International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Stress Detection in IT Employees using Machine Learning

Naralasetti Prudhvi Kalyan[1], Nemala Ravi Kumar[2], Palaparthi Niranjan Reddy[3], Osetti Naveen Sai Ram[4]

[1234]*Computer Science and Engineering, Raghu Engineering College.*

---***---

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*:** Random Forest , AdaBoost Classifier, Extra Tree Classifier, Decision Tree, Stacking etc.

1. INTRODUCTION

In a pandemic, the people's outlook of health- care constraints and lifestyles is completely switched. Since then, covid 19 had 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 the 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 technique.

2. LITERATURE SURVEY

[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] Pramod Bobade, Vani M. Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data, (ICIRCA-2020).

Stress is a common part of everyday life that most people have to deal with on various occasions. However, having long-term stress, or a high degree of stress, will hinder our safety and disrupt our normal lives. Detecting mental stress earlier can prevent many health problems associated with stress. When a person gets stressed, there are notable shifts in various bio-signals like thermal, electrical, impedance, acoustic, optical, etc., by using such bio-signals stress levels can be identified. This paper proposes different machine learning and deep learning techniques for stress detection on individuals using multimodal dataset recorded from wearable physiological and motion sensors, which can prevent a person from various stress-related health problems. Data of sensor modalities like three-axis acceleration (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG) and electrodermal activity (EDA) are for three physiological conditions - amusement, neutral and stress states, are taken from WESAD dataset. The accuracies for three-class (amusement vs. baseline vs. stress) and binary (stress vs. non-stress) classifications were evaluated and compared by using machine learning techniques like K-Nearest Neighbour, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost and Kernel Support Vector Machine. Besides, simple feed forward deep learning artificial neural network is introduced for these three-class and binary classifications. During the study, by using machine learning techniques, accuracies of up to 81.65% and 93.20% are achieved for three-class and binary classification problems respectively, and by using deep learning, the achieved accuracy is up to 84.32% and 95.21% respectively.

[4] A. Kene and S. Thakare, "Mental Stress Level Prediction and Classification based on Machine Learning," 2021 Smart Technologies, Communication and Robotics (STCR), 2021, pp. 1-7, doi: 10.1109/STCR51658.2021.9588803.

Stress is becoming an important factor in a person's life today. According to the World Health Organization, stress is a type of mental illness that affects the health of citizens. There is no one in the world who does not suffer from stress or depression. Everyone gets some amount of stress. Stress is a major symptom for mental health. Stress affects every aspect of a person's life such as emotions, thoughts, and behaviors. This paper presented the study on previous research on stress detection based on machine learning algorithms. presented a stress level classification framework using the PhysioBank dataset to analyze the stress level. The statistical analysis was used for feature selection and extraction, and it found that stress level classification has been successfully implemented based on the proposed gradient boost algorithm. The evaluated results showed that the proposed model achieved accuracy (83.33%), specificity (75%), Sensitivity (75%), Positive Predictive Value (90%), Negative Predictive Value (90%), Error Rate(16.66%), F1_Score (83.33%), Recall (75%). The proposed gradient boost algorithm performed well as compared to other machine learning algorithms, namely, KNN, Random Forest and, support vector machine. The proposed model was shown to be effective in classifying stress level prediction.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

[5] 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.

Today, psychological stress is a huge problem. Mental stress, such as anxiety, overthinking, melancholy, and emotional imbalance, was common during pandemics. Pandemics are mostly caused by a mismatch between labour market demands and slowing economic development. The paper's main goal is to use machine learning algorithms to estimating the levels of stress and that can be detected by grouping together many measurements such as pulse rate, body temperature, heart rate, and systolic blood oxygen saturation (spo2). When a person is under stress, their bio-signals, such as thermally, electrical, impedance, acoustic, and optical, change noticeably. These bio-signals can be used to measure stress levels. Accelerometer, body temperature, respiration, blood volume pulse electrocardiogram, (BVP), electro dermal activity, and other sensor modalities (EDA). In Machine Learning Classification methods such as Kernel Support Vector Machine, K-Nearest Neighbour, AdaBoost, Random Forest and Decision Tree methods used to evaluate and compare the classifications. The Random Forest model beat the other approaches with F1-scores of 93.77 and 70.03 for classification model and three-class classification, respectively.

[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].

The depth-profiles of amorphous TbFeCo films sputtered onto polycarbonate substrate were studied by X-ray photoelectron spectroscopy. Oxidized metals, oxides and hydroxides for example, and adsorbed impurities were found to exist mainly in the vicinity of the film surface and film/ substrate interface.

Proposed System

In the proposed system, we leverage ensemble machine learning techniques like Random Forest , AdaBoost Classifier, Extra Tree Classifier, Decision Tree, Stacking 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.

Advantages

1. **Early Intervention**: Machine learning models can analyze a wide range of physiological and behavioral data, allowing for the early detection of stress indicators. This enables IT professionals and their employers to intervene promptly and implement stress-reduction strategies before the stress escalates to severe levels, potentially preventing burnout and health issues.

2. **Objective Assessment**: Machine learning provides an objective and data-driven approach to stress detection. It removes bias and subjectivity from the assessment process, ensuring that stress levels are evaluated consistently and accurately based on empirical evidence rather than human judgment.

3. **Real-Time Monitoring**: Machine learning algorithms can continuously monitor IT professionals' stress levels in real time. This dynamic monitoring allows for immediate feedback and alerts when stress indicators surpass predefined thresholds, enabling rapid responses and adjustments to workloads or environments.

4. **Personalized Support**: Machine learning models can be tailored to individual IT professionals, taking into account their unique stress profiles. This personalization enables the delivery of targeted interventions and support strategies, such as suggesting stress-relief activities or providing resources specific to each person's needs.

5. **Long-Term Trends Analysis**: Machine learning can also analyze historical data to identify long-term stress trends within an organization. This data-driven insight helps employers make informed decisions about workload management, resource allocation, and workplace policies to reduce stress and improve overall job satisfaction among IT professionals.

Work Flow of Proposed system

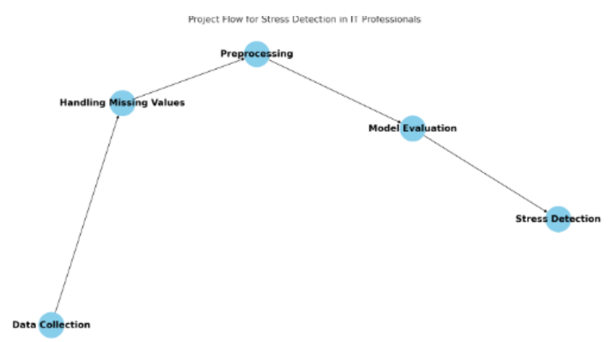


Figure 1: Work flow Diagram

3. ALOGRITHM

3.1 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).

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

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.

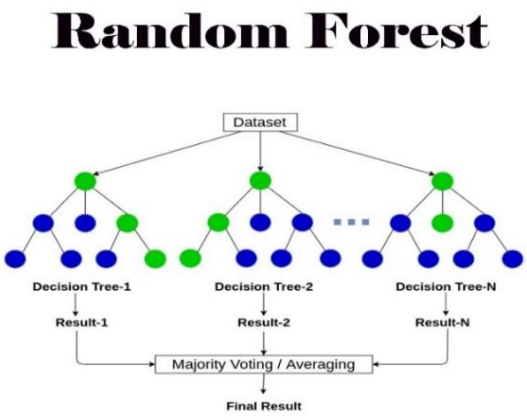


Figure 2: Random Forest

3.2 Extra Tree Classifier:

Extremely Randomized Trees Classifier (Extra Trees Classifier) is a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output it's classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest.

Each Decision Tree in the Extra Trees Forest is constructed from the original training sample. Then, at each test node, Each tree is provided with a random sample of k features from the feature-set from which each decision tree must select the best feature to split the data based on some mathematical criteria (typically the Gini Index). This random sample of features leads to the creation of multiple de-correlated decision trees.To perform feature selection using the above forest structure, during the construction of the forest, for each feature, the normalized total reduction in the mathematical criteria used in the decision of feature of split (Gini Index if the Gini Index is used in the construction of the forest) is computed. This value is called the Gini Importance of the feature. To perform feature selection, each feature is ordered in descending order according to the Gini Importance of each feature and the user selects the top k features according to his/her choice

3.3 ADABOOST

AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. Boosting is used to reduce bias as well as variance for supervised learning. It works on the principle of learners growing sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. The AdaBoost algorithm works on the same

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

principle as boosting with a slight difference. Let's discuss this difference in detail.

First, let us discuss how boosting works. It makes 'n' number of decision trees during the data training period. As the first decision tree/model is made, the incorrectly classified record in the first model is given priority. Only these records are sent as input for the second model. The process goes on until we specify a number of base learners we want to create. Remember, repetition of records is allowed with all boosting techniques.

This figure shows how the first model is made and errors from the first model are noted by the algorithm. The record which is incorrectly classified is used as input for the next model. This process is repeated until the specified condition is met. As you can see in the figure, there are 'n' number of models made by taking the errors from the previous model. This is how boosting works. The models 1,2, 3,…, N are individual models that can be known as decision trees. All types of boosting models work on the same principle.

Since we now know the boosting principle, it will be easy to understand the AdaBoost algorithm. Let's dive into AdaBoost's working. When the random forest is used, the algorithm makes an 'n' number of trees. It makes proper trees that consist of a start node with several leaf nodes. Some trees might be bigger than others, but there is no fixed depth in a random forest. With AdaBoost, however, the algorithm only makes a node with two leaves, known as Stump.

The figure here represents the stump. It can be seen clearly that it has only one node with two leaves. These stumps are weak learners and boosting techniques prefer this. The order of stumps is very important in AdaBoost. The error of the first stump influences how other stumps are made. Let's understand this with an example.

Here's a sample dataset consisting of only three features where the output is in categorical form. The image shows the actual representation of the dataset. As the output is in binary/categorical form, it becomes a classification problem. In real life, the dataset can have any number of records and features in it. Let us consider 5 datasets for explanation purposes. The output is in categorical form, here in the form of Yes or No. All these records will be assigned a sample weight. The formula used for this is 'W=1/N' where N is the number of records. In this dataset, there are only 5 records, so the sample weight becomes 1/5 initially. Every record gets the same weight. In this case, it's 1/5.

Learn AdaBoost Model from Data

Ada Boosting is best used to boost the performance of decision trees and this is based on binary classification problems.

4. RESULTS AND DISCUSSON SCREEN SHOT

Home Page:

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

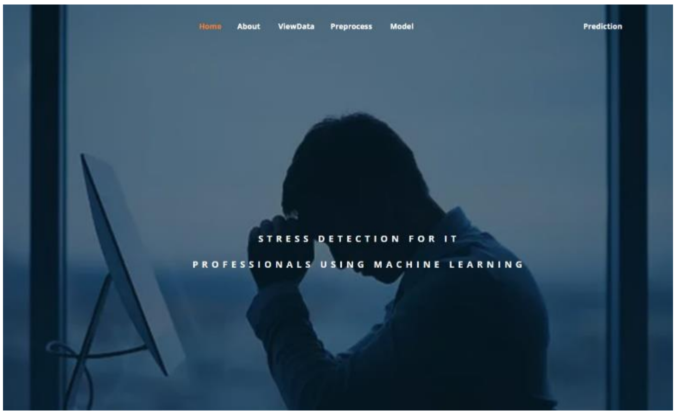


Figure 3: Home page

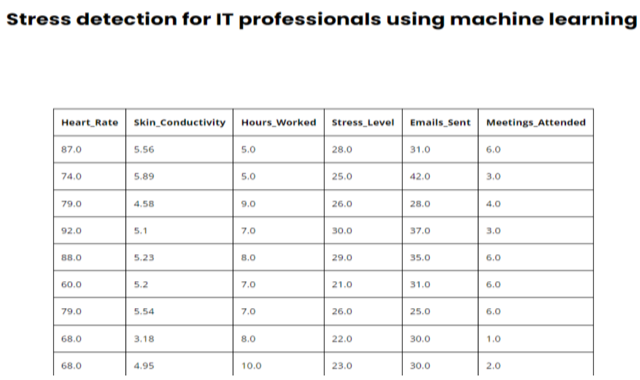
ABOUT:

Here we can read about our project.

Figure 4: About page

View Data:

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



Stress detection for IT professionals using machine learning

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

Figure 5: ViewData

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Pre-process:

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

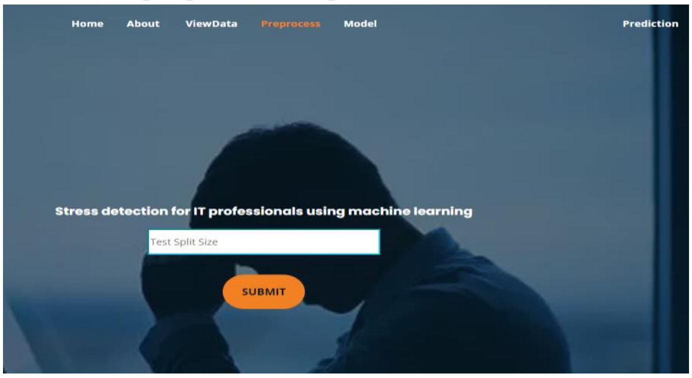


Figure 6: Preprocess

Prediction:

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

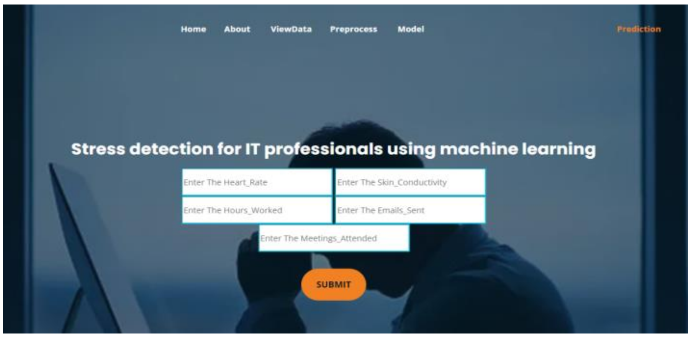


Figure 7: Prediction

5. CONCLUSION

The conclusion 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 Paradigms (ICACCP)