

Machine Learning – Based Early Prediction of Employee Burnout Using Behavioral and Work Pattern Analysis

Muskan S² Mrs. Roopa R¹

¹ Assistant Professor, Department of MCA, BIET, Davanagere

² Student, 4th Semester MCA, Department of MCA, BIET, Davanagere

Abstract

Burnout among employees is a major problem in today's fast-paced and stressful business world. It hurts both the health of the employees and the productivity of the company. Stress at work that lasts for a long time, too much work, not enough help from management, and a bad work-life balance can all make people feel emotionally drained, lower their performance, and cause them to leave their jobs more often. This study proposes a web-based Burnout Risk Prediction System created utilizing machine learning approaches to proactively assess and predict the possibility of employee burnout. The suggested approach assesses a variety of personal and professional parameters, such as work hours, stress levels, job satisfaction, sleep quality, and workplace assistance, to categorize employees as burnout risks. Classification technique like random forest and logistic regression are applied and evaluated to determine the most accurate predictive model. The system employs data preprocessing methods. Feature encoding model training and performance evaluation to guarantee correct predictions. Employees may enter essential information and receive real-time burnout risk assessments through a user-friendly web interface, while administrators can monitor users, provide mental wellness advice, and respond to questions. Experimental evaluation shows that the RF model outperforms Logistic Regression in terms of predicted accuracy, making it appropriate for practical deployment. The suggested system provides a data-driven, scalable, and proactive approach to promote mental health management and a better, more productive workplace.

Keywords: Burnout Risk, Random Forest, Machine Learning, Prediction, and Logistic Regression, Employee Mental Health, Predictive Analytics, Web-Based System, Workplace Stress Analysis.

1. INTRODUCTION

1. In light of the contemporary stressors that involve the need for competition, time constraints, and professionalism in the workplace, employee burnout has emerged as a significant problem. Burnout is perceived by employees as a state whereby an individual experiences physical and psychological exhaustion because he or she is under stress and works hard. This results in reduced efficiency on the part of the employees, as well as reduced satisfaction and productivity. Employee fatigue will mean that employees will be absent from work, lack commitment, and turnover rates will be very high.

The traditional methods used for the analysis of the burnout problem are mostly based on self-reports, continuous evaluations, and assessments that take place through manual processes. While all these measures provide important information, they usually

become very subjective and tedious, as well as reactive rather than proactive. The employees may decide not to report their issues related to stress due to the negative attitude to the problem or because of the consequences that may occur professionally, and therefore the problem can only be detected when it becomes critical. Lastly, there is no tool in the traditional methods of evaluation that

allows continuous assessment of various personal and professional factors.

Machine learning and data analytic technologies provide exciting possibilities when solving the issue. The ability of machine learning to reveal implicit connections and trends that link workplace attributes to outcomes of burnout is one such possibility. Algorithms for classification such as Logistic Regression and Random Forest have a highly accurate predictive capability in determining categorical results from several input variables. The model utilizes structured data including work time, stress, satisfaction

level, sleep pattern, and management help among others to determine burnout probability with high precision.

This study offers a web-based Burnout Risk Prediction System that uses machine learning and an interactive user interface to provide real-time burnout assessments. The system takes pertinent personal and professional information from users and uses trained classification models to assign them to risk ratings such as "Low Risk" or "High Risk." In addition to prediction skills, the platform has administrative features for managing users, providing mental relaxation suggestions, and responding to employee inquiries, resulting in a supportive digital environment for workplace wellness.

The major goal of the proposed system is to give employers a proactive, data-driven tool for monitoring employee mental health and implementing timely interventions. The solution promises to promote workplace well-being, increase productivity, and eliminate burnout-related organizational difficulties by merging predictive analytics with an easy-to-use online platform. The remainder of the paper is structured as follows: Section II focuses at relevant studies on burnout prediction and machine learning-based mental health evaluation; Section III discusses the suggested approach and model building process; Section IV describes the system architecture and technologies used; section VI Gives a summary of the experimental results and performance evaluation, whereas section V focuses the paper's important conclusions and future directions.

II. RELATED WORK

1. A survey titled "using Machine Learning in Burnout Prediction" was written by Malgorzata Grunzielewska. This survey paper provides a comprehensive review of the application of ML techniques in predicting occupational burnout across various domains. The study highlights that burnout is influenced by multiple psychological, social, and workplace-related factors, making data-driven predictive approaches more effective than traditional assessment methods. The author discusses the function of supervised and unsupervised machine learning models in analyzing large-scale datasets to identify patterns associated with stress and burnout. The findings emphasize that machine learning-based prediction systems can support early detection, improve decision-making, and assist organizations in

implementing preventive mental health strategies. The survey also identifies research gaps in burnout monitoring systems and suggests the need for advanced predictive models to enhance workplace well-being and reduce employee turnover.

2. Employee Burnout Prediction Using Machine Learning, Author: Pratiksha Vaijanath Lagad. The study proposes a ML-based approach to predict employee burnout by evaluating various workplace and psychological factors that influence stress and mental exhaustion. The study utilizes structured employee data and applies preprocessing methods including features selection, and data cleaning and encoding before training classification models to identify burnout risk levels. The research emphasizes the limitations of traditional manual and survey-based burnout assessment methods and demonstrates how machine learning algorithms can provide more accurate and early predictions. The results show that predictive models can effectively detect high-risk employees, helping organizations implement timely preventive measures, improve employee well-being, and enhance overall workplace productivity.

3. Predicting Employee Burnout Using Machine Learning Models, Authors: B. W. G. Teodosio, M. J. O. T. Lira, P. H. M. Araújo, L. R. C. Farias this paper describes the ML method for approach predicting employee burnout by analyzing various occupational and behavioral factors such as workload, stress levels, and job satisfaction. The authors implemented multiple supervised learning algorithms and evaluated their performance using validation techniques to identify the most effective predictive model. The research highlights the significance of early burnout detection through data-driven analytics and demonstrates that machine learning models can accurately classify burnout risk levels. Furthermore, the study emphasizes the integration of predictive systems with user-friendly interfaces to encourage the mental health of employees and help corporate decision-making in modern workplaces.

4. Artificial Intelligence-Based Analysis for Burnout Prediction and Human Performance Assessment, Authors: Y. Zhang, H. Liu, X. Chen, and J. Wang.

This study presents an intelligent computational framework for analyzing employee performance and psychological stress using advanced machine learning techniques. This study is aimed at detecting behavioral and occupational determinants that cause the development of burnout in employees based on predictive analytics and data-based modeling. Using

artificial intelligence algorithms for analysis of structured data helps show the benefits of burnout prediction in terms of developing organizational strategies to prevent burnout. Overall, the findings prove that implementation of predictive systems that use artificial intelligence into organizational operations can benefit both employees and the organization as a whole.

5. Authors Chao Liu, Yen-Ching, Lifan Qin, Lijie Ren, Ching-Wen Chien, and Tao-Hsin Tung have developed an ML approach to analyze and predict burnout-related issues for healthcare workers. The paper is dedicated to creating an ML model to analyze and detect determinants of burnout for healthcare professionals. Based on demographic data, occupation-related information, and metrics for assessing the risk of burnout, the researchers propose multiple predictive models based on ML algorithms such as Random Forest, Decision Tree, K-Nearest Neighbor, and Logistic Regression.

The findings show that ensemble models—Random Forest in particular—offer greater prediction accuracy and robustness in identifying high-risk burnout cases. The study further emphasizes that job satisfaction, workload, and workplace environment significantly influence burnout levels, highlighting the importance of early predictive systems for mental health monitoring and organizational well-being management.

5. Liu et al. (2025) created a model based on machine learning to analyze and predict burnout among healthcare workers using survey data from 1125 hospital staff. The study utilized the Maslach Burnout Inventory, job satisfaction metrics, and demographic information to categorize burnout into four levels. Results showed revealed 61.2% of employees had at least one burnout symptom, with doctors, surgeons, and professionals with fewer than ten years of experience showing higher levels of burnout. The Random Forest model outperformed the other models—Logistic Regression, KNN, Decision Tree, and Random Forest—with an accuracy of around 80%. proving to be the most effective for burnout prediction.

6. A comprehensive review and meta-analysis titled "Machine Learning for Predicting Burnout Among Healthcare Workers" was carried out by Shi et al. in 2025. The study analyzed 22 research articles that developed or validated machine learning models for predicting burnout employing reliable evaluation

instruments like the Maslach Burnout Inventory (MBI). With an overall AUC of 0.72, sensitivity of 0.63, and specificity of 0.84, the combined results demonstrated a reasonable predictive performance. The findings indicated better model performance when self-reported data, nurse-focused samples, and MBI-based assessments were used. However, the study highlighted methodological weaknesses, high risk of bias, and lack of external validation, emphasizing the need for rigorous design and ethical safeguards in future machine-learning-based burnout prediction models.

7. Grzędzielewska (2021) presented a survey titled *"Using Machine Learning in Burnout Prediction: A Survey"* examines how machine learning methods are used to forecast occupational burnout. The study discusses the limitations of traditional statistical models such as LR and highlights the advantages of machine learning methods in handling large, complex, and nonlinear datasets. It explains supervised and unsupervised learning approaches, including clustering, decision trees, neural networks, and regression-based models, in the context of burnout analysis. The paper emphasizes that machine learning can improve prediction accuracy, support data-driven interventions, assist in recruitment and monitoring systems, and help reduce employee turnover. The study concludes that there is a research gap in applying advanced machine learning methods in social work burnout studies, encouraging further development of predictive and preventive models.

8. Yeskuatov et al. (2025) explored the feasibility of using supervised machine learning to detect academic burnout among undergraduate computing students by relying exclusively on university administrative records. The study addressed three primary dimensions of burnout: fatigue, pessimism, and a lack of professional effectiveness, utilizing a dataset of 565 student records. Five algorithms—Naïve Bayes, LR, SVM, Decision Tree, and XGBoost—were evaluated using features such as GPA, credit hours, and demographic data. The results revealed that although automated diagnosis is possible, its efficacy varies; models were most successful in identifying fatigue, with Logistic Regression achieving a peak F1 score of 68.4%. In contrast, professional efficacy was the most challenging to detect, suggesting that psychological constructs are difficult to capture through administrative data alone. This research provides a foundation for unobtrusive mental health screening in educational settings without the

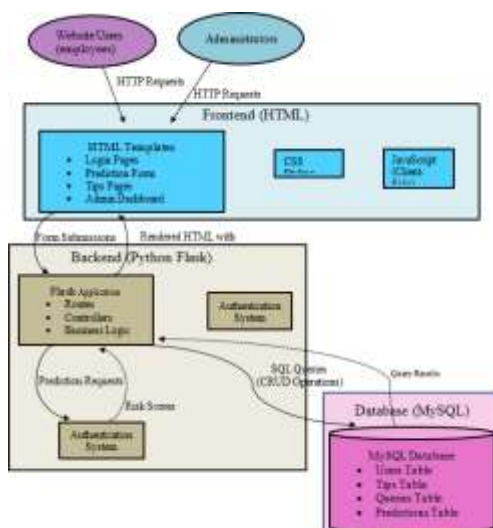
administrative overhead of traditional surveys.

9. Jeong et al. (2025) used data from the Korean Occupational Stress Scale (KOSS) to determine whether machine learning could be used to predict the danger of Burnout Syndrome (BOS). The research evaluated five algorithms: Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR) utilizing a dataset comprising 1,205 individuals from 40 enterprises. To fix the imbalance in the classes, the researchers employed the SMOTEENN resampling technique. The RF model emerged as the optimal predictor, achieving an ROC-AUC of 0.904 and high overall accuracy (0.874). Furthermore, "Lack of Reward" and "Job Instability" were found to be the most significant occupational stress factors contributing to burnout risk. This research demonstrates that KOSS—a mandated assessment in some regions—can serve as an effective, non-stigmatized proxy for identifying employees at high risk of burnout.

III. METHODOLOGY

The methodology of the Burnout Risk Prediction System outlines the systematic approach used to design, develop, and implement the machine learning model. It describes the processes involved in data collection, preprocessing, model training, evaluation, and system integration. The objective of this methodology is to ensure accurate burnout risk prediction using reliable classification techniques. This structured approach guarantees efficiency, accuracy, and seamless integration with the web-based application.

Architecture Diagram



1. Collect data and preparation

Firstly in the methodology involves collecting employee-related survey data that contains both personal and professional attributes influencing burnout. The dataset includes important factors such as work hours, stress levels, job satisfaction, sleep quality, managerial support, and work-life balance. These parameters act as predictive indicators that help in identifying patterns associated with burnout. The collected data provides the framework for creating and honing the ML models.

2. Data Preprocessing

To get the dataset ready for model training, preparation is done after data collection. Unnecessary columns such as employee identification numbers are removed to avoid redundancy. Missing values and duplicate records are identified and handled appropriately to ensure data quality. Label encoding techniques are used to transform categorical data into numerical form since machine learning algorithms require numerical input. Next, the dataset is separated into the dependent variable (BurnoutRisk) and independent variables (features). In order to properly assess the model, the data is finally divided into training and testing sets in a 70:30 ratio.

3. Model Training

In this stage, two classification algorithms, namely The Classifier for Random Forests and Logistic Regression are used. A baseline model for binary classification is logistic regression. Several decision trees are combined in Random Forest, an ensemble learning technique, to increase prediction accuracy and decrease overfitting. The training dataset is used to train both models in order to determine the correlation between input features and burnout risk levels.

4. Model Evaluation and Selection

After training, both models are assessed using performance indicators like F1-score, accuracy score, precision, recall, and confusion matrix. These measurements aid in assessing how successfully the models classify burnout risk levels. A comparison of the performance results is conducted, and the model with higher accuracy and better overall performance is selected. In this case, The Random Forest classifier showed better performance and was chosen as the final

prediction model.

5. System Integration and Deployment

The chosen machine learning model is included into a web application's backend to developed using Python and Flask. The system allows users to enter their personal and professional details through a web form, and the trained model processes this input to generate real-time burnout risk predictions. The prediction output classifies users as either "Low Risk" or "High Risk," along with a probability score. The system includes an Admin module for managing users and content, and a User module for accessing predictions and mental wellness resources, ensuring complete functionality and interaction.

IV. TECHNOLOGIES USED

The proposed Burnout Risk Prediction System uses modern technologies to provide accurate predictions and a smooth user experience. It integrates machine learning, web development frameworks, and database systems for efficient performance and secure data handling. The following are the key technologies used in the development and deployment of the system.

1. Python is a programming language.

The main programming language utilized in the system's development is Python. It is selected for to its readability, ease of use, and broad support for machine learning and data analysis packages. Python handles backend logic, model training, data processing, and integration of the prediction model into the web application.

2. Machine Learning Libraries

Several Python libraries are utilized to implement the machine learning workflow. Pandas is used for activities including data handling and preparation such as cleaning and organizing the dataset. Numerical calculations and array operations are supported by NumPy. Classification methods like Random Forest and Logistic Regression can be implemented with Scikit-learn. along with evaluation metrics. Matplotlib and Seaborn are used for visualizing data distributions and comparing model performance through graphical representations.

3. Flask Framework

NLTK (Natural Language Toolkit): Utilized for basic NLP tasks such as tokenization, stopword Flask is used as the backend web framework to connect the machine learning model with the web interface. It manages routing, user authentication, form handling, and communication between the frontend and database. Flask enables seamless integration of prediction functionality within the web application.

4. HTML

HTML is used to design the structure of the web pages, including login forms, registration pages, prediction forms, dashboards, and tips pages. It ensures that the content of the application is well-organized and accessible to users.

5. CSS

CSS is used to make the visual presentation of the application. It improves layout, styling, colors, and responsiveness of the web interface, making the system more user-friendly and visually appealing.

6. MySQL

The relational database management system (RDMS) used to store user data is MySQL credentials, employee data, relaxation tips, prediction records, and user queries. It guarantees the safekeeping and effective retrieval of organized data required for system operation.

7. The Visual Studio Code

The integrated development environment (IDE) for writing is called Visual Studio Code. debugging, and managing the project code. It provides features such as syntax highlighting, extensions, and version control support, which improve development efficiency.

V Result

Prediction Results



Prediction Results page

VI. CONCLUSION

The Burnout Risk Prediction System successfully combines machine learning methods with a web-based application to offer accurate and timely prediction of employee burnout risk. By analyzing important personal and professional factors such as stress levels, work hours, job satisfaction, and lifestyle patterns, The system helps find people who are at high risk early on. Using Random Forest and other classification methods and Logistic Regression ensures reliable performance, with improved accuracy and stability. The web interface allows users to easily access predictions and mental wellness resources, while administrators can efficiently manage users and content. Overall, the system offers A workable, expandable, and efficient way to enhance employee well-being and enhancing organizational productivity.

REFERENCES

[1] M. Grządzielewska, "Using Machine Learning in Burnout Prediction: A Survey," *Child and Adolescent Social Work Journal*, vol. 38, no. 2, pp. 175–180, 2021, doi: 10.1007/s10560-020-00733-w.

[2] P. V. Lagad, "Employee Burnout Prediction Using Machine Learning," *International Scientific Journal of Engineering and Management (ISJEM)*, ISSN: 2583-6129, 2024.

[3] B. W. G. Teodosio, M. J. O. T. Lira, P. H. M. Araújo, and L. R. C. Farias, "Predicting Employee Burnout Using Machine Learning Models," *Proceedings of the ACM*, 2025, doi: 10.1145/3757110.3757210.

[4] "Artificial Intelligence-Based Analysis for Burnout Prediction and Human Performance Assessment," *Results in Engineering*, Elsevier, 2025, doi: 10.1016/j.rineng.2025.100733

[5] C. Liu, Y.-C. Chuang, L. Qin, L. Ren, C.-W. Chien, and T.-H. Tung, "Machine-learning-based model for analysing and accurately predicting factors related to burnout in healthcare workers," *BMJ Public Health*, vol. 3, no. 2, e000777, 2024.

[6]. Liu, C., Chuang, Y.-C., Qin, L., et al. (2025). *Machine-learning-based model for analysing and*

accurately predicting factors related to burnout in healthcare workers. BMJ Open Quality, 3(2). ISSN: 2753-4294. Published by BMJ Publishing Group.

[7]. H. Shi, J. Liu, C. Yang, J. Shang, and Y. Zeng, "Machine learning for predicting burnout among healthcare workers: A systematic review and meta-analysis," *Contemporary Nurse*, Nov. 2025. doi: 10.1080/10376178.2025.2593294.

[8]. M. Grządzielewska, "Using machine learning in burnout prediction: A survey," *Child and Adolescent Social Work Journal*, vol. 38, pp. 175–180, 2021. doi: 10.1007/s10560-020-00733-w.

[9]. Detecting Burnout Among Undergraduate Computing Students with Supervised Machine Learning, Authors- Eldar Yeskuatov, Lee Kien Foo, Sook-Ling Chua, Publisher- 2025 | Healthcare (MDPI) | Volume 13 Issue 23 | DOI: 10.3390/healthcare13233182

[10]. Predicting the Risk of Burnout Syndrome Using Korean Occupational Stress Scale (KOSS): A Machine Learning Approach, Authors- Hyeonju Jeong, Seong-Cheol Yang, Shin-Goo Park, Inho Hong, Hyung Doo Kim, Publisher- 2025 | Safety and Health at Work (Elsevier) | Volume 16 Issue 4 | DOI: 10.1016/j.shaw.2025.08.006