International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

STRESS PREDICTION USING MACHINE LEARNING

[1] Geetha M , [2] Sneha M S

[1] *Assistant Professor, Department of Master of Computer Applications, BIET, Davangere*
[2] Student,Department of Master of Computer Applications, BIET, Davangere

Abstract: Every year tens of millions of people suffer from depression and few of them get proper treatment on time. So, it is crucial to detect human stress and relaxation automatically via social media on a timely basis. It is very important to detect and manage stress before it goes into a severe problem. A huge number of informal messages are posted every day in social networking sites, blogs and discussion forums. This paper describes an approach to detect the stress using the information from social media networking sites, like tweeter.This paper presents a method to detect expressions of stress and relaxation on tweeter dataset i.e. working on sentiment analysis to find emotions or feelings about daily life. Sentiment analysis works the automatic extraction of sentiment related information from text. Here using TensiStrengthframework for sentiment strength detection on social networking sites to extract sentiment strength from the informal English text. TensiStrength is a system todetect the strength of stress and relaxation expressed in social media text messages. TensiStrength uses a lexical approach and a set of rules to detect direct and indirect expressions of stress or relaxation. This classifies both positive and negative emotions based on the strength scale from -5 to +5 indications of sentiments. Stressed sentences from the conversation are considered &categorised into stress and relax. TensiStrength is robust, it can be applied to a widevarietyofdifferent social web contexts. Theeffectiveness of TensiStrength depends on the nature of the tweets.In human being there is inborn capability to differentiate the multiple senses of an ambiguous word in a particular context, but machine executes only according to the instructions. The major drawback of machine translation is Word Sense Disambiguation. There is a fact that a single word can have multiple meanings or "senses." In the pre-processing part- of- speech disambiguation is analysed and the drawback of WSD overcomes in the proposed method by unigram, bigram and trigram to give better result on ambiguous words. Here, SVM with Ngram gives better resultPrecision is65% and Recall is 67%

.But, the main objective of this technique is to find the explicit and implicit amounts of stress and relaxation expressed in tweets. Keywords: Stress Detection, Data Mining, TensiStrength, wordsense disambiguation.

1.INTRODUCTION

Relaxation gives health whereas stress affects your health[1]. In the present day, situation Stress is most rapidly increasing. This is why people are not happy despite prosperity. Stress is a pressurized feeling. The pressure can be either emotional, physical or even mental. When one is under pressure can be either emotional, physical or even mental. When one is under pressure s/he feels s/he would fail not to succeed. Stress is discovered everywhere. It is discovered in each and every personality. Any thought, any situation can cause stress. But nature is different. Stress is multifold. It is physical, mental, financial. Any thought, any situation can cause stress[2]. When one denies in under stress, it increases in multiples gradually.

In such situations, personality totally changes. Superiority complex and. On the contrary, destructive Inferior complex also cause stress. A disturbed man can hardly do anything well. Stress can be constructive or many times destructive. Constructive stress makes you work, it keeps you active, busy and motivatedstress makes you dull, inactive, scared and you feel lonely. By so far, one can divide stress into twotypes, Short-term and the other Long-term or chronic stress. Short-term stress is situational as soon as the situation changes, the stress is gone or over. Long-term stress is a long term problem. That's why it is dangerous. Sometimes chronic stress is mostly hereditary or genes related also. It can make to endsanyone‟s life. Excessive stress leads one to death. Accepting stress is the first step to lead stress-free happy life. It is very important to detect and manage stress before it turns acute easy remedies are available at an affordable cost.

Research has been going on, since long to detect stress or stressed people. A lot of literature is available to study stress detection. There are traditional as well as scientific methods to detect people under stress. A) Questionnaire: Psychiatrist provides a big questionnaire and based on the answers, they decide whether one is under stress or not. This method has its own limitations and drawbacks because many times the answers are not factual. Sometimes some of the questions in the questionnaire are not appropriate. B) The other method is the sensor measuring method. The limitation of this method is, it is time-consuming and a bit expensive. The other and the recent method to stress detection is social media[3].

Stress detection is possible through social media, Based on one's write-up on social media, one's reacting on a particular problem on social media, one's liking and disliking help to detect stress. Through the continuous write up on social media, a psychiatrist can find people under pressure, crazy, mad after a typical subject. This also is an indication of stress. Use of social networking sites analyses the state of one's mind and thinking[4].

Twitter and Facebook have a wide number of users. People share their own thoughts, ideas, ideology, mood etc. Comments on Twitter or Facebook definitely manifest whether the writer is normal or abnormal. And abnormal is surely under stress. One's comment exhibits his own personality. The property of tweeter is short, informal and limited characters. Twitter is a beyond imagination 58 million tweets a day are posted. Tweets bring out one's own inner, personality core, emotions and anxiety. Tweets indicate what they are made of, and what their make is like. One's tweet can easily indicate whether one is relaxed orlike liberty or under pressure. This is early detection ofstress which in the long term will be helpful to avoid the severe problem.

So this indicates the prevention and it's a fact "Prevention is better than cure." And the minimum use of tweeter also indicates either under pressure or running away from the problem itself[5].

WSD is a technique to discover the exact sense or to findout the right applicable meaning of a word.In WSD the word sense is word meaning. Sometimes, a word doesn't give accurate meaning. It creates some kind of ambiguity. The meaning is obscure. A word may reflect two or more meanings at that time the ambiguity is there, and the opposite term Disambiguity means word gives meaning but it reflects many meanings of the same word. In this case, have to choose the right or appropriate meaning. WSD means automatically identifying the applicable correctmeaning and the meaning is based on the context of the text.Very often, tweets contain such words that have multiple meanings[6].Natural Language Processing(NLP) extracts the information from human language and WSD is required for machine translation.

In human being, there is an inborn capability to differentiate between the multiple senses of an ambiguous word in a particular context, but the machine only works as per the instruction given e.g. In English there are some common words which reflect different meaning as per sentences. For example, the word „bank" may have different senses as "Financial institution" or "River side". Such words with multiple meaning or senses are called ambiguous word and the process of finding the exact sense of an ambiguous word for a particular context is called word sense disambiguation. When working on sentiment analysis, WSD is applied and assigns the polarity to a particular word sense. Some standard NLP methods for WDS works on standard grammar and spelling which is rarely possible in social web. The WordNet is used for the WSD research. WSD improves the performance of lexicon based Stress/Relaxation detection algorithm TensiStrength.

TensiStrength is a lexicon based sentiment analysis algorithm. Main focus of this is to work on real time twitter content analysis. This is a newly automated system to detect strength of stress and relaxation expressed in social media text messages.This gives strength of a sentiment in the form of positive and negative ranging from 1 to 5 from short informal text. E.g. The scale is -1 (no stress) to -5 (extremely stressed) and parallelly for relaxation 1(not relaxing) to 5 (relaxed). Advantage of this method is, it works without considering standard grammar and spelling. This works on non-standard emotions or expressions from social web data

i.e. tweets. TensiStrength is an adaptation ofthe sentiment strength detection software SentiStrength[7]. TensiStrength works more accurate than sentiment analysis program SentiStrength. This algorithm labels each and everytweet or sentence as stressed or relaxed.

Sentiment analysis system either use a lexicon or machine learning algorithm. Compared with generic machine learning algorithm, lexicon based algorithm works well, but the decision about which to use depends on the nature of the text which to use depends on the nature of the text which have to analysed and the purpose of the task. The main drawback of the machine learning is domain specific i.e. it works well on the types of text that is trained by the system, whereas lexicon works on its own rules. But, in the sentiment analysis the coders are not having access of any context that may be topic of tweet or name of the tweeter account. Sometimes performance depends on the corpus used for the evaluation or collection of texts on a specific topic rather than twitter. In machine learning approach, it is hard to predict tweets in which stress will occur. The drawback of machine learning model, it works on labelled data, and it is extremely difficult to get sufficient and correctly labelled data, If the sentence from same domain at that time, the standard machine learning algorithm givesbetter result than TensiStrength for stress or relaxation detection.

In the proposed system, sentiment analysis is done by applying both lexicon based TensiStrength and machine learning algorithm to detect stress.

Stress can be reduced by some mediation or yoga programs and relaxation technique may include breathing exercise, listening to music and watching some relaxing videos or film or work on favourite subject or get involved into hobby. Sleep also reduces level of stress.

2.LITERATURE REVIEW:

2.1 Traditional Methods of Stress/Relaxation Measurement:

Stress has been traditionally measured by some indicative parameters such as heartrates, galvanic skin response, pupil diameter and another method are Questionnaires that help to find a person prone to stress, some life events also helps to detect stress.

But, these traditional methods require continuous observation or assessments or some expensive sensors and also have to believe that the person is sharing correct answers of his or her mindset. And also not lying to make a better image in front of psychiatrists. There is a relationship between personality and psychological stress of a person

.

2.2 Identification of Stress and Relaxation from Social Media Content Moving to Research on Social Media:

As now we are in the upscaling generation of technology, always try to share our ideas and thoughts on social media by posting a status or comments to the particular or current topic of the day to day activities which helps to judge a person is "Stressed" or "Relaxed".

Tweet level emotion detection reflects the instant emotions expressed by a particular tweet which also reflect mental health disorder such as depression or post-traumatic stress disorder (PTSD).

In paper the author Mike Thelwall implemented a WSD technology as a pre-processing stage and further works on lexicon based stress or relaxation method which improves the accuracy of TensiStrength. In this used the dataset of „1000' tweets with word "Fine" which reflects ambiguous meaning in different situation of sentences. This paper also eliminates redundant words such as preposition,conjunction, interjections and articles from the tweet and then work on remaining words

+5 based on dictionary indication.

The paper studies the co-relation between user psychological stress states and their social interactions defined by a unified hybrid model that integrates a factor graph model with convolutional neural networks. This is because CNN is capable of learning a unified latent feature from multiple modalities and FGM is good at modelling the co-relations. In this designed a CNN with cross autoencoder (CAE) to generate user-level content attributes from tweet level attributes and a partially labelled factor graph to combine user level social interactions to detect stress. Used the dataset of Twitter and SinaWeibo to compare the result for better accuracy also work on some comparison methods such as logistic regression, SVM, gradient boosted decision tree, deep neural network. So, like this presented aframework for detecting stress psychological stress states from user weekly social media data but the limitation is to find users who are stressed but less connected to socialmedia.

The author Mike Thelwall emphasise the TensiStrength: Lexicon based algorithm. TensiStrength uses a lexical approach and a set of rules to detect stress and relaxation expressed in social media text messages. It is slightly more effective than a comparable sentiment analysis program. When TensiStrength is compared with a generic machine learning approach, it works well. And the decision about which to use depends on the nature of the text analysed and the purpose of the task. In this paper twitter, dataset and dictionary are used which gives numeric strength rating from -5 to +5. The task of sentiment and stress or relaxation detection are related but are not equivalent. The result of TensiStrength in tweets are at a reasonable level of accuracy compared to human coders but are more accurate than sentiment analysis program and less accurate with compared to machine learning method optimised and trained on the same data. As stress and relaxation are very important aspects of everyone's life so have to design software that can accurately identify and help to design a smart future application.

The paper[8] focuses on social media to detect and diagnose depression in an individual, social media has characterising features of a person that may be positive or negative. In this twitter is used for measuring and predicting depression. Firstly they collected data from crowdsourcing to detect user and report they have been diagnosed as in a depression. Crowd workers can also have an option into their Twitter profile, with an agreement that their data could be mined and analysed using the computer program. Here questionnaire as the primary tool to determine the depression levels of the crowd workers. Here compare the behaviour of the depressed user class and standard user classand put some conclusions that individuals with depression show lower social activities, greater negative emotions, high self-attentional focus, etc. Worked on SVM classifier to predict depression of an individual user showing depression signs tend to be active during evening and night.

In paper sentiments strength detection method is used. Sentiment analysis works on the automatic extraction of sentiment related information from the text. In this proved that SentiStrength is not always better than machinelearning. Machine learning methods are normally domain specific but SentiStrength is robust, can be applied over a wide variety of different social web data. It gives an indication that female gives and receive stronger positive sentiments than male on social networks. The sentiment is accessed for polarity that whether it is positive or negative. It goes difficult to process when there is a large and combined data set. It also takes a large amount of RAM to load the data. The accuracy can also be analysed by 10-fold cross-validation in supervised data set so it gives better accuracy for negative sentiment strength on all type of dataset. So it is proved that SentiStrength is a robust algorithm for sentiment strength detection on social media data.

This paper emphasises on the newly invented algorithmof SentiStrength, which identify sentiment and sentiment strength from the informal English text. In social media, a complicated factor sentiment detection is spelling and grammar is ignored with its abbreviations, emotions and truncated sentences. So spelling correction is must rather than grammar check. Have to include some rules as spelling correction algorithm, a booster word list, a negative word list (i.e very happy and not very happy), repeated letters, repeated punctuations emotion list and negative emotions were ignored in questions. In this, the author has proved the standard version of SentiStrength is better than standard machine learning methods. Also mentioned in future research can apply the sentiment strength detection technique to automatically identify and classify positive and negative sentiments from informal web communication.

It is observed that in paper[9] the use of social media is for self-expression in the form of daily life events or updates, information sharing on favourite subject and comments of thoughts on a discussed topic and all these things are replacing face to face communication of real life. By using the conversation on twitter anyone can work on stress detection domain. In this survey paper, various comparisons are stated on current work. Various methods are available to detect stress but very few works on stress prediction. And as on the ongoing work, there is a greathope for sentiment analysis and prediction of stress.

The author in paper has made a survey on WSD. The main field of application of WSD is a machine translation. This WSD is applied on various Indian as well as on some international languages to measure its accuracy. WSD is categorized mainly into 3 types, Knowledge-based, Supervised and unsupervised method. By applying this method various algorithm gets some accurate result but the accuracy mainly depends on the language dictionary, corpus and wordnet etc. And as in Asian languages, the large scale of morphological inflictions are present, some development of wordnet, corpus and other resources have been going on.

The paper[10] focuses on a unified model for both word sense representation and disambiguation. In a word, representation has to assume each word owns a single vector because many words have different homonymy and polysemy. But in reality, many large knowledge bases have been constructed with word senses available for example WordNet and Wikipedia. By utilising these knowledgevases a word can be represented differently on the

basis of Word Sense and that sense representation is a natural choice. This unified model present Word Sense disambiguation based on vector representation and learn distributed vector representation for sense combined In this using a concept of WordNet: In reality large word senses available online, by using these knowledge base word representation and sense representation can be presented on users choice WordNet is most widely used computational lexicon of English based on the concept of synonym set or synset. In this, each synset has a textual definition or gloss.
Bank He sat on the bank of the river

Bank He cashed a cheque at the bank

Above it shows synset and the corresponding glosses of two common senses of the word bank. In this case, working is hard so using the most popular k nearest neighbor algorithm for classification method. The k nearest neighbor works on words and their senses showing that our senses vector can capture the semantic and syntactic information between words to present better sense vector, have to exploit the semantic relation that is hypernym and hyponym defined in wordnet.The majordrawback is senses of
word change over time based on the sentences.
In the author emphasizes on detecting stress in spoken English. This mainly works on vowel quality features. To measure the vowel quality must have to know, what vowel the speaker actually said and how they pronounced it. The vowel quality feature was calculated using HMM vowel models. Used the English sentences of 6 adult female of New Zealand speaker as a part of New Zealand spoken English databases. Hand labelled these vowels as stressed or unstressed and then tested. On the mentioned datasets both machine learning methods SVM and decision tree are applied. SVM gives a better result than a decision tree. Finally, each and every vowel segments are classified as stressed or unstressed.
In paper focused on the major issue in the process of machine translation is WSD, in which have to choose the correct meaning from multi-meaning word-based on the context in which it occurs this paperwork on cosine similarity of supervised learning method for WSD. In this working is based on two major parts, Feature extraction and convert each paragraph into feature values. Applying the cosine similarity matric for WSD which is based on the inner product of vectors. Also applying 10fold cross validations on a dataset by dividing training and test part for a context similarity-based classifier.
The author Andrew Trask [11] proposed a new model for WSD which uses supervised NLP labelling of data. This uses supervised labelling method instead of unsupervised clustering method which analyse the ambiguous work and assign or label the specific work as per its senses. Becauseof this accuracy, can be increased and also reduces computational complexity. In the sense2vec model different subjective evaluation techniques are applied like:

1) Subjective baseline:- To explain this there is a word "bank" which can be used in 3 different ways as a proper noun, verb and noun

2) Part of speech disambiguation:- For part of speech disambiguation, there is a word "apple" which is disambiguated on the basis of the noun "apple" referring to the fruit and the proper noun "apple" referring to the company.

3) Sentiment disambiguation:- In this by using part of speech tagging adjectives are labelled with positive or negative sentiment associated with each comment. i.e bad is a negative version of "good".

By analysing these different techniques, disambiguated embedding can increase the accuracy of a variety of syntactic dependency parsing of a language.

3. PROPOSED SYSTEM ARCHITECTURE

Fig3.1 explains the architecture of stress categorization.

Stage I : Upload Input Data

System works on textual data, User will upload the textual data in the form of conversations, Tweets, Sentences or blogs.Stress can be detected by the persons conversations. To give input , .txt file, a

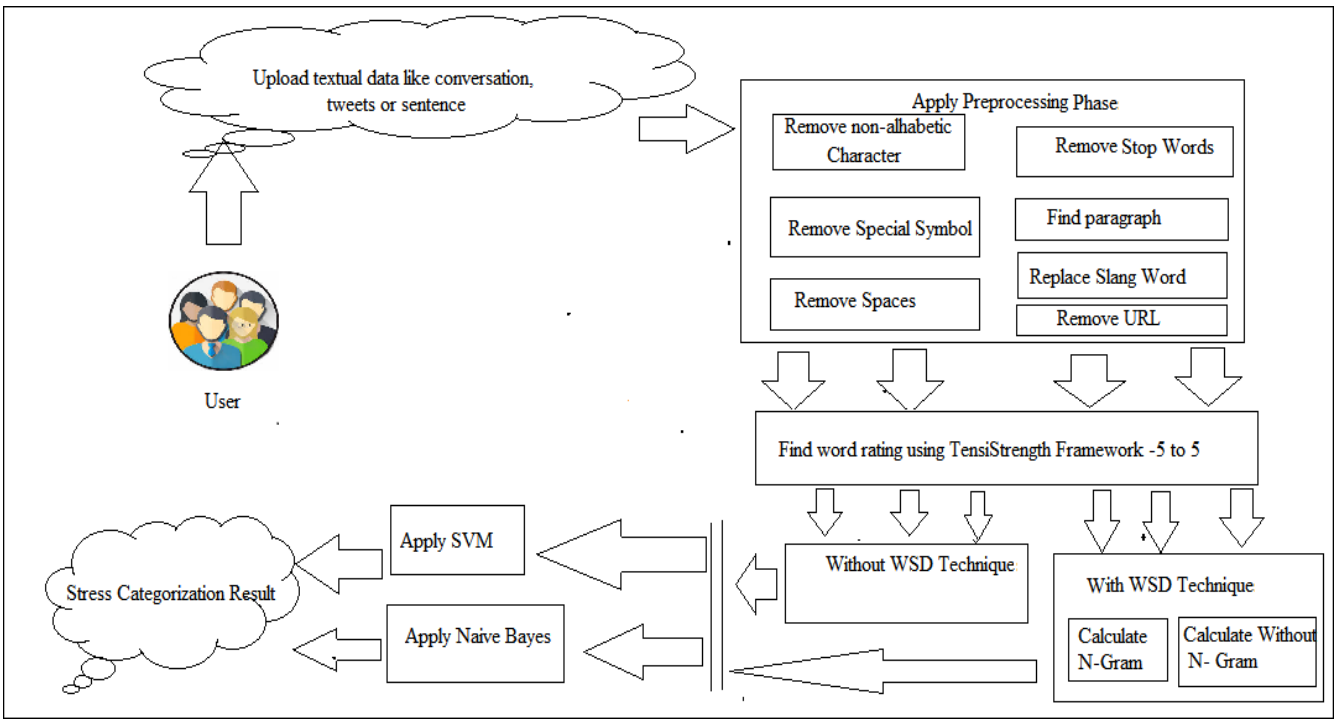


Figure 1: Architecture of System

single sentence on wall of input or a blog in the form of paragraph is uploaded to the system. Now, on this data pre-processing phase is applied.

Remove Non-alphabetic character:

In computers vision a person can read only „A-Z", „a-z", and some special symbol. i.e. In this phase, the symbols which doesn"t specify a alphabetic meaning is removed like „0-9" numbers are removed. This step removes all unreadable format character. And also these symbols will not help to identify the stress.

Remove Special Symbol : In twitter dataset, unwanted special symbols come frequently on which algorithmic rule can"t work. Though the special symbol give a lot information in short, but this works very hard to analyse by this implemented system. This phase removes the symbols like "!,@,#,$,%,&" are removed

Remove Spaces: This implemented system works on hand written data (or human coded data) so many times, by mistake some unwanted spaces or lines which is added by default in the text. So have to remove these spaces to work accurately on given data.

Remove Stop Words: For sentiment analysis, In English language there is a list of some common stop words, which will not change the meaning of sentences and also not indicate any stress related emotion, but it comes along with sentences. There is a stopword.txt to remove all common words in a text like "I, am, she" etc.

1) **Replace Slang Words :** In a sentence some slang words are available, which creates / gives a shortcut of words i.e. couldn"t is replaced by could not. These words gives negation / opposition of sentences. So, there is a necessity to replace all shortcut words into fullform.
2) **Remove URL :** In Pre-Processing, to get clear data for further analysis here removing the websites or links "http", "https".
3) **Find paragraph :** When the pre-processing phase starts its working, this can identify whether there is a tweet, sentence or may be a blog or paragraph. This paragraph can be identified by opening "<! And closing !> special symbol". Some ideas or thoughts are shared in blog contain more than one sentence i.e. collection of sentences so calling this as a paragraph. In such situations have to conclude the paragraph. In this system the paragraph is identified by a special symbol "<!.....!>". So by reading these start and end of the symbols the machine can get the paragraph is started. By applying these various operations in pre-processing , very much clear and limited data can be identified to give input for next phase of architecture.

STAGE II: Word Sense Disambiguation

The Pre-Processed data is given to this phase in the form of input to work.

This stage works in two different category on the basis of Disambiguation concept.

Without WSD Technique:

When the input sentence is specific and single meaning words in it at that time this technique is

applied. In this no any complicated work to do in processing, only skip-gram concept is applied.

Skip gram:

Skip gram works with the help of WordNet dictionary(2.1 version). Here, only synonyms of a word is considered to find the word with similar meaning in the dictionary. So that a word with similar meaning will get rating though it is not present in a dictionary

2) With WSD Technique:

When the Pre-processed data contains some ambiguous words or a word which changes its meaning as per the position in a sentence at that time the data will go through this phase of the architecture.

This phase works on POS tagging, n-gram and skip-gram model.

POS Tagging:

This step categories the sentence in various parts of the sentence in different categories which is necessary for n-gram. Only NOUN, VERB, ADVERB and ADJECTIVES are taken into consideration to process the data further , other part of a sentence has been removed by this stage. Here, will get clear and absolute words to compare with dictionary words to give rating.

N-Gram:

Once POS tagging is employed have to work on the position of the word i.e. Uni, Bi, Tri gram concepts are analysed. Here from POS tagging will fetch NOUN and with the help of noun will calculate the Uni-gram, Bi-gram and Tri-gram. If Uni-gram compares only oneword beside to NOUN, and same as for Bi-gram two words besides and same as for Tri-gram. If Uni-gram will not work then Bi-gram will work otherwise Tri- gram will work.

By implementing this ambiguous word with correct meaning in sentence to remove disambiguation because a same word also make changes with change in itsusage as a verb or may be as adjective also. Here will get a unique meaning of a word though it has more thanone meaning.

Skip-Gram:

This step is same as previous explained in without WSD technique.

This stage is very important to get a specific word for classifying the tweets/ sentences as stressed or relaxed.

STAGE I11: Find Word Rating

WSD technique is applied directly to TensiStrength framework. This step extracts the rating from TensiStrength Framework from -5 to +5.This includes a standard Dictionary AFINN with more than 2500 words with its rating from -5 to +5. -5 indicates extremely stressed / depressed and +5 is highly relaxed . This works on words selected from previous stage of the architecture.This step gives input for the next consecutive stage.

STAGE IV: Apply Algorithm

The extracted ratings are directly applied to this stage and have to categorise them into different classes like Happy, Depressed etc. This classification and prediction process is done in this stage. Here, both SVM and Naïve Bayes algorithms are used.

Naïve Bayes: Naïve Bayes is mostly used in sentiment analysis. This classifier works on Bayes theorem. This algorithm works on categorisation or prediction of the text by the frequency of words present in the document. This classifier works well in many real-word situations particularly on document classification. It requires small amount of training data to estimate the necessary parameters. It is extremely fast compared to others.

SVM: Support Vector Machine is used for both classification and regression analysis. It is a discriminative classifier to separate the hyperplane. It is a linear classifier, but additionally efficiently performa non-linear classification, by mapping the points implicitly into high dimensional feature spaces. Here, require a dataset to train the system and then test it, and the results are totally depends on the trained data. This gives very much accurate result as compared to other machine learning algorithms. This stage gives final categorised result of the system.

Stage V: Output

This stage is final and last step of the architecture. It concludes (gives final result of a) tweet, sentence or blogs in category like Happy, Stressed, Depressed and Non-identified.

3. PERFORMANCE ANALYSIS:

3.1 Experimental Setup:

System is tested on Standard Twitter Sentiment Analysis dataset from Kaggle to train and test the system. To Train the system more than 99999 records are used,which contains positive or negative

The categorization is as:

Table 1. Overall Results of category wise classification for all five Test Cases

Algorithm Applied	Depression	Stress	Normal	Relax	Other
SVM	7	47	20	19	1
NB	16	37	13	26	1
SVMWSD	7	39	22	17	10

done whether the user is depressed or not. Here also using the rating dictionary AFFINN contains more than 2500 words, a stop-word list is inserted to remove the unwanted data. WordNet 2.1 is used for WSD skip gram model. The algorithm may take a blog or sentences as an input and givesprobabilistic values as output. Depending upon the probabilistic values it determines whether the blog or sentence is in Depression,stress, normal, happy or relax.

This application of sentiment analysis can be used in many fields. Data generated on social media sites are different from classic data mining. On Social media data generated in huge amount, and very feasible to use. Social media sites gives more accurate analysis than any traditional method.

3.2 Experimental Results :

Machine Learning algorithms SVM and NB with and without WSD and Ngram are applied on test data. Tweets are classified as Stress, Depression, Normal, Relax, Happy and other i.e.non identified by the system. Results can be expressed with the help of different Testcases. Here five different test cases are analyzed on different data set and the average of the five cases can be summarized withthe help of

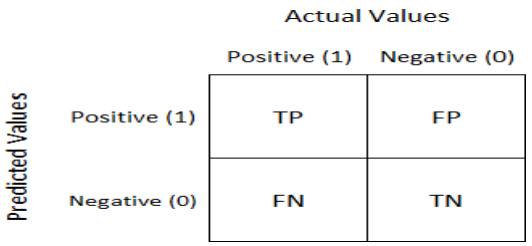
Table 4.1

3.3 Performance Analysis:

By applying Both SVM and NB analysing the test files also WSD and Ngram technique is applied to get more accurate results.

3.1.1 Measuring Performance:

Confusion Matrix



1. Precision and Recall:

Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved.

Precision can be seem as measure of completeness or quality. Recall is nothing but the true positive rate for the class.

In information retrieval, Performance Measures can be calculated with the help of Precision and Recall. These are important for evaluation matrices. Where, Precision is positive predictive value and recall is a sensitivity. Precision is a measure of result relevancy and recall is a measure of how many truly relevant results are returned. It measures relevant percentage of results whereas Recall gives percentage of total relevant results which are correctly classified by the algorithm.

Precision (P) is defined as the number of True Positives (TP) over the number of true positives plus

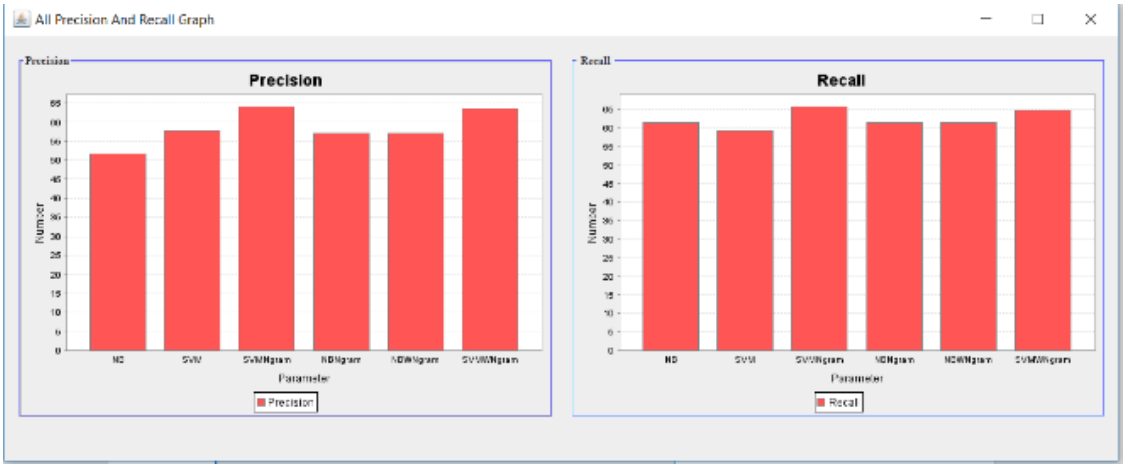


Figure 2: Overall classification for all five Test Cases

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

the number of False Positives (FP). The formula is as follow

: Recall (R) is defined as the number of True Positives (TP) over the number of true positives plus the number of false negatives (FN).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision and Recall for five test cases can be summarised as in Fig. 4.2 Like this the large dataset is divides intodifferent parts of 100 records, and applied SVM, NB, SVMWSD NBWSD, SVMNgram and NBNgram on the same data. So the results are discussed here in this Performance Analysis.

Conclusion:

Psychological stress is injures to health. In existing system stress is identified in face to face interview, communication or any other activities. where two or more people are analyzedbyanother. In this proposed a system framework for detecting users‟ psychological stress states by using users‟ weekly social media data, leveraging tweets‟ content as well as users‟ social interactions.

Here uses a words dictionary rating from -5 to +5 each words. To Classify and predict the data applied the SVM and NB algorithm. To improve result accuracy implemented the Word Sense Disambiguation by using n- gram and Skip-gram model. Support VectorMachine with WSD and Ngram gives 65% precision and 67% recall.

This implemented system will be helpful to detect Stress by using their daily conversations on social media data, without knowing to the user and categories the user as stressed or relaxed. In future work, smiley, like and dislike symbols can be considered for categorization of collected data, as it has major contribution to expresses feelings.

REFRENCES

1. Reshmi Gopalakrishna Pillai, Mike Thelwall, Constantin Orasan, "Detection of Stress and Relaxation Magnitudes for Tweets", International World Wide Web Conference Committee ACM, 2018.
2. Huijie Lin, Jia Jia, JiezhonQiu, " Detecting stress based on social interactions in social networks", IEEE Transactions on Knowledge and Data Engineering, 2017.
3. Mike Thelwall TensiStrength, "Stress and relaxation magnitude detection for social media texts", Information Processing & Management Science direct, 2017.
4. Jiawei Han, Michelin Kamber, Morgan Kaufman "Data Mining Concepts and Techniques",2nd Edition, Elsevier, 2003.
5. Munmun De Choudhury, Michael Gamon, Scott Counts, Eric Horvitz, "Predicting Depression via Social Media", Seventh International AAAI Conference on Weblogs and Social Media, 2013.
6. Thelwall Mike, Kevan Buckley, Paltoglou, "Sentiment strength detection for the social Web", Journal of the American Society for Information Science and Technology, 2013
7. Thelwall M., Buckley, Paltoglou, Cai, Kappas, "Sentiment strength detection in short informal text". Journal of the American Society for Information Science and Technology, 2010.
8. Dr. G V Garje, Apoorva Inamdar, Harsha Mahajan, "Stress Detection And Sentiment Prediction: A Survey", International Journal of Engineering Applied Sciences and Technology, 2016.
9. Alok Ranjan Pall, DigantaSaha, "Word Sense Disambiguation: A Survey", IJCTCM Vol.5 No.3, July 2015.
10. Xinxiong Chen, Zhiyuan Liu, Maosong Sun, "A Unified Model for Word Sense Representation and Disambiguation", Association for Computational Linguistics, 2014.
11. Huayang Xie, Peter Andreae, Mengjie Zhang, Paul Warren, "Detecting Stress in Spoken English using Decision Trees and Support Vector Machines", The New Economy Research Fund, New Zealand, 2004.
12. M. Nameh, S.M. Fakhrahmad, M. ZolghadriJahromi, " A New Approach to Word Sense Disambiguation Based on Context Similarity", Proceedings of the World Congress on Engineering, London 2011.
13. Andrew Trask, Phil Michalak, John Liu, "Sense2vec - A Fast And Accurate Method for Word Sense Disambiguation In Neural Word Embeddings", Under review as a conference paper at ICLR 2016.