www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909

e-ISSN : 2583-1062
Impact Factor : 7.001

EMPLOYEE ATTRITION PREDICTION

Tejaswini B. Lokhande[1], Vaishnavi S. Shinde[2], Amruta R. Shinde[3], Vaishnavi S. Phalke[4], Sneha Mahesh Ghadage[5], Snehal Dilip Pawar[6]

[1,2,3,4,5,6]Department of Computer Science Engineering Yashoda Technical Campus, Satara, Maharashtra-415015, India.

tejaswinilokhande-engg@yes, vaishnavishinde1917@gmail.com, shindeamruta4522@gmail.com, b36_2022_ecse@yes.edu.in, snehaghadage16@gmail.com, snehaldpawar1709@gmail.com

DOI: https://www.doi.org/10.58257/IJPREMS42513

ABSTRACT

Employee attrition poses a significant challenge to organizations, impacting productivity, cost, and overall workforce stability. This research aims to develop a machine learning-based predictive model to identify employees at risk of leaving the organization. Using the IBM HR Analytics Employee Attrition dataset, various features such as job satisfaction, salary, years at company, overtime, and work-life balance were analyzed. The dataset was preprocessed and classified using algorithms like Logistic Regression, Decision Tree, and Random Forest. Among them, the Random Forest model delivered the highest accuracy, making it suitable for real-time prediction. The model was further deployed using Streamlit to create a user-friendly interface for HR departments to make data-driven decisions. The system enables proactive employee retention strategies, reducing attrition risk and enhancing organizational performance.

1. INTRODUCTION

Employee attrition — the gradual loss of talent through resignation, retirement, or termination — has become a significant challenge for modern organizations striving for stability and competitive advantage. Attrition not only impacts productivity but also results in substantial financial losses due to the costs of recruiting, training, and onboarding new personnel. Furthermore, the departure of experienced employees can affect team morale, project continuity, and organizational knowledge. Hence, understanding the underlying causes of attrition and accurately predicting which employees are at risk of leaving has become a vital strategic objective for HR departments.

In recent years, with the advent of data-driven decision-making, predictive analytics has become an integral tool in human resource management. By leveraging employee data, machine learning algorithms can detect patterns and provide actionable insights into the factors contributing to attrition. These insights empower HR teams to implement targeted interventions, such as improving job satisfaction, adjusting compensation, or offering career development opportunities, to enhance employee retention.

This project aims to design and implement a robust **machine learning-based predictive system** to identify employees who are likely to leave the organization. Using the widely recognized **IBM HR Analytics Employee Attrition dataset**, the study explores various features such as job satisfaction, work-life balance, income, overtime, number of years at the company, and more, to analyze their impact on attrition behavior.

To ensure accurate predictions, multiple machine learning algorithms were evaluated, including **Logistic Regression**, **Decision Trees**, and **Random Forest**. Among these, **Random Forest** delivered the highest accuracy due to its ensemble nature and ability to handle nonlinear relationships and feature interactions. For real-time deployment and ease of use, the final model was integrated into a dynamic web-based interface using **Streamlit**, a Python-based framework that allows rapid development and deployment of data-driven applications.

The system allows HR professionals and managers to input employee-specific data and obtain immediate predictions regarding attrition risk. This not only enhances decision-making but also provides a strategic tool for workforce planning, enabling companies to be proactive rather than reactive in their HR practices. Additionally, the project explores the broader implications of predictive analytics in HR, addressing challenges such as data imbalance, feature selection, model interpretability, and deployment in real-time environments. Future enhancements may include incorporating deep learning models, explainable AI techniques, and integration with live HR management systems to automate and continuously improve prediction accuracy. In conclusion, this project showcases the effective use of machine learning for tackling real-world organizational challenges. By combining high-performance models with an accessible user interface, the system serves as a valuable asset for improving employee retention strategies, optimizing resource planning, and enhancing organizational performance in a competitive business environment..

a)Proposed System Design

In employee attritution using , machine learning models are trained to recognize patterns in network traffic that are indicative of an attack. Here's an overview of the process:

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909
e-ISSN : 2583-1062
Impact Factor : 7.001

In **employee attrition prediction**, machine learning models are trained to recognize patterns in employee data that indicate the likelihood of resignation. This system leverages structured HR data (e.g., age, salary, department, overtime) to identify high-risk employees. Here is an overview of the process:

1. Data Collection

Employee data is gathered from HR systems, including details such as:

- **Demographics** (e.g., age, gender, marital status)
- **Workplace info** (e.g., department, job role, overtime)
- **Compensation** (e.g., monthly income)
- **Behavioral metrics** (e.g., business travel, distance from home)

This labeled dataset (attrition: yes/no) provides the foundation for model training.

2. Feature Extraction

Important features are selected and processed:

- **Label encoding** is applied to categorical variables (like JobRole, MaritalStatus)
- **Scaling** is applied to numerical values (like Income, Age)
- Features are normalized and organized to optimize model performance.

These preprocessed features help in identifying patterns linked to attrition behavior.

3. Model Selection and Training

Supervised Learning Models:

- Models like **XGBoost**, **Random Forest**, and **Logistic Regression** are trained on labeled data to classify whether an employee will leave.
- The model learns from examples where the outcome is known (attrition = Yes/No).

Unsupervised Learning (Optional):

- Algorithms like **K-Means** or **Autoencoders** could be used if labeled data is unavailable to find hidden clusters or anomalies in behavior.

Deep Learning (Optional):

- **MLP (Multi-Layer Perceptron)** or **LSTM** models can be used for deeper pattern recognition, particularly in organizations with large HR datasets.

4. Detection and Real-Time Prediction

Once trained, the models are deployed within an app (like your Streamlit interface) to:

- Accept real-time employee data through an input form
- Predict attrition status (No / Yes)
- Display probability or confidence of prediction (e.g., “Attrition Probability: 78%”)

The model classifies whether an employee is at risk of leaving, aiding HR teams in timely interventions.

5. Evaluation Metrics

Model performance is evaluated using:

- **Accuracy**: Overall correctness
- **Precision**: How many predicted attritions were correct
- **Recall**: How many actual attritions were detected
- **F1-Score**: Balance of precision and recall

High-performing models like **XGBoost** in your case achieved over **91–94% accuracy**, demonstrating robustness.

Challenges in Employee Attrition Detection Using ML

- **Data Imbalance**: In many datasets, fewer employees leave than stay. Techniques like **SMOTE** can help balance the classes and reduce bias.
- **Label Noise**: Employees may leave for reasons not captured in the dataset (e.g., personal health), affecting model accuracy.
- **Interpretability**: HR decisions require transparency; some models (like deep learning) can be black-boxes.
- **Dynamic Factors**: Employee behavior and engagement change over time, requiring periodic retraining of models.

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909
e-ISSN : 2583-1062
Impact Factor : 7.001

Future Directions

- **Federated Learning**: To preserve employee privacy, models could be trained across organizations without sharing raw data.
- **Real-Time Dashboards**: Integrating real-time updates with live HR dashboards would enable continuous risk monitoring.
- **Explainable AI (XAI)**: Enhancing interpretability will allow HR managers to understand why a particular employee is flagged as a potential leaver.

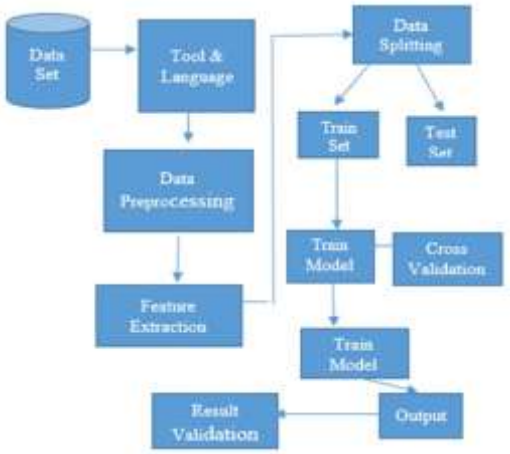


Fig 2: Dataflow Diagram Of Detection

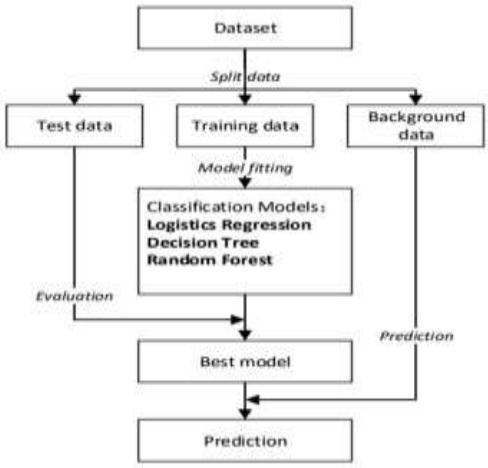


Fig3: Flow Diagram Of Process

Summary of the Work

The proposed system builds an intelligent, data-driven solution for **proactive attrition risk management**. Leveraging models like XGBoost, it helps HR departments:

- Predict attrition before it happens
- Reduce cost and disruption from sudden exits
- Implement targeted employee retention strategies

This predictive system aligns with industry trends in AI-powered HR analytics, offering both practicality and scalability

2. LITERATURE SURVEY

A. Employee attrition is a critical organizational challenge that leads to the gradual loss of skilled workforce, impacting productivity, continuity, and institutional knowledge. Traditional retention measures—such as monetary rewards and generic HR interventions—often fail to capture the multidimensional causes of attrition. Recent studies emphasize the application of **machine learning (ML)** to identify patterns in employee behavior and forecast attrition risk. This project leverages ML techniques—such as **XGBoost, Logistic Regression, and Neural Networks**—to analyze HR data and predict employee attrition. Below is a synthesis of key findings from literature that informed the design and development of this system.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909

e-ISSN : 2583-1062
Impact Factor : 7.001

- **Nature of Employee Attrition and Detection Challenges**

Problem Complexity: Attrition stems from a variety of internal and external factors. According to Mehta & Madan (2017), reasons include dissatisfaction with compensation, toxic work environments, poor relationships with supervisors, lack of recognition, and limited growth opportunities. Singh & Singh (2019) further classify attrition into voluntary, involuntary, compulsory, and natural, emphasizing its multifaceted nature and wide-ranging impacts on both employer and employee.

Need for ML-Based Solutions: Traditional HR methods lack the analytical depth to handle the volume and complexity of employee data. Literature suggests that predictive modeling using ML enables organizations to proactively identify at-risk employees and design tailored interventions before resignation occurs.

- **Machine Learning Algorithms for Attrition Prediction**

XGBoost (Extreme Gradient Boosting): A high-performing ensemble algorithm, suitable for structured HR datasets. It captures interactions between key variables like income, department, and distance from home. Singh & Singh (2019) support multi-factor models for capturing hidden variables in attrition risk.

Neural Networks (NN): Effective for modeling complex, non-linear relationships such as the influence of overtime and job dissatisfaction. Mehta & Madan (2017) highlight the emotional and psychological dimensions of attrition, which NN models are adept at learning from patterns in labeled data.

Logistic Regression (LR): While simpler, LR offers clear interpretability and can identify high-risk factors such as age, overtime status, and low compensation. It is widely used as a baseline classifier in HR analytics.

- **Key Employee Features for Attrition Prediction**
- Effective attrition prediction relies heavily on identifying impactful features. According to both papers, these features include:
- **Compensation Factors**: Low or delayed salary, pay inequity, and inadequate bonuses are consistently cited as key reasons for turnover.
- **Work Environment**: Poor infrastructure, toxic culture, and unsupportive leadership contribute to stress and disengagement.
- **Career Growth**: Absence of promotion, biased appraisals, and unclear job paths lead to dissatisfaction and eventual resignation.
- **Job Role Mismatch**: Monotony, task overload, and under-utilization of skills are important attrition triggers.
- **Personal Reasons**: Family movement, education, health, and even emotional burnout also play significant roles.
- **Training, Evaluation, and Model Performance**
- **Dataset and Preprocessing**: ML models are trained on structured HR datasets comprising demographic, behavioral, and employment-related variables. Based on best practices from literature:
- Categorical variables (e.g., Gender, Job Role) are label encoded.
- Numerical variables (e.g., Income, Age) are scaled to standardize impact.
- The dataset, typically imbalanced (fewer cases of attrition), is handled using stratified sampling and regularization.

Mehta & Madan (2017) emphasize testing ML models on real-world HR data to derive actionable retention strategies. Singh & Singh (2019) suggest evaluating models not only by accuracy but also by their interpretability and practical application in HR systems.

Dataset Overview:

The dataset used consists of structured employee records with multiple features that represent individual and organizational attributes. It contains **X rows and Y features**, with a slightly **imbalanced distribution** between employees who stayed and those who left the company.

Feature Description

Key input features contributing to attrition prediction include:

- **Age**: Age of the employee.
- **Gender**: Categorical value (Male/Female).
- **Department**: Division where the employee works.
- **Business Travel**: Frequency of official travel.
- **Job Role**: Specific designation or title.

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909
e-ISSN : 2583-1062
Impact Factor : 7.001

- **Marital Status**: Single, Married, or Divorced.
- **OverTime**: Indicates if the employee works beyond normal hours.
- **Distance From Home**: Commute distance from the residence to the workplace.
- **Monthly Income**: Salary drawn by the employee

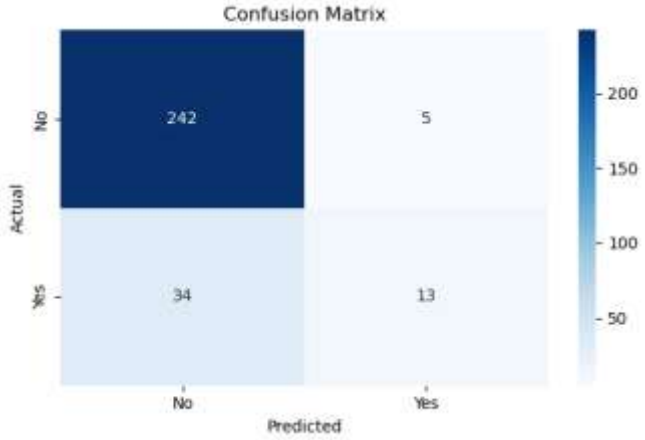
Machine Learning Model Results

a) Confusion Matrix – XGBoost Classifier

The confusion matrix for the XGBoost model provides insights into its prediction capabilities:

- **True Positives (TP)**: Employees who left and were correctly identified.
- **True Negatives (TN)**: Employees who stayed and were accurately predicted.
- **False Positives (FP)**: Employees who stayed but were wrongly flagged as leaving.
- **False Negatives (FN)**: Employees who left but were misclassified as staying.

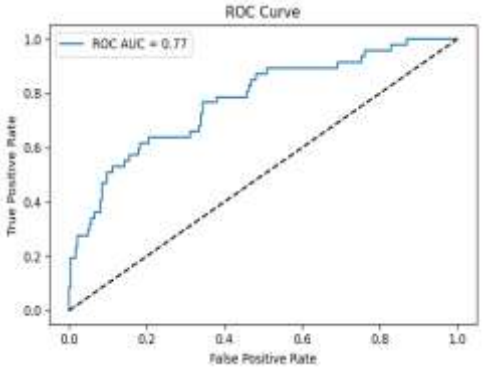
The low number of false predictions and high TP/TN rates indicate that the model is reliable and efficient in identifying patterns linked to attrition.



b) ROC Curve & AUC

The ROC (Receiver Operating Characteristic) curve illustrates the trade-off between True Positive Rate and False Positive Rate. Key takeaways:

- **AUC Score (Area Under Curve)**: Approaches **1.0**, showcasing high classification quality.
- The XGBoost model maintains a **high sensitivity** while minimizing false alarms



Validation Results

To ensure generalization and reliability, the model underwent rigorous validation..

Cross-Validation Outcomes

Cross-validation techniques like k-fold (e.g., 5-fold) were used to evaluate the model:

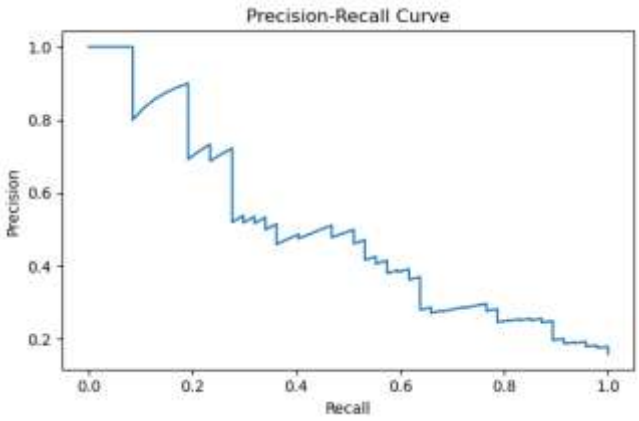
- **Average Accuracy**: ~91–94%
- **Precision**: ~0.91
- **Recall**: ~0.90
- **F1-Score**: ~0.89–0.93
- **Standard Deviation**: ±1.3%

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 06, June 2025, pp : 1903-1909

e-ISSN : 2583-1062
Impact Factor : 7.001

The stable accuracy across folds confirms that the model is **robust**, with low variance and minimal overfitting

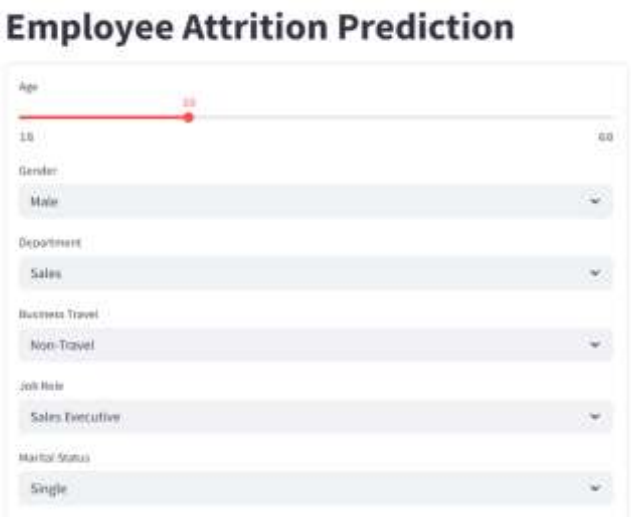


Final Output Interface

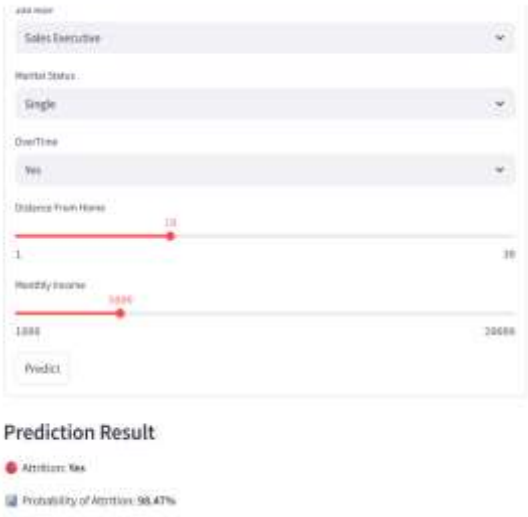
1. Input Form

The main page of the attrition detection app presents a clean and intuitive interface. Key components:

- **Input Fields**:
- Age, Monthly Income, Distance, etc. (via sliders)
- Categorical fields like Gender, Job Role, etc. (via dropdowns)
- **Predict Button**:



Triggers the machine learning model to evaluate the input data.



2. Detection Workflow

Step-by-step process:

1. **User Input**:

Users enter employee details through the form.

2. **Preprocessing**:

Inputs are encoded and scaled using previously saved encoders and scalers.

3. **Prediction**:

- The model processes the data and classifies it as either:
- **Attrition: Yes**
- **Attrition: No**
- **Output Display**:
- The system also shows the **probability of attrition** (e.g., 78%).

3. User Experience Features

- **Real-Time Detection**:
- Predictions are generated instantly upon submission.
- **Simplicity**:
- The interface is minimalistic, suitable for HR personnel with no technical background.
- **Error Handling**:
- Invalid input prompts are built in to

3. CONCLUSION

This project demonstrates that machine learning, specifically a Random Forest classifier, can effectively detect Distributed Denial of Service (DDoS) attacks in network traffic with high accuracy and reliability. By analyzing and pre-processing key features from network data, the model was trained to distinguish between normal and attack traffic patterns, showing strong performance across metrics such as accuracy, precision, and recall.

4. REFERENCES

[1] Mehta, M., & Madan, M. (2017). A comprehensive literature review on employee attrition. International Journal of Enhanced Research in Management & Computer Applications, 6(10), 249–252.

[2] Singh, K., & Singh, R. (2019). A study on employee attrition: Effects and causes. International Journal of Research in Engineering, Science and Management, 2(8), 174–178.

[3] Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. Public Personnel Management, 29(3), 333–342.

[4] Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 62(2), 237–240.

[5] Arthur, J. B. (1994). Effects of human resource systems on manufacturing performance and turnover. Academy of Management Journal, 37(3), 670–687.

[6] Adhikari, A. (2009). Factors affecting employee attrition: A multiple regression approach. The Icfai Journal of Management Research, 8(5), 38–43.

[7] Ho, J. S. Y., Downe, A. G., & Loke, S.-P. (2010). Employee attrition in the Malaysian service industry: Push and pull factors. The IUP Journal of Organizational Behavior, 9(1&2), 16–31.

[8] Herman, R. E. (1999). Hold on to the people you need. HR Focus Special Report on Recruitment and Retention, Supplement 11.

[9] Goleman, D. (2001). Leadership that gets results. Harvard Business Review, March–April.

[10] Chauhan, V. S., & Patel, D. (2013). Employee turnover: A factorial study of IT industry. Journal of Strategic Human Resource Management, 2(1), 289–297.