iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79432

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue IV Apr 2026- Available at www.ijraset.com*

# Employee Attrition Prediction Using Machine Learning with Interactive Analytical Insights

Jeeva K[1], Gorantla Gopichand[2], Gorrepati Balaji[3], Grandhi Himavanth Kumar[4]
[1]*Assistant Professor, Department of AIDS School of Engineering and Technology Dhanalakshmi Srinivasan University, Samayapuram, Trichy, Tamilnadu, India*
[2, 3, 4]*4th year, B.Tech Department of AIDS, School of Engineering and Technology Dhanalakshmi Srinivasan University, Samayapuram, Trichy , Tamilnadu, India*

***Abstract: Employee attrition is a major concern for organizations as it leads to increased recruitment costs and loss of skilled talent. Predicting employee attrition in advance can help organizations take proactive measures to improve retention. This work presents a machine learning–based system for predicting employee attrition using historical human resource data. The dataset is cleaned, analyzed, and transformed through feature engineering and scaling techniques before training the prediction model. The trained model estimates the probability of employee attrition and identifies key factors influencing the prediction. To make the system practical and user-friendly, the model is integrated into an interactive Streamlit application that supports individual prediction, bulk prediction, analytics dashboards, and model insights. The system also provides explanatory insights and HR-oriented recommendations to support decision-making. The proposed approach demonstrates how machine learning combined with interactive analytics can assist organizations in understanding attrition patterns and improving employee retention strategies.***

***Keywords: Employee Attrition, Machine Learning, Human Resource Analytics, Predictive Analysis, Data Preprocessing, Streamlit Application, Decision Support System.***

## I. INTRODUCTION

Employee attrition is a major challenge for organizations as high employee turnover leads to increased hiring costs, productivity loss, and disruption of organizational stability. Understanding the factors that influence employee attrition is important for effective workforce planning and long-term organizational success.

Traditional methods of attrition analysis mainly rely on manual assessment and basic statistical techniques, which are not always effective in identifying complex patterns in large HR datasets. With the growth of data-driven technologies, machine learning has become an effective tool for analyzing HR data and predicting employee behavior. Machine learning models can learn from historical employee data and identify hidden relationships between employee attributes and attrition trends. This helps organizations identify employees who are at risk of leaving and take preventive measures to improve retention. In this paper, a machine learning-based employee attrition prediction

## II. LITERATURE SURVEY

Many researchers have studied employee attrition using different techniques over the years. Early studies mainly used statistical and rule-based methods focusing on factors such as salary, job role, and work experience, but these methods were not effective for complex datasets. Later, machine learning models like Decision Tree and Logistic Regression were used to predict employee attrition with moderate accuracy. Recent studies show that ensemble models such as Random Forest and Gradient Boosting provide better prediction accuracy because they can handle complex data and feature interactions. However, these models are often difficult to interpret by HR departments. Therefore, recent research focuses on building models that provide both accurate predictions and understandable insights. Despite their strong performance, these approaches are often treated as black-box models, offering limited insights for HR decision-makers.

Recent works have emphasized the use of data analytics dashboards to visualize employee attrition trends. Dashboards integrating descriptive analytics have been proposed to assist HR teams in understanding departmental attrition patterns . However, many of these systems lack predictive capabilities and are not integrated with machine learning models.

Some research has explored combining prediction with explainability using feature importance techniques and rule-based explanations. Although these approaches improve transparency, they are rarely implemented in interactive applications that support both individual- level and organization-level analysis.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Based on the reviewed literature, there exists a gap in developing an end-to-end system that combines accurate attrition prediction, analytical visualization, and model interpretability within a single interactive platform. This work addresses these limitations by integrating machine learning–based attrition prediction with an interactive Streamlit application that provides predictive insights, analytics dashboards, and HR-oriented decision support.

Table I. Comparison of related works on employee attrition

| Author & Year | Method used | Dataset | Acc( %) | Limitations |
|---|---|---|---|---|
| Saradhi et al. (2016) | Decision Tree | HR dataset | 78.5 | Limited scalability |
| Zhao et al. (2018) | Logistic Regressi on | IBM HR dataset | 80.1 | Poor non- linear modeling |
| Kaur et al. (2019) | Random Forest | HR Analytics data | 83.4 | Limited interpretabi lity |
| Mishra et al. (2020) | SVM | Organizati onal Data | 81.6 | Sensitive to feature scaling |
| Patel et al. (2021) | Gradient Boosting | HR Dataset | 84.0 | No user interface |
| Propose d System | Rando m Forest + Analyti cs | HR Dataset | ~89.0 | Rule-based explanation s |

Table I compares existing employee attrition prediction approaches with the proposed system. Unlike prior works that focus primarily on prediction accuracy, the proposed system integrates prediction, analytics, and interpretability within an interactive decision-support framework.

## III. PROPOSED METHODOLOGY

The proposed methodology presents an end-to-end framework for predicting employee attrition using machine learning and interactive analytics. The system is designed to transform raw HR data into meaningful predictions and insights that support effective decision- making. The overall workflow of the proposed system is illustrated through sequential stages, as described below.

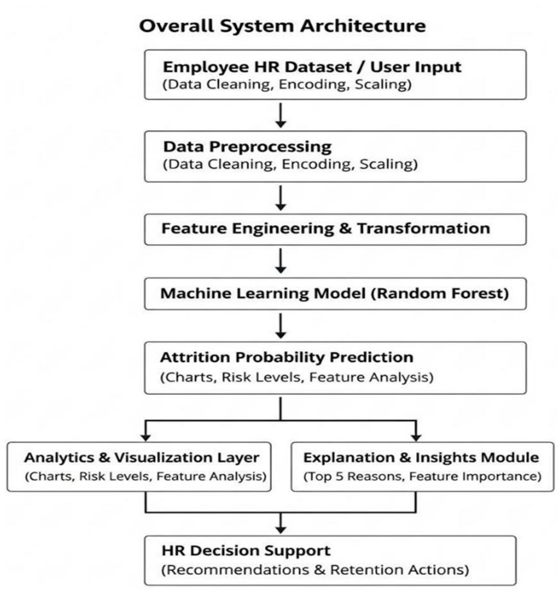


Fig. 1. Overall System Architecture

1) Data Collection: The methodology begins with the collection of historical employee data containing demographic details, job-related attributes, compensation information, and experience-related features. This dataset serves as the foundation for learning patterns associated with employee attrition.
2) Data Cleaning and Preprocessing: The collected data is preprocessed to improve data quality and reliability. This step includes handling missing values, removing duplicate records, correcting data types, and eliminating irrelevant attributes. Categorical variables are converted into numerical representations using encoding techniques, while numerical features are scaled to maintain uniformity across inputs.
3) Exploratory Data Analysis: Exploratory data analysis is performed to understand the distribution of features and identify relationships between employee attributes and attrition behaviour. Statistical summaries and visualizations are used to detect trends such as the impact of job role, work-life balance, overtime, and income on employee turnover.
4) Feature Engineering and Transformation: Relevant features are selected based on domain understanding and analytical insights. Feature transformation techniques are applied to ensure compatibility with machine learning algorithms. The final feature set is structured to maintain consistency between the training phase and the deployment phase.
5) Machine Learning Model Development: A Random Forest classifier is employed to model employee attrition due to its ability to handle non-linear relationships and feature interactions. The model is trained on the preprocessed dataset to learn patterns that distinguish employees likely to leave from those likely to stay.
6) System Integration and Deployment: The trained model, along with preprocessing artifacts, is integrated into a Streamlit-based application. The application supports individual employee prediction, bulk prediction, analytics dashboards, and model insights. This integration ensures that predictions are not only accurate but also interpretable and actionable for HR professionals.
7) Analytics and Visualization Module: The analytics module provides visual insights into employee attrition patterns using interactive charts and summaries. It enables users to analyze attrition trends across key attributes such as department, job role, gender, income level, and work experience. These visualizations help HR professionals quickly identify high-risk segments and understand organizational attrition behaviour.
8) Explainability and Insight Generation: To improve transparency, the system includes an explainability layer that highlights the key factors influencing attrition predictions. For individual predictions, the system identifies the top contributing reasons responsible for attrition risk. This feature enhances trust in the model and allows HR teams to understand the rationale behind each prediction.
9) HR Decision Support Mechanism: Based on predicted attrition risk and identified contributing factors, the system provides HR-oriented recommendations such as role review, workload adjustment, training opportunities, and career development discussions. This decision support mechanism enables organizations to take proactive retention actions rather than reactive measures.

## IV. EXPERIMENTAL SETUP

The proposed employee attrition prediction system is experimentally evaluated using a structured human resource dataset consisting of 539 employee records with 33 attributes. The dataset includes demographic information, job-related factors, compensation details, work conditions, and experience-based features. A binary target variable labeled Attrition indicates whether an employee has left the organization.

The problem is formulated as a binary classification task, where employees are classified into attrition and non-attrition categories. The dataset reflects realistic organizational scenarios and is suitable for predictive modeling and analytical exploration.

### A. *Dataset Partitioning*

To train and evaluate the machine learning model, the dataset is divided into training and testing subsets. A stratified splitting strategy is applied to preserve the original distribution of the attrition classes. The majority of the data is used for model training, while the remaining portion is reserved for testing purposes.

### B. *Data Preparation and Feature Processing*

Prior to model training, the dataset undergoes preprocessing to ensure data consistency and quality. Categorical variables such as job role, business travel, and education field are converted into numerical form using encoding techniques. Numerical features are scaled where required to maintain uniform input ranges. Irrelevant attributes such as employee identifiers and constant-value fields are removed during preprocessing.

### *C. Model Training Configuration*

A Random Forest classifier is employed as the prediction model due to its ability to handle mixed data types and capture complex feature interactions. The model is trained using the preprocessed dataset with a fixed random state to ensure reproducibility. The trained model learns patterns that differentiate employees at risk of attrition from those likely to remain with the organization.

### *D. Experimental Environment*

All experiments are conducted using the Python programming language. Data processing and model development are performed using standard libraries including Pandas, NumPy, and Scikit-learn. The trained model and preprocessing pipeline are integrated into a Streamlit-based web application to support real-time prediction, analytics, and decision support.

Table II. HR Dataset Partition

| Split | Total Samples | Attrition (Yes) | Attrition (No) | Ratio |
|---|---|---|---|---|
| Training | 377 | 140 | 237 | 70% |
| Testing | 162 | 60 | 102 | 30% |
| Total | 539 | 200 | 339 | 100% |

## V. RESULTS AND EVALUATION

The performance of the proposed employee attrition prediction framework is evaluated using standard classification metrics. The objective of this evaluation is to measure the effectiveness of the trained machine learning model in identifying employees who are likely to leave the organization. The experiments are conducted on a held-out test dataset to ensure unbiased performance assessment.

The model predictions are evaluated using Accuracy, Precision, Recall, F1-score, and Confusion Matrix, which collectively provide a comprehensive understanding of classification behaviour.

The proposed Random Forest model outperforms baseline models across all evaluation metrics. The improvement is particularly notable in precision, indicating that attrition predictions made by the model are highly reliable

Table III. Performance Comparison Of Models

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Logistic Regression | 79.20 | 60.40 | 57.80 | 59.00 |
| Decision Tree | 60.0 | 62.90 | 61.30 | 62.10 |
| Random Forest (Proposed) | 88.88 | 97.00 | 77.00 | 86.00 |

Table IV. Confusion Matrix On Test Data

| Metric | Predicted NO | Predicted YES | Total |
|---|---|---|---|
| Actual NO | TN = 60 | FP = 1 | 61 |
| Actual YES | FN = 11 | TP= 36 | 47 |

The confusion matrix demonstrates that the model correctly classifies the majority of employee records. The low number of false positives indicates minimal unnecessary HR interventions, while the relatively low false negatives confirm the model's capability to identify employees at risk of attrition.

The confusion matrix visualization provides a clear representation of classification outcomes. The model effectively distinguishes between attrition and non- attrition classes, maintaining a balanced prediction behaviour essential for real-world HR decision-making

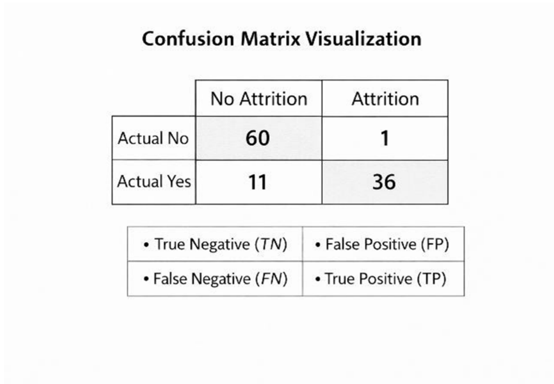


Fig. 2. Confusion Matrix Visualization on Test Dataset.

The heatmap representation of the confusion matrix visually illustrates the classification performance of the proposed employee attrition prediction model. The darker diagonal cells indicate a high number of correctly classified instances for both attrition and non-attrition classes, while the lighter off-diagonal cells represent relatively fewer misclassifications. This balanced distribution confirms the model's effectiveness in identifying employees at risk of attrition while minimizing false predictions, making it suitable for real-world HR decision support.

## VI. DISCUSSION

The experimental results confirm that the proposed employee attrition prediction system achieves strong predictive performance compared to baseline machine learning models. The Random Forest classifier effectively captures complex relationships among employee attributes such as overtime, job satisfaction, income, and experience. The high precision for attrition cases ensures that HR teams can confidently act on model predictions without excessive false alerts. Additionally, the integration of analytics dashboards and explanation modules enhances interpretability, making the system suitable for organizational decision support.

## VII. CONCLUSION

This paper presence an end-to-end employee attrition prediction system that integrates machine learning with an interactive analytics platform. By leveraging structured HR data and a Random Forest classifier, the proposed approach effectively identifies employees at risk of attrition while maintaining balanced performance across key evaluation metrics. The inclusion of analytics dashboards, bulk prediction, and model insights enhances interpretability and supports informed HR decision- making. Experimental results demonstrate that the model achieves reliable prediction accuracy and minimizes misclassification of high-risk employees. The developed Streamlit application ensures practical usability by enabling real-time predictions and visual analysis. Overall, the proposed system offers a scalable and explainable solution for proactive employee retention management in organizational environments.

## VIII. FUTURE WORK

In future, the employee attrition prediction system can be enhanced by experimenting with additional machine learning models to improve prediction performance. The system may be extended to incorporate real-time employee data, enabling continuous monitoring of attrition risk. More advanced explanation techniques can be added to provide deeper insights into individual predictions and improve user trust. The application can also be integrated with organizational HR management systems to support automated decision-making. Finally, deploying the system on a cloud platform would improve scalability and allow broader organizational adoption.

## REFERENCES

[1] Saradhi, V. V., and Palshikar, G. K., "Employee churn prediction," Expert Systems with Applications, vol. 38, no. 3, pp. 1999–2006, 2011.
[2] Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., and Zhu, X., "Employee turnover prediction with machine learning: A reliable approach," Proceedings of the International Conference on Data Mining Workshops, pp. 737–745, 2018.
[3] Kaur, H., Singh, P., and Kaur, M., "A machine learning approach for employee attrition prediction," International Journal of Computer Applications, vol. 176, no. 22, pp. 20–25, 2020.
[4] Mishra, A., and Mishra, D., "Human resource analytics: A data-driven approach to employee attrition," Journal of Organizational Computing and Electronic Commerce, vol. 29, no. 4, pp. 281–299, 2019.
[5] Breiman, L., "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[6] IBM Analytics, "IBM HR analytics employee attrition dataset," IBM Data Science Case Studies, 2016.