

# HR Analytics – Employee Attrition Project Report

## Introduction

Employee attrition is one of the most critical challenges faced by organizations, directly impacting workforce stability, productivity, and recruitment costs. By leveraging data analytics, businesses can identify the factors driving attrition, predict high-risk employees, and implement targeted retention strategies. This project focuses on analyzing employee data, applying machine learning models, and creating an interactive Power BI dashboard to deliver actionable insights into attrition trends.

## Abstract

The objective of this project was to conduct an end-to-end data analytics workflow for employee attrition analysis. The project began with Exploratory Data Analysis (EDA) in Python to uncover attrition patterns across departments, salary bands, job roles, and promotions. Predictive models (Logistic Regression and Decision Tree) were then developed to estimate attrition probability and identify high-risk employees. SHAP explainability was used to understand the contribution of each feature to the model's predictions. Finally, results were integrated into a Power BI Dashboard, providing HR executives with an easy-to-use visualization tool to monitor attrition drivers and employee risk levels.

## Tools Used

- Python (Pandas, Matplotlib, Seaborn, Scikit-learn, SHAP) → Data preprocessing, EDA, model building, explainability.
- Power BI → Dashboard development, KPIs, interactive analysis.
- Jolibr → Model saving and deployment.
- MS Excel/CSV → Data handling and transfer between stages.

## Steps Involved in Building the Project

### 1. Data Collection & Cleaning

- IBM HR Employee Attrition dataset with 1470 employees and 35 features.
- Missing values checked and categorical variables encoded.

### 2. Exploratory Data Analysis (EDA)

- Attrition Rate: 16% overall.
- Highest attrition in Sales (21%) and Low Salary Band (29%).
- Overtime workers attrition = 31% vs 10% for non-overtime.
- Attrition peaks for employees with 5–10 years since last promotion (23–24%).

### 3. Predictive Modeling

- Built Logistic Regression and Decision Tree models.
- Best model: Decision Tree with ~83% accuracy.
- High-risk employees identified (Sales Reps, Lab Technicians).

### 4. Explainability (SHAP)

- Top attrition drivers: Overtime, Salary, Promotion Gap, Job Role.
- SHAP summary plots provided transparency for model decisions.

### 5. Power BI Dashboard

- Page 1: HR Attrition Overview – KPIs, Department & Salary insights.
- Page 2: Workforce Insights – Attrition by Job Role, Promotion, Overtime.
- Page 3: Risk Analysis – Employee-level risk predictions & department risk view.

## **Conclusion**

This project demonstrates how data analytics and visualization can transform HR decision-making. Key insights reveal that Sales employees, overtime workers, and low-salary staff face the highest attrition risks. By focusing on these groups, HR can design targeted retention strategies, such as workload balance, career growth opportunities, and compensation adjustments. The integration of EDA + Predictive Modeling + Explainability + Dashboarding makes this a complete Data Analytics project, ready for real-world application and stakeholder presentation.