International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

PREDICTING EMPLOYEES UNDER STRESS FOR PRE-EMPTIVE REMEDIATION USING MACHINE LEARNING ALGORITHM

[1]MARIPEDHA MANISHA,[2]MANDADAPU DHARMA TEJA,[3]D.S.P TANEEYASI,[4]PULI MANOJ GOUD,[5]S.CHAITANYA

[1,2,3,4]Students, Department of computer Science And Engineering,Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

[5]Assistant Professor, Department of computer Science And Engineering,Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

ABSTRACT

The COVID-19 pandemic has compelled businesses and organizations to adopt unconventional working models, such as remote work and limited on-site staffing. As this "new normal" persists, employees have had to adapt to new work environments and routines, often leading to increased psychological stress and fatigue. This study employs data visualization techniques and machine learning algorithms to predict employee stress levels. By analyzing relevant data, a predictive model has been developed to determine the likelihood of an employee experiencing stress. The XGBoost (XGB) classifier is utilized for this task, demonstrating reliable model performance. The analysis reveals that factors such as working hours, workload, age, and role ambiguity significantly and negatively impact employee performance. Other factors were found to be comparatively less influential. The study concludes that increased working hours, ambiguous roles, and excessive workload are major contributors to reduced employee well-being and overall productivity.

Keywords: Employee Stress Prediction, Machine Learning, XGBoost Classifier, Data Visualization, Workload, Working Hours, Role Ambiguity, Employee Performance, Psychological Stress, Remote Work.

I.INTRODUCTION

On March 11, 2020, the World Health Organization (WHO) officially declared COVID-19 a global pandemic, marking the beginning of a public health crisis that affected nearly every aspect of daily life. Caused by a novel strain of the coronavirus, COVID-19 quickly spread across continents, reaching 202 countries by the end of March 2020. The virus, known for its highly contagious nature, has triggered unprecedented disruptions in healthcare, economic systems, education, and workplaces. As countries enforced lockdowns and travel restrictions to control the spread of the virus, organizations had to rapidly shift from traditional office setups to remote working environments. The pandemic not only posed a threat to physical health but also significantly impacted mental well-being. Employees across sectors faced challenges such as job insecurity, changing work patterns, increased workloads, and the struggle to balance professional responsibilities with personal life at home. This shift, while necessary, led to a rise in stress, anxiety, and burnout among the workforce. The uncertainty brought by the pandemic has become a major contributor to psychological strain, with long-term implications on employee performance and organizational productivity. The rapid integration of technology in business operations has opened new possibilities for understanding and addressing these challenges. Machine learning (ML) and artificial intelligence (AI) have shown promising potential in various domains, including human resource management and behavioral analysis. These technologies can be used to analyze patterns in employee behavior, identify stress indicators, and assist in early intervention strategies. By applying data-driven approaches, organizations can gain deeper insights into the factors contributing to employee stress and take proactive measures to support their well-being. In the context of this study, we aim to explore the impact of workplace stress during the COVID-19 pandemic and build a predictive model to assess employee stress levels using machine learning techniques. Specifically, the XGBoost (Extreme Gradient Boosting) classifier is utilized for its high performance and accuracy in classification problems. The model is trained on relevant employee data, including variables such as working hours, workload, age, and role clarity, to determine the likelihood of stress. The primary objective of this research is to investigate how these factors contribute to stress and how they ultimately affect employee performance. Prolonged stress can lead to physical health problems, emotional exhaustion, and disengagement from work. It can also hinder productivity, innovation, and job satisfaction—elements that are vital for organizational success. Hence, it becomes essential for businesses to identify at-risk employees and provide timely support.

Through this work, we also aim to emphasize the significance of understanding mental health in the workplace, especially in crisis situations like a pandemic. By combining machine learning techniques with data visualization, this research offers a comprehensive approach to stress prediction, enabling better decision-making and resource allocation for employee wellness programs. As organizations adapt to the evolving work landscape, such tools can play a crucial role in building a resilient and healthy workforce.

In conclusion, this study contributes to the growing body of research on workplace stress and mental health by leveraging machine learning for predictive analysis. The insights derived from this model can assist managers, HR professionals, and policymakers in developing effective strategies to mitigate stress and improve employee satisfaction and performance during challenging times like the COVID-19 pandemic and beyond.

II.LITERATURE SURVEY

Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. The Economic Times

Bhattacharyya and Basu (2018) discuss the growing concern of workplace stress in India, particularly within the corporate sector. The article emphasizes how rising pressure and long working hours are affecting employee health and productivity. In response, organizations are increasingly adopting measures to reduce stress, such as offering mental health support programs, improving work-life balance, and implementing stress management workshops. The authors highlight the role of leadership in fostering a supportive work culture and the importance of employee well-being for maintaining a productive workforce. This research is crucial for understanding the broader impact of stress on employee performance, and it serves as a foundation for the development of more targeted workplace interventions, such as those utilizing machine learning for stress prediction and mitigation.

OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle

The OSMI Mental Health in Tech Survey Dataset, hosted on Kaggle, provides valuable insights into the mental health challenges faced by professionals in the tech industry. It offers a collection of responses from employees, revealing how factors such as workload, work culture, and job satisfaction contribute to mental health issues like stress, burnout, and anxiety. The dataset includes information on employee demographics, job roles, and their experiences with mental health support. This dataset is instrumental for research in understanding how the tech industry, known for its high-pressure environment, impacts employee well-being. It is widely used for machine learning applications that predict employee stress levels and develop strategies to address mental health in the workplace.

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

Van den Broeck et al. (2005) focus on the critical process of data cleaning in the context of medical research, discussing techniques to detect, diagnose, and correct data abnormalities. They explore various methods of identifying errors or inconsistencies in datasets, such as missing values, outliers, and data entry mistakes. The paper underscores the importance of high-quality data for ensuring the validity of statistical analyses and machine learning models. In the context of employee stress prediction, this research is relevant as it emphasizes the need for clean, reliable data to develop accurate predictive models. Proper data cleaning ensures that factors influencing employee stress, such as workload and working hours, are correctly represented, improving the model's effectiveness.

Shwetha, S., Sahil, A., Anant Kumar J. (2017). Predictive analysis using classification techniques in healthcare domain, International Journal of Linguistics & Computing Research

Shwetha, Sahil, and Kumar (2017) explore the application of predictive analytics in the healthcare sector, particularly using classification techniques to forecast outcomes such as disease progression, patient recovery, and hospital readmission. They emphasize the effectiveness of machine learning algorithms like decision trees, random forests, and support vector machines in predicting healthcare outcomes. The paper also discusses challenges such as data imbalance and the need for high-quality feature selection. This research is applicable to employee stress prediction, where similar classification techniques can be used to predict stress levels based on employee data such as working hours, workload, and job satisfaction. The insights gained can help organizations develop proactive strategies to improve employee health and performance.

Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for healthcare. International Journal of Bioscience and Biotechnology, 5(5), 241-266

Tomar and Agarwal (2013) provide an extensive review of data mining techniques used in healthcare applications. They explore the potential of machine learning, data mining, and statistical models for predicting health outcomes, identifying disease patterns, and personalizing treatment plans. The authors highlight various data mining algorithms, such as clustering, classification, and regression, and discuss their applicability in medical diagnostics. In relation to employee stress, the methodologies discussed in this paper can be directly applied to analyze patterns in employee behavior, identify stress indicators, and predict potential mental health issues. This research is valuable for organizations using data mining techniques to understand and manage employee well-being.

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

Pedregosa et al. (2011) introduce Scikit-learn, an open-source machine learning library for Python that has become widely popular due to its simplicity and robust functionality. The library provides a wide array of tools for classification, regression, clustering, and dimensionality reduction, making it an essential tool for both novice and expert data scientists. Scikit-learn is often applied in various domains, including business analytics and healthcare. In the context of predicting employee stress, Scikit-learn can be used to develop and evaluate machine learning models using features such as working hours, workload, and job satisfaction. This paper has been instrumental in facilitating the development of predictive models in multiple industries.

Gender and Stress. (n.d.). Retrieved from APA press release 2010

The American Psychological Association (APA) press release (2010) discusses the relationship between gender and stress, exploring how men and women experience and respond to stress differently. The article highlights research showing that women are more likely to report experiencing higher levels of stress than men, particularly in the workplace. This research is crucial for understanding gender-based differences in stress responses, which can have implications for employee well-being and productivity. By examining these differences, organizations can develop more tailored interventions and support systems to reduce stress and improve employee performance. This understanding is essential for stress prediction models that take gender differences into account when analyzing factors that influence employee stress levels.

III.PROPOSED METHODOLOGY

The proposed system emphasizes the importance of data cleaning as a critical first step in preparing the dataset for model execution. Incomplete or erroneous data can drastically reduce model performance, which is why it is essential to handle missing values appropriately. In this case, the dataset contains 3,895 null values, which need to be addressed before applying any machine learning algorithms. For categorical data, missing values are imputed using the mode, while numerical data is handled using techniques like mean, median, and floor methods. While removing null values entirely is an option, it can lead to the loss of valuable data, potentially impacting the model's effectiveness. Therefore, the system adopts a strategy of efficient imputation, ensuring that data integrity is maintained, and model performance is optimized.

The dataset being used is imbalanced, which can negatively affect the accuracy and reliability of the model's predictions. To address this issue, several techniques such as resampling, oversampling, and undersampling can be employed. The Synthetic Minority Over-sampling Technique (SMOTE) is another viable method to balance the dataset. These techniques aim to create a more balanced representation of both classes within the data, which is essential for accurate prediction. Furthermore, to enhance model performance, advanced algorithms such as Support Vector Machines (SVM), logistic regression, and tree-based methods can be utilized. These algorithms have proven effective in handling imbalanced datasets. Additionally, ensemble techniques like bagging and boosting can further improve the model's accuracy and robustness.

The proposed system aims to achieve high accuracy in predicting employee stress levels through the use of machine learning algorithms. By testing and training large datasets, the system can provide meaningful insights into employee stress and its impact on organizational performance. This method helps in predicting potential stress in employees, which can be used to implement early interventions. The use of data visualization and clear performance metrics such as accuracy, precision, and recall ensures that the system's predictions are transparent and reliable. The ability to process and analyze large datasets is vital in offering accurate and actionable predictions that organizations can use to improve employee well-being.

In terms of system functionality, the service provider module plays a central role. Service providers log in using their credentials and gain access to various functions such as browsing datasets, training and testing the model, and viewing accuracy results. The system displays these results in an intuitive bar chart format for easy interpretation. Additionally, service providers can download the predicted datasets, view employee stress prediction types and their corresponding ratios, and monitor the activity of all remote users. This module ensures that service providers can manage and operate the system efficiently, with access to all essential features.

The "View and Authorize Users" module is designed to enable the system administrator to monitor and manage user access. Administrators can view a list of registered users along with their details, such as username, email, and address. The admin has the responsibility of authorizing users before they can access the system's functionalities. This module ensures that only authorized personnel can interact with the system, enhancing security and maintaining data integrity.

The Remote User module caters to the end-users, who must register before using the system. After registration, user details are stored in the database, and users can log in using their authorized credentials. Once logged in, users can perform several operations, such as predicting employee stress types and viewing their profiles. This module provides a seamless experience for users to interact with the system, ensuring that they can access the system's predictive features and benefit from the insights provided.

Together, these modules form a comprehensive methodology that ensures the efficient and secure use of the system. By combining data cleaning, model optimization, and intuitive user interfaces, the proposed system provides an effective solution for predicting employee stress and supporting better organizational decisions.

IV.CONCLUSION

The performance of our model has been optimized using the XGB classifier, which is one of the most effective techniques for improving predictive accuracy. This decision tree-based algorithm leverages the gradient boosting framework to analyze data and evaluate model performance. The confusion matrix is employed to assess how many predictions made by the model are correct, providing insight into its effectiveness. XGBoost stands out due to its remarkable predictive power, being approximately 10 times more efficient than other gradient boosting methods. It also incorporates various regularization techniques that help reduce overfitting and improve overall model performance. As a result, XGBoost is often referred to as the "regularized boosting" technique. By utilizing key metrics such as true positives, true negatives, false positives, and false negatives, the model's classification performance can be thoroughly evaluated and optimized for accurate results.

V.REFERENCES

[1] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from 'The Economic Times'

[2] OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle

[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 Bioscience and Biotechnology, 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] Gender and Stress. (n.d.). Retrieved from APA press release 2010

[8] Vidit Laijawala, Aadesh Aachaliya, Hardik Jatta, Vijaya Pinjarkar. Classification Algorithms based Mental Health Prediction using Data Mining (2020).

[9] U Srinivasulu Reddy, Aditya Vivek Thota, A Dharun. Machine Learning Techniques for Stress Prediction in Working Employees (2020).

[10] Monisha S, Meera R, Vijay Swaminath.R, Dr. Arun Raj L. Predictive Analysis of Student Stress Level Using Naive Bayesian Classification Algorithm (2020).

[11] Fang Li. Research on the College Students Psychological Health Management based on Data Mining and Cloud Platform (2016).

[12] Jigang Zheng, Jingmei Zhang. The Application of Association Rules Mining in the Analysis of Students Test Scores (2016).

[13] Jenny K. Hyun, Brian C. Quinn, Temina Madon, Steve Lustig. Graduate Student Mental Health: Needs Assessment and Utilization of Counseling Services (2014).

[14] Shivendra Jena, Harish Chandra Tiwari. Stress and mental health problems in 1st year medical students: a survey of two medical colleges in Kanpur, India (2014).

[15] Honglei Zhu, Zhigang Xu. An Effective Algorithm for Mining Positive and Negative Association Rules (2008).