International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Predicting Employee Attrition Using Behavioral Data

Mr. Aditya D. Jathar, Mr. Prayag G. Kad, Mr. Anzer Kaleem Mohd.Ali Khan, Mr. Mohd. Murtuza Mohd. Mustafa, Mrs. Kalpana D. Sonval

Mr. Aditya D. Jathar, Comp. dept., ZCOER
Mr. Prayag G. Kad, Comp. dept., ZCOER
Mr. Anzer Kaleem Mohd.Ali Khan, Comp. dept., ZCOER
Mr. Mohd. Murtuza Mohd. Mustafa, Comp. dept., ZCOER
Mrs. Kalpana D. Sonval, Comp. dept., ZCOER

Abstract - Employee attrition prediction has become a vital area of research within human-resource analytics as organizations seek to retain top talent and reduce recruitment costs. Traditional HR systems depend mainly on static data such as salary, tenure, and department, offering limited insight into dynamic factors that influence employee decisions to leave. Recent advances in Machine Learning (ML) and behavioural analytics have enabled proactive identification of at-risk employees by examining digital behaviour patterns, engagement trends, and workplace activity. This paper surveys current approaches for predicting employee attrition, emphasizing behavioural feature extraction, classification models, evaluation metrics, and visualization dashboards.

Key Words: Employee Attrition, Behavioural Analytics, Machine Learning, Predictive Modelling, Human Resource Analytics, Data-driven Retention.

1.INTRODUCTION

Employee retention plays a crucial role in sustaining an organization's performance and culture. High turnover rates result in increased hiring costs, reduced morale, and disrupted operations. Traditional HR-based prediction systems analyze static attributes salary, designation, or performance ratings but often overlook behavioural indicators such as working hours, meeting load, communication frequency, and digital engagement. Machine Learning (ML) models can uncover patterns hidden in employee behaviour, enabling HR professionals to predict attrition risk and take timely action. This survey examines current research and system architectures developed for predicting employee attrition using behavioural data, focusing on feature engineering, learning algorithms, and visualization methods that transform HR management from reactive to proactive.

2. FUNCTIONAL & NON-FUNCTIONAL REQUIREMENTS OF EMPLOYEE ATTRITION PREDICTION

2.1 Functional Features

- Feature Engineering: Derives predictive indicators tenure trends, absenteeism, and digital participation from raw data.
- Model Training & Prediction: Uses algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost to classify employees as "likely to stay" or "likely to leave."
- Visualization Dashboard: Presents prediction outcomes, feature importance, and employee risk scores in real time for HR users.
- Data Balancing Using SMOTE: Applies the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance between "attrition" and "non-attrition" cases, ensuring that machine learning models learn effectively from both categories.

2.2 Non-Functional Features

- Accuracy: the model achieves high predictive performance through parameter tuning and cross-validation, ensuring consistent results and reduced overfitting.
- Scalability: built for large organizations, the system efficiently handles big employee datasets using cloud-based processing and modular architecture.
- Usability: provides an intuitive, cross-platform dashboard that allows HR teams to easily interpret predictions and monitor attrition insights.
- Data Integrity: maintains the accuracy and consistency of HR and behavioural data throughout preprocessing, model training, and visualization stages.
- Interpretability: the system includes explainable AI features such as feature importance visualization, enabling HR teams to understand why a prediction was made and build trust in the model's decisions.

3. METHODOLOGY

The methodology for this project involves collecting anonymized HR data such as salary, tenure, job role, department, performance and behavioural data like login/logout times, activity frequency, communication trends, engagement levels, followed by preprocessing steps such as cleaning, normalization, feature engineering, using SMOTE. Exploratory Data Analysis (EDA) is conducted to identify key patterns and correlations, after which multiple machine learning models such as Logistic Regression, Decision Tree, Random Forest, XGBoost are trained and evaluated using accuracy, precision. The system consists of a Python backend (Pandas, Scikit-learn, XGBoost), React.js frontend, and MySQL/PostgreSQL database, providing a dashboard that displays attrition and alerts for HR managers.

3.1 Data Collection

The dataset combines both traditional HR information and dynamic behavioural indicators. HR data includes employee details such as salary, tenure, job role, department, and performance ratings. Behavioral data is collected from activity logs, which capture login and logout times, communication frequency, work hours, project involvement, and engagement levels. Collecting both static and behavioural information helps in building a holistic view of employee performance and well-being, which is essential for accurate attrition prediction.

3.2 Data Preprocessing

Before model training, the collected data undergoes several preprocessing steps to ensure consistency and accuracy. Missing values are handled using imputation methods, and duplicate or irrelevant records are removed. Categorical variables such as department or job role are encoded, while numerical features are normalized to bring them to a common scale. To handle class imbalance where the number of employees staying far exceeds those who left the Synthetic Minority Oversampling Technique (SMOTE) is applied. This technique generates synthetic samples of the minority class, ensuring that the model learns equally from both attrition and non-attrition cases.

3.3 Exploratory Data Analysis (EDA)

EDA is performed to identify significant patterns, correlations, and key attributes that influence employee turnover. Visual tools such as heatmaps, histograms, and pair plots are used to study relationships between variables like salary, tenure, satisfaction level, and behavioural activity. The analysis helps in selecting the most relevant features and understanding the distribution of attrition within different departments or job roles.

3.4 System Architecture

The system architecture is built using a modular and scalable design. The backend is implemented in Python, leveraging libraries like Pandas, Scikit-learn, and XGBoost for data manipulation, model building, and prediction. The frontend uses React.js, providing an interactive user interface where HR managers can view insights and predictions. The database layer is developed using MySQL or PostgreSQL to store employee data, behavioural metrics, and model outputs securely. This three-tier architecture ensures smooth integration between data processing, visualization, and storage components.

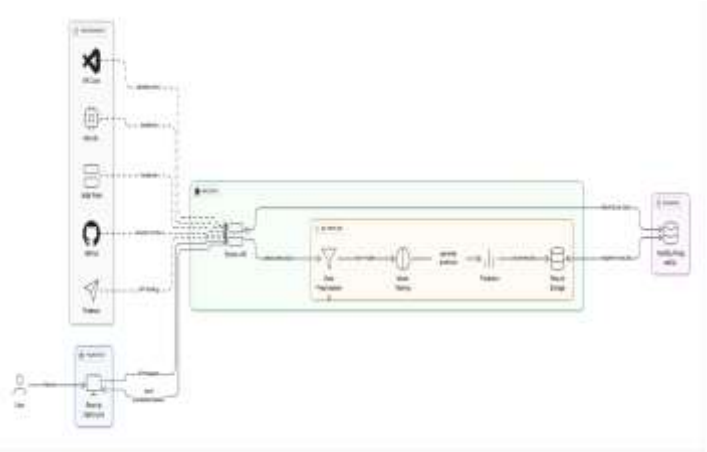

Fig 1. System Architecture

5. CONCLUSIONS

The study demonstrates that integrating behavioural data with traditional HR information greatly enhances the accuracy of employee attrition prediction. By combining factors such as login activity, communication trends, engagement levels, and performance metrics, the proposed system provides a deeper understanding of employee behaviour and early signs of disengagement. Machine learning algorithms like Logistic Regression, Random Forest, and XGBoost have proven effective in identifying at-risk employees, while techniques such as SMOTE help balance datasets and improve model reliability. The developed dashboard allows HR managers to visualize predictions, analyze key factors, and take timely actions to retain valuable talent.

ACKNOWLEDGEMENT

The authors thank Prof. Kalpana D. Sonval for her consistent guidance and valuable insights. The team also appreciates the Department of Computer Engineering, Zeal College of Engineering & Research, Pune, for providing the infrastructure and technical support throughout this project.

REFERENCES

[1] Maya John, Samia Abu Theeb, Madawi Alharbi, and Norah Bin Daham. "Analyzing Employee Attrition: A Visual Perspective". 7th
International Women in Data Science Conference at Prince Sultan University (WiDS PSU 2024).

[2] Umang Garg, Neha Gupta, Mahesh Manchanda, and Kamlesh Purohit. "Classification and Prediction of Employee Attrition Rate using Machine Learning Classifiers". Proceedings of the 7th International Conference on Inventive Computation Technologies (ICICT 2024).

[3] Pooja Nagpal, Avinash Pawar, and Sanjay H. M. "Predicting Employee Attrition through HR Analytics: A Machine Learning Approach". 4th International Conference on Innovative Practices in
Technology and Management (ICIPTM 2024).

[4] Dilina Muthgala, Samantha Mathara Arachchi. "Predicting Employee Attrition & Employee Retention Period using Supervised Learning". 6th International conference on advancements in computing (ICAC 2024).

[5] Chandana Mellachervu, Bhargavi Madiru. "Predicting Employee Attrition with Deep Learning and Ensemble Techniques for Optimized Workforce Management". International Conference on
Sustainable Communication Networks and Application (ICSCNA 2024).

[6] G Raja Rajeswari, Rajkumar Murugesan, R Aruna. "Predicting Employee Attrition through Machine Learning". International Conference on Smart Electronics and Communication (ICOSEC
2022).

[7] Kartik sekaran ,Shanmugam S "Interpreting the Factors of Employee Attrition using Explainable AID", 2022 International Conference on Decision Aid Sciences and Applications (DASA)"

[8] Rohit, Shreya Rai, Riya Sharma, Prakhar Singh, Vijay Kumar Dwivedi, Vimal Mishra "Prediction of Employee Attrition",2024 International Association for Computer Information System.

[9] Shikha Arora, Ashish Kr jha, Sunil Upadhyay "Predicting a Rise in Employee Attrition Rates Through the Utilization of People Analytics". 2023 International Conference on System Modeling & Advancement in Research Trends (SMART).

[10] Devadarshan K. Sardar, Sidhant Chourasiya, and V. Vijyalakshmi. "Employee Turnover Prediction by Machine Learning Techniques". 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT 2023), IEEE.