International Journal on Emerging Research Areas (IJERA)

ISSN :*2230-9993*

MindPulse: Employee Mental Health Detection and Attrition Prediction App

Sreyas George
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
sreyasgeorge2025@cs.ajce.in

Gregan George
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
gregangeorge2025@cs.ajce.in

Ruth Tennyson
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
ruthtennyson2025@cs.ajce.in

Rishil Shajan
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
rishilshajan2025@cs.ajce.in

Dr. Juby Mathew
Dept. of Computer Science
Amal Jyothi College of Engineering
(Autonomous)
Kottayam, India
jubymathew@amaljyothi.ac.in

***Abstract*—Employee mental health problems and excessive turnover are challenges that impact organizations at levels beyond work quality, including productivity, workforce stability, and overall morale. Conventional methods tend to be ineffective in pre-emptively predicting and managing these problems with any consistency because there are no accurate predictive tools available. MindPulse is a machine learning (ML) and natural language processing (NLP) application that uses AI to analyze the well-being of employees and attrition risk. With BERT-based sentiment analysis, it analyzes social media information to identify initial indicators of mental distress, and gradient boosting models analyze employee-specific m etrics t o f orecast a ttrition patterns. By combining these insights, MindPulse allows organizations to make timely interventions, creating a healthier workplace and minimizing turnover. This new methodology improves workforce retention efforts by delivering actionable, data-driven insights, making it an essential tool for contemporary businesses.**

***Index Terms*—Keywords - Employee Attrition, Mental Health Prediction, Machine Learning, Natural Language Processing, Sentiment Analysis, Bidirectional Encoder Representations from Transformers (BERT), Gradient Boosting, Workforce Retention, Predictive Analytics.**

1. INTRODUCTION

The well-being and retention of employees are vital components that determine the success of an organization. Employee mental health issues may contribute to loss of productivity, lower employee engagement, and overall workplace morale. Likewise, high turnover rates may interfere with workforce stability, raise recruitment and training expenses, and deter the effectiveness of working teams. Companies across the globe have been unable to effectively respond to these matters, as conventional solutions tend to lack precise, data-based recommendations that support early intervention. The lack of strong prediction tools makes it challenging for employers to determine which employees are at risk of suffering from mental illness or quitting work.

MindPulse is a cutting-edge machine learning and natural language processing-based software that is developed to address such challenges. MindPulse employs Bidirectional Encoder Representations from Transformers (BERT) to perform sentiment analysis of social media data, detecting minute emotional signals expressing stress, discontent, or deteriorating mental health. Through analyzing patterns of language and emotional undertones, MindPulse is able to identify early warning signals of mental distress, enabling organizations to take pre-emptive measures to care for employees.

Along with mental health tracking, MindPulse uses gradient boosting models to scan employee-specific data such as performance indicators, levels of job satisfaction, and demographic characteristics. The system determines patterns and correlations among this data and predicts the risk of employee attrition, allowing organizations to put in place effective retention strategies. This two-functionality model gives a holistic solution that allows businesses to build a healthy work culture while reducing turnover of workforce.

Through predictive analytics using artificial intelligence, MindPulse enables organizations to make sound decisions that foster employee well-being, increase job satisfaction, and secure workforce stability. The groundbreaking solution is one that fills the lacuna between managing employee mental health and preventing attrition, providing a data-driven approach to enable sustainable workplace practices.

2. Literature Survey

The swift growth of online platforms and sophisticated machine learning methodologies has brought forth novel prospects of analyzing mental health patterns and anticipating employee turnover. Social media sentiment analysis and employee analytics have been comprehensively investigated to detect mental health problems and risks of employee turnover. Although prior studies have significantly advanced in these domains, most studies are constrained by issues of data quality, interpretability of the model, and usability in real-world settings. This survey of literature discusses the methods used in mental health detection and attrition prediction, their shortcomings, and the major research gaps that need to be addressed in the future.

A. Employee Mental Health Detection

Yuan [1] introduced a Deep Belief Network (DBN) for sentiment analysis to predict mental health conditions from text data. Yet, DBN models are also known to face challenges related to explainability, where it becomes challenging to understand the rationale of predictions. Text data itself might not always give a complete picture of mental health conditions since other modalities such as audio and physiological data are untapped. The gap in research is unifying multimodal sources of data to achieve more complete detection.

Jain et al. [2] designed a machine learning system to categorize social media users on the basis of their online activities. The shortcoming of the technique is its reliance on social media content posted openly, which is not a reflection o f one's real state of mind. Privacy also prevents collection of personal sensitive information, hence curtailing the efficiency o f the model. There is a need for future research in terms of privacy-conserving methods like federated learning.

Santos et al. [3] introduced a Mixture of Experts framework for mental health prediction. Despite its effectiveness in improving prediction accuracy, the model is computationally expensive and difficult t o d eploy i n r eal-time a pplications. M oreover, it does not account for evolving linguistic patterns, which may reduce performance over time. Future work should explore lightweight models with adaptive learning capabilities.

Lin et al. [4] used deep neural networks to identify psychological stress based on user-generated content. Nevertheless, the model was mostly text-based, disregarding behavioral information like sleep habits and physical exercise, which are significant predictors of mental health. The gap in research is integrating wearable sensor data for a complete approach.

Sun and Wu [5] utilized deep learning for early disorder identification. While the model showed high accuracy, its black-box nature made it unsuitable for clinical use. Healthcare professionals require interpretable AI solutions to make informed decisions. Future research should focus on explainable AI techniques for mental health applications.

Briand et al. [6] compared social media data to evaluate mental health interventions. The limitation of the study is the use of self-reported symptoms, which can be subject to biases. Additionally, the method does not take into account real-time intervention. The research gap is to formulate proactively monitoring systems that offer real-time intervention.

Mali and Sedamkar [7] used machine learning to augment conventional diagnostic methods. Yet, the dataset was limited and homogeneous, and hence may introduce biases in model predictions. Future research must utilize larger and more diverse datasets to enhance generalizability.

Shivaiah et al. [8] suggested an ensemble-based sentiment classification m odel f or i dentifying m ental h ealth disorders. While ensemble methods enhance performance, they consume a lot of computational power and might not be feasible for real-time use. Future studies need to optimize ensemble models for efficiency.

Lin et al. [9] presented a deep neural network-based approach to user-level psychological stress detection using social media data. The proposed model demonstrated potential in stress level identification b ut f ocused m ainly o n t ext-based f eatures while ignoring other key indicators like sleep patterns and physical activity. In the future, there is a need to integrate wearable sensor data for an overall detection system.

Zhang et al. [10] investigated a multimodal method based on text, speech, and physiological signals for detecting depression. Although the method showed high accuracy, the multimodal integration raised issues of data fusion and synchronization. Future work should examine stronger fusion methods for enhancing reliability.

B. Employee Attrition Prediction

Yadav et al. [11] used data mining methods to predict early attrition. The method, however, did not include external factors like economic climate and industry trends, making it less applicable. Future studies need to incorporate macroeconomic indicators to make the prediction model more holistic.

Zhang et al. [12] performed turnover trends analysis with machine learning classifiers. The limitation of the model is that it is static and fails to accommodate abrupt organizational changes like leadership change or company mergers. The area of research is to create dynamic models that make real-time updates of the prediction.

Sisodia et al. [13] contrasted predictive models of employee turnover. While the research tested several methods, it did not examine unstructured data sources such as employee feedback. Sentiment analysis of exit interviews and survey answers could improve predictive power.

Alduayj and Rajpoot [14] applied supervised learning methods for attrition prediction. One of the limitations was the emphasis on past HR data without the inclusion of workplace culture and job satisfaction. Organizational sentiment analysis should be included in future studies for improved understanding.

Maharana et al. [15] machine-learned turnover prediction with sophisticated machine learning models. Still, the study did not address ethical considerations such as the potential for prediction outcome bias. Future studies ought to prioritize fairness-aware AI models to avert discrimination.

Ganthi et al. [16] compared attrition detection classification models. Although the research compared various algorithms, it did not test model interpretability, which is important for HR decision-making. Future studies should create transparent AI systems for HR use.

Douaidi and Kheddouci [17] also suggested a graph-based approach to employee attrition analysis. While graph-based techniques have the capability to model intricate relationships, they are computationally intensive and hence not practical for small companies. Scalable graph-based techniques should be considered in future work.

Roy et al. [18] described social media analysis as a potential method for early detection and characterization of mental illnesses. The role of real-time tracking was brought into focus but utilized self-reports of symptoms, which is susceptible to various biases. Researchers need to make systems proactive with an aim of early intervention.

Tounsi et al. [19] proposed a deep learning-based method for identifying mental disorders using user-generated content on social media. The paper showcases the capabilities of deep learning models in diagnosing mental health conditions. However, the model is constrained by text-based data only, failing to account for non-verbal indicators like tone and facial expressions that are important for holistic mental health evaluation. Multimodal integration of data needs to be researched in future work to further improve detection robustness and accuracy.

Alaskar et al. [20] have suggested a deep learning-based attrition detection model. Nonetheless, deep learning models are in need of big datasets with labeled data, which are not always accessible in HR scenarios. In the future, transfer learning methods should be investigated to alleviate data paucity problems.

Through these challenges being overcome, AI-based mental health detection and attrition prediction can be more effective, transparent, and actionable solutions for individuals and organizations.

C. *Conclusion and Research Gaps*

Substantial advancement has been realized in employee mental health detection and attrition forecasting, but current models continue to struggle with multimodal integration, explainability, and adaptability. Bridging these gaps will improve the validity and usability of AI-based solutions in workplace analytics.

Research Gaps:

- Limited multimodal integration in mental health detection models.
- Weak explainability of deep learning models in clinical and HR domains.
- Static characteristic of employee attrition forecasting models.
- Ethical issues with AI-based predictions and bias.
- Limited availability of large labeled datasets to train deep learning models.

Through these challenges being overcome, AI-based mental health detection and attrition prediction can be more effective, transparent, and actionable solutions for individuals and organizations.

3. METHODOLOGY

A. *System Architecture*

The system architecture, as depicted in Figure 1, has two primary components:

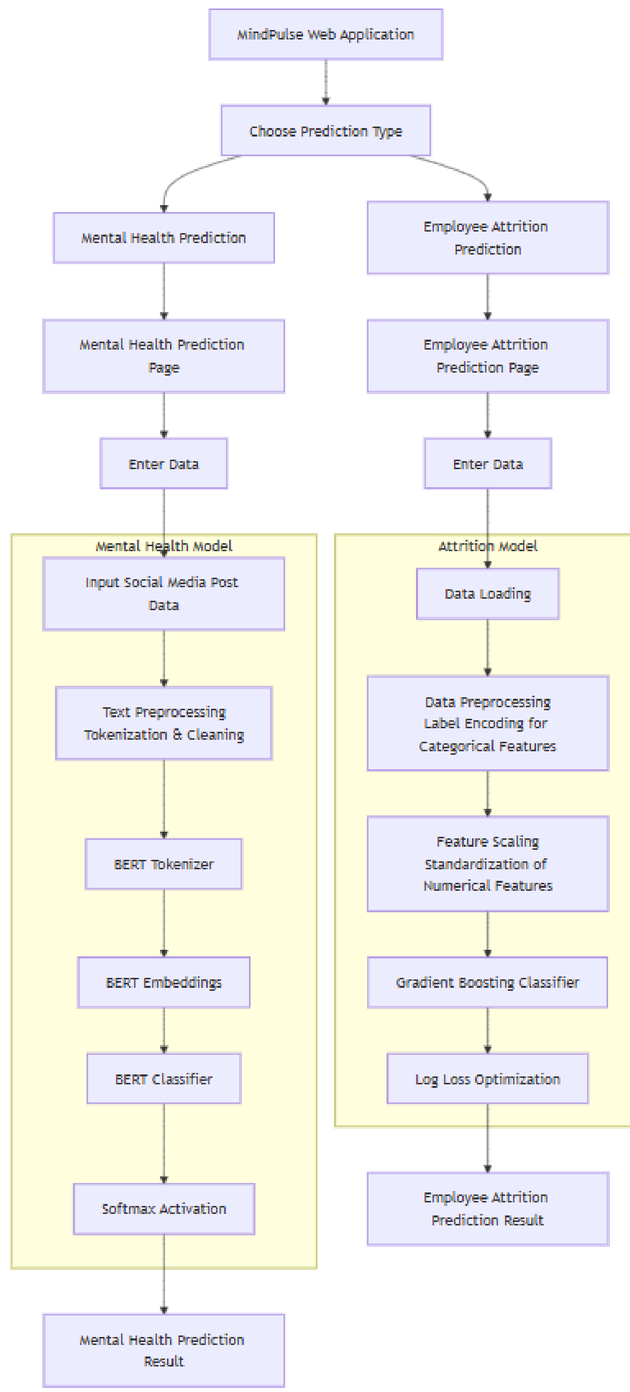

Fig. 1: System Architecture Diagram

ISSN : *2230-9993*

Employee Mental Health Detection and Employee Attrition Prediction. Both components follow a systematic process consisting of data preprocessing, model training, and evaluation.

B. Employee Mental Health Detection

BERT (Bidirectional Encoder Representations from Transformers) is employed in detecting the mental health status of employees using classification. This approach utilizes several libraries, such as pandas for data cleaning, scikit-learn for machine learning functionality, torch for deep learning, transformers for the BERT model, datasets for convenient access to data, and matplotlib and seaborn for visualization. The steps followed are:

1) Data Loading & Cleaning: The process begins by reading the dataset from a CSV file called data_to_be_cleansed.csv. Columns like Unnamed: 0 and title are dropped as they are not required. Missing values are detected and eliminated to maintain data consistency.
2) Data Preprocessing: The text data is tokenized using BERT's tokenizer. Tokenization transforms the text into numerical tokens that can be processed by the BERT model, ensuring the model comprehends the data effectively.
3) Model Training: The BertForSequenceClassification model is trained on the preprocessed data. This step involves multiple iterations where model parameters are fine-tuned to enhance classification accuracy.
4) Evaluation: The model's performance is assessed using classification metrics such as classification_report and confusion_matrix to evaluate its precision, recall, and overall effectiveness.

This methodology ensures efficient and precise detection of employee mental health status using advanced machine learning techniques.

C. Employee Attrition Prediction

The employee attrition prediction model utilizes a Gradient Boosting Classifier. This method leverages powerful libraries such as pandas for data manipulation, numpy for numerical computations, seaborn and matplotlib for visualization, scikit-learn for machine learning functions, and joblib for model serialization. The steps followed are:

1) Data Loading: The IBM HR Analytics Attrition dataset from Kaggle is loaded into a pandas DataFrame for analysis.
2) Data Preprocessing: Categorical variables are encoded using LabelEncoder() from scikit-learn, converting categorical data into numerical format for better model interpretation.
3) Feature Scaling: Numerical values are standardized to have a mean of 0 and standard deviation of 1. Standardization improves convergence rates and model performance in gradient boosting.
4) Model Training: The Gradient Boosting Classifier is trained using ensemble learning, where weak decision trees combine to form a strong predictive model for employee attrition.
5) Evaluation: Model performance is assessed using accuracy and a confusion matrix, ensuring it effectively predicts employee attrition.

This methodology provides a strong and accurate approach for predicting employee attrition using machine learning techniques.

4. Results and Discussion

Dataset: The *Employee Mental Health Dataset* was sourced from the Reddit Mental Health Data on Kaggle, which includes text-based responses from social media platforms. This dataset captures sentiments and contextual language used by employees, providing insights into their mental health status. The *Employee Attrition Dataset* was sourced from the IBM HR Analytics Attrition dataset on Kaggle, which includes structured HR records encompassing various factors such as job satisfaction, work-life balance, salary, job role, tenure, and promotions.

A. Employee Mental Health Detection

The mental health prediction model, using BERT, achieved an **accuracy of 81.40%**, demonstrating the effectiveness of deep learning techniques in analyzing text-based employee responses. Sentiment analysis and contextual understanding played a crucial role in prediction accuracy, with false positives observed in cases where sarcasm or humor was misinterpreted as negative sentiment.

Table I presents the comparative performance of different models used for employee mental health detection. Among the models evaluated, BERT achieved the highest accuracy of **81.4%**, outperforming traditional machine learning methods. The confusion matrix in Figure 2 provides insights into the model's classification performance, illustrating the number of true positives, true negatives, false positives, and false negatives.

TABLE I: Performance of Different Employee Mental Health Detection Models

Model	Accuracy	Precision	Recall
Logistic Regression	76.2%	75.1%	74.9%
Random Forest	79.3%	78.0%	77.5%
SVM	78.6%	77.4%	77.0%
BERT	**81.4%**	**80.5%**	**80.0%**

B. Employee Attrition Prediction

The attrition prediction model achieved an **accuracy of 88.78%**, highlighting the effectiveness of structured HR data in predicting employee turnover. Work-life balance and job satisfaction were identified as the most critical factors influencing attrition, with higher job involvement and career growth opportunities significantly reducing attrition rates.

As summarized in Table II, the Gradient Boosting model emerged as the best-performing classifier, achieving an accuracy

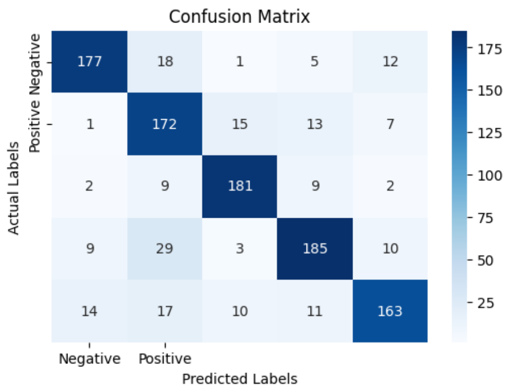

Fig. 2: Confusion Matrix for Employee Mental Health Detection Model

of **88.8%**, surpassing other models like Decision Tree, SVM, and Logistic Regression. Figure 3 illustrates the confusion matrix for the Gradient Boosting model, showcasing its effectiveness in distinguishing between employees likely to leave and those likely to stay.

TABLE II: Performance of Different Employee Attrition Prediction Models

Model	Accuracy	Precision	Recall
Decision Tree	84.5%	83.7%	83.5%
Support Vector Machine	86.1%	85.3%	85.1%
Logistic Regression	84.9%	84.0%	83.8%
Gradient Boosting	**88.8%**	**88.0%**	**87.8%**

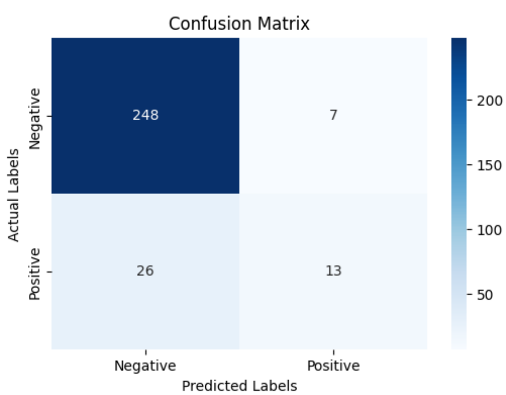

Fig. 3: Confusion Matrix for Employee Attrition Prediction Model

5. Conclusion

Application of deep machine learning algorithms, like BERT in mental health detection and Gradient Boosting in attrition prediction, has exhibited promising findings, indicating the potential of these models to yield insightful information regarding the well-being of employees and employee turnover. Contextual comprehension and sentiment analysis were important factors in the accuracy of the mental health model prediction. The attrition prediction model was able to make good use of structured HR data to find significant factors driving employee turnover. Yet, there are challenges and limitations that have to be met to further increase these models' efficacy.

A major challenge to mental health detection was the misunderstanding of sarcasm and humor, resulting in false positives. Furthermore, data sparsity and brevity in some of the responses restricted the model's capability to render useful insights. The attrition model, while maintaining high accuracy levels, would also gain from using real-time behavior measures like performance patterns and communication trends to enhance its predictability. Fixing these issues will be important to enhancing the reliability and robustness of such models.

In spite of all these challenges, the potential to enhance these models is enormous. With better and more extensive datasets and improvements in natural language processing and machine learning algorithms, organizations can use these technologies to proactively track and aid employee mental health and more effectively predict attrition. This way, this approach can ultimately help create a healthier and more productive workplace environment, allowing timely interventions and improved employee support.

References

[1] X. Yuan, "Social Media Sentiment Analysis and Mental Health Prediction using Deep Belief Network (DBN)," 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2024, pp. 1-5, doi: 10.1109/NMITCON62075.2024.10699157.

[2] S. Jain, M. Jadhav, N. Kolambe, and P. Bhavathankar, "Detection of Mental Disorder on Social Platform," IET Conference Proceedings, Apr. 12, 2022. doi: 10.1049/icp.2022.0612.

[3] W. Santos, S. Yoon, and I. Paraboni, "Mental Health Prediction from Social Media Text Using Mixture of Experts," in IEEE Latin America Transactions, vol. 21, no. 6, pp. 723-729, June 2023, doi: 10.1109/TLA.2023.10172137.

[4] H. Lin et al., "User-Level Psychological Stress Detection from Social Media Using Deep Neural Network," Proc. 22nd ACM Int. Conf. Multimedia, New York, NY, USA, 2014. doi: 10.1145/2647868.2654945.

[5] A. Sun and Z. Wu, "Early Detection of Mental Disorder Via Social Media Posts Using Deep Learning Models," in Gokhale, A., and Grant, E. (Eds.), Proceedings of Asia Pacific Computer Systems Conference 2021, Lecture Notes in Electrical Engineering, vol. 978, Springer, Singapore, 2023. doi: 10.1007/978-981-19-7904-0_13.

[6] A. Briand, H. Almeida, and M.J. Meurs, "Analysis of Social Media Posts for Early Detection of Mental Health Conditions," in Bagheri, E., and Cheung, J. (Eds.), Advances in Artificial Intelligence, Canadian AI 2018, Lecture Notes in Computer Science, vol. 10832, Springer, Cham, 2018. doi: 10.1007/978-3-319-89656-4_11.

[7] A. Mali and R.R. Sedamkar, "Prediction of Depression Using Machine Learning and NLP Approach," in Balas, V.E., Semwal, V.B., and Khandare, A. (Eds.), Intelligent Computing and Networking, Lecture Notes in Networks and Systems, vol. 301, Springer, Singapore, 2022. doi: 10.1007/978-981-16-4863-2_15.

[8] S. Shivaiah et al., "Prediction of Mental Illness Using Ensemble Model and Grid Search Hyperparameter Optimization," Bulletin of Electrical Engineering and Informatics, vol. 13, pp. 3726–3736, Oct. 2024. doi: 10.11591/eei.v13i5.7309.

[9] H. Lin et al., "User-Level Psychological Stress Detection from Social Media Using Deep Neural Network," in Proc. 22nd ACM Int. Conf. Multimedia (MM '14), Orlando, FL, USA, 2014, pp. 507–516. doi: 10.1145/2647868.2654945.

ISSN :2230-9993

[10] W. Zhang, K. Mao, and J. Chen, ”A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio, and Video,” Phenomics, vol. 4, pp. 234–249, 2024. doi: 10.1007/s43657-023-00152-8.

[11] S. Yadav, A. Jain, and D. Singh, ”Early Prediction of Employee Attrition Using Data Mining Techniques,” 2018 IEEE 8th International Advance Computing Conference (IACC), Greater Noida, India, 2018, pp. 349-354. doi: 10.1109/IADCC.2018.8692137.

[12] H. Zhang et al., ”Analysis and Prediction of Employee Turnover Characteristics Based on Machine Learning,” 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, Thailand, 2018, pp. 371-376. doi: 10.1109/ISCIT.2018.8587962.

[13] D.S. Sisodia, S. Vishwakarma, and A. Pujahari, ”Evaluation of Machine Learning Models for Employee Churn Prediction,” 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 1016-1020. doi: 10.1109/ICICI.2017.8365293.

[14] S.S. Alduayj and K. Rajpoot, ”Predicting Employee Attrition Using Machine Learning,” 2018 International Conference on Innovations in Information Technology (IIT), Al Ain, UAE, 2018, pp. 93-98. doi: 10.1109/INNOVATIONS.2018.8605976.

[15] M. Maharana, R. Rani, A. Dev, and A. Sharma, ”Automated Early Prediction of Employee Attrition in Industry Using Machine Learning Algorithms,” 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, India, 2022, pp. 1-6. doi: 10.1109/ICRITO56286.2022.9965017.

[16] L.S. Ganthi et al., ”Employee Attrition Prediction Using Machine Learning Algorithms,” in Saraswat, M., Roy, S., Chowdhury, C., and Gandomi, A.H. (Eds.), Proceedings of International Conference on Data Science and Applications, Lecture Notes in Networks and Systems, vol. 288, Springer, Singapore, 2022. doi: 10.1007/978-981-16-5120-5_44.

[17] L. Douaidi and H. Kheddouci, ”A New Approach for Employee Attrition Prediction,” in Braun, T., Cristea, D., and Jäschke, R. (Eds.), Graph-Based Representation and Reasoning, ICCS 2022, Lecture Notes in Computer Science, vol. 13403, Springer, Cham, 2022. doi: 10.1007/978-3-031-16663-1_9.

[18] M. Roy, H. Chhibber, P.V. Illavarasan, and A.K. Kar, ”Detecting and Characterizing Mental Health Using Social Media Analytics,” in Sharma, S.K., Dwivedi, Y.K., Metri, B., Lal, B., and Elbanna, A. (Eds.), Transfer, Diffusion and Adoption of Next-Generation Digital Technologies, TDIT 2023, IFIP Advances in Information and Communication Technology, vol. 698, Springer, Cham, 2024. doi: 10.1007/978-3-031-50192-0_31.

[19] M. Tounsi, H. Ameur, and S. Jamoussi, ”Deep Learning for Detecting Mental Disorder from User Content on Social Media,” SN Computer Science, vol. 6, p. 182, 2025. doi: 10.1007/s42979-025-03684-0.

[20] L. Alaskar, M. Crane, and M. Alduailij, ”Employee Turnover Prediction Using Machine Learning,” in Alfaries, A., Mengash, H., Yasar, A., and Shakshuki, E. (Eds.), Advances in Data Science, Cyber Security and IT Applications, ICC 2019, Communications in Computer and Information Science, vol. 1097, Springer, Cham, 2019. doi: 10.1007/978-3-030-36365-9_25.