HR Analytics – Employee Attrition Prediction

Executive Summary

This project analyzes HR data of 1,470 employees to understand the main causes of employee attrition and predict future resignation risks. Using machine learning models (Logistic Regression and Decision Tree), the project identifies key factors leading to attrition and provides actionable recommendations for HR teams.

Objectives

- Perform Exploratory Data Analysis (EDA) to identify attrition patterns
- Build classification models to predict employee attrition
- Explain model predictions using SHAP analysis
- Visualize attrition insights with Power BI dashboards
- Recommend preventive measures to reduce attrition

Tools & Technologies

Python (Pandas, Scikit-learn, Matplotlib, Seaborn, SHAP), Power BI

Key Insights

- Overtime and low salary are the strongest predictors of attrition.
- Employees with fewer promotions are more likely to resign.
- Research & Development employees show higher attrition than Sales.
- Young employees (25–35) are most likely to leave the organization.

Model Results

Two models were tested: - Logistic Regression (after scaling) achieved good accuracy with balanced precision and recall. - Decision Tree (max_depth=5) provided interpretable results and highlighted key attrition drivers. Confusion matrices confirmed that both models performed reasonably well in predicting employee attrition.

Recommendations

- Introduce flexible work policies to reduce overtime stress.
- Provide salary adjustments, especially for lower income bands.
- Offer timely promotions and career growth opportunities.
- Enhance employee engagement in high-risk departments such as R&D.;
- Conduct retention interviews to understand employee concerns proactively.

Project Deliverables

- Python notebook with data analysis and predictive modeling
- Power BI dashboard with interactive HR attrition insights
- PDF report summarizing findings and recommendations