International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Mood and Vulnerability Prediction through Natural Language Processing

Debabrata Datta, Srijita Majumdar, Olie Sen, Aparna Sen

Abstract*: Analyzing various phases of mood using the verbal form of writing can serve as advancement in the field of psychology. The research work highlighted in this paper focuses on the use of sentiment analysis to predict the emotional state and vulnerability of written statements, as per the most generic perceptions, in the English language, with the help of an algorithm. The text pre-processing step discussed in this work involves cultivating and analyzing each word of user input, analyzing their literal and emotional essences to sum up the mood inclination of the statements and other parts-of-speech, to determine the specific mood and the vulnerability of the writing itself. The vulnerability level of the document is also determined, in order to extent out the purpose towards medical treatments where a vulnerable mindset, suffering from mental illness, depression, perceives the capability to inflict harm upon oneself or others can be given proper help and counseling.

***Keywords* : Mood Inclination, Text Mining, Natural Language Processing, Connotation Analysis, Porter Stemmer Algorithm.**

I. INTRODUCTION

Mental healthcare is one of the biggest unmet needs of the modern era. A huge population of people suffer from depression, anxiety disorders or any another mental health ailment. Young people are especially prone to these troubles. Yet millions of people living with these conditions do not particularly recognize this as a serious issue.

Predicting a person's sentiment by surveying his/her written statements could have a number of beneficial clinical applications. This technique can be further extended out towards collecting written statements from the person's social media account over time [1], to conclude upon the genuine problems of the person's mental state as users nowadays routinely share their thoughts, opinions, feelings as well as their daily activities on various social networks.

Emotions evoked by words (either explicitly, implicitly or none), would be considered in a sentiment analysis procedure through the medium of text mining – a multidisciplinary field extracting interesting and non-trivial knowledge or information about words and their denotative and connotative meaning from unstructured text data.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Debabrata Datta*, Department of Computer Science, St. Xavier's College(Autonomous), Kolkata, India. Email: debabrata.datta@sxccal.edu
Srijita Majumdar, Department of Computer Science, St. Xavier's College(Autonomous), Kolkata, India. Email: subho1219@gmail.com
Olie Sen, Sprinriver Technology Private Limited, Kolkata, India. Email: olie4u7@gmail.com
Aparna Sen, XCD HR Private Limited, Kolkata, India. Email: senaparna530@gmail.com

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

The system if developed further can be used to identify any kind of suicidal thoughts or motives involving revenge or hate attacks and thereby to predict any upcoming damage, not only self harming possibilities.

The method described in this work can be used to develop a quick desktop application on a smart phone application to produce instant access and prediction, with or without a proper internet connection.

Sentiment analysis or the mood detection in this case, is done by analysing each word of the user input and categorizing them into certain kind of positive, negative and neutral forms and then their categorical and subjective placements in the sentence with the help of the algorithm would result into the final output of the sentiment. This method, thereby, eases out the whole process.

This work resembles a stark deviation from the traditional counseling techniques that involved people visiting a psychiatrist. This traditional procedure involved several drawbacks as firstly, the person may be hesitant to open up completely. Secondly, treatments and counseling may vary across practitioners and their subjective methods and finally and most evidently, it would be extremely time consuming to come to a conclusion after doing an extensive background checks upon a patient's medical history and other subjective issues.

II. RELATED WORK

Sentiment analysis in any domain basically starts off by analyzing the sentimental polarity of the source elements by assigning various standards and quantities to it, having developed several categories, whether it is a recommendation for a movie or an emotion expressed or a genre of literature. The participation of several parts-of-speech of the respective language(s), is evident, given that they impart a significant deal of sense and emotion in the way a particular fact or a speech is capable of being put or expressed in more than one manner. Certain other related work in this field takes up little varying measures to a certain degree, retaining the approach to be as generic as ever.

One approach has been undertaken where POS-tagging and dependency trees have been used to analyze the sentence constructs [2]. Various parts of speech like adjectives, verbs, nouns, adverbs, conjunctions and prepositions which have been mostly dealt with, act as feeling words or affect the sentiment of the phrase, especially, the effects of conjunctions in detail on the overall semantic orientation of the sentence is analyzed. Also, the positive and negative polarities of words are determined without any limitations of a word list.

Instead WordNet has been used to find the semantic orientation of the words which are not found in the General Inquirer word list.

A similar approach was taken up by Matsumoto, et al. in [3] to recognize the word order considering syntactic relations between words to be an extremely important aspect in the area of sentiment classification, and therefore it was imperative that they were not discarded. They had used text mining techniques to extract frequent word sub-sequences and thereby constructed dependency sub-trees from sentences in a document dataset to use them as features of support vector machines in the experiments on movie review datasets.

Document level sentiment classification [4] has assumed the whole document to have a single overall sentiment e.g. a simple unsupervised learning algorithm was used for recommendation percentage in reviews and classifying them predicted by the overall average semantic orientation of the phrases that contain adjectives or adverbs. The phrases were categorized based upon their type of associations (good/bad/neutral or factual) that they made as per the general perception was concerned. A review was classified as recommended if the average semantic orientation of its phrases was positive and vice versa. Similar approaches were taken up in [5][6].

A large subset of sentiment expressions was identified in [7] by foremostly determining whether an expression was neutral or polar and then disambiguating the polarity of the polar expressions.

However, the foundation on sentiment analysis was put forward by Pang and Lee in their research work [8] where Naive Bayes, Maximum Entropy, and Support Vector Machine approaches were used to compare sentiments of movie reviews.

Also, Natural Language Processing or NLP along with text mining is extensively used in the field of sentiment analysis [9][10][11][12][13]. With the advancement of technology, more and more data are available in the digital form, among which most of the data are unstructured. So it has become essential to extract useful and interesting information from this large amount of textual data. Hereby, text mining comes in, to use the new useful information that are generated from a large collection of existing information to obtain meaningful structured data.

A search using text mining is efficient as it reads and does text analysis of the document on the user's behalf. It can understand the real meaning with the help of NLP which allows it to recognize similar concepts, inherent patterns, even if things are expressed in a variety of ways and forms e.g. in different languages or spellings. With this facility in hand, searching of any kind of data becomes easier as well as more systematic.

To be able to understand human speech, a technology must also understand the grammatical rules, meanings and contexts, slangs, tones and acronyms used in the language and thereby the goal of NLP had always been to make interactions between computers and humans feel exactly like interactions between two human beings. Without NLP, artificial intelligence can only understand the meaning of language and answer simple questions, but would not able to understand the meaning of words in context. However, in its presence AI can account for communicating with intelligent systems using natural language processing, which helps computers read and respond by using the human ability to understand the everyday language that people use to communicate with each other.

III. PROPOSED METHODOLOGY

The proposed research work sentiment analysis has been developed on the concepts of natural language processing.

Because of this, the following paragraph briefly discusses about NLP:

There are many approaches used for processing the natural language input, these approaches include-

- Symbolic approach
- Statistical approach
- Connectionist approach

In the proposed work, emphasis has been given primarily on the Symbolic approach which is based on the rules and regulations of the particular input language. Linguistic experts materialize and record these rules, which are then followed by the computer systems. The method for mood and vulnerability prediction to be illustrated hereafter, has utilized the symbolic approach, to categorize text based upon the essence that they are known to provide, popularly, by following certain prevalent linguistic constraints and parts-of-speech axioms of the English Language.

Any traditional text categorization framework comprises of the basic steps of preprocessing, feature extraction, feature selection and classification.

A. Analysis through Text Pre-processing

The stage of 'Preprocessing' usually consists of the tasks such as tokenization, filtering, lemmatization and stemming. In the following, these methods have been explained in context with the connotation analysis and prediction of mood inclination used in the present research work-

Tokenization:

Tokenization is basically the task of breaking up a character sequence/ string up into pieces (words/phrases) called tokens, and perhaps at the same time throw away certain characters such as punctuation marks. The list of tokens then is used to further processing. In this context of the proposed work, the connotations and certain special category of words, words that impart a temperamental essence to a statement, have been considered to be the desired tokens.

Filtering:

Filtering is usually associated with the removal of some words that appear in the text without having much content information. As in the case of the proposed algorithm, meaningless stemmed words, neutral connotations, numbers and successively appearing punctuations and sentence terminators have been filtered away as they essentially incremented the processing time of the algorithm. Removing them had made it more convenient to analyze the relative placements of various connotations, in order to make faster and correct prediction.

Stemming:

Stemming methods aim at obtaining the stem (root) of derived words. The stem needs not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

There are several types of stemming algorithms which differ in respect to performance and accuracy and how certain stemming obstacles are overcome. In this work, Porter's Stemmer Algorithm [14] has been used for performing several checks upon etymologically related words.

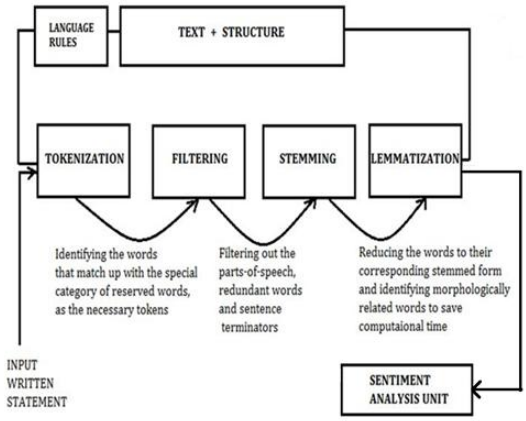

Figure I: Text pre-processing unit for sentiment analysis

Lemmatization:

Lemmatization is the task that considers the morphological analysis of the words, i.e. grouping together the various inflected forms of a word so they can be analyzed as a single item trying to map various forms to a singular form. It is related to stemming, differing in that it is able to capture canonical forms based on a word's lemma. In the present context, the aforementioned connotations have been considered as the lemmas. Words falling under a single connotation would emit out similar disposition around its adjacent words and phrases. For example, stemming the word "worse" would fail to return its citation form or lemma. However, lemmatization would identify the nature of the word.

The major steps of the proposed method are a) emotion analysis and b) ultimate mood prediction and vulnerability detection. These are described below:

B. Emotion Analysis

Emotions play an important role in our daily activities as quite a significant amount of time is spent witnessing the emotions of others, interpreting what these signals might mean, determining how to respond to and deal with the complex emotional experiences. Emotions are expressed in a number of different ways including both verbal communication and nonverbal communication e.g. body language, facial expressions, voice modulations, method of speech, writings etc.

The main target of the proposed method was to predict and interpret the mood of an individual through his/her written expressions/writings as a whole. Several aspects of writing, primarily in the English language have been analyzed and focus has been given mainly upon elements that have provided a temperamental polarity to the statement. The most primary aspect serving this cause would be a connotation, which is a temperamental association evoked by a certain word or a set of words. Depending on how a word has been used over time, it may have a positive, negative or neutral connotation. For example, '*childish*','*childlike*' and '*youthful*' have the same denotative, but different connotative, meanings. *Childish* and *childlike* have negative connotation, as they refer to immature behavior of a person. Whereas, *youthful* implies that a person is lively and energetic. Also, the aforementioned basic types of connotations (positive, negative and neutral) are classified further into certain subtypes based upon several studies and rigorous observations of English texts.

Similarly, the words like steadfast, tenacious and stubborn are the three words with similar dictionary meanings but varying connotations i.e. positive, neutral and negative respectively.

These three basic types have been further broken down into subtypes depending upon the impact they make upon the polarity of a sentence as we proceed through it. This has been shown in figure II.

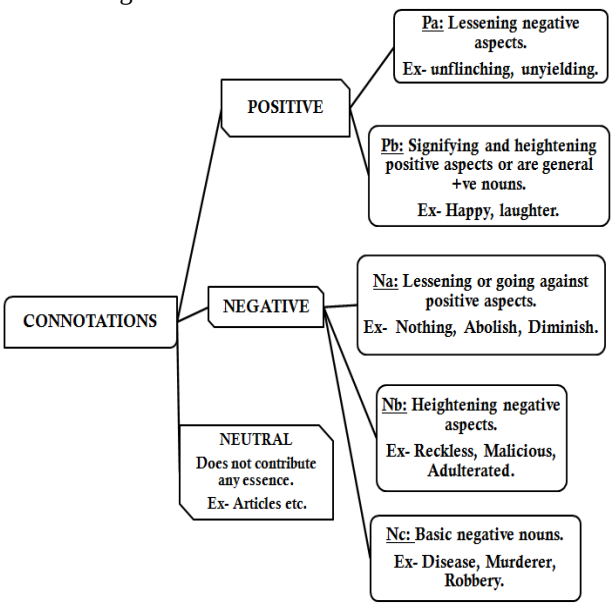

Figure II: Connotation types and sub-types

Such connotations and their mutual placements need to be observed to create a sentiment-analyzer or mood-predictor application by analyzing the usage of various parts of speech from a certain piece of writing or message. Also, the other various parts-of-speech taken into consideration include –

Words deciphering intensity of other words, including both comparatives and superlatives:

These words were collectively termed as 'Intensity Words' in the present work. Here the mood inclination of a writing is heightened in their presence, e.g. words like "more", "extreme", "extremely", "too", "further", "farther", "high", "low", "less", "much", "more", "many", "very", "best", "better", "worse", "worst" or any sort of comparative or superlative terms. They could be used to predict the level of vulnerability of a person's mood and were thereby needed to be tracked down separately.

Interjections (Positive and Negative):

An interjection is a part of speech that shows the emotion or feeling of the author. These words or phrases can stand alone or be placed before or after a sentence. Many times an interjection is followed by a punctuation mark, often an exclamation point. Examples include "Wow!", "Bravo!", "Hurrah!", "Phew!", "Yeah!" etc. as positive feelings and "Oh!", "Oops!", "Argh!", "Eww!" etc. as negative feelings.

Conjunctions, Punctuation Marks and Contradictory clauses:

These are useful for marking token boundaries and making decisions with respect to the placement of connotations, as will be demonstrated hereafter in the proposed algorithm.

Neutral connotations:

The neutral tone is said to be "light" or "de-emphasized," meaning one does not have to give it the same amount of stress, and it should actually be a bit shorter than the other tones.

In the proposed prediction algorithm, a separate categorization for the neutral connotation has not been framed. Instead, the ones that do not fall under the previous categories have been eliminated, during training, considering them to be neutral or having no significant effect on the result. Also, the database for such an application, requires containing the necessary categorization on the basis of the several variety of connotations (positive, neutral, negative) and other necessary parts of speech that would also take up a significant role in determining the essence of an given text (ex- comparative and superlative words, contradictory classes, punctuation marks, interjections etc.) for the convenience of scanning and prediction for algorithms, to work on deriving the mood/sentiment of a particular piece of writing.

C. Raw Data Analysis

The pathway for the basic algorithm development was performed in the following manner:

a. Determining the various basic categories of connotations

b. Performing stemming on all the words to lessen the cost of connotation derivation

c. Performing analysis on a given stemmed data set to detect the following:

i. Words falling under the aforementioned categories of connotations, intensity words, conjunctions etc.

ii. Their relative positioning and the alteration of temperaments [15]

iii. Deriving relationships between various categories of connotations and determining the resultant form.

iv.Utilizing a greater amount of data set to come across newer derivations.

Given below is an analysis of texts and categorizing the ones falling under the positive and negative categories of connotations, from a given data set taken from various status updates on a renowned social media platform. The underlined words have been considered-

[P]- Positive Connotation

[N]- Negative Connotation

i. My uncles had told me stories about how famous (P) my dad had been on campus for having simultaneously raised hell (N) and aced (P) all his classes.

ii. Although I do not (N) expect (P) much to come out of this ruling (P) at this stage, I hope (P) this is the first step in something much, much bigger.

iii. England did well (P), and will probably be happy (P) as semi-finalists, but surely they would not (N) get much better chance (P) to reach the finals than this time.

iv. Belgium had a brilliant (P) world cup, but France seemed more balanced (P) and just did the job this match. This France team is scarily (N) talented (P), and seems that they can pull off another gear if required.

v. It's nice (P) to see that the turmoil (N) in both the teams was not (N) a hindrance (N) for a good (P) game of football.

Resultant observations from the analysis:

i. The literal essence of a sentence is primarily dependent upon the placement of negative vocabulary/connotations, than much on the positive ones. The positive vocabulary signifies only the positive essence, while the essence varies with the placement of negative vocabulary with respect to each other or with the positive vocabulary [7].

ii. The literal essence of a word is partially dependent upon the connotative words present around it in the text [10].

iii. Intensity words do not signify a positive or negative essence until they are associated with yet another positive or negative connotation; therefore they need to be tracked down separately for predicting the critical level rather than actually predicting whether the mood or the opinion is positive or negative. Comparatives are often used with a conjunction or other grammatical means to indicate with what the comparison is being made.

iv. Case of positive connotations: One or more positives placed adjacent to each other, or in a single sentence always signify a positive essence.

v. Case of negative connotations: If two negatives are adjacent to one another, some of the following cases may arise -

a. <Negative verb or adjective><negative noun> leads to a negative connotation.
   Examples include: reckless murderer, conspiring treachery etc.
b. <Negative verb or adjective><negative noun> leads to a positive connotation.
   Examples include: abolish child labour etc.

Also the essence keeps on varying with the placement of various forms of negative alongside various kinds of positive.

vi. If two connotations (positive or negative) are separated by a conjunction or a punctuation mark, then they mostly result in the form of the last connotation derived [2].

vii. If a sentence is preceded by an interjection; it may then save the time and effort for scanning through the entire sentence. If the interjection is negative then that unit sentence may be derived to possess a negative essence and vice versa. Examples of positive interjections include: Wow!, Amazing!, Wonderful!, Cheers!, Congratulations! etc. whereas negative interjections include: Oops!, My Goodness!, uh-oh! etc.

viii. If two connotations are separated by 'but' the later connotation is taken into consideration.

ix. If two connotations are separated by 'or' then either one of them can be taken into consideration.

x. Further detailing considering other popular clauses like - Either…or, Neither…nor, Not only…but also, Scarcely...when may be put into axioms for further precision in analysis.

xi. When predicting a sentiment from a group of statements has been dealt with, usually the connotation derived from the very last statement has provided for the major sentiment of the group unless they began with a contradictory clause, in which case the connotation derived from the former part has been taken into consideration.

xii. The Neutral Connotation (Nt) serves as the identity factor for all other connotations.

< Neutral connotation (Nt)> <positive or negative connotation > = < same connotation >

xiii. Operations on connotations may or may not be associative in nature.

Table I depicts the way the essence of a particular sentence is derived with the help of few examples.

Table I: A generic placement pattern to derive the essence of a sentenc

1st Placement	2nd Placement	Resulting Connotation	Example
N_a	N_a	P_b	No(N_a) other deaths (N_a) occurred today for Chicago fire.
N_b	N_b	N_a	Ever since my brother got that car for his birthday, he's been motoring around at night with reckless (N_b) abandon (N_b).
N_a	N_c	P_a	Abolishing (N_a) slavery (N_c)
N_c	N_a	-do-	-do-
N_b	N_c	N_c	Malicious (N_b) intruder (N_c)
N_c	N_b	-do-	-do-
N_a	N_b	P_a	Diminishing (N_a) adulterated (N_b) foods in the market
N_b	N_a	N_a	The prime suspect behind the series of bomb blasts was sentenced with a critical (N_b) punishment (N_a).
N_a	P_b	N_a	Diminishing (N_a) prosperity (P_b)
P_b	N_a	-do-	Punishing (P_b) an innocent (N_a)
N_b	P_b	P_b	Contagious (N_b) happiness (P_b)
N_a	P_a	N_b	obody (N_a) surpasses unrivaled (P_a) in his life.
P_a	N_a	-do-	-do-
P_a	N_b	N_b	Supporters of the two parties posses an unflinching (P_a) hatred (N_b) for one another.
N_b	P_a	N_b	-do-
P_a	N_c	N_c	e used to be quite an unyielding (P_a) criminal.
N_c	P_a	N_c	-do-
N_c	N_c	N_c	The eeriness (N_c) of his behavior caused me severe discomfort (N_c).

Since the assignment of the connotations totally depends on human interpretations of their meanings, the aforementioned table therefore needs to contain a significant amount of words placed by user intervention and understanding, in order to determine the connotation
for groups of statements.

D. The ultimate Mood Prediction and Vulnerability Detection

After having predicted the connotation using the above method, intensity words can help specifying the right mood of the text. They are usually of two types: comparatives and superlatives.

While comparative words usually describe subtle transitions from the current towards a corresponding extremity, superlative words describe the extremities in mood predicaments.

Examples of comparatives: better, worse, lesser, greater, poorer etc.

Examples of superlatives: best, least, extremely, greatest, worst etc.

Vulnerability level here specifies the instability of one's mood. Since, over here the greater objective of this work is to consider whether a person with his corresponding mood, have the suitable help or treatment or not, this category was considered to form a suitable parameter for or objective.

The following operations may be performed when a predicted connotation, total count of connotations and a separate count of comparatives and superlatives are given:

i. To calculate the percentage of comparatives and superlatives from the given numbers.

ii. In the final mood prediction, therefore, three basic elements are needed for prediction- the resulting connotation, count of comparative, count of superlatives.

The expressions to calculate the percentages of comparatives and superlatives are given below:

comparative% = (number of comparative words) / (total number of intensity words) * 100

superlative% = (number of superlative words) / (total number of intensity words) * 100

So the required mood and vulnerability level = connotation (mood inclination) + % of intensity words.

Table II depicts the mood and vulnerability specifications from all possible combinations of the desired parameters. The connotations have been derived from an algorithm called Mood_Sense ( ) which is described next.

Table II: Mood and vulnerability level prediction

Connotation derived from the algorithm- Mood_sense ( )	Comparative %	Superlative %	Corresponding mood	Vulnerability Level
Positive	0	0	Satisfied	None
	>50	<50	Happy	None
	<50	>50	Cheerful	None
	100	0	Cheerful	None
	0	100	Exuberant	None
	50	50	Jovial	None
Negative	0	0	Stressed	Mild
	>50	<50	Melancholic	Intermediate
	<50	>50	Depressed	High
	100	0	Angry	Critical
	0	100	Extremely shocked and angry (or in special cases, suicidal)	Extreme
	50	50	Worried	Mild
Static	×	×	×	×

E. Algorithm for deriving the connotation of a given input: MOOD_SENSE()

Input: The sentence/group of sentences, given by the user, called 'Input'.

Output: The specific mood inclination, mood and the vulnerability level as derived.

Used data structures: Lists – Stemmed [], Statement [], Connotation [].

BEGIN MOOD_SENSE( )

1. The 'Porter's Stemmer algorithm' is used to reduce each word in the statement to its respective stem, and this reduced form is kept aside in a variable, called 'Stem'.

2. Input is checked separately, again, for any consecutive occurrences of spaces & punctuations (i.e. sentence terminators or pauses) e.g. '.', ',' , '!', '?', ';' , ':' , and the spaces in such occurrences are reduced.

3. The sentences from Stem and Input are extracted w.r.t the aforementioned sentence terminators ('.','?','!')/

Punctuations and are stored separate lists, say Stemmed [] & Statement [], respectively; the length of the lists being equal to each other and the number of full-length statements given as input.

4. A list, called Connotation [], is kept for storing the resultant connotation derived for each statement of Stemmed [] / Statement [], the length of the list also being equal to Statement []. The initial value of each of the list is initialised with the symbol of the NEUTRAL connotation.

5. Every word from Stemmed [i] (where, i <= the total number of statements) is extracted and is checked with the following categories:

a. The pre-existing database of the aforementioned connotations.

b. Certain conjunctions like 'but', 'yet' etc.

c. Superlatives, Comparatives (Collectively referred to as 'Intensity words').

d. Positive and Negative Interjections.

e. Contradictory clauses like 'although', 'inspite of', 'either or', 'neither nor', 'despite'.

Since the Porter's Stemmer algorithm does not ensure providing a 100% meaningful stemming for all sorts of words, a check with the original word i.e. word from Statement [i] has been done, in case the stemmed word i.e. word from Stemmed [i], does not match any of the above mentioned categories.

5.1. If the word from Statement [i] does not match any of the above categories, then it is either not considered or is placed in the neutral category during the training process.

5.2. For each of the words from Stemmed [i] / Statement[i] satisfying a category of connotation base, the respective connotation symbol is appended to Connotation[i], thereby deriving a string of connotations for each of the statements.

5.3. A single connotation from a single statement can be derived by the following procedure mentioned below:

The connotations derived from the words are placed subsequently and then using table 1, the resultant connotation of that single statement is derived. As an example, for the sentence 'For years, he had driven (N_a) and criticized (N_a) and condemned (N_b) his employees without (N_a) stint (P_b) or discretion (P_b).',

The connotations are traced out and are placed as follows keeping both the punctuation marks and the conjunctions in track-

N_a 'and' $N_a = N_a \rightarrow N_a$ 'and' $N_b = N_b \rightarrow N_b + N_a = N_a \rightarrow N_a + P_b = N_a \rightarrow N_a + P_b = N_a$

Thus, Na is the resultant connotation.

Similarly, later this procedure is repeated for all the constituent sentences and the placement is done in a similar fashion to derive the final connotation of the writing as a whole.

5.4. A separate count of words falling in the categories - Superlative, Comparative is kept.

5.5. Words falling under the suicidal category are kept in count, called 'Suic', inorder to be made useful during predicting the vulnerability level of the input statement.

5.6. If the first word is a negative interjection, then the resulting connotation for the entire sentence Statement [i] is classified under category N_a and is stored in Connotation[i].
5.7. If the first word is a positive interjection, then the resulting connotation for the entire sentence Statement [i] is classified under category P_a and is stored in Connotation[i].
5.8. If the first word extracted from a sentence is a contradictory clause e.g. 'but', 'yet', 'although', 'Inspite of' etc. then the connotation that follows it, without any punctuation marks in between, is not taken into consideration. However, if there is a punctuation mark, then an operation is performed on the derived connotation.
6. Steps 5.1 to 5.8 are repeated, till the end of Statement [i].
7. 'i' is incremented by 1.
8. Steps 5 to 8 are repeated until the very end of input statements when all the statements are traversed and their respective connotations (strings) have been stored.
9. Every connotation string in each of Connotation[i] can be further compressed to a single connotation symbol for each statement, using the axioms described in table 1.
10. Using the similar set of axioms, all the elements of Connotation[] are appended into yet another string from where a single connotation/mood inclination is derived for the entire input as a whole.
11. The percentages of comparative and superlative words are calculated using the following formulae mentioned earlier.
12. Using the above intensity word (%), count of suicidal words (i.e. Suic) and the derived Mood Inclination from algorithm Mood_Sense (), we can conclude upon one specific mood and its associated vulnerability level, as illustrated in Table 2.
END MOOD_SENSE( ).

IV. RESULTS OBTAINED AND FURTHER ANALYSIS

A. Training

The procedure for training mainly comprises of developing the database with newly categorized words, with some amount of user intervention. If a word or its stemmed form does not match up with any of the pre-existing categories of connotations or 'special category' words (e.g. interjections, intensity words etc.), it is then left upon the user to categorize it, most generically, else it is left in the neutral connotation.
A sufficient amount of time given to feed a significant amount of sample input statements into the databases to categorize the newly found words/ stems enables the algorithm to give even more precise results in terms localizing the specific sort of mood inclination and the vulnerability level.
The algorithm has been designed in such a way that a word falling in two or more category of connotations would be adjusted with respect to the placement of the other connotations around it and the percentage calculation of intensity words. Also, with the help of the contradictory clauses, the algorithm gets a sense of which portion of an input statement accounts for the major mood inclination of the document.
Thus, with a combinatory effect of grammatical, etymological rules and some amount of a generic user perception, a certain level of effectiveness has been achieved. Therefore, after extracting each word and its stem form, the following algorithm has been incorporated.

B. Algorithm for training with user intervention: USER_INPUT( )

BEGIN USER_INPUT ( )
Step 1. Initialize categorized = 0.
Step 2. Enquire whether STEM may be categorized as –
i. A positive or negative connotation.
ii. A positive or negative interjection.
Step 2.1. If (User_Decide (STEM) = True)
Step 2.1.1. Place STEM in category.
Step 2.1.2. Set categorized = 1.
Step 2.1.3. Check for STEM for superlative, comparative and suicidal category of words.
Exit.
Step 2.2. Else
Step 2.2.1. Enquire whether WORD maybe categorised as -
i. A positive or negative connotation
ii. A positive or negative interjection.
Step 2.2.1.1. If (User_Decide (WORD) = True
Step 2.2.1.1.1. Place WORD in category.
Step 2.2.1.1.2. Set categorised = 1. Step 2.2.1.1.3. Check for WORD for superlative, comparative and suicidal category of words.
Exit.
Step 2.2.1.2. Else
Step 2.2.1.2.1. Place WORD in neutral category.
Exit.
END

In the above algorithm, the following identifiers have been used:

categorised – A flag variable that indicates that a word has been made to fall within a particular category by the user himself.

User_Decide ( ) – A method that accepts a word or its stem and returns 'True' when the user comes upon with the category and sets the newly found word to it.

C. Effects of training on a sample data

Demonstrated below is a set of statements given as input to algorithm MOOD_SENSE () before and after the training process.

Table III: Before training

No.	Input Statement	Basic mood inclination	Specific mood	Vulnerability level	Validity
1	It's not good to criticize someone unnecessarily.	+ve	Satisfied	None	✗
2	Often, parents are tempted to criticize their children.	-ve	Stressed	Mild	✓
3	The sun sets in the west.	Neutral	Static	None	✓

4	Count your age by friends, not years. Count your life by smiles, not tears.	+ve	Stressed	Mild	×
5	s, this is the reality of Indian Politics. No one is ready to take the responsibility. Only they will allegate each other for the Amritsar Tragedy.	-ve	Depressed	High	✓
6	After pouring out all my anger, I felt relieved.	-ve	Stressed	Mild	×
7	When time comes, I will die like hero.	-ve	Stressed	Mild	×
8	We run from things that scare us.	-ve	Stressed	Mild	✓
9	Things that scare us the most, should be done first.	-ve	Extremel-y shocked and angry (or in special cases, suicidal).	Extreme	×
10	To think only the best, to work only for the best and to expect only the best. To be just as enthusiastic about the success of others as you are about your own.	+ve	Cheerful	None	✓
11	To the well-organised mind, death is but the next great adventure.	+ve	Cheerful	None	✓

With the above input statements, before training, the result has got an accuracy of approximately 55%.

Table IV: After training

No.	Input Statement	Basic mood inclination	Specific mood	Vulnerability level	Result
1	It's not good to criticize someone unnecessarily.	-ve	Stressed	Mild	✓
2	Often, parents are tempted to criticize their children.	-ve	Stressed	Mild	✓
3	The sun sets in the west.	Neutral	Static	None	✓
4	ount your age by your friends, not years. Count your life by smiles, not tears.	-ve	Stressed	Mild	✗
5	es, this is the reality of Indian Politics. No one is ready to take the responsibility. Only they will allegate each other for the Amritsar Tragedy.	-ve	Depressed	High	✓
6	After pouring out all my anger, I felt relieved.	+ve	Satisfied	None	✓
7	When time comes, I will die like hero.	+ve	Exuberant	None	✓
8	We run from things that scare us.	-ve	Stressed	Mild	✓
9	Things that scare us the most, should be done first.	-ve	Extremely shocked and angry (or in special cases, suicidal).	Extreme	✗
10	To think only the best, to work only for the best and to expect only the best. To be just as enthusiastic about the success of others as you are about your own.	+ve	Cheerful	None	✓
11	To the well-organized mind, death is but the next great adventure.	+ve	Cheerful	None	✓

Resultant accuracy of the sample now becomes almost 82%, which is a significant improvement on the previous accuracy level of almost 55%. However, the greater the amount of training, the more surplus the amount of data, the accurate will be the results over time. Thus, this approach enables the algorithm to achieve a long-term fidelity, rather than promising a constant and direct precision.

D. Testing

The testing of the proposed algorithm has involved providing a certain amount of random English statements as inputs and observing the results, and the accuracy. Once this model has passed the validation process, and the used database has been sufficiently updated, it is ready for data without any particular targets. This will show whether it is able to deal with real world data. If the algorithm does well during testing, it is ready to be used for the purpose it was designed for. The following statements describes the results obtained during the testing the proposed method with sample input texts.

The proposed algorithm was implemented with the help of the object oriented features and tools available with jdk8.0 and jvm8.0 and the inbuilt packages from JAVA programming language, in the most generic manner. The application was passed through a series of tests. The obtained results are shown here, in the form of a table given below:

Table V: Results obtained after testing

No.	Input Statement	Basic mood inclination	Specific mood	Vulnerability level	Result
1	I want you to tell me about all the wonderful places that you have visited and the sights that you have seen.	+ve	Satisfied	None	✓
2	Keeping religion immune from criticism is both unwarranted and dangerous. Unless we are willing to expose religious irrationality whenever it arises, we will encourage irrational public policy and promote ignorance over education for our children.	-ve	Angry	Critical	✓
3	Two things are infinite: the universe and human stupidity; and I'm not sure about the universe.	-ve	Worried/ Annoyed/ Confused	Mild	✓
4	I went to the worst of the bars hoping to get killed but all I could do was get drunk again.	-ve	Suicidal	Most extreme	✓
5	Sometimes it may be easy to say a lie and get away from a situation. But whatever may be the consequences of being honest, we must always be truthful in our words and deeds. It shows integrity of our character and strength of our morals and ethics. When we are honest we earn others respect.	+ve	Cheerful	None	✓
6	The use of power for personal interest or gains is termed as corruption. It is a common problem that leads to various other serious issues. Corruption occurs at various levels in our country. From the general public to the political leaders to the big businessmen, everyone is contributing to this problem and making it bigger by the day.	-ve	Worried/ Annoyed/ Confused	Mild	✓
7	Pollution is something harmful substances are added to the environment. Polluted water or garbage in the water bodies is a type of pollution. It adds germs and viruses.	-ve	Worried/ Annoyed/ Confused	Mild	✓
8	Infants born in poverty are weak and have less birth weight. This leads to many physical and mental disabilities. Further, the health and growth of infant is affected due to lack of resources and unhealthy surroundings. The infants in poor economic conditions are not only more likely to be sick and unhealthy they are also more likely to die due to undernourishment and malnutrition.	-ve	Melancholic	Intermediate	✓
9	I can prove that I am not lying.	+ve	Satisfied	None	✓
10	Impartially, shrewdly, I considered suicide, though not in my worst moments. The bottle of pills. The note: 'No hard feelings, everyone, but I've thought about it and it's just not on, is it? It's nearly on, but not quite. No? Anyway, all the best.	-ve	Suicidal	Extreme	✓
11	To the well-organized mind, death is but the next great adventure.	Static: Not +ve or -ve	×	×	✓
12	Today, many people are living in poverty all over the world. It means insecurity, helplessness, and marginalization of individuals and families. It means being exposed to vulnerability and humiliation.	-ve	Worried/ Annoye/Confuse-d	Mild	✓
13	The walls of her room are screaming out loud all of the secrets that she's held onto for all these years, every tear that broke her heart and slashed her wrist, and every memory that rips apart her soul. And to this day, she still can't breathe.	-ve	Suicidal	Extreme	✓
14	After all, life's better when we're happy, healthy and successful	+ve	Jovial	None	✓

15	You will never be happy if you continue to search for what happiness consists of. You will never live if you are looking for the meaning of life.	-ve	Stress-ed	Mild	✓
16	I am both happy and sad at the same time, and I'm still trying to figure out how that could be.	-ve	Worried/ Annoyed/ Confuse-d	Mild	✓
17	If he speaks again without me knowing who he is, I will thrash him and throw him out of the window. And I won't open it first.	-ve	Angry	Critical	✓
18	Most of the time I wish I was dead. I don't want to be here anymore. I'm tired of this. I think to much. I'm never okay. I'm always faking a smile. I always care and get hurt. My very own thoughts are suffocating me.	-ve	Suicidal	Extreme	✓
19	The anger welled inside me, with nowhere to go. I could feel it eating away at me. I knew if i didn't find a way to release it, it would destroy me.	-ve	Angry	Critical	✓
20	We are sometimes dragged into a pit of unhappiness by other's opinions that make us unhappy.	-ve	Depre-ssed	High	✓
21	Success is not the key to happiness. Happiness is the key to success. If you love what you are doing, you will be successful.	Static: Not +ve or -ve	×	×	✓
22	Here is a gnawing and unfaltering dream, and the rare individual who honestly satisfies this dream will always be held by people in their heart.	+ve	Satisfi-ed	None	✓
23	The misfortune of a young man who returns to his native land after years away is that he finds his native land foreign; whereas the lands he left behind remain forever like a mirage in his mind.	-ve	Stress-ed	Mild	✓
24	My purse just got stolen!	-ve	Stress-ed	Mild	✓
25	Sometimes, hate is just confused love!	-ve	Worried/ Annoyed/ Confuse-d	Mild	✓

As observed from table 5, after testing, the proposed algorithm could accurately predict the results.

V. LIMITATIONS

What is defined as a natural language is a human spoken language, in opposition to computer languages such as C or FORTRAN or PASCAL. The main goal of NLP is to make human languages automatically recognizable and to find techniques to convert an utterance which can be either spoken or written, into formal data which is a representation of that utterance that can be processed using a computer and with no or minimal supervision.

Part-of-speech tagging consists in corresponding grammatical categories (e.g. noun, verb, adjective, adverb, etc.) to newly extracted tokens i.e. words. The system relies upon the results of the most basic categories for processing. However, it still results in a lot of ambiguities and can hardly be resolved without the context of each word.

For instance, with the sentence 'I saw her drawing board':

- *I* can either be a noun or a pronoun,
- *saw* can either be a noun or a verb,
- *her* can either be an adjective, a pronoun or a noun,
- *drawing* refers to forms of verbs,
- *board* can either be a noun or a verb.

Also, the syntax analysis, the study of the rules, is used to govern the correctness of a sentence in a language i.e. to decide if a sentence is grammatical; it must also allow to define clearly if it is a sentence or not. It implies anything that is recognized by the grammar is a correct sentence and anything that is not recognized is incorrect.

This idea, however, has not been reached yet because a natural language is so irregular that no system has successfully achieved that goal.

In the semantics portion of the language, on the other hand, there is an issue of lexical ambiguity.

For instance, the word 'bank' can be a noun but as a noun it has two distinct meanings: one is a financial institution; the other is an edge of a river.

The problem is even more complicated in translation because it combines the ambiguities of the two languages at once. In the present research work, a problem was faced regarding the part-of-speech tagging. For instance, the word *bank* can be either a verb or a noun. It matters because it can lead to very different translation. Some of the syntactic problems, like an improper formalization, irrespective of the proper part-of-speech tagging, may persist i.e. with the wrong category on a certain word, it is clearly not possible to produce the right translation. Also, even with the correct categorization, some syntactic structures are not well formalized yet.

Techniques for proper formalization of parts-of-speech, their syntax and semantics still remain as a void in the process of semantic analysis in lemmatization.

VI. CONCLUSION AND FUTURE SCOPE

Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings, tone of voice, and emotions regarding their brand. Up till recent times, the greater part of sentiment analysis projects undergone, have been conducted almost exclusively by corporations, brands and in medical treatments, through the use of social media data, survey responses and other hubs of user-generated content.

By investigating and analyzing client sentiments, these brands are able to get a perspective at consumer behaviors and thereby being able to cater to their needs in a much efficient manner. The future of sentiment analysis mainly depends on the activity of a human being in his/her virtual world in the form of the number of likes, comments, sarcasm detection techniques, usage of emojis and shares, and aim to reach, and truly understand, the significance of social media interactions. These things may have been incorporated in any future endeavor related to the present research work to obtain a more accurate result.

REFERENCES

1. Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas By, "Sentiment Analysis on Social Media", In proceedings of the International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 919 – 926.
2. Arun Meena, T.V. Prabhakar, "Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis", In proceedings of European Conference on Information Retrieval, 2007, pp. 573 – 580.
3. Shotaro Matsumoto, Hiroya Takamura, Manabu Okumura, "Sentiment Classification using Word Sub-Sequences and Dependency Sub-trees", Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, Volume 3518, 2005, pp. 301 – 311.
4. Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", In proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002, pp. 417 – 424.
5. Kushal Dave, Steve Lawrence, David M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", In proceedings of the 12th International Conference on World Wide Web, 2003, pp. 519 – 528.
6. Bo Pang, Lillian Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts", In proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, 2004, Article number 271.
7. Theresa Wilson, Janyce Wiebe, Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", In proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 347 – 354.
8. Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", In proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002, pp. 79 – 86.
9. Oskar Ahlgren, "Research on Sentiment Analysis: The First Decade", In proceedings of IEEE 16th International Conference on Data Mining Workshop, 2016.
10. Kiser, M., "Introduction to Natural Language Processing",2016, https://blog.algorithmia.com/introduction-naturallanguageprocessing-nlp, last accessed on 12.10.2019 at 12:35 pm.
11. "AI – Natural Language Processing", https:// www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm, last accessed on 12.10.2019 at 12:40 pm.
12. "What is Natural Language Processing?", http://www.expertsystem.com/natural-language-processing/, last accessed on 12.10.2019 at 12:45 pm.
13. "How does AI training work?", https://lionbridge.ai/articles/how-does-ai-training-work/, last accessed on 12.10.2019 at 12:45 pm.
14. "The Porter Stemming Algorithm", http://snowball.tartarus.org/algorithms/porter/stemmer.html, last accessed on 12.10.2019 at 12:50 pm.
15. Rudy Prabowo and Mike Thelwall, "Sentiment Analysis: A Combined Approach" Journal of Informatics, Volume 3, Issue 2, April, 2009, pp. 143 – 157.
16. Vinayak Sinha, "Sentiment analysis on java source code in large software repositories", Youngstown State University,2016.
17. Shahbaz Anwar, "The Future of Sentiment Analysis", 2017.
18. G.Vinodhini, R.M.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June, 2012, pp. 282 – 292.
19. "SentimentAnalysis", https://en.wikipedia.org/wiki/Sentiment_analysis, last accessed on 12.10.2019 at 12:30 pm.
20. Jimmy Ma, "Automata in Natural Language Processing", Technical Report no0834, December 2008.
21. Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining: Concepts and Techniques", 3rd Edition, 2011.

AUTHORS PROFILE

Mr. Debabrata Datta pursued Master of Technology from University of Calcutta, India and he is currently pursuing his Ph.D. in Technology from the same university. He is an Assistant Professor in the department of Computer Science, St. Xavier's College (Autonomous), Kolkata, India. He is a life member of IETE. He has published more than twenty five research papers in different reputed international journals and conferences mainly in the field of data analysis which is his primary area of research interest. He has more than eleven years of teaching experience both in the undergraduate and post graduate level of Computer Science. He has more than seven years of research experience.

Miss Srijita Majumdar did her B.Sc. with honours in Computer Science from Bethune College, Kolkata and subsequently pursued her M.Sc. in Computer Science from St. Xavier's College (Autonomous), Kolkata, India.

Miss Oile Sen pursued her M.Sc. in Computer Science from St. Xavier's College (Autonomous), Kolkata, India. She is currently working as a trainee software developer at Sprinriver Technology Private Limited, Kolkata.

Miss Aparna Sen pursued her M.Sc. in Computer Science from St. Xavier's College (Autonomous), Kolkata, India. She is currently working as a support associate at XCD HR Private Limited, Kolkata.