Project Report

Title: HR Analytics – Predict Employee Attrition

---

1. Introduction

In the modern workforce, employee retention has become a critical success factor for organizations. High attrition rates can lead to talent loss, operational disruptions, and increased recruitment costs. To address this issue, companies are now leveraging data analytics to understand the root causes of attrition and implement proactive strategies. This project focuses on using machine learning and data visualization techniques to predict employee attrition and provide actionable insights for HR decision-making.

---

2. Abstract

This project analyses employee data from a multinational consultancy firm to identify patterns leading to employee attrition and build predictive models. The goal is twofold: first, to understand the factors contributing to attrition through Exploratory Data Analysis (EDA), and second, to create a machine learning model that predicts which employees are at risk of leaving. Further, SHAP value analysis was used to interpret the model outputs, and a Power BI dashboard was developed to visualize critical attrition trends and help HR teams take timely action.

---

3. Tools and Technologies Used

- **Python**: Data manipulation and analysis using Pandas, visualization with Seaborn and Matplotlib
- **Scikit-learn**: Building and evaluating machine learning models (Logistic Regression, Decision Tree)
- **SHAP**: Interpretability of model predictions
- **Power BI**: Dashboard for visualizing attrition trends and risk categories

---

4. Steps Involved in Building the Project

a. Data Pre-processing

- Loaded and cleaned the dataset
- Handled missing values and outliers
- Encoded categorical variables
- Scaled numerical features

b. Exploratory Data Analysis (EDA)

- Analysed department-wise attrition
- Evaluated attrition across salary bands and promotions
- Identified patterns in overtime, travel frequency, income, and tenure

c. Model Development

- Built Logistic Regression and Decision Tree models
- Achieved reliable performance with a balanced confusion matrix
- Evaluated models using accuracy, classification reports, and ROC curves

d. SHAP Value Analysis

- Used XGBoost for model training and SHAP to interpret predictions
- Found that key drivers of attrition included: Monthly Income, Overtime, Age, and Years at Company

e. Power BI Dashboard

- Developed an interactive dashboard to visualize:
  - High-risk roles and departments
  - Attrition by income levels, tenure, and promotions
  - Individual employee risk flags using model results

---

5. Conclusion

The HR Analytics project successfully demonstrated the potential of data-driven approaches in predicting and mitigating employee attrition. Through in-depth analysis and machine learning, we uncovered actionable insights such as high attrition risk among younger, lower-paid, and overworked employees. The SHAP analysis provided transparency into model decisions, and the Power BI dashboard enabled HR teams to act in real-time. These tools collectively empower organizations to retain talent effectively, reduce costs, and enhance workforce stability.

Model Implementation Summary (Python)

1. Conducted EDA and pre-processing using Pandas, Seaborn
2. Trained Logistic Regression and Decision Tree models using Scikit-learn
3. Evaluated models using accuracy, confusion matrix, ROC curve
4. Used SHAP values to interpret feature impact on attrition
5. Visualized results to support HR decisions

HR Analytics – Predict Employee Attrition ([Dataset](#))

Python ([File](#))

Power BI ([Dashboard](#))

Dashboard Snapshots ([HR Analytics - Snapshots](#))

Presentation ([PPTX](#))