STRESS DETECTION AND ANALYSIS IN WORKING ENVIRONMENT USING SVM CLASSIFIER

Mrs. V. Swathi[1], V. Aasritha[2], K. Lakshmi Tulasi[3], U. Bhavika[4], P. Arun[5]
Assistant Professor :- 1 , Students :- 2,3,4,5
Computer Science and Engineering, Dhanekula Institute of Engineering and Technology

ABSTRACT: - Stress is commonly defined as a feeling of strain and pressure which occurs from any event or thought that makes you feel frustrated, angry, or nervous. Stress at work is common in many professions. While some stress is a normal part of work, excessive stress or high levels of stress over prolonged periods of time can affect the employee's productivity. It will have serious implications on the physical and mental health of a person. Stress increase nowadays leads to many problems like depression, suicide, heart attack, and stroke. There are many ways to detect stress. Some of the instruments are used to detect stress while there is a medical test to know the stress level. We can also use machine learning techniques. We are using the SVM algorithm in our project to detect stress. A dataset that contains different parameters is used. By applying this SVM technique to the dataset we also evaluate performance metrics like accuracy, sensitivity, and specificity.

I. INTRODUCTION

In this buzzing world, professionals are haunted constantly by occupational trauma caused by hectic schedules. This hardship of occupational traumas can trigger stress levels in professionals which can lead to catastrophic circumstances that can't be handled at a Stretch. An increase in stress levels in individuals can lead from short to prolonged detrimental states that can blowback their life in both personal and professional ways. Stress can lead to habitual problems like depression, anxiety, sleeplessness, social withdrawal, apathy, trouble in concentration, headaches, and fatigue, and it can also be the unspecified cause behind the dreadful diseases like heart attacks, increase in blood pressure, cancers, and many other. So, in order to avoid these obstacles and improve the quality of life, our project STRESS DETECTION AND ANALYSIS IN WORKING ENVIRONMENT USING SVM CLASSIFIER mainly aims to help the organization in identifying the stress caused by employees at different levels, so that the administrator can monitor their crew members, and recast them to improve the job satisfaction which enthuses them to be more productive and organization can have fruitful accomplishments.

II. LITERATURE SURVEY

Being affected by the stress in the working environment can lead to a queue cause to unidentified problems. So as to reduce the stress-related problems caused by working professionals in place of work, we identified salient points (e.g.: long hours, heavy workload, job insecurity, work pressure, etc) and framed questionary and collected the responses from our college(Dhanekula Institute of Engineering and Technology) formed a dataset ascertained to detect stress using SVM Classifier with better convincing results previous implementations.

In the paper [1] "Prediction of Mental Disorder for employees in IT industry" authored by Sandhya P, Mahek Kentesaria they used a survey-based dataset collected among the IT Professionals from disparate regions. The survey consists of various work-related questions dissimilar to the questionary we are using. They applied several Machine Learning algorithms but Random Forest stood as the best model trained for their dataset with an accuracy of 81.2%.

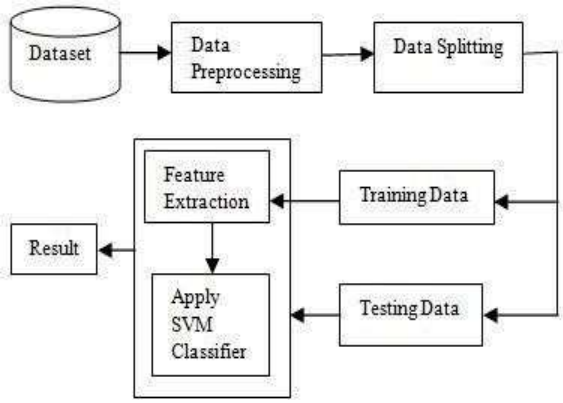
In paper [2] "Prediction of mental stress based on machine learning" penned by Akshada Kene, Dr. Shubhada Thakare the implementation is done in two ways where the stress levels are analyzed by using Single Processing and another way by using Image & Video Processing. They took a Stress Dataset where the data is a combined entry from various sensors or online stress scale questionnaires etc. In this paper [2] they took a comparison of SVM and RF the SVM stood as the best suitable model for their implementation.

In paper [3] "Detection and Analysis of Stress using Machine Learning Techniques" drafted by Reshma Radheshamjee Baheti, Supriya Kinariwala they took a textual dataset that can be of any form of text from tweets, sentences, or a blog can be entered by the user. Here to classify and predicted they applied Naïve Bayes and SVM, but to make it more accurate it is implemented by using the Word Sense Disambiguation by using n-gram and Skip-gram model.

III. METHODOLOGY

Architecture:-

Architecture describes each and every stage of the project. It contains various phases like data collection, data preprocessing, data splitting, feature extraction, and visualizing the result.



Data Collection

A questionnaire form is prepared based on the working environment stress factors. The questions are like the situations and they have to answer how they will feel in that particular situation. Each question contains three options such as (1) Never (2)Sometimes (3) Always. All the questions are marked as mandatory so that we will get the responses for each and every question.

Data Preprocessing

From the collected responses, we don't have any missing values or NULL values. Even though we don't have missing or NULL values, we should check it once either by manual preprocessing or by using some python functions. It displays whether the data contains any missing values or NULL values. If found, they should be handled in such a way that the data should be in our required format.

Data Splitting

The data should be divided into two: Training data and Testing data. In our project, the training data is 80% and the remaining 20% of the data is used for testing. We can divide the training and testing data into sub-training and testing data again. The training data is used for the machine to learn according to the data. Whereas, the testing data is used to test the data whether the predictions are correct or not.

Feature Extraction

In feature extraction, we will select a feature or the combination of features that are required to detect and analyze the stress. In our project, we have considered the questions as the features for which we have received the highest responses or the questions which have more number of similar responses.

IV. IMPLEMENTATION

Support Vector Machine (SVM)

Support Vector Machine (SVM) classifier is used to detect the stress in the working professionals with the help of the responses dataset. The classification problems are mainly solved by using the SVM classifier. It can be either binary classification or multiclass classification. Here, we have two classes. Therefore, our problem is a binary classification problem. It is applied to the dataset by taking the features as input.

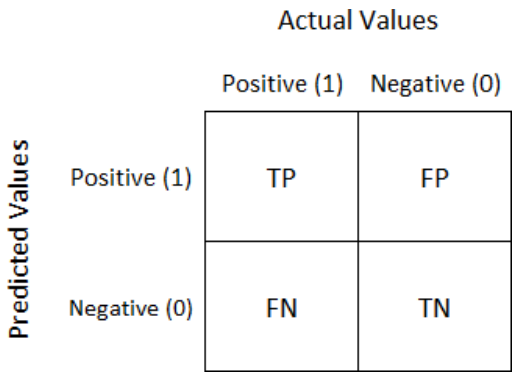
Performance Metrics

There are various types of performance metrics to measure and evaluate the performance of machine learning algorithms. The various performance metrics are accuracy, precision, recall, and many more. The suitable performance metrics should be selected based on our selected algorithm and predictions. Here we have selected accuracy, precision, recall, and f1-score as our performance metrics.

Confusion Matrix

The confusion matrix is used to measure the performance of classification problems in machine learning. The output of this confusion matrix may be of two or more classes. For a problem of two classes, there will be 4 different combinations of actual and predicted values. They are TP, FP, FN, andTN.

- **True Positive (TP)**
  If the actual value is positive and the value predicted by the model is also positive. Then if the predicted value matches the actual value, it is said to be True Positive (TP).
- **True Negative (TN)**
  If the actual value is negative and the value predicted by the model is also negative. Then if the predicted value matches the actual value, it is said to be True Negative (TN).
- **False Positive (FP)**
  If the actual value is negative and the value predicted by the model is positive, the predicted value is falsely predicted. Then it is said to be a False Positive (FP).
- **False Negative (FN)**
  If the actual value is positive and the value predicted by the model is negative, the predicted value is falsely predicted. Then it is said to be a False Positive (FP).



1. Accuracy

Accuracy is one of the most important and common performance metrics which is used to define the correct number of predictions made by our model. A model is considered the best model if that model has a high accuracy score.

Accuracy=TP+TN/TP+FP+FN+ TN

2. Precision

Precision gives the predictions which are actually correct. It is used to identify all the positive predictions that are predicted by our model. It helps us to ensure that weare getting the correct results or not. It is the ratio between true positives and all thepositives.

Precision=TP/TP+FP

3. Recall

Recall gives the correct predicted values. It is the ratio of all the true positives to thetrue positives and false negatives. It is used to identify the number of positive predictions that are predicted by our model.

Recall=TP/TP+FN

4. F1-Score

F1-Score is the mean of precision and recall. In other words, it is the weighted average of precision and recall. It is alsoknown as the balanced F-score or F-measure. The value of the F1-Score is high only when the precision and recall values are high.

F1 Score = 2*(Recall * Precision) /(Recall + Precision)
=TP/ (TP+((FP+FN)/2))

5. Sensitivity

Sensitivity is also called as True Positive Rate (TPR) is defined as the people with stress who will have a positive result. It correctly identifies the people with stress. Itis calculated by TP and FN

TPR = TP / (TP+FN)

6. Specificity

Specificity is also called as True Negative Rate (TNR) is defined as the people without stress who will have a negative result. It correctly identifies the people without stress. It is calculated by TN and FP.

TNR = TN / (TN+FP)

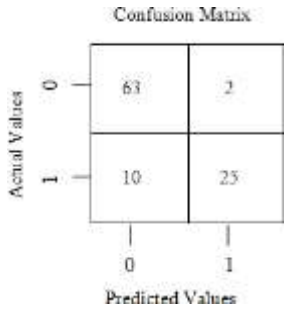
V. RESULTS

By applying the SVM algorithm, we have evaluated our model by various performance metrics.

The confusion matrix evaluates the TP, FP, TN, andFN. The results obtained will be in a matrix form. Here, we have 2 classes, so our confusion matrix is of 2x2 dimension. Here, we have applied the SVM algorithm to our dataset and

calculated performance metrics such as accuracy, precision, recall, and f1-score, sensitivity and specificity from the confusion matrix.

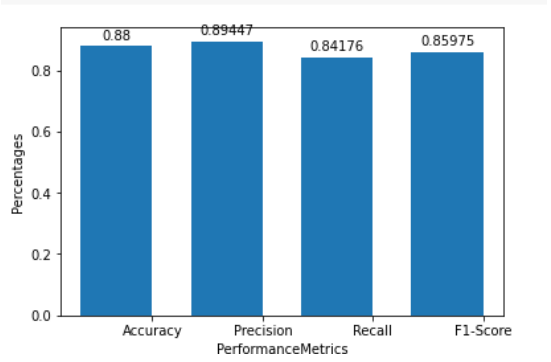
The metrics which are used to evaluate the performance of our model are accuracy, precision, recall, and f1-score. The results that are obtained after applying the Support Vector Machine show that SVM is performing well on our dataset.



The obtained percentages for performance metrics are:

Performance Metric	Percentage
Accuracy	88%
Precision	89%
Recall	84%
F1-Score	85%

We have also plotted a bar graph which shows the comparison between different performance metrics and their scores obtained. The x axis of the graph represents the performance metrics and the y axis represents the percentages.

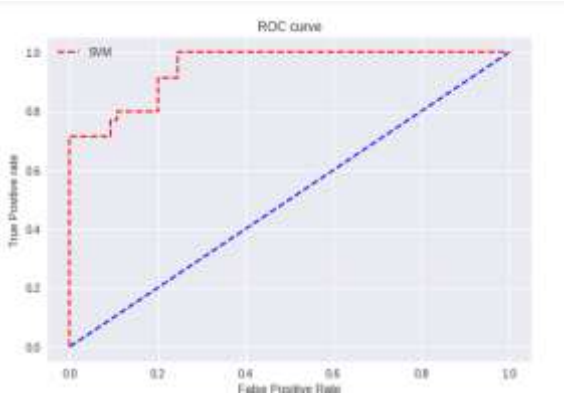


The sensitivity and specificity percentages obtained are 75% and 95% respectively.

ROC Curve

Receiver Operating Characteristic (ROC) Curve is a graphical plot which is created by plotting TPR against FPR for varying threshold values. It is a plot of FPR on x axis and TPR on y axis.

Area Under Curve (AUC) is calculated to measurethe area under the ROC curve. The AUC value obtained for our ROC curve is 94.



VI. CONCLUSION AND FUTURE ENHANCEMENTS

If the survey is taken at the end of a heavy workday then the results will be more accurate and as predicted. Here, we have considered the working environment to identify stress in the employees. We can also take different types of working environments and prepare questionnaires according to them. We can apply different machine algorithms to that dataset to detect the stress at a particular organization. For further implementation, we can use the image and video processing dataset and apply CNN to it. By doing so, we can identify the individual people who are feeling stressed.

VII. REFERENCES

[1] From Sandhya P, Mahek Kantessaria “Prediction of Mental Disorder for employees in IT Industry”. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6S, April 2019

[2] From Akshada Kene, Dr. Shubhada Thakare “Prediction of Mental Stress based on Machine Learning”.

[3] From Reshma Radheshamjee Baheti, Supriya Kinariwala “Detection and Analysis of Stress using Machine Learning Techniques”. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019

[4] G. Azar, C. Gloster, N. El-Bathy, S. Yu, R. H. Neela and I. Alothman, "Intelligent data mining and machine learning for mental health diagnosis using genetic algorithm," 2015

IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, 2015,
pp. 201-206.
[5] B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online
Mental Health-Related Communities with an Interest in Depression," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 4, pp. 1008-1015, July 2016.
[6] T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez and C. Giraud-Carrier, "Detecting Cognitive Distortions Through Machine Learning Text Analytics," 2017 IEEE
International Conference on Healthcare Informatics (ICHI), Park City, UT, 2017, pp. 508-512.
[7] 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.
[8] R. M. Khalil and A. Al-Jumaily, "Machine learning based prediction of depression among
type 2 diabetic patients," 2017 12 th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-5.
[9] Srividya, M & Subramaniam, Mohanavalli & Natarajan, Bhalaji. (2018). Behavioural Modelling for Mental Health using Machine Learning Algorithms. Journal of Medical Systems. 42. 88.10.1007/s10916-018-0934-5.

[10] Ashar Khan, Mohd Shahid Husain & Anam Khan. (2018). Analysis of Mental State of Users using social media to predict Depression! A Survey. International Journal of Advanced Research in Computer Science, II International Conference on “Advancement in Computer Engineering and Information Technology” Volume 9, Special Issue No.2, pp.100-106.