

1. Introduction

Employee attrition is a critical concern for HR departments and organizations aiming to retain top talent and reduce operational costs. Predicting which employees are likely to leave helps HR teams take proactive steps, such as targeted engagement strategies or compensation adjustments. This project leverages data analytics and machine learning to identify patterns and risk factors contributing to employee attrition using the IBM HR dataset.

2. Abstract

This project aims to analyze employee attrition data and build predictive models to forecast potential resignations. It includes extensive exploratory data analysis (EDA), model training (Logistic Regression and Decision Tree), feature importance evaluation using SHAP, and an interactive Power BI dashboard for stakeholder insights. The goal is to transform raw HR data into meaningful business recommendations to help reduce voluntary turnover.

3. Tools Used

Tool/Library	Purpose
Python (Pandas, NumPy)	Data manipulation and preprocessing
Seaborn & Matplotlib	Exploratory data analysis and visualization
Scikit-learn	Machine learning modeling and evaluation
SHAP	Model explainability and feature importance
Power BI	Business dashboard and interactive analysis
Colab	Code development and experimentation

4. Steps Involved in Building the Project

4.1. Data Cleaning & Preprocessing

- Removed irrelevant columns (e.g., `EmployeeNumber`, `Over18`, `StandardHours`)
- Handled categorical columns using `LabelEncoder`
- Scaled numerical features using `StandardScaler`

4.2. Exploratory Data Analysis (EDA)

- Analyzed attrition rates by **department**, **job role**, **monthly income**, **age**, and **promotion**
- Correlation heatmap revealed relationships between factors like **OverTime**, **Age**, and **MonthlyIncome**

4.3. Model Building

- Trained two models: **Logistic Regression** and **Decision Tree Classifier**
- Evaluated models using **accuracy**, **confusion matrix**, **classification report**, and **ROC-AUC score**

4.4. Model Comparison

- Decision Tree performed slightly better in terms of recall and ROC-AUC
- ROC curves plotted to compare both models visually

4.5. Feature Importance with SHAP

- Used SHAP (SHapley Additive exPlanations) to interpret model predictions
- Identified top influential features: **OverTime**, **MonthlyIncome**, **JobRole**, **Age**, **YearsAtCompany**

4.6. Power BI Dashboard

- Created interactive bar charts for attrition by: Department, Job Role, Salary Band, Age Group, Gender
- Enabled filters to allow users to drill into specific attrition trends

5. Conclusion

This HR Analytics project successfully combined data exploration, predictive modeling, and explainability techniques to uncover actionable insights into employee attrition. With SHAP, we achieved model transparency, identifying the most influential features behind resignations. The Power BI dashboard allows HR teams to visually monitor attrition risk and filter by department or role. These insights can inform better workforce planning, compensation strategies, and employee engagement programs to reduce future attrition.