ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

An improved method for Detection of Employee Stress Using Machine Learning.

MUSUNURU RATNAKAR[1], SWATHI TERLI[2], PILLEM RANI[3], G VENKATA PRADEEP KUMAR[4]

[1] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.
[2] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.
[3] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.
[4] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.

mrk@sircrreddycollege.ac.in[1], tswathi@sircrreddycollege.ac.in[2], pr@sircrreddycollege.ac.in[3], gpk@sircrreddycollege.ac.in[4]

ABSTRACT

Disorders of stress are very casual thing among the employees who are working in corporate sectors. As with changing work of people and their living lifestyle, we can see the increment of stress in the working employees. Even many corporate sectors are providing variety of schemes related to mental health and trying to reduce the disorders of stress in the working environment, the disorder is very far from stopping. In our paper, we are going to make use of two techniques of machines to determine the amount of stress the employee is having who is working in corporate sectors and try to narrow down the issues that identify the stress levels. We are going to apply two techniques of machine learning (i.e. SVM and Random Forest) when the data preprocessing and the cleaning of data is once finished. The correctness of our trained model was clearly read and analyzed. By using these two techniques of machine learning, the main features that result in disorders of stress are found to be as sex, background of family and ease of benefits of health in the working place of employee. With these results, corporate industries can now narrow down the stress and can establish a very friendly working place for the corporate sectors employees.

Keywords: Machine Learning, SVM and Random Forest, corporate sectors employees

I. INTRODUCTION

Disorders of stress which are related to mental health are not rare for the employees working in corporate sectors. Some analysis done earlier have created some concern on the very same. Based on the work done by Association of Industry, Assocham, we come to know that above 42% of the professional working employees in the corporate private sectors of India are suffering from stress or common disorders of anxiety because of late night working hours and also due to fixed timings. This part of singles are growing as mentioned in the Economic Times of 2018 article which is dependent on the survey that was managed by the Optum[4].There is a survey that considers the replies of nearly eight lakh working employees who are working from more than seventy huge companies, with each single company having its employees more than 4,500 working professionals. The

workplace which is free form stress must be given at most importance for higher productivity and happy living for the working employees. There are many steps which we can take to help the employees come up with the disorder of stress for well-being of the mental health like assistance for counselling, guidance given for the career, sessions for management of stress, and creating an awareness of health identification of working employees who will need such kind of help will definitely improve the rates of such kind of measures for becoming victorious. We try to make this happen by using our machine learning techniques to overcome with a model that predicts the rate of the stress that is accomplished. This approach is not only going to help company HR managers to know better about their working professionals, it will also help in taking proper precautions to reduce the chances of stress in their working employees.

Disorders of stress which are related to mental health are not rare for the employees working in corporate sectors. Some analysis done earlier have created some concern on the very same. Based on the work done by Association of Industry, Assocham, we come to know that above 42% of the professional working employees in the corporate private sectors of India are suffering from stress or common disorders of anxiety because of late night working hours and also due to fixed timings

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

II. METHODOLOGY

Existing System

This system uses Naive Baye's algorithms, Support vector Machine, Gaussian Classifiers etc and but being they need lot of Data for training which apparently takes more execution time and also yields less efficient outputs

Drawbacks

which apparently takes more execution time and also yields less efficient outputs.

Proposed System

we propose an efficient system which is uses Machine learning algorithms SVM and Random Forest it give the best accuracy compare to Existing system. In this system we use the Employee Twitter data for predicting the employee is stress are not.

Advantages

Twitter data for predicting depression and stress**.**

SYSTEM DESIGN

UML DIAGRAMS

The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces.

Global Use Case Diagrams:

Identification of actors:

Actor: Actor represents the role a user plays with respect to the system. An actor interacts with, but has no control over the use cases.

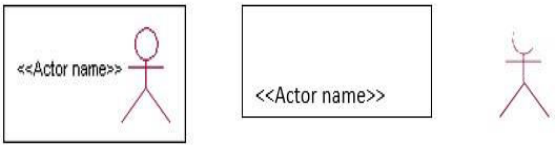

Figure 1: Proposed system design

Process:

An actor is someone or something that:

Interacts with or uses the system.

Provides input to and receives information from the system.

Is external to the system and has no control over the use cases.

Actors are discovered by examining:

- Who directly uses the system?
- Who is responsible for maintaining the system?
- External hardware used by the system.
- Other systems that need to interact with the system. Questions to identify actors:
- Who is using the system? Or, who is affected by the system? Or, which groups need help from the system to perform a task?
- An actor is someone or something that:
- Interacts with or uses the system.
- Provides input to and receives information from the system.
- Is external to the system and has no control over the use cases. Actors are discovered by examining:
- Who directly uses the system?
- Who is responsible for maintaining the system?
- External hardware used by the system.
- Other systems that need to interact with the system. Questions to identify actors:
- Who is using the system? Or, who is affected by the system? Or, which groups need help from the system to perform a task?

SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- **Python idel 3.7 version (or)**
- **Anaconda 3.7 (or)**
- **Jupiter (or)**
- **Google colab**

The actors identified in this system are:

a. **System Administrator**
b. **Customer**
c. **Customer Care**

Flow of Events

A flow of events is a sequence of transactions (or events) performed by the system. They typically contain very detailed information, written in terms of what the system should do, not how the system accomplishes the task. Flow of events are created as separate files or documents in your favorite text editor and then attached or linked to a use case using the Files tab of a model element.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

A flow of events should include:

- When and how the use case starts and ends
- Use case/actor interactions
- Data needed by the use case
- Normal sequence of events for the use case
- Alternate or exceptional flows

Construction of Usecase diagrams:

Use-case diagrams graphically depict system behavior (use cases). These diagrams present a high level view of how the system is used as viewed from an outsider's (actor's) perspective. A use-case diagram may depict all or some of the use cases of a system.

Relationships in use cases:

1. Communication:

The communication relationship of an actor in a usecase is shown by connecting the actor symbol to the usecase symbol with a solid path. The actor is said to communicate with the usecase.

2. Uses:

A Uses relationship between the usecases is shown by generalization arrow from the usecase.

3. Extends:

The extend relationship is used when we have one usecase that is similar to another usecase but does a bit more. In essence it is like subclass.

SEQUENCE DIAGRAMS

A sequence diagram is a graphical view of a scenario that shows object interaction in a time-based sequence what happens first, what happens next. Sequence diagrams establish the roles of objects and help provide essential information to determine class responsibilities and interfaces. There are two main differences between sequence and collaboration diagrams: sequence diagrams show time-based object interaction while collaboration diagrams show how objects associate with each other. A sequence diagram has two dimensions: typically, vertical placement represents time and horizontal placement represents different objects.

Object:

An object has state, behavior, and identity. The structure and behavior of similar objects are defined in their common class. Each object in a diagram indicates some instance of a class. An object that is not named is referred to as a class instance.

The object icon is similar to a class icon except that the name is underlined: An object's concurrency is defined by the concurrency of its class.

Message:

A message is the communication carried between two objects that trigger an event. A message carries information from the source focus of control to the destination focus of control. The synchronization of a message can be modified through the message specification. Synchronization means a message where the sending object pauses to wait for results.

Link:

A link should exist between two objects, including class utilities, only if there is a relationship between their corresponding classes. The existence of a relationship between two classes symbolizes a path of communication between instances of the classes: one object may send messages to another. The link is depicted as a straight line between objects or objects and class instances in a collaboration diagram. If an object links to itself, use the loop version of the icon.

III. RESULTS & DISCUSSION

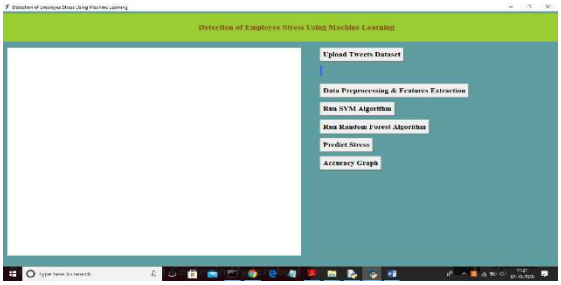

Figure 2: Upload Tweets Dataset

In above screen click on 'Upload Tweets Dataset' button to load dataset

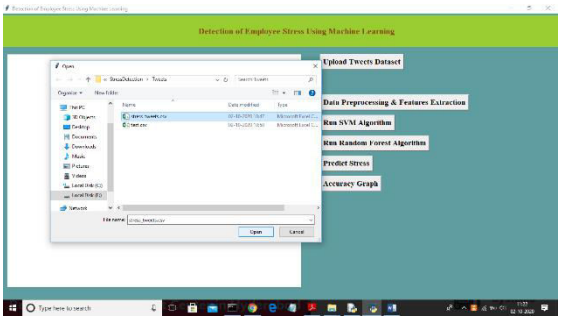

Figure 3: stress_tweets.csv

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

In above screen select 'stress_tweets.csv' dataset and then click on 'Open' button to load dataset and to get below screen

Upload Tweets Dataset

Data Processing and features extraction

Run SVM

Run Random Forest

Predict stress

Accuracy graph

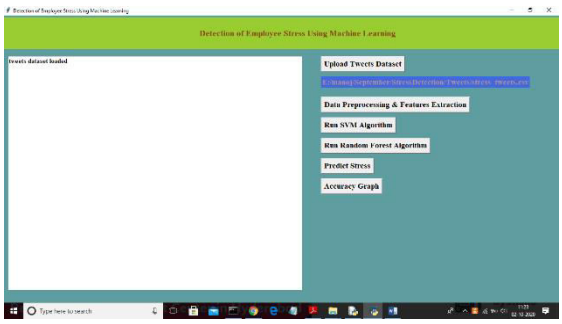

Figure 4: Data Preprocesing & Features Extraction

In above screen click on 'Data Preprocesing & Features Extraction' button to read dataset and to clean and extract features such as words from dataset and find total records in dataset, total words and application using how many records fro training and testing

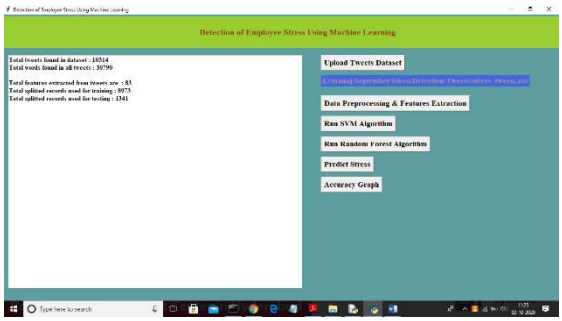

Figure 5: screen dataset contains total 10314 tweets

In above screen dataset contains total 10314 tweets and all tweets contains 30790 words and total unique words are 83 and application using 8973 records for training and 1341 for testing. Now both train and test data is ready and now click on 'Run SVM Algorithm' button to trained data using SVM machine learning algorithm.

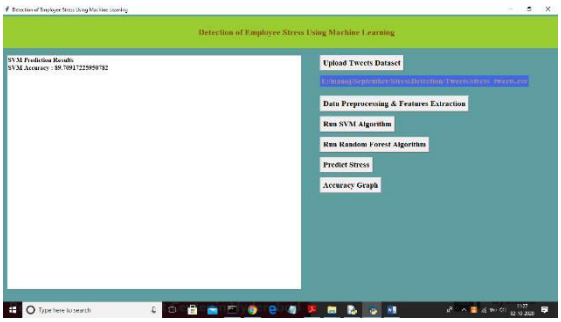

Figure 6: SVM got 89.70 correctly predicted accuracy from test data

In above screen SVM got 89.70 correctly predicted accuracy from test data and now click on 'Run Random Forest Algorithm' button to calculate its accuracy

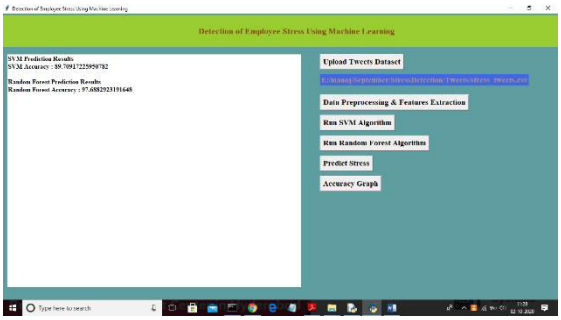

Figure 7: random forest got 97.68 correctly

In above screen random forest got 97.68 correctly prediction accuracy and now click on 'Predict Stress' button and upload test file which contains tweets and by analysing those tweets machine learning algorithm will predict whether tweets contains any stress data or not. Below is the screen shots of test tweets which we upload in next screen

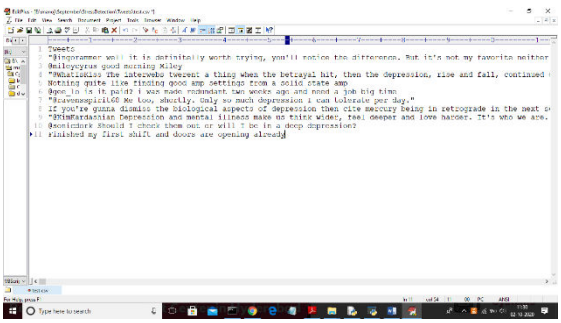

Figure 8: some tweets are there in test

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

In above screen we can see some tweets are there in test .csv file and now we upload this tweets to predict stress by clicking on 'Predict Stress' button

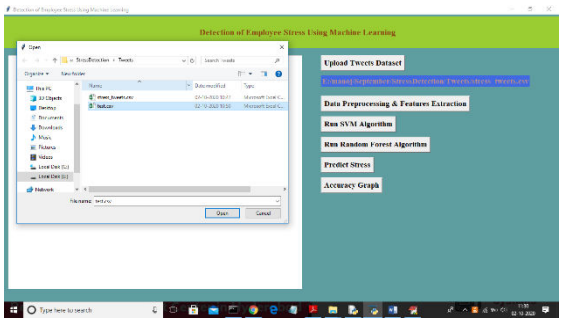

Figure 9: uploading 'test.csv'

In above screen uploading 'test.csv' file and now click on 'Open' button to predict stress

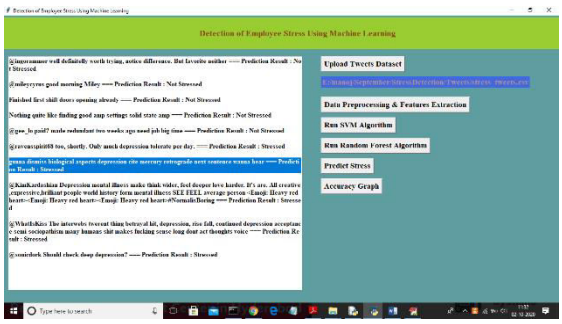

Figure 10: beside each tweet

In above screen beside each tweet we can see predicted result as Stressed or Not stressed. From above screen we can see application detecting stress successfully from messages and now click on 'Accuracy Graph' button to get below comparison graph

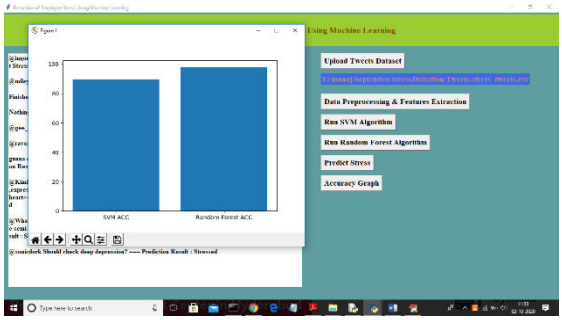

Figure 11: random forest is better than SVM

In above x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can say random forest is better than SVM

IV. CONCLUSION

Gender, also the family background which has the illness, and considering whether a single employer provides the conceptual benefits of health for their employees was having more significance compared to the other factors for determining whether an employee can obtain conceptual health associated issues. From our study, we were able to find that the people who are working in the tech companies are at more risk of obtaining stress , even though their job role was not based on tech. These perceptions could be successfully used by business companies to make more desirable HR strategies for the working employees. A 75% correctness shows that the application of two Machine Learning techniques(i.e. SVM and Random forest) for predicting the stress and conceptual health conditions provides worthy results and could be searched further, and thus meets the aim of this paper.

REFERENCE

[1] Detecting and characterizing Mental Health Related Self-Disclosure in Social Media. SairamBalani and Munmun De Choudhury. 2015.In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems -CHI EA "15, pages 1373–1378.

[2] Measuring Post Traumatic Stress Disorder in Twitter.Glen Coppersmith, Mark Dredze, and Craig Harman. 2014.

[3] Role of Social Media in Tackling Challenges in Mental Health.Munmun De Choudhury. 2013.

[4] Bhattacharyya, R., &Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from „The Economic Times"

[5] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011).Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2

[6] OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle.

[7] Van den Broeck, J., Cunningham, S. A., Eeckels, R., &Herbst, K. (2005). Data cleaning:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

detecting, diagnosing, and editing data abnormalities. PLoS medicine, 2(10), e267.
[8] Relationship between Job Stress and Self-Rated Health among Japanese Full Time Occupational Physicians Takashi Shimizu and Shoji Nagata 2007 Academic Papers in Japanese 2007.
[9] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.
[10] Gender and Stress. (n.d.). Retrieved from APA press release 2010
[11] Julie Aitken Harris, Robert Saltstone and Maryann Fraboni.(2000)An Evaluation of the Job Stress Questionnaire with a Sample of Entrepreneurs"2000 JSQ scale Entrepreneurs.
[12] "Demographic and Workplace Characteristics which add to the Prediction of Stress and Job Satisfaction within the Police Workplace" ,Jeremy D. Davey, Patricia L. Obst, and Mary C. Sheehan 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCICC). 2015.
[13] Mario Salai, Istv an Vass anyi, and Istv an Kosa, "Stress Detection using low cost Heart rate sensors", Journal of Healthcare Engineering, pp.1-13,Hindawi Publishing corporation , 2016
[14] Shwetha, S, Sahil, A, Anant Kumar J, (2017) Predictive analysis using classification techniques in healthcare domain, International Journal of Linguistics & Computing Research, ISSN: 2456-8848, Vol. I, Issue.I, June-2017.
[15] O.M.Mozos et al, "Stress detection using wearable physiological and sociometric sensors". International Journal of Neural Systems,vol 27,issue 2, 2017.