HR ANALYTICS - PREDICTING EMPLOYEE ATTRITION

INTRODUCTION

Employee attrition is a critical challenge for many organizations. High turnover can significantly impact productivity, morale, and overall organizational performance. With the rise of data analytics, companies now have the capability to understand the root causes of employee resignation and take proactive steps to retain talent. This project aims to analyze HR data to identify the key factors leading to attrition and build a predictive model to forecast future resignations.

ABSTRACT

This project focuses on using data science to understand and predict employee attrition. Exploratory Data Analysis (EDA) techniques were applied to uncover patterns in department-wise attrition rates, salary bands, and promotion history. A Decision Tree classifier was developed using scikit-learn to predict whether an employee is likely to leave the organization. Furthermore, SHAP (SHapley Additive exPlanations) analysis was performed to interpret model predictions, and a Power BI dashboard was built to present the findings in an interactive and visual format. The combined approach of machine learning and explainability supports data-driven decision-making in HR departments.

TOOLS USED

- **Python (Jupyter Notebook):** Data processing, visualization, and model building
  - ➤ Libraries: Pandas, Seaborn, Scikit-learn, SHAP, Matplotlib
- **Power BI:** Visual analytics and dashboard design
- **SHAP:** Model explainability and interpretation of decision paths

STEPS INVOLVED IN BUILDING THE PROJECT

1. Data Preprocessing and EDA

- Loaded HR dataset using pandas.
- Explored attrition trends across departments, salary bands, and promotions.
  - Used seaborn plots for visual insights:
    - Stacked bar charts for department-wise attrition
    - Countplots and boxplots for salary and promotion features

2. Feature Engineering

- Converted categorical features using Label Encoding.
- Created a SalaryBand feature by binning MonthlyIncome into quartiles.

3. Model Building

- Trained a **Decision Tree Classifier** with max_depth=5 to predict attrition.
- Model inputs excluded redundant columns (Over18, EmployeeNumber, etc.).

4. Model Evaluation

- Generated a confusion matrix and classification report.
- Evaluated accuracy, precision, recall, and F1-score.

5. Model Explainability with SHAP

- Used SHAP to interpret feature impact on individual predictions.
- Summary plots and interaction plots showed:
  - Features like Age, YearsSinceLastPromotion, and MonthlyIncome had notable effects.
  - SHAP interaction plots showed mostly clustered values around zero, indicating limited interaction effect of Age in isolation.

6. Power BI Dashboard

- Designed a comprehensive Power BI dashboard:
  - Department-wise attrition
  - Attrition by job role, gender, and salary bands
  - Key KPIs and filters for interactivity.

CONCLUSION

This HR analytics project demonstrates how data science can effectively guide HR strategy. Using a combination of visual analytics, predictive modeling, and explainability tools, the organization gains insight into why employees leave and which segments are most at risk. Notably:

- Employees with low salaries or limited promotions were more likely to leave.
- Departmental trends revealed high attrition in specific teams like Sales.
- SHAP analysis helped validate model logic and built trust in predictions.