



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: <https://doi.org/10.22214/ijraset.2022.41510>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Psychological Analysis Using Social Media Data

Shriya Gawade¹, Riya Sawant², Aakash Rathod³, Prof. Chhaya Dhavale⁴

^{1, 2, 3, 4}Department of Information Technology, Xavier Institute of Engineering, Mumbai University, Maharashtra

Abstract: *Mental Stress is an important aspect of our life that is given the least importance. We tend to ignore the fact that we need to be emotionally stable along with physical stability. To keep your mental state sound, we proposed this system where the psychological state of a person is being predicted. One such place where a person comes up and shares his/her thoughts, through texts is on social media with their friends. To detect such a state, we made use of NLP techniques accompanied by a reliable scale, the Perceived Stress Scale (PSS) developed by Cohen, Kamarck and Mermelstein. The huge texts were cleaned using text processing methods. In Machine Learning, there are many ways for sentimental analysis such: decision-based systems, Bayesian classifiers, support vector machines, neural networks and sample-based methods. We have performed sentimental analysis and in order to give the severity of the condition we made use of the Perceived Stress Scale (PSS). The model will be predicting whether the given text indicates stress or not and further classifies it as low, medium or high-level stress.*

Keywords: *TF-IDF, Natural Language Processing (NLP), Stress, WordCloud, Perceived Stress Scale*

I. INTRODUCTION

People use social media to express what they are feeling, why they are feeling it. They tend to express their feelings through images, comments, or sometimes even to their friends, family, and well-wishers over social media platforms [19] such as Instagram, Facebook, WhatsApp, Reddit, Quora, etc. Often, the opposite party tends to misinterpret their feelings. They tend to take it lightly or many times neglect it. This may result in the person feeling even more lonely and depressed, which further adds to disturbing the psychological health. So, it is necessary to point out why they are posting such things and then after understanding social media behavior. Generally, the reasons why people post and consume updates can vary [15]. It might be because sometimes they post for the sake of the well-being of their close ones, sometimes because of acceptance from society, sometimes to share their success or achievements so that people could look up to someone for inspiration and other times for self-centered content.

Generally, the reasons why people post and consume updates are as follows:

- 1) *Physiological Needs:* People sometimes post to benefit the health or well-being of their friends and family.
- 2) *Safety:* Physical, mental, and financial security are important for people when they choose to post some material on their social media.
- 3) *Love/belonging:* Users generally want to post to feel some kind of social acceptance from a group or a particular individual.
- 4) *Esteem:* People want to quell the rewards-oriented parts of their brains, which helps explain why people post “me-centric” content regularly.
- 5) *Self-Actualization:* As the most important facet of the hierarchy of human needs, this aspect of social media posting manifests when people share their successes – getting a new job, completing an arduous project, or graduating from school, to name a few examples

II. BACKGROUND

The significant rise in complexity of technology accompanied by usage of social media has led to many psychological health issues like anxiety, stress. It has become of utmost importance to analyze them and find prevention for the same. Traditionally, psychologists used the questionnaire and interview method but these methods seem to be tedious and hysteric. Some papers have used the Effective Stress Detection Framework (ESDF) to utilize the hybrid ontology for detecting stress. While some researchers analyzed tens of thousands of posts on social media websites by sorting, creating a distribution of individuals arranged according to the similarities or differences in their vocabularies. We haven't yet come across a scenario where pronouns and other very common words were not used to identify the array of users. Surprisingly, users had strong variations in pronoun use. The new analysis also associates these personality markers with topics of conversation as the way the people refer to themselves and others have previously been correlated with personality. Now, concerning stress, the linguistic features of event-related stress have been predicted from social media posts about experiences such as travel and work; however, these findings cannot be applied to improve the psychological understanding of stress, because people suffering from chronic stress do so irrespective of stressful events.

For instance, trait-related stress like preparation for an exam can be a stressful event, because you might feel overburdened with responsibilities. Another research gap is that the previous work has focused on known stressors collected using search keywords. However, the labels thus acquired likely have personality confounds, emphasizing the need for using stronger ground truth. Instead, we anticipate that insights into psychological stress could help in the following ways. Designing social-media-based interventions to enable a low-stress lifestyle. Developing a better understanding of regional variations in stress. Smart devices are ubiquitous and we can use the internet from any part of the world which means that social media is a very powerful tool that can be used for measuring the psychological states and behaviors of people.

III. LITERATURE REVIEW

1) *Dreaddit: A Reddit Dataset for Stress* -Elsbeth Turcan , Kathleen McKeown

Year of Publication: 2019

This paper aims to use Reddit data to examine a variety of mental health conditions such as depression and other clinical diagnoses such as general anxiety, but to our knowledge. and train the corpus to focus on stress as a general experience, not only a clinical concept

2) *Stress Detection Methodology based on Social Media Network*-S. M. Chaware, Chaitanya Makashir, Chinmayi Athavale, Manali Athavale, Tejas Baraskar

Year of Publication: 2020

In this paper, the authors have provided a user interface wherein the user has to register in the system only once. Then he /she has to perform Token Generation and Token Extraction so that they can get access to users' social media interactions

3) *Depression detection from social network data using machine learning techniques*-Md. Rafqul Islam1*, Muhammad Ashad Kabir2, Ashir Ahmed3, Abu Raihan M. Kamal1, Hua Wang4 and Anwaar Ulha

Year of Publication: 2018

In this paper authors have mentioned four types of factors such as emotional process, temporal process, linguistic style and all (emotional, temporal, linguistic style) features together for the detection and processing of depressive data received as Facebook posts.

4) *Detection and Analysis of Stress using Machine Learning Techniques*-Reshma Radheshamjee Baheti, Supriya Kinariwala

Year of Publication: 2019

In this paper, the researcher has used WSD technology as a pre-processing stage and further works on lexicon-based stress or relaxation method which improves the accuracy of TensiStrength

5) *Mental Stress Detection in University Students using Machine Learning Algorithms* -Ravinder Ahujaa, Alisha Banga

Year of Publication: 2019

The research paper aims to identify the increasing stress level in the students and to predict the stress and be able to stop the major damage to their life before it happens.

IV. PROPOSED SYSTEM

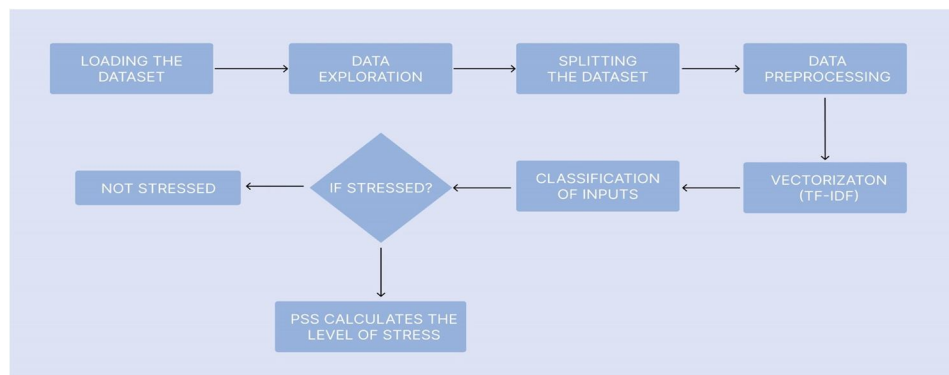


Fig. 1 Flow chart

- 1) *Importing Dataset:* We have taken the messages from twitter. Our Dataset contains post numbers, text messages and labels. The dataset contains conversations, sentences that people have written to someone, conveying their feelings of what they are going through under certain situations. There are in total 10315 messages taken from twitter for analysis purposes for predicting stress.
- 2) *Data Exploration:* Data exploration is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. So for that purpose we have used Word Cloud. A Word cloud, or tag cloud, is a textual data visualization which allows anyone to see in a single glance the words which have the highest frequency within a given body of text. Word clouds are typically used as a tool for processing, analyzing and disseminating qualitative sentimental data. Input any text into word cloud generator and you'll see a visual representation of the most frequently used words, according to their relative size. The larger words in the word cloud are the most frequently repeated. WordCloud enables anyone to quickly visualize any text.



Fig. 2 Representation of depressive words using wordcloud



Fig. 3 Representation of positive words using wordcloud

- 3) *Data Preprocessing*: Text Preprocessing is the first step in the pipeline of Natural Language Processing (NLP), with potential impact in its final process. Text Preprocessing is the process of bringing the text into a form that is predictable and analyzable for a specific task. A task is the combination of approach and domain. Psychology of the person from the text. So the irrelevant words, as well as the punctuation, should be omitted from the text. The text pre-processing will guarantee you a consistent result from their NLP applications only to realize that they were not preprocessing their text or were using the wrong kind of text preprocessing for their project.

The steps performed in the Data preprocessing phase:

- a) *Remove Stop Words*: Stop words are a set of widely used words in a language. The common stop words are “a”, “the”, “is”, “are” etc. The reason behind using stop words is that, by removing low information words from the text, we can focus on the important words instead. If Stop words are removed, this can prevent all words from your stop word list from being analyzed. Stop words are commonly applied in search systems, text classification applications, topic modeling, topic extraction, and others.
 - b) *Remove Punctuation*: Removing the punctuation is important, otherwise, the same words will be treated differently because of punctuation marks. For example, if we don't remove the punctuation, then yeah. yeah, yeah! will be treated separately. If punctuation is removed it will treat yeah and yeah! Equally.
 - c) *Stemming*: Stemming is useful for dealing with sparsity issues as well as standardizing vocabulary. This process reduces the word to its root stem for example run, running, runs, run derived from the same word as run. Stemming is removing the prefix or suffix from words like ing, s, es, etc. For example, the word ‘loving’ and ‘loves’ is reduced and map to ‘love’. NLTK library ‘Wordnet’ is used for stemming the words. The stemming technique is not used for production purposes because it is not an efficient technique so to overcome this, lemmatization is used.
 - d) *Lemmatization*: Lemmatization & Stemming are almost similar. The objective is to remove inflections and map a word to its root form. What makes lemmatization different from stemming is that lemmatization tries to remove inflections but it doesn't match the actual word to the root word. For example, the word “better” would map to “good”. It may use a dictionary such as WordNet for mappings or some special rule-based approaches.
- 4) *Splitting the Dataset*: After preprocessing our complete dataset, we divided it into two sets: a training set and a test set. Data is used for training in 80% of cases and testing in 20% of cases. The photos are then divided into test and training sets.
 - 5) *TF-IDF Vectorization*: Vectorization is the concept of an optimization algorithm that tries to reduce the error and computes to get the best parameters for the machine learning models most commonly used traditional text vectorizers is- Bag of words (BOWs) and Term frequency inverted document frequency (TF-IDF) [9]. Text vectorization techniques namely Bag of Words and tf-idf vectorization. It helps in converting text to numeric feature vectors. The Machine learning model uses words with only numeric input, but the text has words, alphabets, and other special characters. So to apply machine learning to the text you will use the method TF-IDF [9] to convert the text as the numeric table representation. In TF-IDF table rows comprise each document in a corpus, and the columns represent the words. Each value in a cell has the count-value that determines the relevance of the words in that particular document. If in case if a particular word is present in every document then the count-value of such word is low. Whereas, if the word is unique for the document then the strength of the word for a particular document will be higher than any other word.

V. METHODOLOGY

In order to make the machine learn, we need to train the model to be able to classify between stressful and non-stressful text. For making classification possible, we have created a dictionary of stressful and no stressful bigrams using TF-IDF (Term Frequency - Inverse Document Frequency). So for each sentence that is taken as input it undergoes pre-processing and then the words in the sentence are checked with the dictionaries created.

In case, a new word turns up then it undergoes TF-IDF again. Further, the final classification is done on the basis of the count of stressful and non stressful words. If the count of stressful words is more than non stressful words then it is classified as Stressful text and vice-versa. Additionally, if the result comes out as Stressful text then it also gives the severity of the stress, this is done by using Perceived Stress Scale (PSS) [22]. The accuracy for this model was around 86%. Furthermore, The model is deployed locally using flask in the form of a webpage.

VI. RESULTS

#Get Started

Let us see what you're feeling

Enter Your Message Here

Extreme sadness, lack of energy, hopelessness

Compute Results

Results for Comment

Up for a small Questionnaire?

- Q1 - In the last month, how often have you been upset because of something that happened unexpectedly?
 Never Almost Never Sometimes Fairly often Very often
- Q2 - In the last month, how often have you felt that you were unable to control the important things in your life?
 Never Almost Never Sometimes Fairly often Very often
- Q3 - In the last month, how often have you felt nervous and stressed?
 Never Almost Never Sometimes Fairly often Very often
- Q4 - In the last month, how often have you felt confident about your ability to handle your personal problems?
 Never Almost Never Sometimes Fairly often Very often
- Q5 - In the last month, how often have you felt that things were going your way?
 Never Almost Never Sometimes Fairly often Very often
- Q6 - In the last month, how often have you found that you could not cope with all the things that you had to do?
 Never Almost Never Sometimes Fairly often Very often
- Q7 - In the last month, how often have you been able to control irritations in your life?
 Never Almost Never Sometimes Fairly often Very often
- Q8 - In the last month, how often have you felt that you were on top of things?
 Never Almost Never Sometimes Fairly often Very often
- Q9 - In the last month, how often have you been angered because of things that happened that were outside of your control?
 Never Almost Never Sometimes Fairly often Very often
- Q10 - In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?
 Never Almost Never Sometimes Fairly often Very often

Calculate

Results for Comment

It's a Stressful text

Your Score is:

18

Scores ranging from 0-13 would be considered low stress.

Scores ranging from 14-26 would be considered moderate stress.

Scores ranging from 27-40 would be considered high perceived stress.

Fig. 4 Results of Computation

VII. CONCLUSION

As a result of all this research, it is certainly noted that the psychological condition of a person can make him/her their worst enemy by affecting their mental state of being. Mental state which is meant to be the most valuable aspect in a person, which once disturbed, can be very difficult to bring back to normal state. This condition has both long-term chronic affect and short-term chronic affect. So, in order to prevent the person from going into long-term effect, we have created this research model as a result of which, anyone suffering from any of these psychological problems can be detected with the way he/she converses with other people and shares their feelings with them. Prevention is better than cure is what makes us believe that this research work may help a few people initially, but as time passes, it may be of utmost help to a greater mass and for their good and well-being. With that purpose in mind, we have made this model which will predict whether the person is under any kind of stress or not, along with that it will also predict the level of stress using PSS [22] with the accuracy of 86% by simply analyzing the text sent on social media.

REFERENCES

- [1] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R., Wang, H., and Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6(1). doi:10.1007/s13755-018-0046-0
- [2] Turcan, E., and Mckeown, K. (2019). Dreddit: A Reddit Dataset for Stress Analysis in Social Media. *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*. doi:10.18653/v1/d19-6213
- [3] S. M. Chaware, Chaitanya Makashir, Chinmayi Athavale, Manali Athavale and Tejas Baraskar. *Stress Detection Methodology based on Social Media Network: A Proposed Design*. (2020). *International Journal of Innovative Technology and Exploring Engineering Regular Issue*, 9(3), 3489-3492. doi:10.35940/ijitee.b7537.019320
- [4] Reshma Radheshamjee Baheti, and Supriya Kinariwala. *Detection and Analysis of Stress using Machine Learning Techniques*. (2019). *International Journal of Engineering and Advanced Technology Regular Issue*, 9(1), 335-342. doi:10.35940/ijeat.f8573.109119
- [5] Ahuja, R., & Banga, A. (2019). *Mental Stress Detection in University Students using Machine Learning Algorithms*. *Procedia Computer Science*, 152, 349-353. doi:10.1016/j.procs.2019.05.007
- [6] Shen, C. (n.d.). *Text Analytics of Social Media: Sentiment Analysis, Event Detection and Summarization*. doi:10.25148/etd.fi14110776
- [7] Shafi, J., & Rahman, M. S. (2021). *Application of Machine Learning Models for Detecting Mental Stress during COVID-19 Pandemic*. *SSRN Electronic Journal*. doi:10.2139/ssrn.3778839
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. May 2019.
- [9] *TF-IDF from scratch in python on a real-world dataset ...* (n.d.). Retrieved from <https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089>
- [10] Yla R. Tausczik, J. W. (1970, January 01). *The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods* - Yla R. Tausczik, James W. Pennebaker, 2010. Retrieved from <https://journals.sagepub.com/doi/10.1177/0261927X09351676>
- [11] Qasrawi, R., Polo, S. V., Al-Halawah, D. A., Hallaq, S., and Abdeen, Z. (2021). *Schoolchildren' Depression and Anxiety Prediction Using Machine Learning Algorithms (Preprint)*. doi:10.2196/preprints.32736.
- [12] Tobin, R. M. (2005). *Measuring Emotions With the Linguistic Inquiry and Word Count (LIWC)*. *PsycEXTRA Dataset*. doi:10.1037/e525752006-001
- [13] Quadir, Ryana and Hossain, Md. Fokhray, *Mental Anxiety and Depression Detection during Pandemic using Machine Learning*, Volume 11, November-2020.
- [14] Nijhawan, T., Attigeri, G., and T, A. (2021). *Stress Detection using Natural Language Processing and Machine Learning over Social Interactions*. doi:10.21203/rs.3.rs-994868/v1
- [15] *Using Social Media to Study Human Behavior*. (2018, February 05). Retrieved from <https://www.aau.edu/research-scholarship/featured-research-topics/using-social-media-study-human-behavior>
- [16] Buffone, A. (2018). *Language of Stress in Social Media [Transcript]*. *PsycEXTRA Dataset*. doi:10.1037/e514912018-001.
- [17] Staff, S. X. (2017, August 09). *What social media reveals about your personality*. Retrieved from <https://phys.org/news/2017-08-social-media-reveals-personality.html>
- [18] *Leveraging on NLP to gain insights in Social Media, News ...* (n.d.). Retrieved from <https://towardsdatascience.com/leveraging-on-nlp-to-gain-insights-in-social-media-news-broadcasting-ca89752ef638>
- [19] *The Psychology of Social Media*. (n.d.). Retrieved from <https://online.king.edu/news/psychology-of-social-media/>
- [20] Nijhawan, T., Attigeri, G., and T, A. (2021). *Stress Detection using Natural Language Processing and Machine Learning over Social Interactions*. doi:10.21203/rs.3.rs-994868/v1
- [21] *The German Version of the Perceived Stress Scale – Psychometric Characteristics in a Representative German Community Sample | BMC Psychiatry | Full Text*



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)