HR Analytics - Predict Employee Attrition

Introduction

Employee attrition is one of the most critical challenges faced by organizations. High attrition leads to increased recruitment costs, loss of knowledge, and reduced productivity. This project aims to analyze HR data to understand the key drivers of attrition and build predictive models.

Abstract

The project focuses on identifying factors that contribute to employee attrition and predicting the likelihood of employees leaving. Using analytics and machine learning, we build models to classify employees based on attrition risk and provide actionable insights.

Tools Used

• Python (Pandas, Seaborn, Scikit-learn, SHAP) • Power BI for data visualization • ReportLab for report generation

Steps Involved in Building the Project

1. Data Collection: HR dataset prepared with employee details. 2. Exploratory Data Analysis (EDA): Analyzed department-wise attrition, salary bands, and promotions. 3. Model Building: Logistic Regression, Decision Tree, and Random Forest models were built and evaluated. 4. Explainability: SHAP values and feature importance were used to interpret model predictions. 5. Visualization: Created an interactive Power BI dashboard. 6. Reporting: Summarized findings and suggestions in PDF format.

Conclusion

The analysis highlighted that factors such as overtime, job satisfaction, and years at the company significantly affect attrition. The Decision Tree and Random Forest models achieved 98% accuracy, making them reliable tools for HR planning. This project provides actionable strategies to reduce attrition and improve employee retention.