IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 1, November 2022

Linguistic Stress Prediction System

Tejashree Nikam[1], Aarohi Mohrir[2], Vishakha Zambare[3], Ashwini Manike[4], Dr. Kishor Wagh[5]

BE Students, Department of Computer Science[1,2,3,4]

Professor, Department of Computer Science[5]

AISSMS Institute of Information Technology, Pune, Maharashtra, India

Abstract: *The COVID-19 pandemic has proved to be a challenge for most of the businesses, working professionals and organizations. With the unconventional and different work style patterns being adopted there has been an increased risk of stress among employees. Despite a lot of mental health awareness programs stress continues to creep in everyone's life. Mental Health being a paramount subject for everyone, we have a taken a small step to utilize the various predictive analytic methods to discover and examine stress in employees. For this project twitter data of working employees has been used. Multiple predictive analytic methods were used to train the model post required data sanitation and preliminary filtering. Natural Language Processing methods have been used to clean the tweets from the dataset and then using predictive analytics methods like Random Forests and Support Vector Machines, sentiment of the tweet of that employee is analyzed to predict whether the employee is in stress or not. Using these results the employers can spot early signs of stress among employees and address the scenario.*

Keywords: Natural Language Processing, Tweets, Random Forests, Support Vector Machines

I. INTRODUCTION

Stress has become very common these days, be it children or adults they all are prone to the changing weathers of emotional and mental pressures. Today, more than three hundred million people suffer from stress and depression. Undoubtedly, the working population is highly impacted due to stress. With COVID-19 being the ultimate game changer, the corporate employees have faced a lot of issues in adapting to new work patterns, lower salaries, lack of opportunities, unmanageable workload, prolonged working hours, and lack of communication with team.

Traditionally, psychiatrists and psychologists are the ones who help anyone spot reasons of stress and maintain stress ratios. Unfortunately, as a result of costly fees that too in the era of pandemic, there are a very few who can afford the luxury of treating themselves. Using the escalated stress rates as a condition and inability to get treated for the general masses we decided to focus on the source that are the employers. Our approach consumes the textual expressions of the working employees as a measure to predict stress. Most importantly social media provides an ocean of never know opportunity to predict early stress. Every second approximately six thousand tweets are tweeted.

Twitter being a more content unique platform, account holders utilize aliases and are more likely to be linked with physically unacquainted people. Not only does twitter allows getting to know new people, but also gives an opportunity to everyone to freely express themselves without fear of being judged or being prejudiced.

II. MATERIALS AND METHODS

2.1 Dataset Description

We are using the twitter dataset which contains the tweets of 14,410 employees. We have used 2410 stress tweets and 2100 non stress tweets for a biased model and further divided into training and testing datasets.

2.2 Proposed System

In this system we are using Support Vector Machines and Random Forest as the candidate machine learning models to be used for training the processed dataset.Figure-1 shows the system architecture.to collect customer feedback after purchasing a product by asking a number of specific questions from customers that will help improve website performance and provide a better feedback system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 1, November 2022

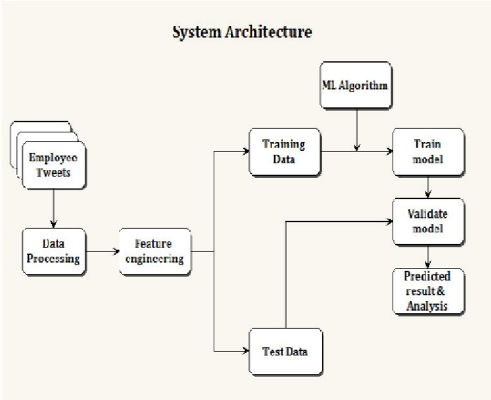

Figure 1: System Architecture

2.3 Data Cleaning

We have pre-processed out twitter data for cleaning the tweets with irregular words and emojis which do not carry any linguistic sense of emotion in predicting stress.

2.4 Feature Engineering

We have used tweets as our inputs, hence in order to bring in right and grammatically correct words into our stress and non-stress linguistic dictionary

2.5 Prediction of Stress

This system majorly serves the intention of doing a very preliminary check of whether the employee is stressed or not by analysing their tweets. This system uses linguistic markers as means to detect the emotion behind the tweet of that employee. The machine learning algorithm output 1 or 0 to predict if the tweet written by the used signifies stress or non-stress respectively.

2.6 Deployment as a Flask Web Application

We have used Flask based framework to create a web application through which our "Linguistic Stress Prediction System" is deployed.

III. ALGORITHMS USED

3.1 SVM

Support-vector machine is an algorithm that represents different classes in a hyper plane in a multi-dimensional space. The hyper plane is generated in an iterative fashion to minimize the error. The aim of SVM is to divide the dataset into two classes to get Maximum marginal hyper plane. Figure -2 shows Support Vector Machine.

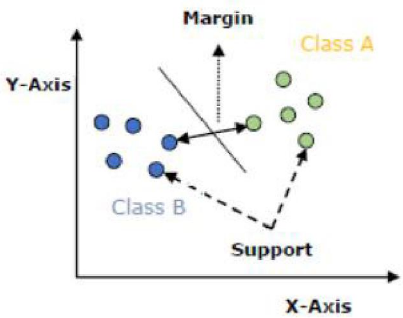

Figure 2: Support Vector Machine

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, November 2022

3.2 Random Forests

Random forest is a machine learning model which combines results of many decision tress to classify data. The output of many decision trees is averaged out to give result of random forest .Figure 3 shows Decision tree.

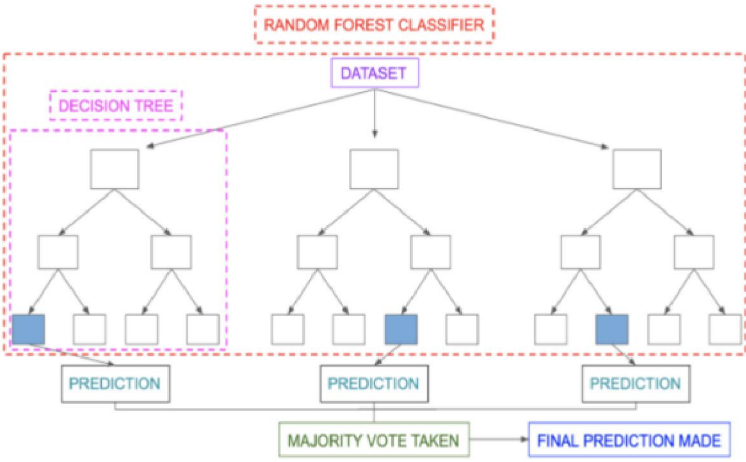

Figure 3: Random Forest

IV. IMPLEMENTATION

4.1 Deployment Setup

For implementing this system, we initially pre- processed the stress and non-stress data individually and then concatenated them together. As next steps we split the dataset into two segments – 80% of data as training set which contains nearly and 20% of data in testing/validation set. Initially we trained our machine using SVM algorithm and then tested using the same. Then, using the same data, we created the Random Forest model and validated it. Accuracy of the models we created has been covered in Table1.

4.2 Model Selection and Comparison

From the below table we can clearly see the model outperforms the.

Algorithm	**Accuracy**
Support Vector Machine	80%
Random Forest	96%

Table 1: Accuracy Comparison

We can see that the Random Forest Model is providing a better accuracy as compared to Support Vector Machine. Hence , we have incorporated the Random Forest model for training and testing the final data and using it in our front-end application.

4.3 Performance Metrics

We have considered the following performance metrics for selection of model –

A . Precision: It is the percentage of records that the machine categorized as helpful that are originally positive.

Precision = TP/ TP+FP

B. Recall: We use recall for completeness as to what percentage of positive record has the classifier labeled positive.

Recall = TP/TP+FN

C. F1 Score: The F1 score is -

Precision = Two x Precision x Recall / Precision + Recall

D. Support: Support is the quantity of records of the true outputs which reside in each category of target .

	Precision	Recall	F1-Score	Support
0.0	0.74	0.82	0.78	391
1.0	0.85	0.77	0.81	488

Table 2: Random Forest Metrics

	Precision	Recall	F1-Score	Support
0.0	0.97	0.93	0.95	391
1.0	0.94	0.98	0.96	488

Table 3: Support Vector Machine Metrics

V. DISCUSSION

As a next step towards enhancing this system a larger corpus of tweets can be used to create the stress and non-stress linguistic marker dictionary. Tweets being a highly subjective matter which changes from person to person is a good way to predict temporary stress among employees . But, on the longer run more and more newer attributes can be incorporated to support the stress prediction among working employees.

VI. CONCLUSION

With the dynamic competitive world moving in all possible directions of progress, mental health has become most important on the longer run. With the amalgamation of mental health data and machine learning algorithms like decision tree and random forest we have successfully been able to develop a system which utilizes linguistic data to predict whether an employee is in stress or not.

ACKNOWLEDGMENT

The success and the outcome of this report required a lot of guidance. We are very grateful to our guide Dr. K. S. Wagh who has provided expertise and encouragement. We thank Ma'am who provided vision and knowledge that was very helpful throughout the research. All that we have done is only due to the great guidance. We also express our gratitude to Dr. S.N. Zaware Head of Computer Engineering Department, AISSMS Institute of Information Technology, for the valuable support. We would also like to extend our sincere thanks to Principal Dr.P.B. Mane, for his dynamic and valuable guidance throughout the project and providing the necessary facilities that helped us to complete our dissertation work.

REFERENCES

[1]. SairamBalani and Munmun De Choudhury ,"Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media." , 2015.

[2]. Ronan Collobert and Jason Weston. "Unified architecture for natural language processing: Deep neural networks with multitask learning.", 2008

[3]. Yiping Li1, Jing Huang, Hao Wang, Ling Feng," Pre-dicting Teenager's Future Stress Level from Micro-blog", IEEE 2015.

[4]. S.K. Yadav, Arshad Hashmi," An Investigation of Occu-pational stress Classification by using Machine Learning Techniques", Vol.-6, Issue-6, 2018.

[5]. Jinghe Zhang, Jiaqi Gong, Laura Barnes, "HCNN: Hetero-geneous Convolution Neural Networks for Comorbid Risk Prediction with Electronic Health Records", IEEE 2017.

[6]. Min Chen, YixueHao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities", IEEE 2017.

[7]. Mario Munoz˜-Organero, Victor Corcoba-Magana, ˜" Pre-dicting Upcoming Values of Stress While Driving", IEEE 2016.

[8]. Thomas Stutz1, Thomas Kowar, Michael Kager, Martin Tiefengrabner, Markus Stuppner, Jens Blechert, Frank H. Wilhelm, Simon Ginzinger," Smartphone Based Stress Prediction", UMAP 2015, LNCS 9146, pp. 240–251, 2015

[9]. Xu, Q., Nwe, T.L., Guan, C.. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE Journal of biomedical and health informatics 2015; 19(1):275–281.

[10]. David Liu, Mark Ulric," Listen to Your Heart: Stress Prediction Using Consumer Heart Rate Sensors", Stanford CS 229: Machine Learning, Autumn 2013-2014.

[11]. SayaliAmbekar, RashmiPhalnikar," Disease Risk Pre-diction by Using Convolutional Neural Network", IEEE 2018.

[12]. Sharath Chandra Guntuku, AnnekeBuffone, KokilJaidka, Johannes C. Eichstaedt, Lyle H. Ungar," Understanding and Measuring Psychological Stress Using Social Media", ICWSM 2019.

[13]. ChrandrasekarVuppalapati, Mohamad S Khan, Nisha Raghu, Priyanka Veleru, Suma Khursheed," A System To Detect Menta

[14]. SamratBisai, Richa Chaudhary, "Stress Among the Students of an Engineering Institution in India: An Empirical Analysis", Jindal Journal of Business Research, Volume: 6 issue: 2, page(s): 186-198 Article first published online: October 30, 2017; Issue published: December 1, 2017

[15]. Huijie Lin, JiaJia, QuanGuo, YuanyuanXue, Qi Li, Jie Huang, LianhongCai, and Ling Feng, "User-level psychological stress detection from social media using deep neural network", 2014, In Proceedings of the 22nd ACM international conference on Multimedia (MM '14). ACM, New York, NY, USA, 507-516

[16]. C. Liao, R. Chen, and S. Tai, Emotion stress detection using EEG signal and deep learning technologies - IEEE Conference Publication, 2018 IEEE Int. Conf. Appl. Syst. Invent., no. 2, pp. 9093, 2018

[17]. Chigerwe et al. BMC Medical Education 2014, Published on: 28 November 2014

[18]. https://www.forbes.com/sites/lucysherriff/2019/ 03/25/how-onecompany-is-using-ai-to-address-mental-health

[19]. Jung, Yuchae, and Yong Ik Yoon. "Multi-level assessment model for wellness service based on human mental stress level." Multimedia Tools and Applications 76.9 (2017): 11305-11317.

[20]. Norizam, Sulaiman. Determination and classification of human stress index using nonparametric analysis of EEG signals. Diss. UniversitiTeknologi MARA, 2015