International Journal for Modern Trends in Science and Technology
Volume 12, Issue 05, pages 255-261
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue05.html
DOI: https://doi.org/10.5281/zenodo.19893174

Employee Attrition Prediction and Workforce Planning Dashboard using ML and Power BI

Vemulapudi Ram Charan, Vudasi Harshitha Singh, Yantrapati Rohith, Yarlagadda Harsha Vardhan, Yedida Sai Praveen, V. Pranav

Department of Computer Science and Engineering, Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh, India

To Cite this Article
Vemulapudi Ram Charan, Vudasi Harshitha Singh, Yantrapati Rohith, Yarlagadda Harsha Vardhan, Yedida Sai Praveen & V. Pranav (2026). Employee Attrition Prediction and Workforce Planning Dashboard using ML and Power BI. International Journal for Modern Trends in Science and Technology, 12(05), 255-261. https://doi.org/10.5281/zenodo.19893174

Article Info
Received: 28 March 2026; Revised: 24 April 2026; Accepted: 26 April 2026.

KEYWORDS
Attrition, Classification, Dashboard, Machine Learning, Workforce Planning

ABSTRACT

In today's competitive business landscape, retaining skilled employees is crucial for organizational success. This project presents an integrated solution for predicting employee attrition and supporting workforce planning through machine learning and data visualization. A comprehensive dataset of HR attributes was utilized, and a Random Forest classifier was developed to identify patterns influencing employee turnover. Key features such as job role, job satisfaction, environment satisfaction, and overtime were encoded and analysed. The trained model achieved strong predictive performance, allowing HR teams to identify at-risk employees proactively. Furthermore, the insights derived from model predictions and feature importance were visualized using Power BI dashboards, providing stakeholders with an intuitive and interactive platform for strategic decision-making. This project not only enhances attrition prediction accuracy but also empowers HR professionals to formulate data-driven retention strategies and workforce optimization plans.

1. INTRODUCTION

Employee attrition, or turnover, refers to employees leaving an organization either voluntarily (resignation) or involuntarily (termination or layoffs).

It has become a major concern for modern organizations due to its impact on productivity, efficiency, and overall performance. High attrition not only disrupts workflow but also leads to financial losses, with replacement costs often ranging from 50% to 200% of an employee's annual salary, including hiring, training, and productivity loss.

Beyond financial costs, attrition affects organizational stability. When experienced employees leave, they take valuable knowledge, skills, and relationships with them, which are difficult to replace. This can weaken teamwork, delay projects, and reduce work quality. Frequent turnover also lowers employee morale and job satisfaction, increasing the chances of further attrition.

Therefore, managing employee attrition is essential for maintaining a stable and competitive organization. With advancements in technology, organizations are increasingly using data-driven approaches to address workforce challenges. HR analytics helps in understanding employee behavior and identifying patterns that lead to attrition. Factors such as job role, salary, work environment, job satisfaction, work-life balance, overtime, and career growth opportunities play a key role in influencing employee decisions.

Machine learning, a branch of artificial intelligence, enhances predictive analysis by learning from historical data to make accurate predictions. In this project, machine learning models are used to predict the likelihood of employees leaving the organization. These models can identify complex relationships and hidden patterns that traditional methods may miss.

The project uses the Random Forest algorithm, an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. It works effectively with both numerical and categorical data. By analyzing attributes like age, job role, job satisfaction, environment satisfaction, and overtime, the model identifies employees at higher risk of leaving.

Data preprocessing is a crucial step, ensuring the dataset is clean and suitable for analysis. Techniques like label encoding are used to convert categorical data into numerical form, and irrelevant features are removed to improve performance. The dataset is then split into training and testing sets, and the model is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

To support decision-making, the project also uses data visualization. Power BI dashboards present attrition trends, risk levels, and key influencing factors in a clear and interactive format. HR managers can analyze patterns, identify high-risk departments, and understand the impact of factors like overtime and job satisfaction. This helps in implementing strategies such as improving work-life balance, offering better compensation, and enhancing career growth opportunities.

Overall, this project combines machine learning and data visualization to provide an effective solution for workforce management. By predicting attrition and identifying its causes, organizations can take proactive steps to reduce turnover, improve employee engagement, and ensure long-term success.

2. LITERATURE REVIEW

Employee attrition has been widely studied in Human Resource Management and Machine Learning due to its strong impact on organizational performance and sustainability. Researchers have explored various methods to understand the causes of employee turnover and to develop predictive models that help organizations reduce attrition. With the growth of data analytics and artificial intelligence, machine learning techniques are increasingly used to analyze employee data and improve prediction accuracy.

Many studies have identified key factors influencing employee attrition, including job satisfaction, salary, work-life balance, job role, organizational environment, and career growth opportunities. Employees with low satisfaction, high workload, and limited growth prospects are more likely to leave. Demographic factors such as age, education, and marital status also contribute to turnover. Understanding these factors is important for building effective predictive models and retention strategies.

Machine learning techniques are widely applied for attrition prediction because they can process large datasets and detect complex patterns. Common algorithms used include Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Random Forest. Logistic Regression is simple and interpretable, but struggles with non-linear data. Decision Trees are easy to understand but prone to overfitting. SVM performs well in complex scenarios but requires careful tuning, while KNN is simple but computationally expensive and sensitive to noise.

Ensemble methods such as Random Forest and Gradient Boosting generally provide better performance. Random Forest, in particular, improves accuracy by combining multiple decision trees and reduces overfitting. It also helps identify important features influencing attrition, making it highly suitable for HR analytics. Studies have shown that models using these techniques can achieve high accuracy, with some reaching around 85% or even above 90% using advanced methods.

Data preprocessing is another critical step highlighted in research. Handling missing values, encoding categorical variables, and selecting relevant features improve model performance. Techniques like Label Encoding and One-Hot Encoding are commonly used. Feature selection and engineering also help in reducing complexity and increasing accuracy.

Class imbalance is a common challenge in attrition datasets, as fewer employees leave compared to those who stay. This imbalance makes the model ignore the rare cases and focus almost entirely on the majority. To address this, techniques such as SMOTE,

oversampling, and undersampling are used to balance the dataset and improve prediction results.

Model evaluation is essential to measure effectiveness. While accuracy is commonly used, it is not sufficient for imbalanced data. Therefore, metrics like precision, recall, F1-score, and ROC-AUC are used to provide a more complete evaluation of model performance.

Visualization tools also play an important role in decision-making. Tools like Power BI and Tableau help present insights through interactive dashboards, enabling HR managers to analyze attrition trends, identify high-risk employees, and understand key influencing factors. This combination of machine learning and visualization improves practical decision-making.

Recent research has also explored deep learning and neural networks for attrition prediction. Although these models can achieve high accuracy, they often lack interpretability. Therefore, traditional models like Random Forest remain popular due to their balance between accuracy and explainability.

In conclusion, employee attrition prediction is a complex problem that requires a combination of machine learning, data preprocessing, and visualization techniques. Ensemble methods, especially Random Forest, have proven to be highly effective. Proper handling of data and the use of suitable evaluation metrics further improve model reliability, while visualization tools enhance decision-making and support effective retention strategies.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

The system architecture describes the overall design and workflow of the employee attrition prediction system. This project combines machine learning techniques with data visualization tools to create an efficient system for workforce analysis. The architecture follows a structured pipeline, starting from raw data input and ending with meaningful insights displayed through interactive dashboards.

The system follows a step-by-step process that includes data collection, preprocessing, model training, evaluation, prediction, and visualization. Each stage is connected and contributes to accurately predicting employee attrition and supporting workforce planning.

3.1 System Overview

The proposed system is designed as a machine learning pipeline that processes employee data to generate predictions and useful insights. The key components of the system include:

- Data Collection
- Data Preprocessing
- Model Training (Random Forest)
- Model Evaluation
- Prediction Generation
- Visualization using Power BI

This modular design ensures flexibility and scalability, allowing individual components to be improved without affecting the entire system.

3.2 Overall System Architecture

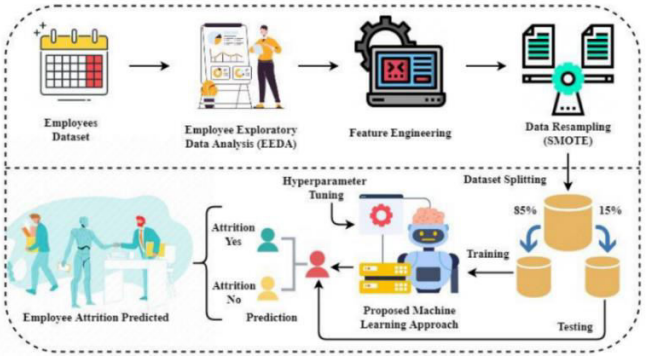

Fig 1: Overall System Architecture of Employee Attrition Prediction System

The figure 1 illustrates the complete workflow of the system, showing how the HR dataset is processed through preprocessing, model training, and evaluation before producing predictions and visual outputs.

3.3 Data Collection

The system begins with collecting HR data containing features such as age, job role, salary, job satisfaction, and overtime. These features serve as input variables for the prediction model, while the target variable is Attrition.

3.4 Data Preprocessing

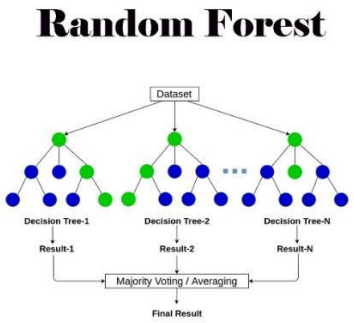

Fig 2: Random Forest Model Structure

The Random Forest algorithm uses multiple decision trees for prediction. Each tree provides an output, and the final result is determined through majority voting, which improves accuracy and reduces overfitting.

3.5 Model Evaluation

After training, the model is evaluated using the following metrics:

- Accuracy

- Confusion Matrix
- Precision
- Recall
- F1-score

These metrics help measure how effectively the model predicts employee attrition.

3.6 Prediction and Output Generation

The trained model generates predictions such as:

- Attrition (Yes/No)
- Probability scores
- Feature importance

These results are stored in CSV files for further analysis.

3.7 Visualization using Power BI

Fig 3: Power BI Dashboard for Attrition Analysis

The Power BI dashboard presents insights such as attrition rates, feature importance, and employee risk profiles using interactive visualizations.

3.8 Workflow of the System

- **Load dataset** – The system imports the employee dataset for analysis.
- **Preprocess data** – Data is cleaned and prepared by handling missing values and inconsistencies.
- **Encode categorical variables** – Non-Numerical data is converted into numerical form for model compatibility.
- **Split dataset** – The data is divided into training and testing sets for proper evaluation.
- **Train Random Forest model** – The model is trained using the prepared training data.
- **Evaluate model** – Performance is measured using metrics like accuracy, precision, recall, and F1-score.
- **Generate predictions** – The trained model predicts employee attrition outcomes.
- **Visualize results** – The results are displayed using visual tools for better interpretation.

4. RESULTS AND DISCUSSION

This chapter presents the results generated by the Employee Attrition Prediction system implemented using machine learning. The outputs are obtained from the prediction and evaluation phases of the model and are stored in CSV files and console outputs. These results help HR teams identify at-risk employees, understand key attrition drivers, and take data-driven actions.

4.1 Overview of Outputs

The system produces the following outputs:

- Console evaluation metrics
- attrition_predictions.csv
- feature_importance.csv
- Power BI dashboard visualization

These outputs collectively provide both prediction accuracy and business insights.

4.2 Console Evaluation Metrics

The model prints key performance metrics directly in the console.

4.2.1 Accuracy Score

Example: Accuracy = 0.87 (87%)

Accuracy indicates the overall success rate of the model's predictions. A score between 85% and 90% is considered good for the IBM HR dataset. However, accuracy alone may not be sufficient due to class imbalance in the dataset.

4.2.2 Confusion Matrix

Example:

[[2505]

[35 4]]

Structure:

[[TN,FP]

[FN, TP]]

- True Negatives (TN): 250 → Correctly predicted stay
- False Positives (FP): 5 → Predicted leave but actually stayed
- False Negatives (FN): 35 → Predicted stay but actually left
- True Positives (TP): 4 → Correctly predicted attrition

The model performs well in predicting employees who stay but fails to identify many employees who leave. This is due to class imbalance.

4.2.3 Classification Report

- Precision (Attrition class): 0.44
- Recall (Attrition class): 0.10
- F1-score: 0.17

The model has low recall for the attrition class, meaning it misses many employees who are likely to leave. This indicates the need for techniques like SMOTE or class weighting.

4.3 Attrition Predictions Output

The file attrition_predictions.csv contains employee-level predictions.

Each row includes:

- Employee features
- Actual attrition value (0 or 1)
- Predicted attrition value (0 or 1)

This output is useful for identifying high-risk employees. Employees with predicted value 1 can be targeted for retention strategies. Rows where the actual value is 1 but the predicted is 0 indicate missed cases that require further analysis.

4.4 Feature Importance Analysis

The file feature_importance.csv shows the importance of each feature in predicting attrition.

Top features include:

- Overtime
- Monthly Income
- Age
- Total Working Years
- Distance from Home
- Job Satisfaction

Overtime is the most significant factor, indicating that employees working excessive hours are more likely to leave. Low salary and younger age also contribute significantly to attrition.

4.5 HR Analytics Dashboard

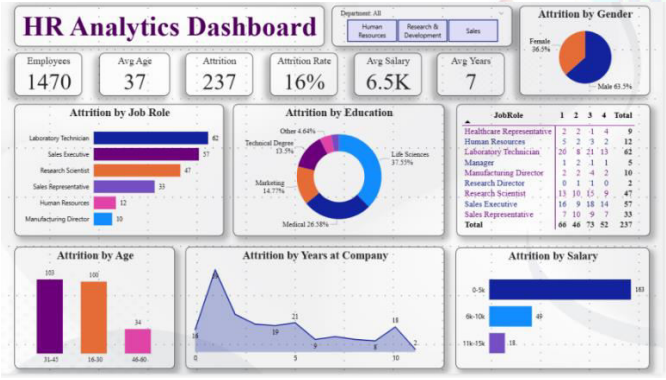

Fig 4. HR Analytics Dashboard

The HR Analytics Dashboard provides an interactive visualization of employee attrition patterns.

Key metrics include:

- Total Employees: 1470
- Attrition Count: 237
- Attrition Rate: 16%
- Average Age: 37
- Average Salary: 6.5K
- Average Years at Company: 7

The dashboard highlights important patterns:

- Higher attrition among male employees
- Most attrition occurs in roles like Laboratory Technician, Sales Executive, and Research Scientist
- Employees aged between 16 and 45 show higher attrition
- Majority of employees leaving earn less than 5K
- Highest attrition occurs within the first year of employment

These insights help HR teams identify problem areas and take corrective actions.

4.6 Interpretation of Results

The results indicate that employee attrition is influenced by multiple factors. Employees working overtime or earning low salaries are more likely to leave. Younger employees and those in early stages of their careers show higher turnover rates. Certain job roles also experience higher attrition.

The model successfully identifies patterns in employee behavior, but it struggles to detect all employees who leave due to class imbalance.

5. Discussion

The results of the employee attrition prediction model provide useful insights into workforce behavior and demonstrate the effectiveness of machine learning in HR analytics. The Random Forest model achieved an accuracy of around 85–87%, showing its ability to capture general patterns in employee data. However, a closer look at evaluation metrics such as precision, recall, and F1-score highlights certain limitations, especially in predicting employees who are likely to leave.

The confusion matrix indicates that the model performs well in identifying employees who stay but has difficulty detecting those who leave. This issue is mainly due to class imbalance, where the number of employees staying is much higher than those leaving. As a result, the model is biased toward the majority class, leading to more false negatives. In practical terms, this is a significant drawback because failing to identify at-risk employees limits the effectiveness of retention strategies.

Despite this limitation, feature importance analysis offers valuable insights into the key factors affecting attrition. Variables such as overtime, monthly income, age, and job satisfaction were found to be major contributors. These results align with real-world scenarios, where heavy workload, low pay, and poor job satisfaction often lead to employee turnover. Such insights help organizations focus on improving workload management, compensation, and employee engagement.

The use of Power BI dashboards improves the understanding of model outputs. Visualizations clearly present attrition trends across factors like job roles, age groups, salary levels, and experience. For instance,

higher attrition among younger employees and those with lower salaries suggests the need for better career growth opportunities and competitive pay. Identifying high-risk job roles also allows organizations to take targeted retention measures.

Another key finding is the presence of a high-risk period during the early stages of employment, especially within the first year. This indicates that onboarding and initial work experience play an important role in employee retention. Improving training programs, mentorship, and early engagement can help reduce early attrition. Although the current model provides meaningful insights, there is room for improvement. Techniques such as SMOTE, class weighting, and hyperparameter tuning can help address class imbalance and improve the model's ability to detect employees who may leave. Additionally, advanced algorithms like XGBoost or Gradient Boosting could further enhance prediction accuracy.

In conclusion, the model effectively identifies general trends and key factors influencing attrition, but improvements are needed to better detect at-risk employees. Combining machine learning with data visualization creates a strong foundation for data-driven decisions and proactive workforce management.

6. CONCLUSION AND FUTURE SCOPE

Employee attrition is a major challenge for organizations, as it affects productivity, efficiency, and overall performance. This project demonstrates the use of machine learning, specifically the Random Forest algorithm, to predict employee attrition using HR data. By analyzing factors such as job satisfaction, overtime, salary, and age, the model identifies patterns linked to employee turnover.

The results indicate that the model achieves a good level of accuracy, showing its ability to capture general trends in employee behavior. Feature importance analysis highlights key factors influencing attrition, including high overtime, low income, and reduced job satisfaction. These findings are consistent with real-world organizational scenarios and provide useful insights for improving employee retention.

The integration of machine learning with Power BI dashboards enhances the system by presenting insights in a clear and interactive way. These dashboards allow HR managers to analyze attrition trends, identify high-risk groups, and make informed decisions. This approach supports a shift from reactive to proactive workforce management.

However, the study also identifies certain limitations, particularly the model's lower performance in detecting employees likely to leave due to class imbalance in the dataset. Despite this, the project provides a strong base for implementing data-driven HR strategies.

There is scope for further improvement. Addressing class imbalance using techniques such as SMOTE, oversampling, undersampling, or class weighting can improve the prediction of at-risk employees. Additionally, advanced algorithms like Gradient Boosting, XGBoost, and deep learning models can enhance overall accuracy and performance.The system can also be expanded by integrating real-time data, enabling continuous monitoring of employee attrition. Adding more features, such as engagement scores, performance trends, promotion history, and employee feedback can provide deeper insights into workforce behavior.

From a visualization perspective, the Power BI dashboard can be improved with advanced analytics, automated reporting, and predictive alerts. Integration with enterprise HR systems would further increase its practical applicability.

In conclusion, this project provides a solid foundation for predicting employee attrition using machine learning and visualization tools. With further enhancements, it can evolve into an effective decision-support system that helps organizations manage employee retention and achieve long-term success.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Pelliccione, P., Kazman, R., Weber, I., and Liu, A., Cervantes, H., and Kazman, R., Designing Software Architectures: A Practical Approach Addison- WesleyProfessional, 2016.

[2] Peng, B. Statistical Analysis of Employee Retention. In Proceedings of the International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021), Nanjing, China, 26–28 November 2021; Volume 12163, pp. 7–15.

[3] 19 Employee Retention Statistics That Will Surprise You. (2022). Available online: https://www.apollotechnical.com/employeeretention-statistics/ (accessed on 6 May 2022).

[4] Here's What Your Turnover and Retention Rates Should Look Like. Available online: https://www.ceridian.com/blog/turnoverand-retention-rates-benchmark (accessed on 6 May 2022).

[5] SHRM Survey: Average Cost Per Hire Is $4129. Available online: https://www.businessmanagementdaily.com/46997/shrmsurvey-average-cost-per-hire-is-4129/ (accessed on 6 May 2022).

[6] Gandomi, A.H.; Chen, F.; Abualigah, L. Machine Learning Technologies for Big Data Analytics. Electronics 2022, 11, 421.

[7] Jia, X.; Cao, Y.; O'Connor, D.; Zhu, J.; Tsang, D.C.W.; Zou, B.; Hou, D. Mapping Soil Pollution by Using Drone Image Recognition

and Machine Learning at an Arsenic-Contaminated Agricultural Field. Environmental Pollution 2021, 270, 116281.

[8] Reshma Ramchandra, N.; Rajabhushanam, C. Machine Learning Algorithms Performance Evaluation in Traffic Flow Prediction. Materials Today Proceedings 2022, 51, 1046–1050.

[9] Aljedani, N.; Alotaibi, R.; Taileb, M. HMATC: Hierarchical Multi-Label Arabic Text Classification Model Using Machine Learning. Egyptian Informatics Journal 2021, 22, 225–237.

[10] Tsai, I.-J.; Shen, W.-C.; Lee, C.-L.; Wang, H.-D.; Lin, C.-Y. Machine Learning in Prediction of Bladder Cancer on Clinical Laboratory Data. Diagnostics 2022, 12, 203.

[11] Aggarwal, S.; Singh, M.; Chauhan, S.; Sharma, M.; Jain, D. Employee Attrition Prediction Using Machine Learning: A Comparative Study. Smart Innovation, Systems and Technologies 2022, 265, 453–466.

[12] Jadhav, A.; Mostafa, S.M.; Elmannai, H.; Karim, F.K. An Empirical Assessment of Performance of Data Balancing Techniques in Classification Tasks. Applied Sciences 2022, 12, 3928.

[13] Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.E.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. Heliyon 2018, 4, e00938.

[14] Ganthi, L.S.; Nallapaneni, Y.; Perumalsamy, D.; Mahalingam, K. Employee Attrition Prediction Using Machine Learning Algorithms. Lecture Notes in Networks and Systems 2022, 288, 577–596.