IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Explainable Machine Learning Framework for Predicting Employee Attrition and Layoffs in HR Management

Mrs Sowjanya Balaga[1], Reddymalla Bhavani[2], Sri Vardhan Jandhyala[3], H Srikanth[4]
Assisstant Professor[1], Student[2], Student[3], Student[4], Department of
Computer Science and Engineering ,
Methodist College of Engineering and Technology, Hyderabad, Telangana, India

Abstract: This study presents a clear machine learning approach to identify employees who might leave an organization or face layoffs. The main goal is to help human resource teams make informed decisions based on data. The model combines predictions of voluntary attrition and involuntary layoffs into one framework, analyzing various employee-related attributes like job performance, salary, and organizational factors. Before applying machine learning methods, the data undergoes careful preparation, including cleaning, transformation, and feature selection. Since real-world HR data often has imbalances, techniques such as SMOTE are used to ensure better model learning. The prediction system uses ensemble algorithms, including Random Forest and XGBoost, which help enhance reliability and performance. To make the system more transparent and trustworthy, it includes explainability methods like SHAP and LIME. These methods highlight the key factors behind each prediction, allowing HR professionals to understand the reasons for employee risk levels instead of relying on black-box outputs. The proposed system also features a user-friendly dashboard that displays predictions, risk scores, and key contributing factors in a visual format. This helps organizations take timely actions, such as improving employee engagement or planning workforce changes more effectively. Overall, the framework aims to balance predictive performance with clarity, helping organizations lower turnover costs and improve workforce stability.

Keywords: Employee Attrition, Layoff Prediction, Machine Learning, Ensemble Learning, Explainable AI (XAI), SHAP, LIME, Human Resource Analytics, Feature Engineering, SMOTE, Workforce Management.

INTRODUCTION

Employee management has become one of the most critical challenges for modern organizations, especially with the increasing uncertainty in workforce stability. Two major issues that significantly impact organizations are employee attrition, where employees voluntarily leave their jobs, and layoffs, which are involuntary decisions made by organizations. Both scenarios lead to financial loss, reduced productivity, and disruption in organizational performance. As a result, there is a growing need for intelligent systems that can predict such events in advance and support better decision-making. Traditionally, human resource decisions have relied on experience, manual analysis, and periodic feedback mechanisms. However, these approaches are often reactive and may fail to identify potential risks early. With the availability of large-scale employee data, machine learning techniques provide an opportunity to analyze patterns and predict employee behavior more accurately. By considering factors such as performance, salary, job satisfaction, and work environment, predictive models can identify employees who are at higher risk of leaving or being laid off. In recent years, advanced machine learning models such as ensemble methods have shown strong performance in classification tasks. Algorithms

like Random Forest and XGBoost are capable of capturing complex relationships within data, making them suitable for workforce analytics. However, one major limitation of such models is their lack of transparency, often referred to as the "black-box" problem. This makes it difficult for HR professionals to trust and act upon the predictions without understanding the underlying reasons. In this paper, we propose an explainable machine learning framework that integrates both attrition and layoff prediction into a unified system. The framework includes data preprocessing, handling class imbalance using SMOTE, and applying ensemble learning techniques for accurate predictions.

BACKGROUND

Managing employees effectively has become increasingly challenging for organizations in today's dynamic work environment. Two major workforce-related issues are employee attrition, where employees voluntarily leave their jobs, and layoffs, where organizations reduce workforce due to business or financial reasons. Both situations can negatively affect productivity, increase operational costs, and disrupt organizational stability. High attrition leads to loss of skilled employees and increased hiring costs, while layoffs can impact employee morale and long-term performance. Therefore, understanding and predicting these events is crucial for better workforce management.

CHALLENGES

Data Imbalance Problem: One of the major challenges in employee attrition and layoff prediction is the imbalance in data. In most organizations, only a small number of employees actually leave or get laid off, while the majority stay. This makes it difficult for machine learning models to correctly identify high-risk employees, as the model may become biased toward the majority class. Even though techniques like SMOTE are used to handle this issue, they may sometimes introduce noise or lead to overfitting if not applied carefully.

Limited and Non-Generalized Datasets: Another challenge is the availability and quality of datasets. Many studies rely on synthetic or organization-specific data, which may not fully represent real-world scenarios across different industries or regions. As a result, models trained on such data may not perform well when applied to other organizations. This limits the generalization ability of the system and requires additional validation and retraining.

Lack of Interpretability in Models: Many high-performing machine learning models, especially ensemble and deep learning models, often act as "black boxes." This makes it difficult for HR professionals to understand why a particular employee is predicted to be at risk. Without clear explanations, it becomes challenging to trust and adopt these systems in real-world decision-making. This creates a need for explainable AI techniques, but integrating them effectively is still a challenge.

Real World Implementation and Adoption Issues: Even if a model performs well technically, implementing it in real-world HR systems is not straightforward. Organizations require user-friendly dashboards, proper data integration, and secure handling of employee information. Additionally, HR decisions involve ethical considerations, and relying completely on automated predictions may not always be acceptable. Ensuring practical usability along with fairness and transparency remains a significant challenge.

LIMITATIONS OF TRADITIONAL METHODS

Traditional approaches to managing employee attrition and layoffs mainly rely on manual analysis, basic statistical techniques, and periodic feedback methods such as exit interviews or performance reviews. These methods are often reactive rather than proactive, meaning organizations usually identify issues only after employees decide to leave. They also fail to handle large and complex datasets effectively, making it difficult to capture hidden patterns and relationships among multiple factors like job satisfaction, salary, performance, and work environment. In many cases, decisions are influenced by subjective judgment rather than data-driven insights, which can lead to inconsistent or biased outcomes. Additionally, traditional methods lack predictive capability, preventing organizations from identifying at-risk employees in advance. This limits their ability to take timely actions such as improving employee engagement or planning workforce changes, ultimately affecting organizational productivity and stability.

SCOPE AND APPLICATIONS

The scope of the proposed Explainable Machine Learning Framework for Employee Attrition and Layoff Prediction (2022–2025) involves developing a unified, data-driven system that integrates advanced preprocessing, class imbalance handling using SMOTE, and ensemble learning techniques such as Random Forest and Gradient Boosting. The framework is designed to analyze multidimensional HR data, including employee performance, compensation, job satisfaction, and organizational factors, to accurately identify high-risk individuals. It also incorporates Explainable AI techniques like SHAP and LIME to enhance transparency and interpretability, enabling organizations to move beyond black-box predictions. The system supports scalable deployment across different organizational environments and can be adapted to diverse workforce structures and datasets. Applications include generating accurate attrition and layoff risk predictions, identifying key factors influencing employee behavior, and enabling proactive workforce planning. The system supports real-time decision-making through interactive dashboards, assisting HR professionals in designing targeted retention strategies and optimizing resource allocation. By leveraging predictive analytics and explainability, the framework can significantly improve organizational efficiency, reduce employee turnover risks, and enhance decision-making accuracy, making it a valuable tool for modern HR management systems.

PROPOSED METHOD

Data Preparation

The data preparation phase plays a crucial role in building the proposed explainable machine learning framework, focusing on converting raw employee data into a structured and model-ready format. In this study, a synthetic HR dataset is utilized to simulate real-world employee attributes such as job role, salary, performance, and satisfaction levels. This approach allows controlled experimentation while preserving data privacy and enabling flexible feature design. The preparation process begins with data cleaning, where inconsistencies, noise, and missing values are handled to ensure data quality. Irrelevant or redundant features are removed to reduce complexity and improve model efficiency. Feature engineering techniques are then applied, including encoding categorical variables using label encoding and one-hot encoding, and scaling numerical features to maintain consistency across different ranges. Since employee attrition and layoff cases typically represent a smaller portion of the dataset, class imbalance is addressed using SMOTE (Synthetic Minority Oversampling Technique). This helps in generating balanced data and improves the model's ability to detect high-risk employees accurately.

Architecture

The proposed system architecture is designed as a multi-stage, explainable machine learning framework that integrates data processing, predictive modeling, and decision-support mechanisms for employee attrition and layoff analysis. The architecture begins with data ingestion and integration, where both public datasets and internal HR data are combined to form a unified dataset. This is followed by a data preprocessing stage, which includes cleaning, normalization, and feature engineering to ensure data consistency and improve model performance. A key component of the architecture is the class imbalance handling mechanism, where the dataset is evaluated and techniques such as SMOTE are applied when necessary to balance minority classes and enhance predictive accuracy. The processed data is then divided into training and testing sets and passed through multiple learning models. These include base learners such as Logistic Regression and KNN, ensemble methods like Random Forest and XGBoost, and deep learning models such as LSTM and DNN. A stacking and model selection process is performed to identify the best-performing model based on evaluation metrics. To ensure transparency, the architecture integrates an Explainable AI (XAI) framework, which provides both local explanations using SHAP and LIME and global insights through feature importance analysis. These explanations are presented through an HR dashboard, enabling users to understand the reasons behind predictions. Finally, the system generates attrition risk and layoff risk scores, which are used to support decision-making. The overall architecture enables a seamless flow from raw data to actionable insights, combining predictive accuracy with interpretability to support effective human resource management.

RESULTS

Data Preparation

The proposed system prepares data by transforming the synthetic HR dataset into a structured and model-ready format through processes such as cleaning, normalization, and feature engineering. Employee-related attributes including performance, salary, job satisfaction, and organizational factors are standardized and encoded to ensure consistency across the dataset. The application of SMOTE effectively balances class distribution by generating synthetic samples for minority classes, reducing bias in prediction. Implementation results demonstrate that the optimized data preparation pipeline significantly enhances model performance, achieving an overall accuracy of 96.28%, with balanced precision, recall, and F1-scores of approximately 0.96 across both classes. The system shows strong capability in identifying employee risk, with a recall of 0.98 for low-risk employees and 0.94 for high-risk cases, ensuring reliable detection. By replacing traditional manual analysis with a data-driven preprocessing approach, the system improves prediction reliability and enables accurate identification of at-risk employees, supporting effective and timely HR decision-making.

Architecture

The system is designed using a simple and flexible architecture that converts employee data into meaningful insights for predicting attrition and layoffs. It combines data processing, prediction logic, and explainable AI within an easy-to-use dashboard, allowing real-time analysis and visualization. Key employee factors such as salary, job satisfaction, and performance are used to assess risk levels, while role-based access ensures that both HR and employees can interact with the system securely.
By moving away from manual analysis and using an automated approach, the system helps improve the speed and reliability of decision-making.

- Achieved an overall accuracy of 96.28% with consistent performance across classes.
- Provides real-time insights through an interactive dashboard.
- Helps in identifying employees who are at higher risk.
- Reduces the need for manual analysis by offering clear and automated insights.

RESEARCH GAPS

A review of existing studies on employee attrition and layoff prediction shows that, although there has been progress in applying machine learning techniques in HR analytics, several important gaps still exist. These gaps affect the reliability, usability, and real-world adoption of such systems and can be grouped into technical, data-related, and practical limitations.

1. Technical and Model-Related Gaps

- **Dependence on single Models:** Many existing approaches rely on a single algorithm such as Logistic Regression or Random Forest. This limits performance, as no single model can handle all types of patterns present in complex HR data.
- **Accuracy vs. Interpretability Trade-off:** While advanced models improve prediction accuracy, they often behave like black boxes. This creates difficulty in understanding the reasoning behind predictions, reducing trust among HR professionals.
- **Limited Handling of Class Imbalance:**

Employee attrition and layoffs are relatively rare events, but many systems do not effectively address this imbalance, leading to biased predictions toward non-risk employees.

2. Data and Dataset Gaps

- **Limited and Non-Diverse Datasets:** Most studies rely on small, synthetic, or organization-specific datasets (such as IBM HR dataset), which do not fully represent real-world workforce diversity.
- **Lack of Multi-Source Data Integration:** Existing systems mainly use structured HR data and ignore other useful inputs such as employee feedback, external job market trends, or behavioral patterns.
- **Static Data Usage:** Many models are trained on static datasets and fail to capture changes over time, such as evolving employee performance or satisfaction levels.

3. Explainability and Decision-Making Gaps

- **Lack of Transparent Predictions:**

Many models provide predictions without clear explanations, making it difficult for HR teams to understand why an employee is considered at risk.

- **Limited Use of Explainable AI:**

Although techniques like SHAP and LIME exist, they are not widely integrated into real-world HR systems.

- **Gap Between Prediction and Action:**

Existing systems often stop at prediction and do not provide actionable insights or recommendations for HR decision-making.

4. System Implementation and Practical Gaps

- **Lack of Interactive Systems:**

Most research focuses on model development but does not provide user-friendly dashboards or tools for HR professionals.

- **Poor Real-Time Adaptability:**

Many systems do not support real-time data updates, making them less useful in dynamic organizational environments.

- **Limited Deployment in Real Scenarios:**

Several models remain at the experimental stage and are not integrated into actual HR workflows, limiting their practical impact.

5. Gaps Overall Gaps Identified

- **Lack of Unified Framework:**

Most existing works focus either on attrition or layoffs, but not both together in a single system.

- **Insufficient Focus on Usability:**

There is limited emphasis on making systems simple, interactive, and easy for HR teams to use.

- **Need for Scalable and Generalizable Models:**

Many models are designed for specific datasets and do not adapt well to different organizations or industries.

CONCLUSION

Employee attrition and layoffs continue to be major challenges for organizations, affecting both productivity and long-term growth. This work presents an explainable machine learning framework that combines data-driven prediction with transparency, allowing organizations to better understand and manage workforce risks. By using advanced preprocessing techniques, handling class imbalance, and applying ensemble models, the system is able to achieve strong predictive performance while maintaining reliability. A key contribution of this study is the integration of Explainable AI methods such as SHAP and LIME, which provide clear insights into the factors influencing employee risk. This helps bridge the gap between complex machine learning models and practical HR decision-making, making the system more trustworthy and usable in real-world scenarios. The results demonstrate that combining predictive analytics with interpretability and interactive visualization can support proactive decision-making. Instead of reacting after employees leave, organizations can now identify potential risks early and take appropriate actions. Overall, the proposed framework highlights the importance of building intelligent, transparent, and user-friendly systems for modern workforce management.

REFERENCES

[1] Baydili, İ. T., & Tasci, B. (2025). Predicting Employee Attrition: XAI-Powered Models for Managerial Decision-Making. Systems, 13(7), 583.

[2] Manafi Varkiani, S., Pattarin, F., Fabbri, T., & Fantoni, G. (2025). Predicting Employee Attrition and Explaining Its Determinants. Expert Systems with Applications, 272, 126575.

[3] Konar, K., Das, S., Das, S., & Misra, S. (2025). Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI. SN Computer Science, 6, 672.

[4] Talebi, H., & Khatibi Bardsiri, A. (2025). Machine Learning Approaches for Predicting Employee Turnover: A Systematic Review. Engineering Reports, 7(8), e70298.

[5] Prasad, T. S. L., Gunda, M., Esargundi, R., & Kandagatla, G. (2025). Predicting Employee Layoffs with Machine Learning: A Social Network and Data Mining Approach. Global Journal of Engineering Innovations & Interdisciplinary Research (GJEIIR), 5(4), 070.

[6] Barman, S., Biswas, M. R., Adnan, M. A., Nahar, N., Imam, M. H., Hossain, M. S., & Andersson, K. (2025). An Explainable Machine Learning Framework for Prediction of Employee Attrition. In Proceedings of the 2nd International Conference on Machine Intelligence and Emerging Technologies.

[7] Sekaran, K., & Shanmugam, S. (2022). Interpreting the Factors of Employee Attrition Using Explainable AI. In 2022 International Conference on Decision Aid Sciences and Applications (DASA).

[8] Mohiuddin, K., Alam, M. A., Alam, M. M., Welke, P., Martin, M., Lehmann, J., & Vahdati, S. (2023). Retention Is All You Need: Explainable AI for Employee Attrition in HR Decision Support. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM ’23).

[9] Joseph, T., Ggaliwango, M., Makanga, C., Mukwaba, D., Agaba, C., & Murindanyi, S. (2024). Explainable Machine Learning and Graph Neural Network Approaches for Predicting Employee Attrition. In 2024 Sixteenth International Conference on Contemporary Computing (IC3-2024).

[10] Marín Díaz, G., Galán Hernández, J. J., & Galdón Salvador, J. L. (2023). Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making. Mathematics, 11(22), 4677.