



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

A futuristic digital interface with a glowing globe in the center. The globe is surrounded by various data visualizations, including bar charts, line graphs, and network diagrams. The background is dark blue with light blue and white digital elements. The text 'INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH' is overlaid in large, bold, white letters with a black outline. Below it, 'IN SCIENCE | ENGINEERING | TECHNOLOGY' is written in smaller white letters. The interface also features a 'LOADING 100%' progress bar and a 'MEDIA' button.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijrset@gmail.com

 www.ijrset.com

Detection of Stress in IT Employees Using Machine Learning Technique

Jenifer R ¹, Samsudeen Shaffi S ²

¹PG student, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

²Assistant Professor, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

ABSTRACT - Stress disorders are a common issue among working IT professionals in the industry today. With changing lifestyle and work cultures, there is an increase in the risk of stress among the employees. Though many industries and corporates provide mental health related schemes and try to ease the workplace atmosphere, the issue is far from control. In this paper, we would like to apply machine learning techniques to analyze stress patterns in working adults and to narrow down the factors that strongly determine the stress levels. We apply Machine-learning algorithms like SVM to find the model with the best accuracy.

KEYWORDS: Stress prediction, Healthcare, Machine Learning

I. INTRODUCTION

Stress-related mental health disorders are not uncommon among the working class. Several studies in the past have raised concerns over the same. According to a study by the industry association, Assocham, more than forty-two percentage of working professionals in the Indian private sector suffer from depression or general anxiety disorder due to long work hours and tight deadlines. This portion of individuals is rising as stated in the 2018 Economic Times article based on the survey conducted by Optum[1] that half of the working professionals in India suffer from stress. The survey considered the responses of as many as 8 lakh employees from over 70 major companies, each with a workforce of 4,500 or above. Maintaining a stress-free workplace must be given a prime importance for greater productivity and well-being of the employees. Several steps can be taken to help working professionals cope up with stress for mental well-being like counselling assistance, career guidance, stress management sessions, and health awareness programs. Early identification of employees who will be needing such a help will improve the chances of such measures being successful.

We hope to ease this process by using machine learning methods to develop a model to predict the risk of stress experienced and if treatment is required by an individual by taking some of his/her professional and personal factors as parameters collected in the form of carefully drafted surveys. Such an approach will not only help HR managers to understand better about their employees but also help in taking preventive measures to decrease the chance of an employee leaving the company or underperforming. We can also perform early prediction if a person requires treatment for his mental health or not.

II. REVIEW OF LITERATURE

USRINIVASULUREDDY Machine Learning Techniques for Stress Prediction in Working Employees(2019) Stress disorders are a common issue among working IT professionals in the industry today. With changing lifestyle and work cultures, there is an increase in the risk of stress among the employees. Though many industries and corporates provide mental health related schemes and try to ease the workplace atmosphere, the issue is far from control. In this paper, we would like to apply machine learning techniques to analyze stress patterns in working adults and to narrow down the factors that strongly determine the stress levels. Towards this, data from the OSMI mental health survey 2017 responses of working professionals within the tech-industry was considered. Various Machine Learning techniques were applied to train our model after due data cleaning and preprocessing. The accuracy of the above models was obtained and studied comparatively. Boosting had the highest accuracy among the models implemented. By using Decision Trees, prominent features that influence stress were identified as gender, family history and availability of health benefits in the workplace. With these results, industries can now narrow down their approach to reduce stress and create a much comfortable workplace for their employees.

Hakan Yekta Yatbaz, Meryem Erbilek Deep Learning Based Stress Prediction From Offline Signatures (2020)

Soft-Biometric measurements are now increasingly adopted as a robust means of determining individual's nonunique characteristics with the emerging models that are widely used in the deep learning domain. This approach is clearly valuable in a variety of scenarios, specially those relating to forensics. In this study, we specifically focus on stress emotion, and propose automatic stress prediction technique from offline signature biometrics using well-known deep learning architectures such as AlexNet, ResNet and DenseNet. Due to the limited number of research that study emotion prediction from offline handwritten signatures with deep learning methods, best to our knowledge this is the first experimental study that presents empirical achievable prediction accuracy around 77%

Ramtin Gharleghi, Gihan Samarasinghe, Arcot Sowmya, Susann Beier Deep Learning For Time Averaged Wall Shear Stress Prediction In Left Main Coronary Bifurcations(2020)Analysing blood flow in coronary arteries has often been suggested in aiding the prediction of cardiovascular disease (CVD) risk. Blood flow induced hemodynamic indices can function as predictive measures in this pursuit and a fast method to calculate these may allow patient specific treatment considerations for improved clinical outcomes in the future. In vivo measurements of these metrics are not practical and thus computational fluid dynamic simulations (CFD) are widely used to investigate blood flow conditions, but require costly computation time for large scale studies such as patient specific considerations in patients screened for CVD. This paper proposes a deep learning approach to estimating the well-established hemodynamic risk indicator time average wall shear stress (TAWSS) based on the vessel geometry. The model predicts TAWSS with good accuracy, achieving cross validation results of average Mean Absolute error of 0.0407 Pa and standard deviation of 0.002 Pa on a 127 patient CT angiography dataset, while being several orders of magnitude faster than computational simulations, using the vessel radii, angles between bifurcation (branching) vessels, curvature and other geometrical features. This bypasses costly computational simulations and allows large scale population studies as required for meaningful CVD risk prediction.

Wenli Gao, Xinming Lu, Yanjun Peng, And Liang Wu A Deep Learning Approach Replacing the Finite Difference Method for In situ Stress Prediction(2020)In the domain of geotechnical engineering analysis, fast Lagrangian analysis based on the finite difference method is the most commonly used numerical analysis method. It is a classical numerical algorithm applied to calculate the in situ stress. In this paper, we propose a deep learning (DL) architecture called the enhance-and-split feature capsule network embedded in fully convolutional neural networks (ES-Caps-FCN) to predict the in situ stress for a strain-softening model when using the finite difference method for the numerical computation. Experiments indicate that this novel approach is stable and convergent. Compared with some classical prediction methods including linear regression analysis and a deep neural network, the mean squared error of our proposed algorithm is as low as 0.059866%, which is lower than the 0.616676% of the deep neural network prediction algorithm and the 0.978495% of the conventional machine learning algorithm. Additionally, the calculation efficiency of fully trained deep learning models is superior to that of the conventional finite difference method. Therefore, DL is a feasible and promising fast and accurate surrogate for the finite difference method for solving the in situ stress.

III. PROPOSED SYSTEM

Nowadays, stress become a major problem in these days. Sometimes it response is positive as well as negative. These days the term stress is considered to be one of the major factor learning to various health problem (1993, Gmelch,). when it crosses certain level, it complex the day by day in our lives and powers people to digress from the typical public activity. The growing pace of life rushed and cantered lifestyles suggest that pressure is a vital piece of human life. A human being in a state of changing in accordance with pressure exhibits direct. There are different types of sources of stress but three are commonly found in everyone that's are external stress, environmental stress, physical stress. The Environment Stressor: The external stress can be occurred because by the environment stressor, when the person is unable to respond to the external and internal stimulus or situation it causes the stress. The examples include disturbance in the environment, crowding, cold and hot weather, traffic, high number of crimes in the societies, pollution and pandemic viruses.

Social Stressor: Every individual lives in the society and interact with many people in their day to day life. The external stressor can become the source of stress in the individual life which includes- Hot and cold climate, natural disasters, criminal offence, contamination, death. This kind stressors are happened by the earth and humans have no control over these kinds of stressors.

Physiological Stressor: Every individual suffers from the different kind of stressful situation in their lives. Every individual also plays a social role in their lives by the playing a multiple character with the different people. Each

person does their jobs to live in a society and with family, such as brother, parents, friends, boss, life partners and many more social roles they play. Because of stress there might be other medical problems like weight, respiratory failure, diabetes, asthma and so on. Every day, many employees commits suicide in the different parts of the country. According to lancet report in 2012 our nation has reported large number of suicide cases age between fifteen to twenty eight. In year 2015, 8,934 number of employee suicide cases was formed. In 2010-2015, 39,775 employees were commits suicide due to under stress The main software used in a typical Python machine learning pipeline can consist of almost any combination of the following tools: 1. NumPy, for matrix and vector manipulation 2. Pandas for time series and R-like DataFrame data structures 3. The 2D plotting library matplotlib 4. SciKit-Learn as a source for many machine learning algorithms and utilities 5. Keras for neural networks and deep learning.

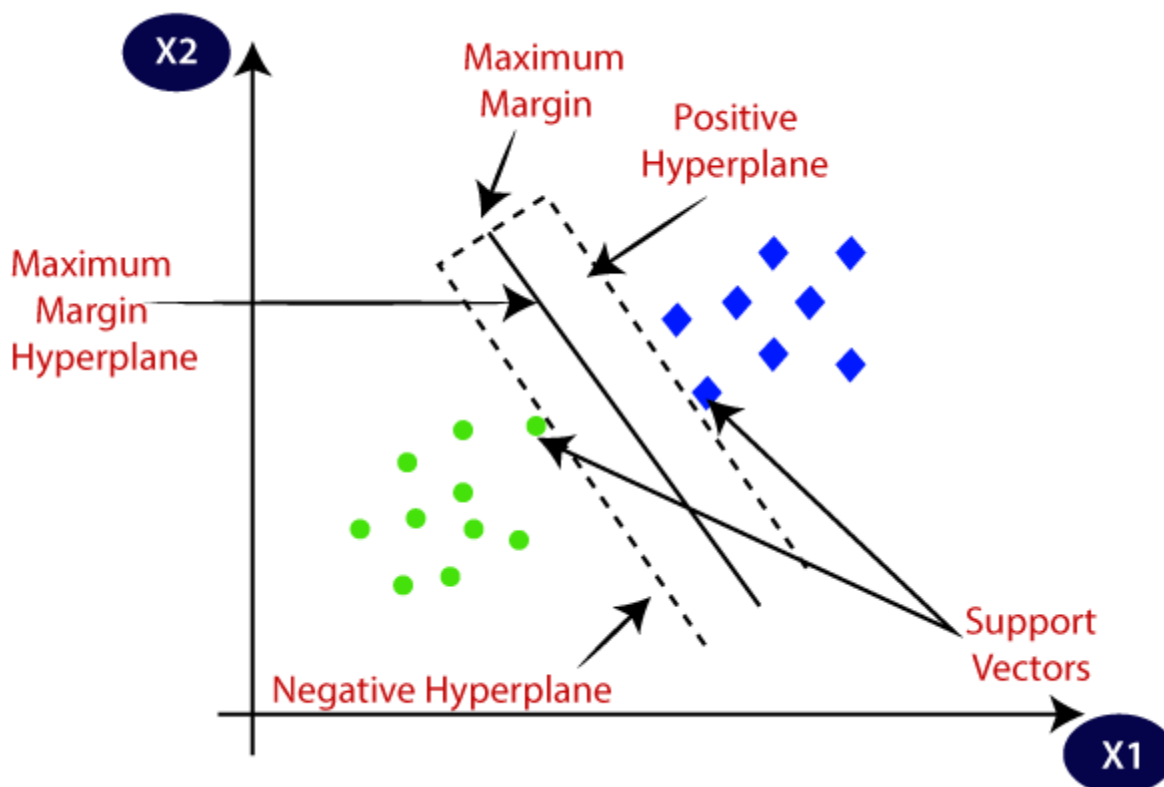
The automation for employee stress prediction in institutes and educational organizations has been very minimal. Observing each employee and his or her profile is a huge task. This responsibility lies under human interaction and that is why our work paves way for the automatic stress prediction of each employee succumbing under various parameters and proposes the solution to each employee rightly. This is done with the help of Machine learning and data science techniques. Keeping a check of each employee's stress levels, and monitoring it closely, helps to heighten their performance in an organization. Employees are classified as to whether they are stress free or stress full and if they are stressful, their range of stress is highlighted. Based on this percentage, the authorities give each individual solutions and advice. This system reports the accurate predictions.

(a) Removal of unnecessary data: The pattern of question inquiry or thinking can be changed so as to remove needless samples. For example, compared with all the parameters the unwanted null values and the relative minority can be removed preferentially when the model is constructed. (b) Data representative: Where appropriate, advanced data processing methods can be added (such as existing data generated from some statistics) to reinforce the rationality of data interpretation and summarization and to make the data abundant. The data can be increased or decreased by means of up-sampling and down sampling process. (c) Processing of data missing: The missing of some data in the database will bring about difficulties in analyzing. Although some methods can be adopted to supplement the data, the data are not all real and may be inaccurate at times. Therefore, sometimes rules for removing or methods and principles for supplement the missing data can be thought about. Login module - Employees can login to the website using their credentials given by the admin. Stress prediction Module - Various parameters are enlisted and is visible to the employees. Employees can input these parameters to predict their stress level prediction. The obtained result is categorized into stressful and stress free and the stressful result is further classified into the percentage of stress and also represented graphically. Speed and very low complexity, which makes it very well suited to operate on real scenarios. Computation load needed for image processing purpose is much reduced, combined with very simple classifiers.. Ability to learn and extract complex image features. With its simplicity and fast processing time, the proposed algorithm is suitable to be implemented in embedded system or mobile application that has limited processing resources

IV. METHODOLOGY

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane. Types of SVM can be of two types: Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier. Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier. Hyperplane and Support Vectors in the SVM algorithm: Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM. The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane. We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points. Support Vectors: The data points or vectors that are the

closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector. PERFORMANCE MATRICES: Data was divided into two portions, training data and testing data, both these portions consisting 70% and 30% data respectively. All these two algorithms were applied on same dataset using Entought Canaopy and results were obtained. Predicting accuracy is the main evaluation parameter that we used in this work. Accuracy can be defied using equation. Accuracy is the overall success rate of the algorithm. CONFUSION MATRIX: It is the most commonly used evaluation metrics in predictive analysis mainly because it is very easy to understand and it can be used to compute other essential metrics such as accuracy, recall, precision, etc. It is an NxN matrix that describes the overall performance of a model when used on some dataset, where N is the number of class labels in the classification problem. to Random Forest classifier.



V. MODULE

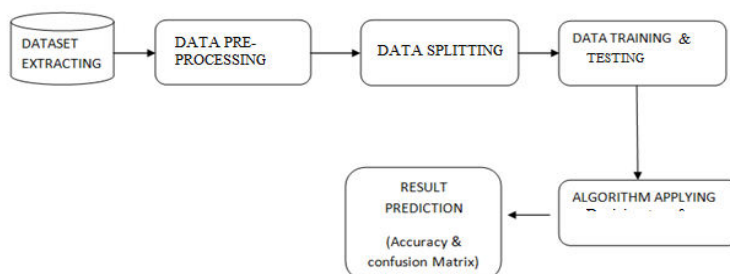


Fig 1. System Architecture

DATA COLLECTION : Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be

filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external. Our data mainly consists of the previous year stock prices. Initially, we will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

FORMATTING :You may not have chosen the details in a format that suits you for working with. The data may also be in in an electronic database and you would like it to be in a spreadsheet, or the information may be in a proprietary file format and you would like it to be in an electronic database or folder. **CLEANING**: Cleaning data is the eradication or restoration of unfinished or empty data. There may also be incomplete occurrences of data which do not carry the information that you think you'd like to lever may need to eliminate these occurrences. In addition, there are attributes which carry sensitive information and that the attributes are likely to be omitted. **DATA-PRE**

PROCESSING : Data pre-processing is a part of machine learning, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the dataset into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs. In this paper we have used is null() method for checking null values and lable Encoder() for converting the categorical data into numerical data

FEATURE EXTRACTION:A new method to detect malicious Android applications through machine learning techniques by analyzing the extracted permissions from the application itself. Features used to classify are the presence of tags uses-permission and usesfeature into the manifest as well as the number of permissions of each application. These features are the permission requested individually and the «usesfeature» tag.the possibility of detection malicious Android applications based on permissions and 20 features from Android application packages.

TRAINING THE MACHINE:Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data.

SPLIT THE DATASET INTO TRAIN AND TEST SET: This step includes training and testing of input data. The loaded data is divided into two sets, such as training data and test data, with a division ratio of 80% or 20%, such as 0.8 or 0.2. In a learning set, a classifier is used to form the available input data. In this step, create the classifier's support data and preconceptions to approximate and classify the function. During the test phase, the data is tested. The final data is formed during preprocessing and is processed by the machine learning module.

DATA TRAINING; Algorithms learn from data. They find relationships, develop understanding, make decisions, and evaluate their confidence from the training data they're given. And the better the training data is, the better the model performs.In fact, the quality and quantity of your training data has as much to do with the success of your data project as the algorithms themselves.Now, even if you've stored a vast amount of well-structured data, it might not be labeled in a way that actually works for training your model. For example, autonomous vehicles don't just need pictures of the road, they need labeled images where each car, pedestrian, street sign and more are annotated; sentiment analysis projects require labels that help an algorithm understand when someone's using slang or sarcasm; chatbots need entity extraction and careful syntactic analysis, not just raw language.In other words, the data you want to use for training usually needs to be enriched or labeled. Or you might just need to collect more of it to power your algorithms. But chances are, the data you've stored isn't quite ready to be used to train your classifiers.Because if you're trying to make a great model, you need great training data. And we know a thing or two about that. After all, we've labeled over 5 billion rows of data for some of the most innovative companies in the world. Whether it's images, text, audio, or, really, any other kind of data, we can help create the training set that makes your models successful.

VI. DISCUSSION

A dynamic model to predict bus arrival time using only limited data sets based on machine learning. Based on the previous study results, the base travel time is predicted using the machine learning model. The performance analysis of KNN and random forest has been denoted as graphical representation in Fig 2.

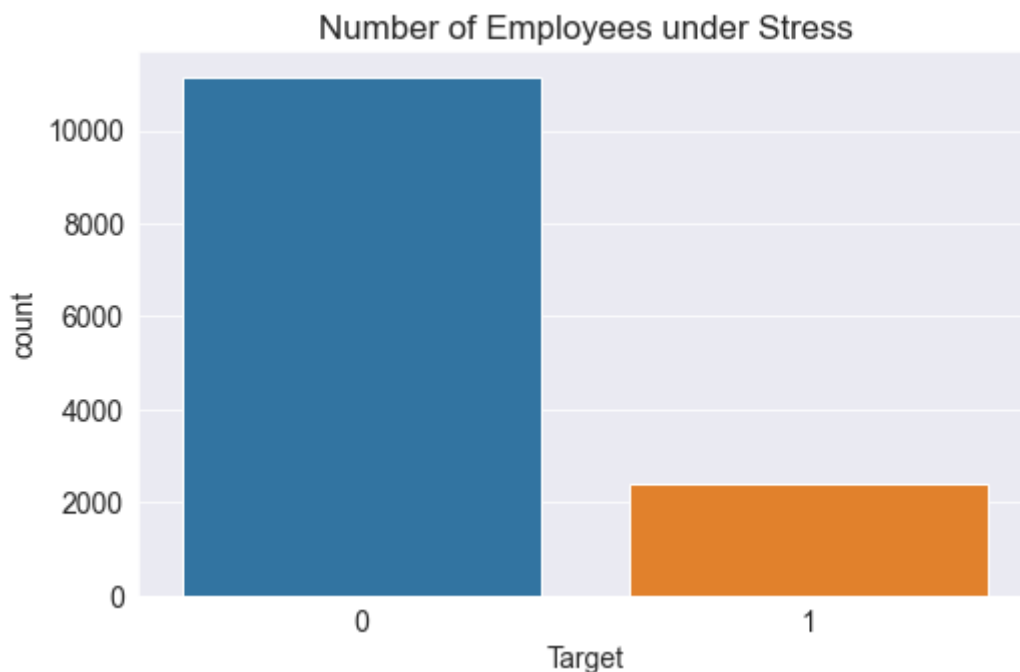
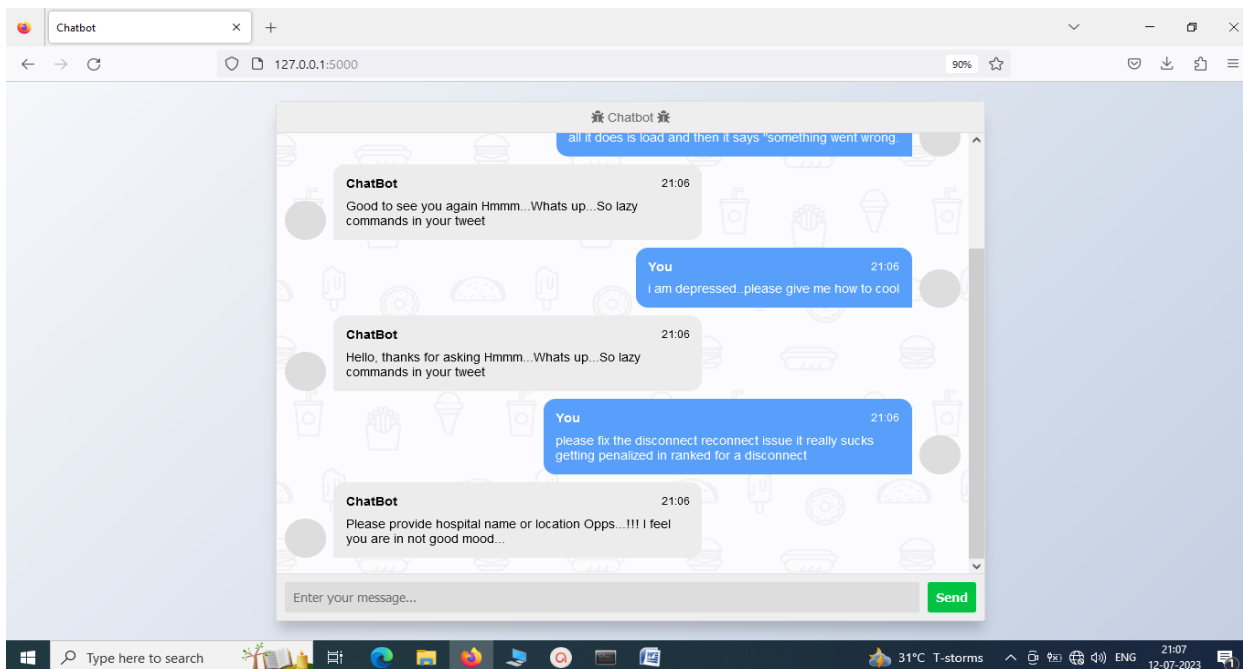
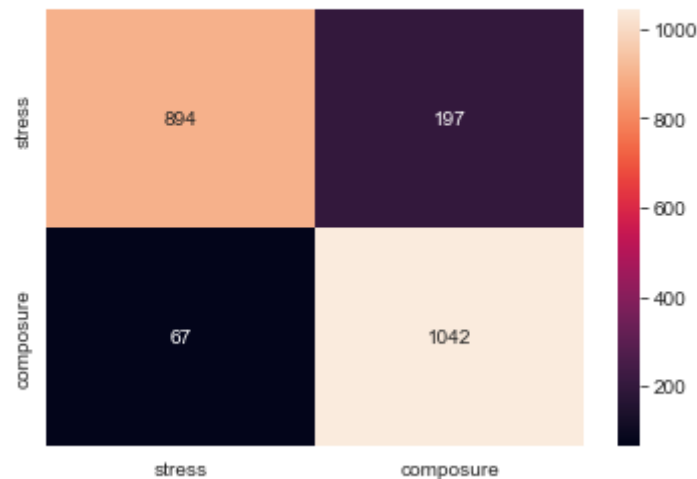


Fig 2. Number of Employees Under Stress

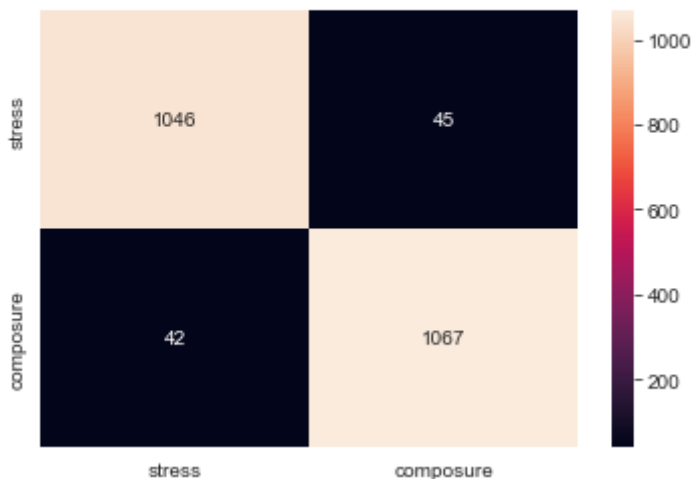




```
*****
the accuracy on testing data 0.88
the accuracy on training data 0.987840909090909
*****
```



```
the accuracy on testing data 0.9604545454545454
the accuracy on training data 1.0
*****
```



VII. CONCLUSION AND FUTURE ENHANCEMENT

The accuracy of existing mental health prediction methods is low because the relationship between the feature variables and the prediction results is non-linear and the prediction dataset contains a lot of irrelevant and redundant features. At the same time, current mental health prediction methods cannot estimate the extent to which the feature variables are important to the prediction results. First introduced to deal with the non-linear problem between prediction results and feature variables, which improves the accuracy of mental health prediction. To calculate the influence weight, which assesses the extent to which feature variables contribute to mental health.

The proposed work can be extended further. Let us consider the case of a university. After an interval of few months, the stress prediction should be taken in college to know how well their employees are. This survey will help to us know that the employees are not facing any high stress levels and studying properly. The future work would be to replicate this project in various universities for the beneficiary of the employees and their mental well-being. The solution module could be further integrated with a Chatbot. A much more vast and detailed dataset can be collected and worked

with other Machine learning techniques. The parameters for stress can further be classified into numerous sub parameters.

VIII. ACKNOWLEDGEMENT

I would like to express my sincere thanks to S.Samsudeen Shaffi for her valuable guidance and support in completing this article. I would also like to express my gratitude towards our College.

REFERENCES

- [1] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from 'The Economic Times'
- [2] Lee MH, Seo MK. Community integration of persons with mental disorders compared with the general population. *Int J Environ Res Public Health*. (2020)
- [3] Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS medicine*, 2(10), e267.
- [4] 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
- [5] 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.
- [6] 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), 2825-2830.
- [7]. Derogatis LR, Melisaratos N. The brief symptom inventory-an introductory report. *Psychol Med*. (1983) 13:595-605. doi: 10.1017/S0033291700048017
- [8]. Hathaway SR, McKinley JC. A multiphasic personality schedule (Minnesota): I. construction of the schedule. *J Psychol*. (1940) 10:249-54. doi: 10.1080/00223980.1940.9917000
- [9]. Yang X, Jin M, Zheng L. The prediction model of college students' life stressor on mental health. *China J Health Psychol*. (2018)
- [10]. Ko HC. A sustainable approach to mental health education: an empirical study using Zhuangzi's self-adaptation. *Sustainability*. (2019) *e Detector System and Automatic Toll Collection System, Transportation Research Record Journal of the Transportation Research Board* 2256:51-59, 2012.
- [15]P. Vidnerová, RBF-Keras: an RBF Layer for Keras Library.,2019, Available at https://github.com/PetraVidnerova/rbf_keras



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details