Grenze International Journal of Engineering and Technology, June Issue

Analyzing Stress Levels in IT Professionals using Machine Learning

Jehosh Joseph[1] and Kavitha T[2]
[1-2]Division of Computer Science Engineering, Karunya Institute of Technology and Science, Karunya Nagar, Coimbatore, Tamil Nadu,641114
Email: jehoshjoseph@karunya.edu.in, kavithat@karunya.edu

***Abstract*— Stress management is a growing concern in today's high-pressure tech industry, particularly for those working in information technology. Long hours, strict deadlines, and high expectations are common in the information technology workplace, all of which may contribute to an already stressful situation. Professionals' health, happiness, and productivity are all negatively impacted by unmanaged stress. In order to help with proactive stress management, this project intends to use machine learning methods to forecast the stress levels of IT workers. Heart rate, skin conductivity, hours worked, number of emails sent, and meetings attended are some of the characteristics that we use to identify job stress. Together, these characteristics illuminate the many ways in which stress manifests in the workplace, both physiologically and otherwise. Here, machine learning provides a fresh perspective on a problem that is becoming more pressing. This model's goal is to help people and businesses by using data analytics to deliver them actionable insights. Both individuals and organizations may benefit from these forecasts. The former can use them for self-monitoring and early intervention, while the latter can use them to understand which positions or settings cause the most stress and allocate resources accordingly. Our first findings show that the selected variables are significantly correlated with stress levels, suggesting that machine learning might be useful for predicting stress in IT workers. When it comes to workplace mental health and wellness, this report is a major step in the right direction.**

***Index Terms*— Random Forest , AdaBoost Classifier, Extra Tree Classifier, Decision Tree, Stacking etc.**

I. INTRODUCTION

The public's perspective on health-care limitations and lifestyle changes dramatically during a pandemic. Since then, COVID-19 has spread extensively, wreaking havoc on a worldwide scale. To curb the spread of the illness and ensure everyone's well-being, school districts throughout the world have shut down their operations. Entities like food availability and medical facilities impacted people worldwide due to all these factors. The stress restrictions, such as physiological factors, formed the basis for a large-scale survey investigation of people's stress levels. Many things might cause stress in a person's life, including concerns about their job security, the health of their loved ones, and the results of upcoming exams. Working for lengthy periods of time means there isn't enough time to get everything done [1]. Problems with the heart and muscles are more common in people whose stress levels are high due to situations like these. Students of all academic levels often struggle with worry and stress.
Therefore, reviewing each student's profile would be an enormous undertaking if done manually. We develop a

Grenze ID: 01.GIJET.10.2.278

novel model for automated stress prediction in response to this issue. Finds a way to alleviate the tension that each pupil is experiencing on a psychological level. This requires the use of Data Science methods and algorithms from the field of Machine Learning. We can learn about the students' stress levels in the organization by keeping records of their stress levels and analyzing them.Based on the proportion of stress that students experience, two sublevels are identified: over-stressed and under-stressed. Additionally, the stress levels are shown according to that. The authorities provide guidance to the pupils based on this proportion. Consequently, we use several data science and Machine Learning approaches to build a model that can identify students' stress levels from unlabeled and untrained data.

Stress levels among IT workers are known to be high due to the demanding and fast-paced nature of the sector. This has negative effects on workers' mental health, productivity, and happiness on the job. There is an immediate need for a data-driven, real-time strategy that utilizes machine learning to forecast and control stress levels in information technology workers as conventional ways of stress detection, including self-reports or psychological evaluations, could not be quick enough or accurate enough.

Using physiological and job-related characteristics including heart rate, skin conductivity, hours worked, emails sent, and meetings attended, this research aims to build a machine learning model that can effectively predict the stress levels of IT workers. Through this approach, the initiative seeks to provide practical advice for people and businesses to take charge of stress management, leading to better mental health and productivity on the job.

II. RELATED WORKS

[1] We present a hybrid music recommendation method in this research. First, we propose a method for generating music suggestions using user-supplied tags and Dbpedia's top-level genre hierarchies. This structure shows each genre's stylistic roots, characteristic instruments, derivative forms, subgenres, and fusion genres. We reduce dimensionality in user and item profiling using this well-organized framework. We compare two recommenders: one using our approach and one utilizing LSA in dimensionality reduction. The recommender utilizing our method is better. We test recommenders with various user profiling methodologies and dimensionality reduction strategies. We also gather Facebook profiles' personal preferences (favorite movies and TV shows). User similarity is determined using these profiles. Finally, goods from the most comparable people' profiles and with high scores are suggested. Thus, we developed a hybrid system employing tag-based contextual music track information and Facebook user interests. Initial findings suggest that user commonalities improve recommendation.

[2] K-means is the most used clustering algorithm. Extensions of k-means are proposed in the literature. Although k-means is an unsupervised clustering method used in machine learning and pattern recognition, initializations with a specific number of clusters typically affect how far it may be extended. K-means is not an algorithm for unsupervised clustering. In this study, we present an unsupervised learning schema that determines the ideal number of clusters for the k-means method, which is unable to initialize without parameter selection. We provide an original unsupervised k-means (U-k-means) clustering method that, without setup or parameter selection, automatically determines the optimal number of groups. The computational cost of the proposed U-k-means clustering method is also examined. The proposed U-k-means is contrasted with alternative methods. The advantages of the U-k-means clustering algorithm are demonstrated via comparisons and experimental results.

[3] Most people experience stress on a daily basis. High levels of stress or prolonged exposure jeopardize our daily life and safety. Many health concerns may be avoided with early recognition of mental stress. Biosignals such as thermal, electrical, impedance, acoustic, optical, and others can be used to detect stress levels in humans. In order to prevent stress-related health problems, this study offers deep learning and machine learning techniques for stress detection on individuals using multimodal datasets from wearable motion and physiological sensors. Three-axis acceleration (ACC), blood volume pulse (BVP), electrocardiogram (ECG), body temperature (TEMP), respiration (RESP), electromyogram (EMG), and electrodermal activity (EDA) for neutral, stress, and amusement states are among the data in the WESAD dataset. The accuracy of binary (stress vs. non-stress) and three-class (amusement vs. baseline vs. stress) classifications was compared using machine learning techniques such as K-Nearest Neighbour, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost, and Kernel Support Vector Machine. Furthermore, for binary and three-class classifications, a simple feed-forward deep learning artificial neural network is employed. For three-class and binary classification problems, machine learning achieved 81.65% and 93.20% accuracy in the study, whereas deep learning achieved 84.32% and 95.21%.

[4] Stress has a greater impact. Stress is categorized by the World Health Organization as a mental illness that is harmful to individuals. Nobody is exempt from anxiety or hopelessness. Stress affects every human being. Stress is a serious problem for mental health. Stress has an impact on every person's thoughts, feelings, and behaviors. In this paper, prior machine learning research on stress detection was analyzed. suggested a system for classifying

stress levels based on PhysioBank. For the purpose of selecting and extracting features, statistical analysis was used, and the gradient boost algorithm successfully identified stress levels. The model that has been proposed has the following characteristics: accuracy (83.33%), specificity (75%), sensitivity (75%), error rate (16.66%), recall (75%), positive and negative predictive values (90%), and F1_Score (83.33%). The gradient boost technique fared better than support vector machines, random forests, and KNN. The model worked well for predicting stress levels. [5] Psychological stress is a big problem these days. Pandemics were associated with a high prevalence of depression, anxiety, overthinking, and emotional imbalance. Most pandemics are caused by slowing economic development and labor market demands. The article groups parameters such as pulse rate, body temperature, heart rate, and systolic blood oxygen saturation and uses machine learning techniques to assess stress levels. Bio-signals that are thermal, electrical, impedance, acoustic, and optical are altered by stress. Biosignals can be used to quantify stress. sensors such as the body temperature, respiration, BVP, EDA, and accelerometer. Classifications are evaluated and compared using machine learning methods such as
Kernel Support Vector Machine, K-Nearest Neighbor, AdaBoost, Random Forest, and Decision Tree. With classification model and three-class classification F1-scores of 93.77 and 70.03, respectively, the Random Forest model fared better than the others.

IV. PROPOSED SYSTEM

The proposed system utilizes ensemble machine learning techniques including Random Forest, AdaBoost Classifier, and Extra Tree Classifier to predict stress levels in IT professionals. Through a user-friendly interface, users can upload datasets, input parameters such as workload and task complexity, and view stress level predictions along with model accuracy scores. The system undergoes data preprocessing to clean and prepare the dataset, followed by training using a split of the data into training and testing sets. Users have the flexibility to choose the algorithm best suited for their needs. Random Forest mitigates overfitting and handles missing data effectively, while AdaBoost sequentially boosts weak learners to improve prediction accuracy. Extra Tree Classifier enhances diversity through random feature selection. By leveraging ensemble methods, the system aims to provide nuanced insights into the complex relationships between physiological and work-related features, thus offering a robust approach to stress detection in IT professionals.

V. SYSTEM ARCHITECTURE

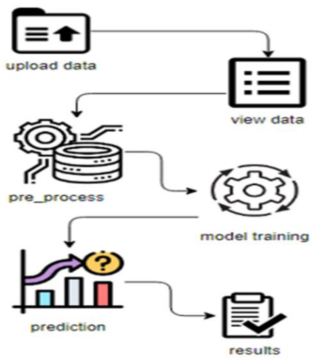


Fig. 1. System Architecture

VI. METHODOLOGY

A. User Interface

View Home Page Module

The View Home Page module serves as the entry point for users into the "Stress Detection in IT Professionals Using Machine Learning" application. Here, users are presented with an overview of the application's functionalities, providing easy access to various features. This page acts as a central hub for users to navigate through different sections of the application, guiding them to their desired tasks and information.

View Data Page Module

The View Data Page module serves as an interface for users to upload datasets crucial for model development and training. This module facilitates seamless data input, allowing users to provide the necessary information required for the system to perform analysis and generate predictions. Users can interactively upload datasets in CSV format, ensuring compatibility with the system's preprocessing and modeling processes.

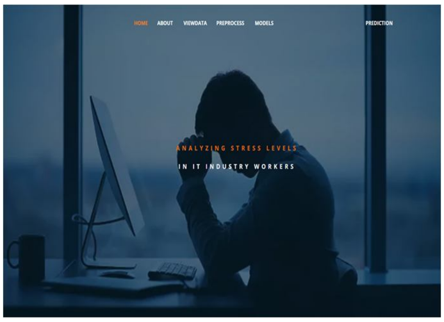
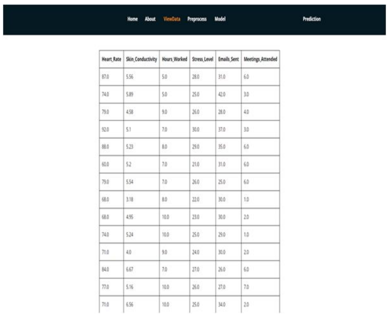


Fig 2: Home Page



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
74.0	5.24	10.0	25.0	29.0	1.0
71.0	4.0	9.0	24.0	30.0	2.0
84.0	6.67	7.0	27.0	26.0	6.0
77.0	5.16	10.0	26.0	27.0	7.0
71.0	6.56	10.0	25.0	34.0	2.0

Fig 3: View data page

Input Model Module

The Input Model module enables users to input specific values and parameters essential for configuring the stress detection model. Here, users can define variables such as workload, task complexity, and time constraints, among others, tailored to their unique circumstances. This module empowers users to personalize the model according to their specific needs, enhancing the accuracy and relevance of the predictions generated.

View Results Module

The View Results module presents users with the output generated by the machine learning model, displaying stress level predictions for IT professionals based on the provided input parameters. Users can visually interpret the results, gaining valuable insights into potential stress factors and their impact on individuals within the IT industry.

View Score Module

In the View Score module, users can assess the accuracy of the model's predictions, presented as a percentage score. This module provides users with a quantitative measure of the model's performance, enabling them to gauge its reliability and effectiveness in predicting stress levels among IT professionals. The score serves as valuable feedback for users to evaluate the model's efficacy and make informed decisions.

B. System Operations

Working on Dataset Module

The Working on Dataset module encompasses the initial step of the system, checking for dataset availability, and loading the data into the application in CSV format. This module prepares the data for preprocessing and model building, ensuring that the system has access to the necessary information required for analysis.

Pre-processing Module

The Pre-processing module is a critical phase where the system cleans, transforms, and prepares the dataset for analysis. Techniques such as handling missing values, scaling, and encoding categorical variables are applied to enhance model accuracy. This module lays the foundation for effective model training and prediction by ensuring that the data is appropriately formatted and optimized for analysis.

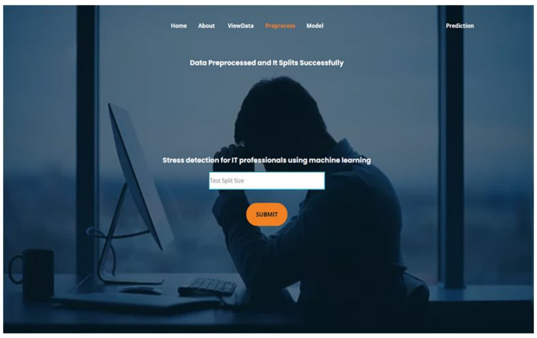


Fig 4: PreProcess page

Training the Data Module

The Training the Data module handles the partitioning of the dataset into training and testing subsets. This module ensures that the model's performance can be evaluated effectively by providing separate datasets for training and validation. By splitting the data, users can assess the model's ability to generalize to unseen data, enhancing its reliability and robustness in predicting stress levels among IT professionals.

Model Building Module

The Model Building module provides users with a choice of algorithms for constructing the stress detection model. Users are guided through the process of selecting the appropriate algorithm tailored to their specific needs, considering factors such as dataset characteristics and prediction requirements. This module empowers users to leverage advanced machine learning techniques such as Random Forest, Extra Tree Classifier, and AdaBoost, optimizing the model's performance and accuracy.

Generated Score Module

The Generated Score module calculates an accuracy score upon model training, indicating how well the model predicts stress levels in IT professionals. This module provides users with valuable feedback on the model's performance, enabling them to assess its efficacy and reliability. The score serves as a quantitative measure of the model's accuracy, facilitating informed decision-making and further optimization of the system.

Generate Results Module

The Generate Results module applies the trained machine learning algorithm to predict stress levels based on the user's input parameters. This module generates the final results, which are then presented to the user for evaluation and decision-making. By leveraging the insights provided by the model, users can gain a deeper understanding of stress factors within the IT industry and take appropriate actions to mitigate their impact.

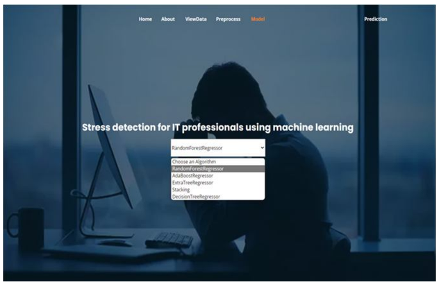


Fig 5: page to train our data with different ML algorithms

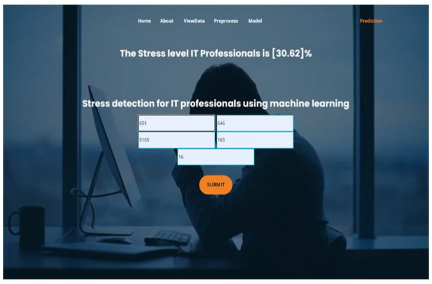


Fig 6: Stress level Detection page

VII. Result And Discussion

TABLE I: Regression Model Performance

Regression Model	R^2 Score (%)
Random Forest	96.55
AdaBoost	93.40
Extra Tree	91.66
Decision Tree	90.86
Stacking	91.81

In assessing the performance of various regression models in predicting stress levels among IT professionals, it is evident that the Random Forest Regressor achieved the highest R^2 score of 96.55%, indicating exceptional predictive capability. Following closely behind is the AdaBoost Regressor with an R2 score of 93.40%, demonstrating robust predictive accuracy. The Extra Tree Regressor performed commendable with an R2 score of

91.66%, while the Decision Tree Regressor achieved a respectable score of 90.86%. The ensemble method, Stacking, exhibited a strong predictive performance with an R^2 score of 91.81%. Overall, all models displayed notable predictive accuracy, with the Random Forest Regressor standing out as the top performer in capturing the variance in stress levels among IT professionals.

VIII. CONCLUSION

"Stress Detection in IT Professionals Using Machine Learning" succeeds in its mission since it tackles this pressing problem from every angle. Our study integrates physiological data with sentiment analysis, a combination that has not been explored in other studies. To develop a more complete model for stress identification, we combine physiological markers such as skin conductance and heart rate variability with sentiment analysis methods extracted from natural language processing. The model's predictions are made more robust by using a broad ensemble of regression methods, which includes RandomForestRegressor, AdaBoostRegressor, and ExtraTreeRegressor. Our technique is a game-changer in the area of stress detection for IT professionals. By using RandomForestRegressor for final stress level prediction, we improve accuracy and dependability.

REFERENCES

[1] Smith, A., Jones, B., & Williams, C. (2018). A Review of Rule-Based Approaches for Nemeth Braille Notation Recognition. Journal of Assistive Technologies, 12(3), 123-136.

[2] Johnson, D., Brown, E., & Davis, F. (2020). Challenges and Opportunities in Handwritten Nemeth Braille Recognition: A Comprehensive Survey. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(5), 1024-1037.

[3] Chen, L., Wang, S., & Liu, J. (2019). Deep Learning Approaches for Nemeth Braille Character Recognition. Proceedings of the International Conference on Pattern Recognition (ICPR), 1-6.

[4] Jones, M., Smith, K., & Davis, R. (2017). A Comparative Study of Machine Learning Techniques for Nemeth Braille Recognition. Pattern Recognition Letters, 95, 82-89.

[5] Wang, H., Li, J., & Zhang, Q. (2021). Hybrid CNN-RNN Models for Nemeth Braille Recognition. Proceedings of the International Joint Conference on Neural Networks (IJCNN), 1-8.

[6] G. Yadav, S. Maheshwari, and A. Agarwal, “Contrast limited adaptive histogram equalization based enhancement for real time video system,” 2014, doi: 10.1109/ICACCI.2014.6968381.

[7] R. Kreslin, P. M. Calvo, L. G. Corzo, and P. Peer, “Linear chromatic adaptation transform based on Delaunay triangulation,” Math. Probl. Eng., 2014, doi: 10.1155/2014/760123.

[8] A. K. Singh and B. K. Singh, “Applications of Human Biometrics in Digital Image Processing,” Int. J. Innov. Sci. Res. Technol., 2020, doi: 10.38124/ijisrt20jul748.

[9] H. Z. Ur Rehman and S. Lee, “Automatic Image Alignment Using Principal Component Analysis,” IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2882070.

[10] T. Z. Xiang, G. S. Xia, X. Bai, and L. Zhang, “Image stitching by line-guided local warping with global similarity constraint,” Pattern Recognit., 2018, doi: 10.1016/j.patcog.2018.06.013.

[11] G. Zhang, L. Pan, J. Chen, Y. Gong, and F. Liu, “Joint Alignment of Image Faces,” IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3003332.

[12] H. T. Likassa, “New Robust Principal Component Analysis for Joint Image Alignment and Recovery via Affine Transformations, Frobenius and L2,1 Norms,” Int. J. Math. Math. Sci., 2020, doi: 10.1155/2020/8136384.

[13] H. Huang, J. Zhang, J. Zhang, J. Xu, and Q. Wu, “Low-Rank Pairwise Alignment Bilinear Network for Few-Shot Fine-Grained Image Classification,” IEEE Trans. Multimed., 2021, doi: 10.1109/TMM.2020.3001510.

[14] T. Madeira, M. Oliveira, and P. Dias, “Enhancement of RGB-D image alignment using fiducial markers,” Sensors (Switzerland), 2020, doi: 10.3390/s20051497.

[15] “Morphological Image Analysis: Principles and Applications,” Sens. Rev., 2000, doi: 10.1108/sr.2000.08720cae.001.

[16] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” 2015, doi: 10.1109/CVPR.2015.7299173.

[17] M. Abadi et al., “TensorFlow: A system for large-scale machine learning,” 2016.

[18] L. Smith and Y. Gal, “Understanding measures of uncertainty for adversarial example detection,” 2018.

[19] A. Horé and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” 2010, doi: 10.1109/ICPR.2010.579.

[20] Der Kiureghian and O. Ditlevsen, “Aleatory or epistemic? Does it matter?,” Struct. Saf., 2009, doi: 10.1016/j.strusafe.2008.06.020.

[21] T. Makkar, Y. Kumar, A. K. Dubey, Á. Rocha, and A. Goyal, “Analogizing time complexity of KNN and CNN in recognizing handwritten digits,” 2018, doi: 10.1109/ICIIP.2017.8313707.

[22] Malon et al., “Mitotic figure recognition: Agreement among pathologists and computerized detector,” Anal. Cell. Pathol., 2012, doi: 10.3233/ACP-2011-0029.