

HR Analytics – Predict Employee Attrition (Final Report)

1. Introduction

Employee attrition is a major challenge for organizations as it increases recruitment and training costs, reduces productivity, and affects overall morale.

Understanding the reasons behind attrition and predicting employees at risk of leaving allows HR teams to design effective retention strategies.

This project applies data analytics and machine learning to analyze HR data, uncover attrition patterns, and build predictive models.

2. Abstract

The objective of this project was to analyze employee attrition patterns and predict whether an employee is likely to leave the organization.

The dataset used was the IBM HR Analytics Attrition Dataset, which contains information about demographics, job roles, income, promotions, and tenure.

The project involved:

- Performing exploratory data analysis (EDA) to identify attrition drivers.
- Building classification models (Logistic Regression, Decision Tree).
- Applying SHAP analysis to interpret model predictions.
- Designing a Power BI dashboard for interactive HR insights.

3. Tools & Technologies

- Python: Pandas, Seaborn, Matplotlib (data cleaning, EDA, visualization)
- Scikit-learn: Logistic Regression, Decision Tree (machine learning models)
- SHAP: Feature importance and explainability
- Power BI: Interactive dashboard for business reporting

4. Methodology

Step 1 – Data Preparation

- Imported and cleaned the dataset.
- Handled missing values and encoded categorical features.
- Split dataset into training (80%) and testing (20%).

Step 2 – Exploratory Data Analysis (EDA)

Key findings from visualizations are shown below.

Model Building

- Logistic Regression: Accuracy = 83%, balanced predictions.
- Decision Tree (max_depth=5): Accuracy = 81%, slightly less accurate but easier to interpret.

Model Explainability with SHAP

SHAP analysis revealed that:

- Monthly Income and Years at Company were the strongest predictors of attrition.
- Age influenced attrition, with younger employees more likely to leave.
- Job Role also contributed, with Sales and R&D showing higher risks.

Power BI Dashboard

An interactive dashboard was created to present insights:

- Attrition by Department → High attrition in R&D and Sales.
- Attrition by Monthly Income → Lower salary bands showed higher attrition.
- Attrition by Years at Company → High risk in early career years.
- Filters allowed HR to drill down by department and role.

5. Key Results

- Logistic Regression Accuracy: 83%
- Decision Tree Accuracy: 81%
- SHAP values confirmed salary, tenure, and age as critical drivers of attrition.
- Dashboard insights showed attrition concentrated among low-income, early-tenure employees in Sales and R&D.

6. Conclusion & Recommendations

This project showed how HR analytics can be applied to predict and explain attrition by combining EDA, machine learning, SHAP interpretability, and dashboard visualization.

Recommendations:

1. Improve salary competitiveness for entry- and mid-level employees.
2. Provide career growth opportunities through promotions and skill development.
3. Focus retention efforts in Sales and R&D departments.
4. Support younger employees with mentorship, onboarding, and work-life balance programs.
5. Engage employees during their first 10 years at the company to reduce early attrition.

By acting on these insights, organizations can reduce attrition, improve employee satisfaction, and retain critical talent.