HR Analytics - Predicting Employee Attrition

Introduction:

Employee attrition is a pressing challenge for many organizations. Understanding why employees leave and predicting future attrition is crucial for talent retention and business continuity. This project focuses on analyzing employee data to uncover attrition patterns and build predictive models.

Abstract:

We performed an end-to-end HR Analytics project to predict employee attrition. The project involved data preprocessing, exploratory data analysis, machine learning modeling (Logistic Regression, Decision Tree, Random Forest), SHAP value analysis for interpretability, and building a dynamic Power BI dashboard to visualize insights and support HR decisions.

Tools Used:

- Python (Pandas, Seaborn, Sklearn, SHAP)
- Power BI
- Jupyter Notebook

Steps Involved:

1. Data Cleaning: Removed irrelevant columns and converted categorical variables using encoding.
2. Exploratory Data Analysis: Analyzed trends in attrition across age, department, marital status, etc.
3. Model Building: Trained Logistic Regression, Decision Tree, and Random Forest models.
4. Model Evaluation: Compared accuracy, recall, and F1 scores.
5. SHAP Analysis: Identified key features influencing attrition using SHAP values.
6. Dashboard Creation: Built interactive Power BI dashboard with dynamic slicers and insights.

Conclusion:

This project provided clear insights into employee attrition trends. Key factors influencing attrition included age, daily rate, job role, and overtime. The predictive models showed high accuracy, and SHAP analysis made them interpretable. The Power BI dashboard allows HR professionals to explore attrition by department, age group, marital status, and more, enabling data-driven retention strategies.

HR Analytics - Predicting Employee Attrition

Model Comparison Summary:

Model	Accuracy	Recall (Class 1)	F1 Score (Class 1)
Logistic Regression	0.86	0.34	0.44
Decision Tree	0.76	0.36	0.32
Random Forest	0.83	0.11	0.17

Prepared by Sithu Sunny, Data Analyst Intern