HR ANALYTICS - PREDICT EMPLOYEE ATTRITION

Introduction

Employee attrition significantly impacts an organization's stability, productivity, and operational costs. It is essential for HR teams to understand the factors influencing employee resignations to implement effective retention strategies. This project leverages Python and machine learning to analyze employee data, identify key factors contributing to attrition, and predict which employees are at risk of leaving. By integrating these insights into interactive dashboards in Power BI, the project enables HR teams to make informed, data-driven decisions to reduce turnover and strengthen workforce stability.

Abstract

This project analyzes HR data to identify key reasons for employee resignation and predict future attrition using machine learning. Python in Jupyter Notebook was used for data cleaning, analysis, and model building, while Power BI was used for creating interactive dashboards to visualize insights for HR decision-making.

Tools Used

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

Steps Involved in Building the Project

1. Data Exploration and Cleaning

- Loaded the dataset in Jupyter Notebook.
- Checked for null values, duplicates, and removed irrelevant columns.
- Cleaned and prepared the data for analysis.

2. Exploratory Data Analysis (EDA)

- Visualized attrition distribution across departments and salary bands.
- Used boxplots to analyze the impact of years since last promotion on attrition.

- Identified trends showing higher attrition in certain departments and among employees with delayed promotions.

3. Data Preprocessing

- Applied Label Encoding to the Attrition target variable.
- Performed One-Hot Encoding on categorical columns.
- Conducted train-test split with stratification and applied feature scaling.

4. Model Building and Evaluation

- Built a Logistic Regression model with class balancing to address data imbalance.
- Evaluated using accuracy (~76%), confusion matrix, classification report, and ROC-AUC.
- Achieved 62% recall for attrition prediction, providing actionable HR insights.

5. SHAP Analysis for Model Interpretation

- Applied SHAP analysis to interpret model predictions.
- Identified YearsSinceLastPromotion, MonthlyIncome, and JobRole as key factors influencing attrition.

6. Power BI Dashboard

- Exported the cleaned dataset to Power BI for visualization.
- Created:
    - Cards for Total Employees, Attrition Count, Attrition %.
    - Pie Chart: Attrition Yes vs No.
    - Bar Charts and column Charts for department and salary band analysis.
    - Line Chart: Attrition % vs Years Since Last Promotion.
    - Slicers for interactive filtering by department, gender, and marital status.
- Captured Power BI screenshots for reporting and submission.

Conclusion

The project successfully identified factors influencing employee attrition and built a predictive model using systematic EDA, machine learning, and visualization. These insights can guide HR teams to implement focused retention strategies, reduce attrition rates, and improve organizational performance.

Prepared by: Akhil Gaddam