IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

MENTAL STRESS DETECTION USING TF-IDF WITH MULTINOMIAL NAIVE BAYES

[1] Aathira S, [2] Reena B, [3] Dr. S. Uma
[1] PG Scholar, [2] Assistant professor, [3] professor
[1, 2, 3] Department Computer science engineering
[1, 2, 3] Hindustan college of engineering and technology,
[1, 2, 3] Coimbatore, TamilNadu, India.

ABSTRACT – Mental stress is a major issue nowadays, especially among youngsters. The age that was considered once most carefree is now under a large amount of stress. Stress increase nowadays leads to many problems like depression, suicide, heart attack, and stroke. In this paper, we are calculating the mental stress of students one week before the exam and during the usage of the internet. Our objective is to analyze stress in the college students at different points in his life. The effect that exam pressure or recruitments stress has on the student which often goes unnoticed. We will perform an analysis on how these factors affect the mind of a student and will also correlate this stress with the time spent on the internet. In this paper, hybrid algorithm is proposed which combines Tf-Idf with Multinomial Naive Bayes and Tf-Idf with SVM. This can be reduced to an extent if such intimidating messages or comments are segregated. The process of classifying a sentence whether it is positive, negative or neutral is known as sentiment analysis. It helps in determining emotional tone behind a sentence. To classify these intimidating messages this paper proposes a Tf-Idf with Multinomial Naive Bayes and Tf-Idf with SVM approach which classifies reviews into positive or negative.

Keywords: [Mental stress, SVM, Navie Bayes, Machine Learning.]

1. INTRODUCTION

Stress is the reaction of a human body marked by great anxiety or duress when faced with a challenging condition. The clinical meaning of stress can be a psycho-physiological condition of extreme inconvenience and distress for a person that can get extrapolated to intense emotional wellness problems like depression or anxiety attacks. A stressor is an event or condition present in or around an individual which may tend to trigger stress. The impact of stress on an individual can be positive and negative (also called as good and bad respectively) depending on the way stressful circumstances are dealt with. That's what this intends though a circumstance can be extremely stressful for one individual it might turn out to be only a gentle response for another. Also, a prior stressful experience provides a defensive mechanism in repeated conditions. For individuals who like to carry on with an existence brimming with difficulties, stress goes about as an adrenalin sponsor. Henceforth they think about stress as an agreed reaction.

Electronic communication networks pervade numerous parts of our regular routines, offering what presently approaches anytime, anyplace admittance to the web, supporting between private communication in a scope of structures, and permitting people to openly make and offer substance through multiple platforms without requiring coding skills. This the substance comprises of tremendous amounts of normal language information, giving an uncommon understanding into human social way of behaving, feelings, conclusions and ability on a global scale, creating a wealth of new opportunities for Natural Language Processing (NLP) research, with immediate implications for business and commerce.

2. LITERATURE SURVEY

1. C. Vuppalapati, M. S. khan, N. Raghu, P. Veluru and S. Khursheed (2018) et.al proposed A System to Detect Mental Stress Using Machine Learning and Mobile Development. Stress can be characterized as an actual reaction to the extreme measure of pressure looked by a person. The stress could be initiated because of any mental or social situation. At the point when unreasonable pressure is incited on an individual, this could prompt multiple mental problems. These could incorporate being discouraged or on the other hand whenever deteriorated, encountering heart failures. The NB Classifier execution depends on the age of contingent probabilities for tests thinking about stress and control. The classifier comes into the image by giving example sets to classes for higher probabilities. This research paper addresses the test by developing and conveying AI empowered information driven and Electroencephalogram biosensor incorporated portable application that proactively gathers User's stressful episodes, mixes cooperative knowledge got from de-distinguished at this point User pertinent demographical, physiological, way of life and conduct datasets and preventive healthcare experiences to counter in any case the drawn out adverse consequences of the stress on the Users health. The paper presents prototyping arrangements as well as their application and certain experimental outcomes.

2. L. Anishchenko and A. Turetzkaya (2020) et.al proposed Improved Non-Contact Mental Stress Detection via Bioradar. Chronic mental stress might prompt different physiological and mental disorders, like hypertension, depression, diabetes, heftiness, and cardiovascular infections. As of now, there are no solid accurate and non-emotional techniques material for schedule regular stress appraisal. In the current work, we proposed a better inconspicuous AI based technique for far off mental stress identification by radar. It was approved utilizing experimental information of 34 basically healthy workers of various fitness states and stress resistance. The work could contribute to the advancement non-contact framework for delayed observing of individual reactions to mental stress in day to day existence suitable for individuals with various degrees of fitness and stress resistance.

3. P. Bobade and M. Vani (2020) et.al proposed Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. The BioNomadix model BN-PPGED from Biopac was a wearable gadget that was utilized to gauge physiological reactions Stress is a typical piece of daily existence that a great many people need to manage on different events. Notwithstanding, having long haul stress, or a serious level of stress, will prevent our security and upset our normal lives. Identifying mental stress prior can forestall numerous health problems related with stress. This paper proposes different AI and profound learning procedures for stress location on people utilizing multimodal datasets recorded from wearable physiological and movement sensors, which can keep an individual from different stress-related health problems. Information of sensor modalities like three-hub speed increase (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), breath (RESP), electromyogram (EMG) and electrodermal activity (EDA) are for three physiological circumstances - entertainment, nonpartisan and stress states, are taken from WESAD dataset. The modalities like facial cuts, logging information, sound/video accounts, FITBIT information, and so on that are utilized in different investigations independently can be converged with physiological information, and a new dataset can be presented.

3. PROPOSED METHODOLOGY

In this paper, we used machine learning (ML) to identify the increasing stress level in the students and to predict the stress beforehand and be able to stop the major damage to their life before happening. In the test, we evaluate students amongst different situations. The level of stress was approved by the undertaking execution. The proposed model includes dataset collection, pre-processing, and applying machine learning algorithm (Tf-Idf with Multinomial Naive Bayes and Tf-Idf with SVM) and comparing them on these performance parameters.

In this paper, hybrid algorithm is proposed which combines Tf-Idf with Multinomial Naive Bayes and Tf-Idf with SVM and calculated accuracy of all these. We found that Tf-Idf with support vector machine is performing well out of all algorithms giving a high accuracy. Thus we can say that SVM is performing well out of all algorithms in this scenario.

Preprocessing

Data preprocessing is a data mining procedure that includes changing crude data into a justifiable arrangement. Genuine data is frequently fragmented, conflicting, or potentially ailing in specific ways of behaving or drifts, and is probably going to contain numerous blunders. After a text is acquired, then start with text normalization. Text normalization includes:

- switching all letters over completely to lower or capitalized
- changing over numbers into words or removing numbers
- removing punctuations, accent marks and different diacritics
- removing White areas

Stopwords Removal

"Stop words" are the most common words in a language like "the", "a", "on", "is", "all". These words do not carry important meaning and are usually removed from texts. It is possible to remove stop words using Natural Language Toolkit (NLTK), a suite of libraries and programs for representative and statistical regular language handling. Stop words are a bunch of regularly involved words in any language. In NLP and text mining applications, stop words are utilized to dispense with unimportant words, permitting applications to zero in on the important words all things considered.

Tokenization

Tokenization is the process of splitting the given text into more modest pieces called tokens. Words, numbers, punctuation marks, and others can be viewed as tokens. Tokenization is the demonstration of breaking up a succession of strings into pieces, for example, words, catchphrases, expressions, images and different components called tokens. Tokens can be individual words, expressions or even entire sentences. In the process of tokenization, a few characters like punctuation marks are disposed of. Tokens themselves can likewise be separators. For instance, in most programming languages, identifiers can be put along with number-crunching administrators without blank areas. In spite of the fact that it appears to be that this would show up as a solitary word or token, the syntax of the language really thinks about the mathematical administrator (a token) as a separator, so in any event, when multiple tokens are grouped up together, they can in any case be isolated by means of the mathematical administrator.

Data Splitting

The data we use is usually split into training data and test data. The training set contains a known output and the model learns on this data in order to be generalized to other data later on. We have the test dataset (or subset) in order to test our model's prediction on this subset.

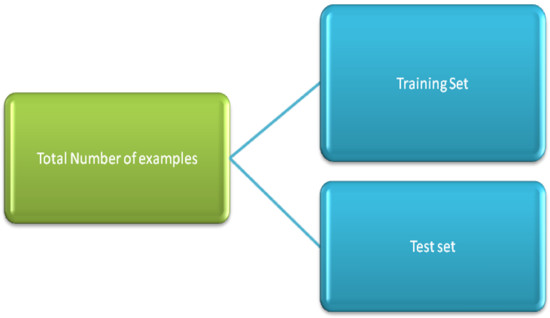

Figure 1. Data splitting Example

TF-IDF Vectorizer

Term Frequency-Inverse Document Frequency implies the measure of importance of a word in a document. A quick run of this will sort all entries in dataset and help in eliminating great deal of inconvenience. Simplest way of representing text in numeric for is count of occurrence of word in entire document. TF-IDF is one of the familiar algorithms used to mutate text into meaningful delineation of numbers. Representation of text in numeric form helps in depicting significant traits and extracting features of text.

Inverse Document Frequency

Term Frequency of certain words (‘a’, ‘the’, ‘of’) that are regular in documents might put down weights of meaningful words. To overcome this problem, term frequency is discounted by factor called inverse document frequency (IDF). It is measure of how isolated a term is. If the IDF score is more, then the term is considered to be more isolated.

$$\mathrm{IDF(x)} = \log_e \frac{\text{total number of document}}{\text{number of document with x in it}}$$

The product of Term Frequency with Inverse Document Frequency is known as TFIDF.

$$(\mathrm{TF\_IDF})\mathrm{score} = \mathrm{TF} * \mathrm{IDF}$$

Stress Classification

Stress is a physiological reaction to mental, close to home, or other actual difficulties that people defy in their genuine exercises, remembering for their workplaces. The objective of this study is to show stress levels from various social variables got from input data and specifically with the restriction that the marked data for an individual is scant. A definitive point is to diminish dependence on self-revealed, abstract data for stress estimation and utilize objectively detected data to permit ceaseless estimation of stress levels. Anticipating the apparent stress of an individual can be demonstrated as a classification issue. We utilized a classifier to demonstrate the subject's stress since this representation can be handily perceived by a human, and this could assist with having a superior comprehension of what causes stress. Likewise, utilizing this representation we can think about various subjects, which is important for move learning. Our most memorable objective is to investigate the way in which subjects are connected with one another as far as how comparative their models are.

Algorithm: Learning the Classifiers
Input: Dataset D, Learning rate, Network
Output: Trained the classifiers
Step 1: Input the dataset
Step 2: The random values -1 to +1 i.e., weight must be assigned to input. Each Input is sent to neural network.
Step 3: Sum of all weight input
Step 4: generate the Output.

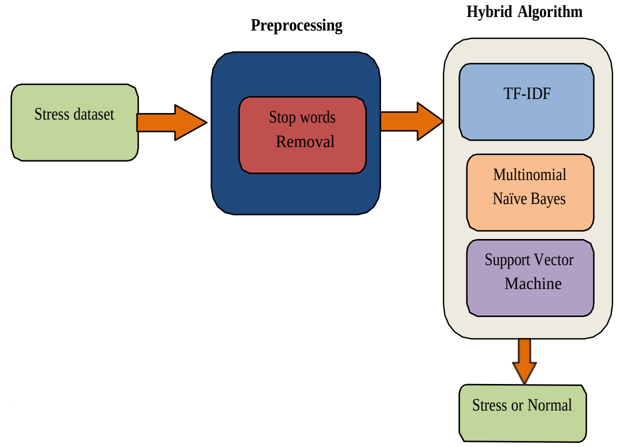

Figure 2. Activity Diagram

Naïve Bayesian Theorem

Naïve Bayesian theorem classifier assumes that all the predictors in the provided dataset are independent of each other, that is the reason why is called “naïve”. The formula for the conditional probability is:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Where:

- $P(A)$: Is the probability that the hypothesis “A” is true (regardless of the data). This is known as the prior probability of “A”. •
- $P(B)$: Data probability (regardless of the data). This is known as the prior probability.
- $P(A|B)$: Is the probability that the hypothesis “A” given the data “B”. This is known as the prior probability.
- $P(A|B)$: is the probability of the data "A" given that hypothesis B was true. This is known as the posterior probability.

In other words, the prior P(A) and the evidence P(B) mean to the probabilities of A and B being independent of each other, while the posterior and likelihood are conditional probabilities of A given B and the other way around.

Naïve Bayes Classifier

In machine learning, naïve bayes classifier is associated to family of probabilistic classifiers that applies bayes theorem with strong independent hypothesis between features. This classifier requires slight number of training set to deem features required for classification. Naïve bayes classifier can be used for both multiclass and binary classification. Performance of naïve bayes classifier is well in case of categorical input variables when compared to that of numerical variables. Naïve bayes classifier predicts probabilities of every class.

SVM Classifier

SVMs are supervised learning models which are related with learning algorithms which analyse data that is required for classification analysis. SVM is a classifier which is formally defined by separating hyperplane. Examples in this model are 7 represented in form of points in space which helps in separating different categories by a gap. SVM not only performs linear classification but also nonlinear classification.

Stress Level

Primarily, the purpose of the user is to calculate the stress index in the system. Physical data and cognitive data were used to calculate the stress index value, after significant data was collected. The threshold value is derived from the user's stress index value. Depending on the user's behavior, the user's EEG data is varied. Further, to calculate the individual stress level, it is necessary to calculate the threshold value of the individual.

4. EXPERIMENT RESULT

Precision

$$Precision = TruePositives \,/\, (TruePositives + FalsePositives)$$

Dataset	NB CLASSIFIER	BN-PPGED	Proposed TF-IDF NB
50	66.94	74.91	87.01
100	69.66	71.77	90.87
150	74.12	67.93	92.48
200	79.09	68.05	95.23
250	86.38	65.39	97.52

Table 1. Comparison tale of Precision

The Comparison table 1 of Precision Values explains the different values of existing NB CLASSIFIER, BN-PPGED and proposed TF-IDF NB algorithm. While comparing the Existing algorithm and proposed TF-IDF NB algorithm, provides the better results. The existing algorithm values start from 66.94 to 86.38, 65.39 to 74.91 and proposed TF-IDF NB values starts from 87.01 to 97.52. The proposed method provides the great results.

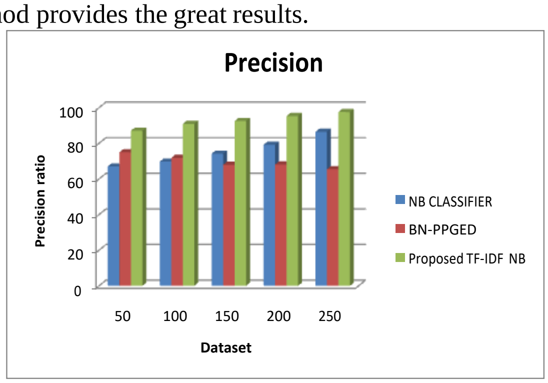

Figure 3 Comparison chart of Precision

The Figure 3 Shows the comparison chart of Precision demonstrates the existing NB CLASSIFIER, BN-PPGED and proposed TF-IDF NB. X axis denote the Dataset and y axis denotes the Precision ratio. The proposed TF-IDF NB values are better than the existing algorithm. The existing algorithm values start from 66.94 to 86.38, 65.39 to 74.91 and proposed TF-IDF NB values starts from 87.01 to 97.52. The proposed method provides the great results.

Recall

$$Recall = TruePositives \,/\, (TruePositives + FalseNegatives)$$

Dataset	NB CLASSIFIER	BN-PPGED	Proposed TF-IDF NB
20	0.625	0.721	0.836
40	0.663	0.654	0.874
60	0.706	0.598	0.905
80	0.728	0.623	0.941
100	0.752	0.591	0.962

Table 2. Comparison tale of Recall

The Comparison table 2 of Recall Values explains the different values of existing NB CLASSIFIER, BN-PPGED and proposed TF-IDF NB algorithm. While comparing the Existing algorithm and proposed TF-IDF NB algorithm, provides the better results. The existing algorithm values start from 0.625 to 0.752, 0.591 to 0.721 and proposed TF-IDF NBvalues starts from 0.836 to 0.962. The proposed method provides the great results.

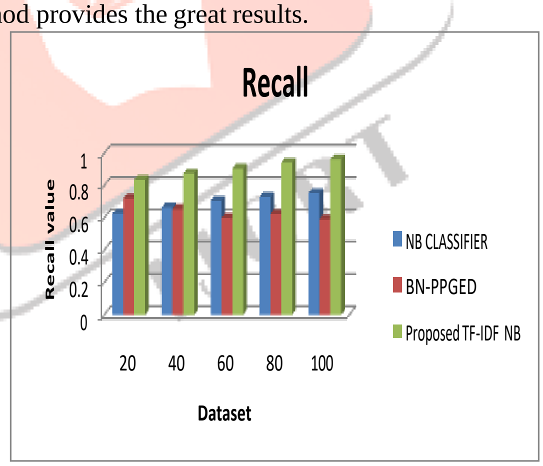

Figure 4 Comparison chart of Recall

The Figure 4 Shows the comparison chart of Recall demonstrates the existing NB CLASSIFIER , BN-PPGED and proposed TF-IDF NB algorithm. X axis denote the Dataset and y axis denotes the Recall ratio. The proposed TF-IDF NBvalues are better than the existing algorithm. The existing algorithm values start from 0.625 to 0.752, 0.591 to 0.721 and proposed TF-IDF NBvalues starts from 0.836 to 0.962. The proposed method provides the great results.

F -Measure

$$F - Measure = (2 * Precision * Recall) \,/\, (Precision + Recall)$$

Dataset	NB CLASS IFIER	BN-PPGED	Proposed TF-IDF NB
100	0.89	0.72	0.98
200	0.85	0.70	0.96
300	0.86	0.67	0.95
400	0.84	0.64	0.93
500	0.82	0.61	0.92

Table 3. Comparison tale of F –Measure

The Comparison table 3 of F -Measure Values explains the different values of existing NB CLASSIFIER, BN-PPGED and proposed TF-IDF NB algorithm. While comparing the Existing algorithm and proposed TF-IDF NB, provides the better results. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed TF-IDF NB values starts from 0.92to 0.98. The proposed method provides the great results.

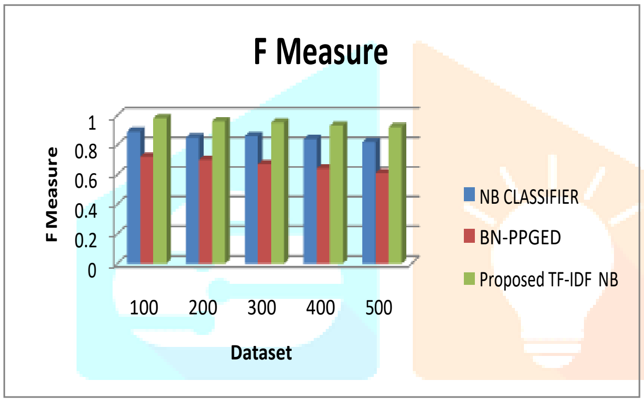

Figure 5 Comparison chart of F –Measure

The Figure 5 Shows the comparison chart of F -Measure demonstrates the existing NB CLASSIFIER, BN-PPGED and proposed TF-IDF NB. X axis denote the Dataset and y axis denotes the F -Measure ratio. The proposed TF-IDF NB values are better than the existing algorithm. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed TF-IDF NB values starts from 0.92to 0.98. The proposed method provides the great results.

CONCLUSION

In this paper, we explored the problem of finding the possible variations and analyzing user's sentiment tendency in dynamic social networks. Micro blogging based Treatment Outcome System, which provides treatment suggestions for depressed persons through social media. Naïve Bayes Classifier with TF-IDF and SVM with TF-IDF are used in this proposed method to classify the depressed from the social conversations. Though digital medium connects people from various locations, the same is taken as advantage by others leading to cyber bullying. Disturbance created and executed by means of digital devices is cyber bullying. It affects mental stability of a person increasing stress, pressure leading to extreme activities. So by detecting the comments which are cyber bullying helps an individual overcome the mental stress.

REFERENCES

[1]. C. Vuppalapati, M. S. khan, N. Raghu, P. Veluru and S. Khursheed, "A System To Detect Mental Stress Using Machine Learning And Mobile Development," 2018 International Conference on Machine Learning and Cybernetics (ICMLC), 2018, pp. 161-166, doi: 10.1109/ICMLC.2018.8527004.

[2]. L. Anishchenko and A. Turetzkaya, "Improved Non-Contact Mental Stress Detection via Bioradar," 2020 International Conference on Biomedical Innovations and Applications (BIA), 2020, pp. 21-24, doi: 10.1109/BIA50171.2020.9244492.

[3]. P. Bobade and M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 51-57, doi: 10.1109/ICIRCA48905.2020.9183244.

[4]. A. Deschenes, et al. "Facial Expression Recognition Impairment fol- ˆ lowing Acute Social Stress," Journal of vision, 2015 , 15(12) , 1383.

[5]. Peake, J. M., Kerr, G., Sullivan, J. P., "A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations," Frontiers in physiology, 2018, 9, 743.

[6]. Etiwy, M., Akhrass, Z., Gillinov, L., Alashi, A., Wang, R., Blackburn, G., Javadikasgari, H., "Accuracy of wearable heart rate monitors in cardiac rehabilitation," Cardiovascular Diagnosis and Therapy, 2019, 9(3), 262.

[7]. L. Anishchenko, et.al, "Application of step-frequency radars in medicine," Proceedings of SPIE 9077 Radar Sensor Technology XVIII, pp. 90771N, 2014.

[8]. J.R.M. Fernandez, L. Anishchenko, "Mental stress detection using bioradar respiratory signals," Biomedical Signal Processing and Control, Volume 43, May 2018, pp. 244–249.

[9]. Gjoreski, M., Gjoreski, H., Lustrek, M., Gams, M. "Continuous stress ˇ detection using a wrist device: in laboratory and real life," In proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct (pp. 1185–1193), 2016.

[10]. A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel and A. S. Malik, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels," in IEEE Access, vol. 5, pp. 13545- 13556, 2017.