Stress Detection in IT Professionals using Machine Learning

Vishwagna Praharshitha[1], Dr. Gifta Jerith G[2], Dr Thayabba Katoon[3]

[1]Student, MTech in ML&DL, Malla Reddy University
[2]Assistant Professor-AIML, Malla Reddy University
[3]HoD-AIML, Malla Reddy University

---***---

Abstract - In today's fast-paced technology landscape, stress management is becoming increasingly important, especially among IT professionals. The work environment in the IT industry is often characterized by long hours, tight deadlines, and high expectations, which can lead to elevated stress levels. Unchecked stress not only impacts the health and well-being of professionals but also affects productivity and job satisfaction. This study aims to predict the stress levels of IT professionals using machine learning techniques, thereby aiding in proactive stress management. We utilize a range of features indicative of work stress, including Heart Rate, Skin Conductivity, Hours Worked, Number of Emails Sent, and Meetings Attended. These features provide a comprehensive view of both the physiological and work-related factors that contribute to stress. The application of machine learning in this context serves as an innovative approach to an increasingly pertinent issue. By leveraging the power of data analytics, this model aims to provide actionable insights for both individuals and organizations. Individuals can use these predictions for self-monitoring and early intervention, while organizations can utilize them to identify high-stress environments or roles, thereby allocating resources or interventions more effectively. Our preliminary results indicate a strong correlation between the chosen features and stress levels, demonstrating the viability of using machine learning for stress prediction in IT professionals. This study stands as a crucial step towards a more data-driven approach to mental health and well-being in the workplace.

***Key Words*: :** Randomforest, adaboost, extratree, Stress Detection, IT Professionals, Machine Learning, Stress Analysis, Mental Health

1. INTRODUCTION

In today's dynamic IT industry, stress is a prevalent issue impacting professionals' mental health and productivity. High job demands, extended working hours, and continuous technological advancements contribute to elevated stress levels. This study aims to leverage machine learning techniques to detect stress levels among IT professionals. Early identification of stress can lead to timely interventions, enhancing the overall well-being and efficiency of the workforce.

1.1 Motivation

Motivation for stress detection in IT professionals using machine learning stems from the pressing need to address the mental health and well-being of individuals in this high-pressure industry. With the relentless pace, tight deadlines, and complex problem-solving, IT workers often face stress-related challenges. Machine learning offers the potential to proactively identify and mitigate stress by analyzing physiological data, sentiment patterns, and communication cues. By doing so, it aims to improve the overall quality of life for IT professionals, enhance productivity, and promote a healthier work environment.

1.2 Problem Statement

The fast-paced and demanding nature of the IT industry often leads to elevated stress levels among professionals, adversely affecting their mental well-being, productivity, and overall job satisfaction. Traditional methods for stress detection, such as self-reports or psychological assessments, may not be timely or accurate enough; hence, there is an urgent need for a real-time, data-driven approach using machine learning to predict and manage stress levels in IT professionals.

1.3 Objective of the Project

The primary objective of this project is to develop a machine learning model that can accurately predict the stress levels of IT professionals based on physiological and work-related features such as heart rate, skin conductivity, hours worked, emails sent, and meetings attended. By doing so, the project aims to provide actionable insights for both individuals and organizations to proactively manage stress, thereby improving mental well-being and overall work performance.

1.4 Scope

The scope of this project extends to the development and validation of a machine learning model that can predict stress levels in IT professionals based on various physiological and work-related parameters. By leveraging real-time data, the model aims to facilitate proactive stress management strategies, both at an individual and organizational level.

1.5 Project Introduction

In a pandemic, the people's outlook of health- care constraints and lifestyles are completely switched. Since then, covid 19 has spread a lot, causing global disturbances. The administration of educational institutions has closed across the globe to prune the growth of the disease and in welfare of all people. Considering all these circumstances the people around all countries were affected by entities like food availability and medical facilities. Many surveys were conducted to study the person's stress level based on the stress constraints like physiological conditions. A person can be stressed out in scenarios like worrying about losing their employment, family health conditions and about the grades in examinations. Because of working for long period of time, limited time to complete task [1]. These kinds of stressful scenarios increase stress levels which affects the increase of heart and muscles related issues. Generally, anxiety and stress are very much common among all the students with a variation of degree. So, by observing each and every student it would be a huge task to go through their profiles. This problem makes us create a new model for predicting stress automatically. For each student who is undergoing various psychological parameters of stress and proposes a solution for that. So for this to be done, some Machine Learning algorithms and Data Science techniques are used. Maintaining track records of each student's stress levels and studying them makes us understand the degrees of stress of the students in organization. Students are categorized into 2 sub levels in regards with the stress percentage they face: i.e., over-stressed or under- stressed. And according to that, the range of stress is highlighted based on the levels. Based on this percentage, the authorities give advice to the students. As a result, we create a model for unlabeled data and untrained data that will determine the stress level of students using different Machine Learning and data science techniques.

2. LITERATURE REVIEW

[1] Bakker, Mykola Pechenizkiy, Natalia Sidorova, What's your current stress level? Detection of stress patterns from GSR sensor data, 2011 11th IEEE International Conference on Data Mining Workshops, Department of Computer Science Eindhoven University of Technology, 978-0-7695-4409-0/11 $26.00 © 2011 IEEE DOI 10.1109/ICDMW.2011.17

In this paper, we propose a hybrid approach for music recommendation. Firstly, we describe an approach for creating music recommendations based on user-supplied tags that are augmented with a hierarchical structure extracted for top level genres from Dbpedia. In this structure, each genre is represented by its stylistic origins, typical instruments, derivative forms, sub genres and fusion genres. We use this well-organized structure in dimensionality reduction in user and item profiling. We compare two recommenders; one using our method and the other using Latent Semantic Analysis (LSA) in dimensionality reduction. The recommender using our approach outperforms the other. In addition to different dimensionality reduction methods, we evaluate the recommenders with different user profiling methods. Moreover, our approach collects personal interests (favorite movies and television series) from the Facebook profiles. These user profiles are then used to find the similarity between users. At the end, items belonging to the most similar users' profiles and having a high score against users' profiles are recommended. Thus, we have focused on a hybrid system using tag-based contextual information of music tracks and user interests acquired from Facebook profiles. Initial results are promising such that using similarities of users affects the recommendation positively

[2] UnsuKristina P. Sinaga and Miin-Shen Yang, pervised K-Means Clustering Algorithm , 2018 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS).

The k-means algorithm is generally the most known and used clustering method. There are various extensions of k-means to be proposed in the literature. Although it is an unsupervised learning to clustering in pattern recognition and machine learning, the k-means algorithm and its extensions are always influenced by initializations with a necessary number of clusters a priori. That is, the k-means algorithm is not exactly an unsupervised clustering method. In this paper, we construct an unsupervised learning schema for the k-means algorithm so that it is free of initializations without parameter selection and can also simultaneously find an optimal number of clusters. That is, we propose a novel unsupervised k-means (U-k-means) clustering algorithm with automatically finding an optimal number of clusters without giving any initialization and parameter selection. The computational complexity of the proposed U-k-means clustering algorithm is also analyzed. Comparisons between the proposed U-k-means and other existing methods are made. Experimental results and comparisons actually demonstrate these good aspects of the proposed U-k-means clustering algorithm.

3. SYSTEM ANALYSIS

3.1 Existing System

In existing methods has been employed to segment IT professionals into different stress categories based on similar behavioral and physiological traits simplify the dataset, making it easier to apply subsequent machine learning algorithms like Logistic Regression. Logistic Regression has been widely adopted as a predictive model for categorizing individuals into stressed and non-stressed groups based on a combination of physiological and work-related factors.

3.2 Proposed System

In the proposed system, we leverage ensemble machine learning techniques like Random Forest, AdaBoost, and Extra Trees to predict stress levels in IT professionals. These advanced models are designed to capture the intricate relationships between various physiological and work-related features, offering a more nuanced understanding of stress factors. By employing ensemble methods, the system aims to achieve higher predictive accuracy and robustness compared to traditional methods.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

3.4 Work Flow of Proposed System

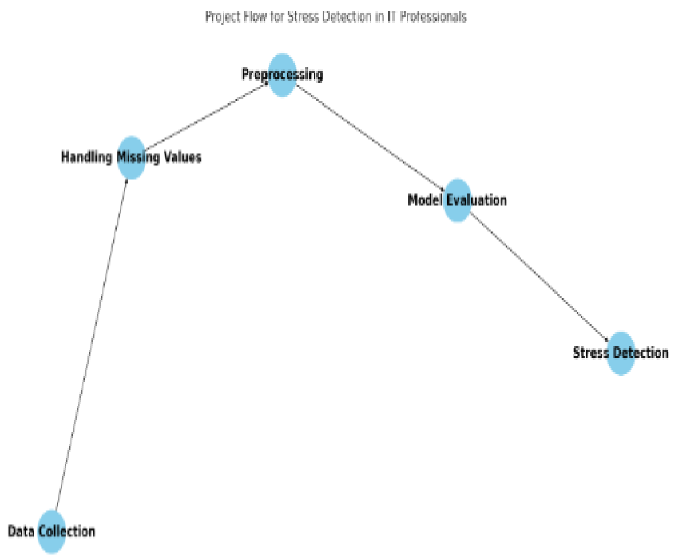

Fig -1: Workflow

4. SYSTEM DESIGN

4.1 Introduction of Input Design

In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

Therefore, the quality of system input determines the quality of system output. Well-designed input forms and screens have following properties −

- It should serve specific purpose effectively such as storing, recording, and retrieving the information.
- It ensures proper completion with accuracy.
- It should be easy to fill and straightforward.
- It should focus on user's attention, consistency, and simplicity.
- All these objectives are obtained using the knowledge of basic design principles regarding −

- What are the inputs needed for the system?
- How end users respond to different elements of forms and screens.

Objectives for Input Design:

The objectives of input design are −

- To design data entry and input procedures
- To reduce input volume
- To design source documents for data capture or devise other data capture methods.
- To design input data records, data entry screens, user interface screens, etc.
- To use validation checks and develop effective input controls.

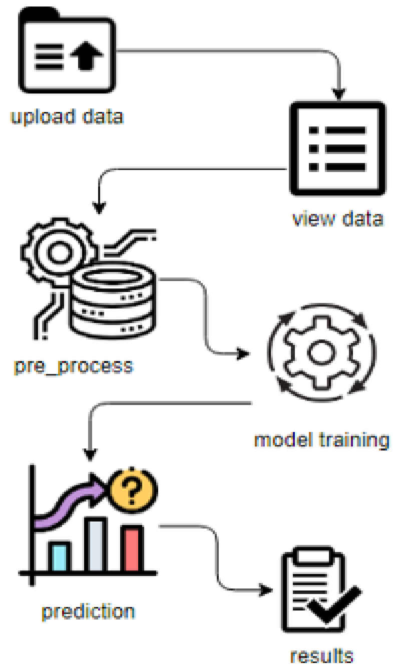

Fig -2: Architecture

Output Design:

The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

Objectives of Output Design:

The objectives of input design are:

- To develop output design that serves the intended purpose and eliminates the production of unwanted output.
- To develop the output design that meets the end user's requirements.
- To deliver the appropriate quantity of output.
- To form the output in appropriate format and direct it to the right person.
- To make the output available on time for making good decisions.

4.2 Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

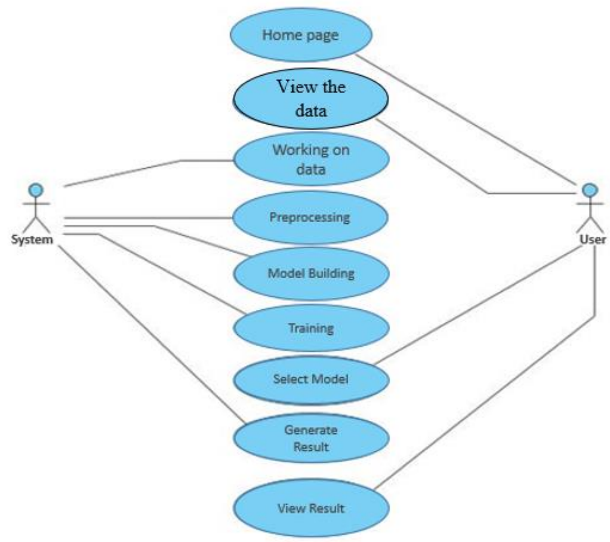

Fig -3: Use Case Diagram

5. IMPLEMENTATION

Algorithms Used.

Random Forest:

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome. A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.
A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further.
The nodes in the decision tree represent attributes that are used for predicting the outcome. Decision nodes provide a link to the leaves. The following diagram shows the three types of nodes in a decision tree.

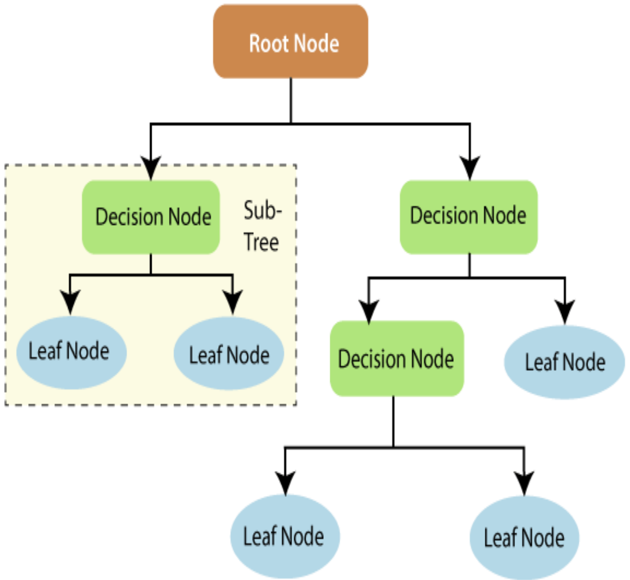

Fig -4: Decision Tree

The Information theory can provide more information on how decision trees work. Entropy and information gain are the building blocks of decision trees. An overview of these fundamental concepts will improve our understanding of how decision trees are built.
Entropy is a metric for calculating uncertainty. Information gain is a measure of how uncertainty in the target variable is reduced, given a set of independent variables. The information gain concept involves using independent variables (features) to gain information about a target variable (class). The entropy of the target variable (Y) and the conditional entropy of Y (given X) are used to estimate the information gain. In this case, the conditional entropy is subtracted from the entropy of Y.

Information gain is used in the training of decision trees. It helps in reducing uncertainty in these trees. A high information gain means that a high degree of uncertainty (information entropy) has been removed. Entropy and information gain are important in splitting branches, which is an important activity in the construction of decision trees.
Let's take a simple example of how a decision tree works. Suppose we want to predict if a customer will purchase a mobile phone or not. The features of the phone form the basis of his decision. This analysis can be presented in a decision tree diagram. The root node and decision nodes of the decision represent the features of the phone mentioned above. The leaf node represents the final output, either *buying* or *not buying*. The main features that determine the choice include the price, internal storage, and Random Access Memory (RAM). The decision tree will appear as follows.

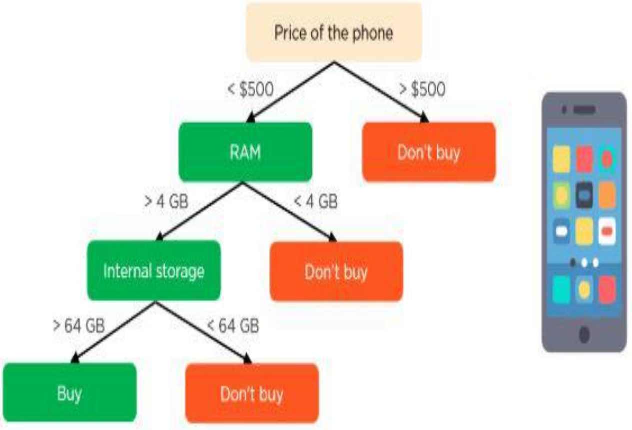

Fig -5: Decision To buy a Mobile

Applying decision trees in random forest
The main difference between the decision tree algorithm and the random forest algorithm is that establishing root nodes and segregating nodes is done randomly in the latter. The random forest employs the bagging method to generate the required prediction.

Bagging involves using different samples of data (training data) rather than just one sample. A training dataset comprises observations and features that are used for making predictions. The decision trees produce different outputs, depending on the training data fed to the random forest algorithm. These outputs will be ranked, and the highest will be selected as the final output. Our first example can still be used to explain how random forests work. Instead of having a single decision tree, the random forest will have many decision trees. Let's assume we have only four decision trees. In this case, the training data comprising the phone's observations and features will be divided into four root nodes. The root nodes could represent four features that could influence the customer's choice (price, internal storage, camera, and RAM). The random forest will split the nodes by selecting features randomly. The final prediction will be selected based on the outcome of the four trees.
The outcome chosen by most decision trees will be the final choice. If three trees predict *buying*, and one tree predicts *not buying*, then the final prediction will be *buying*. In this case, it's predicted that the customer will buy the phone.

6. RESULTS

OUTPUT SCREEN SHOTS WITH DESCRIPTION.

Home Page:

Here user view the home page of Stress Detection In It Professionals Using Machine Learning web application.

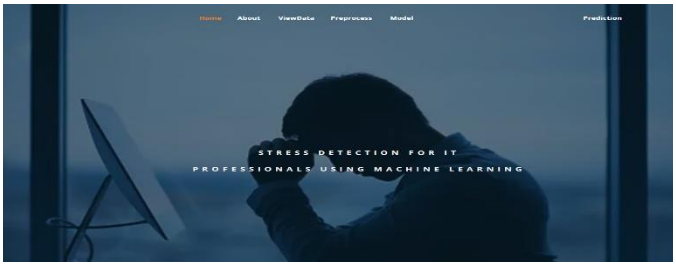

Fig-6: Home Page

View Data:

In the Viewdata page, users can view the stress dataset.

Heart_Rate	Skin_Conductivity	Hours_Worked	Stress_Level	Emails_Sent	Meetings_Attended
87.0	5.56	5.0	28.0	31.0	6.0
74.0	5.89	5.0	25.0	42.0	3.0
79.0	4.58	9.0	26.0	28.0	4.0
92.0	5.1	7.0	30.0	37.0	3.0
88.0	5.23	8.0	29.0	35.0	6.0
60.0	5.2	7.0	21.0	31.0	6.0
79.0	5.54	7.0	26.0	25.0	6.0
68.0	3.18	8.0	22.0	30.0	1.0
68.0	4.95	10.0	23.0	30.0	2.0

Table-1: View Data Page

Pre-process:

Here we can pre-process and split our data into train and test.

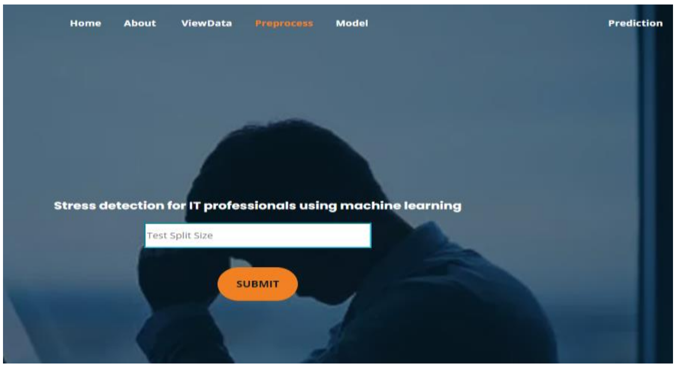

Fig-7: Preprocessing Page

Model:

Here we train our data with different ML algorithms.

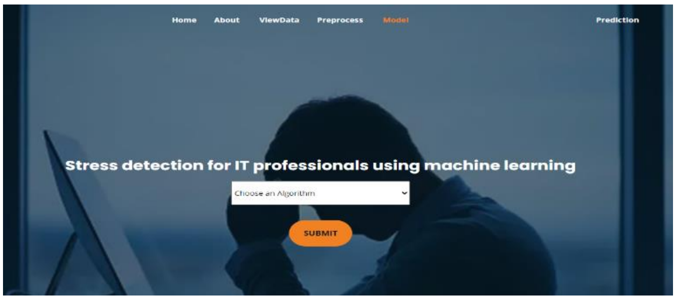

Fig-8: Model Page

Prediction:

This page show the result of the user given input data.

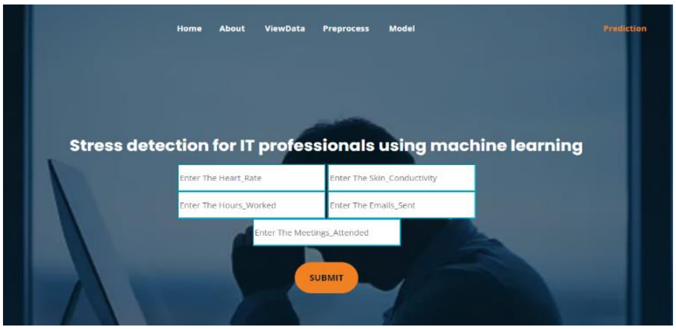

Fig-9: Prediction Page

7. CONCLUSIONS

The novelty in the "Stress Detection in IT Professionals Using Machine Learning" lies in the comprehensive approach taken to address this critical issue. While previous studies primarily focused on physiological data or sentiment analysis alone, our research combines both domains. By integrating physiological indicators like heart rate variability and skin conductance with natural language processing techniques for sentiment analysis, we create a more holistic stress detection model. Additionally, the utilization of a diverse ensemble of regression algorithms, including RandomForestRegressor, AdaBoostRegressor, and ExtraTreeRegressor, adds robustness to the model's predictions. The use of RandomForestRegressor for final stress level prediction enhances accuracy and reliability, making our approach a pioneering solution in the field of stress detection for IT professionals.

REFERENCES

[1] Bakker, Mykola Pechenizkiy, Natalia Sidorova , What's your current stress level? Detection of stress patterns from GSR sensor data, 2011 11th IEEE International Conference on Data Mining Workshops, Department of Computer Science Eindhoven University of Technology, 978-0-7695-4409-0/11 $26.00 © 2011 IEEE DOI 10.1109/ICDMW.2011.17

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references).

[9] UnsuKristina P. Sinaga and Miin-Shen Yang, pervised K-Means Clustering Algorithm , 2018 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS).

[10] Jirong Gu, Jieming Zhou Xianwei Chen Xianwei Chen, An Enhancement of K-means Clustering Algorithm, 2009 International Conference on Business Intelligence and Financial Engineering.

[11] V Sasikala; T Rajeswari; Sk Naseema Begum; Ch Divya Sri; M Sravya, Stress Detection from Sensor Data using Machine Learning Algorithms,16-18 March 2022-21688590.

[12] Prithwijit Mukherjee,Dr.Anisha Halder Roy,Detection of Stress in Human Brain ,2019 Second International Conference on Advanced Computational and Communication Paradigmsbv