Finally, our classification technique Na¨ıve Bayesian classifier is applied which calculates the probability of all words in the document. We train our classification algorithm using this training set and we get trained classifier. By using this trained classifier we can classify the document. It shows the probability of each tweet saying whether the tweet is either positive or negative.

IV ALGORITHM

A. Pre-processing

The Twitter language model has many unique properties. These properties can be used to reduce the feature space:

1.Usernames

In order to direct their messages users often include twitter usernames in their tweets. A de facto standard is to include @ symbol before the username (e.g. @towards humanity). A class token (ATUSER) replaces all words that begin with @ symbol.

2.Usages of links

Users very often include links in their tweets. To simplify our further work, we convert a URL like "http://tinyurl.com/cmn99f" to the token "URL".

3.Stop words

There are a lot of stop words or filler words such as "a", "is", "the" used in a tweet which does not indicate any sentiment and hence all of these are filtered out.

4.Repeated letters

Tweets contain very casual language. For example, if you search "hello" with an arbitrary number of o's in the middle (e.g. helloooo) on Twitter, there will most likely be a nonempty result set. We use pre-processing so that any letter occurring more than two times in a row is replaced with two occurrences. In the samples above, these words would be converted into the token "hello".

A. Feature Vector

After pre-processing the tweets, we get features which have equal weights. Unigram Features which are individually enough to understand the sentiment of a tweet is called as unigram. For example, words like good, happy clearly express a positive sentiment.

B. Classification

For the purpose of classification of tweets, we make use of Na¨ıve Bayes classifier. Na¨ıve Bayes is a probabilistic classifier based on Bayes theorem. It classifies the tweets based on the probability that a given tweets belongs to a particular class. The Bayes theorem is defined as:

$$P(A \mid B) = (P(B \mid A)\, P(A)) / P(B)$$

where A and B are some events and P() is a probability. This equation gives us the conditional probability of event A occurring given B has happened. In order to find this, we need to calculate the probability of B happening given A has happened and multiply that by the probability of A (known as Prior) happening. All of this is divided by the probability of B happening on its own. We have used the Python based Natural Language Toolkit library to train and classify using the Na¨ıve Bayes method.

C. And finally using formula i.e Depression Level = (No. of negative tweets / Total no. of tweets) * 100 we can calculate depression level of a user and if calculated depression level is greater than our given threshold value we send some motivational video links to that fellow by using his/her contact information.

V FUTURE SCOPE

This model can further explore new issues from the perspective of a social network service provider i.e. Facebook or Instagram to improve the well beings of OSN users without compromising the user engagement.

VI RESULT

The sentiment analysis on twitter messages or data is used for analysis. The Twitter API is used to retrieve the tweets. The Naive Bayes classifier will only work on the tweets that are only in English. The result is represented in the form of the probability. The tweets are stored in database.The negative tweet and positive tweet is classified and by using this depression level of the user is calculated. If depression level is greater than the specified value then users is depressed

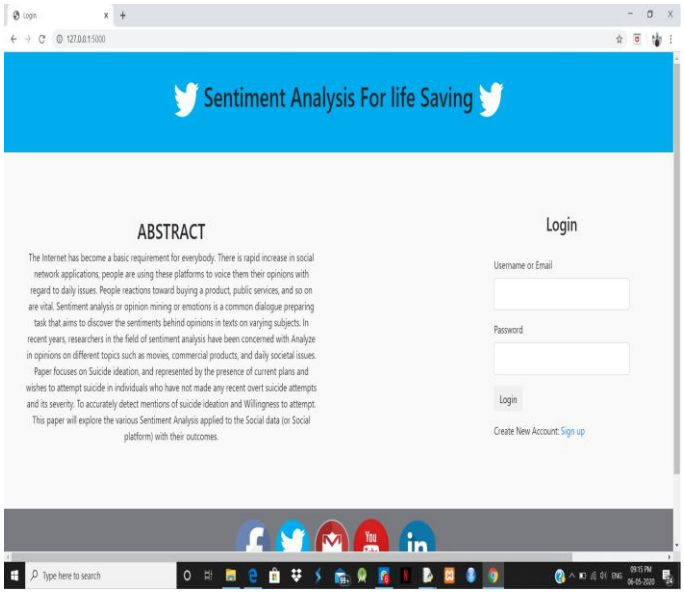

Figure 3 Login

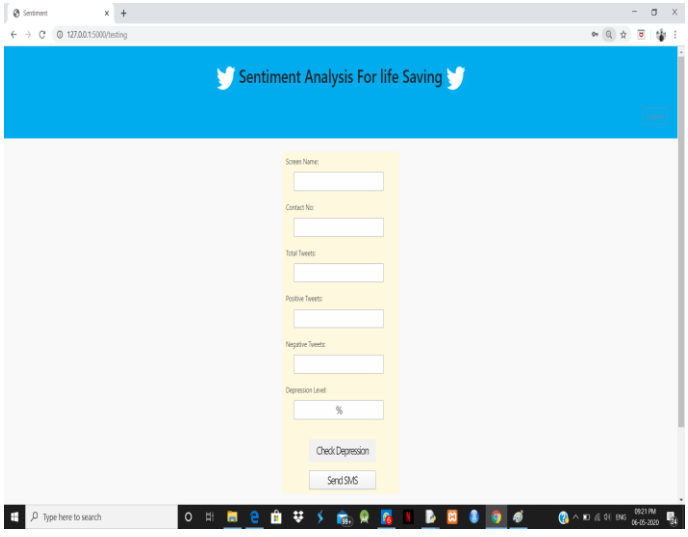

Figure 4 Depression Level Calculation

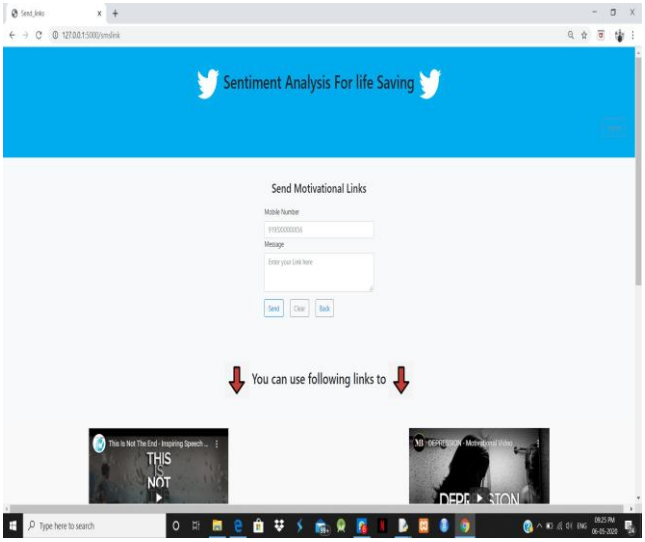

Figure 5 SMS Feature

VII. CONCLUSION

In conclusion, we have developed a model which can extract the Twitter data and sentiment analysis have performed on data using Machine Learning. The natural language processing and Naive Bayes Classifier is used for analysing the tweets. This is very useful to analyse the Mental State of the users and in order to minimize suicidal ideation or suicidal thoughts by send motivational links through SMS. The social media users including their time of usage, their posts and other social activities reflects the mood of the user which is helpful to analyse the state of users. The users can be depressed, stressed, disturbed, sad and emotional status. This model is gives better performance and improves social media data analysis classification.

REFERENCES

[1] Rohith V, D. Malathi, ”Sentiment Analysis on Twitter: A Survey”, published at International Journal of Pune Applied Mathematics, Vol.118, No.22 2018 365-375.

[2] kavya Suppala, Narasinga Rao, ”Sentiment Analysis using Naive Bayes Classifier”, International Journal of Innovative Technology and Explor- ing Engineering (IJITEE) ISSN:2278-3075, Vol.8, Issue - 8 June, 2019.

[3] Sankar. H, Subramaniyaswany.V, ”Investigating Sentiment Analysis us- ing Machine Learning approach”, Conference on Intelligent Sustain- able System (ICISS 2017) IEEE Xplore Compliant - Part Number: CFP17M19 ART, ISBN:978-1-5386-1959-9.

[4] Nurfadhlina Mohd Sharef, Harnanimat Zin, Samnesh Nadali, ”Overview Future Opportunities of Sentiment Analysis for Big Data”, Journal of Computer Sciences, 2016.

[5] Renata L. Rosa, Gisele M. Schwartz, Wilson V. Ruggiero and De- mostenes Z., Senior Member, IEEE, ”A Knowledge Based Recommen- dation System that includes Sentiment Analysis and Deep Learning”, IEEE Transaction of Industrial Informatics Vol.15, April 2019.

[6] Peng Yang, Yunfang Chen ”A Survey on Sentiment

Analysis by using Machine Learning methods."

[7] Pranav Waykar, Kailash Wadhwani, Pooja More, "Sentiment Analysis in twitter using Natural Language Processing (NLP) and classification algorithm", International Journal of Advanced Research in Computer Engineering Technology (IJARCET) Vol.5 Issue.1,January 2016.

[8] Onam Bharti, Mrs. Monika Malhotra, "Sentiment Analysis on Twitter Data", International Journal of Advanced Research in Computer Engi- neering Mobile Computing, IJCSMC, Vol.5, Issue.6,June 2016, pg.601- 609.

[9] Shaoxiong Ji, Ceina Ping Yu, Sai-fu Fung, Shirui Pan, Guodong Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content", Hindawi Complexity Volume 2018, Article ID 6157249.