HR Analytics — Predict Employee Attrition

Abstract

This project analyzes HR data to identify patterns and drivers of employee attrition and builds classification models to predict which employees are at risk. It combines exploratory data analysis, model training (Logistic Regression and Decision Tree), and explainability (SHAP or permutation importance).

Introduction

Employee attrition increases hiring costs and disrupts operations. This analytics project provides a reproducible pipeline to (1) analyze attrition patterns across departments and income bands, and (2) build explainable predictive models that help HR target retention interventions.

Tools Used

- Python (pandas, scikit-learn, matplotlib)
- SHAP (optional) for explainability
- Power BI for interactive dashboards
- GitHub for code and documentation

Steps Involved

1. Data loading & cleaning: handle missing values, convert 'Attrition' to binary.
2. EDA: department-wise attrition, income bands, promotion history.
3. Feature engineering: encode categorical variables and scale numerics.
4. Model training: Logistic Regression and Decision Tree; evaluate using accuracy, precision, recall, F1 and confusion matrix.
5. Explainability: SHAP summary plot or permutation importance fallback.
6. Dashboard: export plots and summarized tables for Power BI ingestion.
7. Recommendations: prioritize retention actions for high-risk groups.

Conclusion & Suggestions

The pipeline helps identify high-risk cohorts (by department, income band, or promotion history). Key interventions include targeted retention bonuses, clearer promotion tracks, manager training, and flexible schedules. Use model explanations to craft personalized interventions and monitor impact via Power BI dashboards.