HR Analytics – Predict Employee Attrition

Introduction

Employee attrition is a critical challenge for organizations, impacting operational efficiency, employee morale, and overall business performance. This project aims to understand the factors contributing to employee resignations and to predict future attrition using data-driven insights. The analysis leverages internal HR data to uncover key drivers, visualize trends, and support strategic retention planning.

Abstract

The project involved exploratory and predictive analysis on a dataset containing employee details such as demographics, job roles, satisfaction scores, and historical attrition data. Through data preparation, visualization, and machine learning techniques, the project identified high-risk attrition factors such as overwork, lack of promotions, and poor work-life balance. An interpretable decision tree model was developed to predict attrition probability, while SHAP analysis provided transparency into model decisions. The final outcome included a three-page interactive Power BI dashboard offering actionable insights for HR leadership to address risk segments proactively.

Tools Used

- **Python (Jupyter Notebook)**: Data cleaning, model building, SHAP analysis
- **Power BI**: Dashboard creation and visualization of trends and risk factors
- **Scikit-learn**: Decision Tree Classifier, confusion matrix
- **Pandas, Seaborn, Matplotlib**: Exploratory Data Analysis (EDA), data handling

Steps Involved in Building the Project

1. Data Understanding & Preparation

- Loaded and reviewed HR_Employee_Attrition.csv dataset.
- Retained important identifiers such as EmployeeNumber.
- Categorical columns were encoded using one-hot encoding.
- Missing/null values were handled — though minimal in the dataset.

2. Feature Engineering

- Created the Attrition_Flag column (1 = Yes, 0 = No) for binary classification.

- Grouped numerical columns (e.g., Age, YearsAtCompany, DistanceFromHome) into logical bins where needed for slicers in Power BI.
- Built new DAX measures for dashboard analysis (Attrition Rate %, Avg Satisfaction, Risk Segment Counts).

🔹 3. Exploratory Data Analysis (EDA)

- Identified trends such as:
  - Employees in roles like Sales Representative and Laboratory Technician had higher attrition rates.
  - High attrition correlated with longer YearsSinceLastPromotion.
  - Employees with frequent business travel and over time were more likely to leave.
  - Low job satisfaction and poor work-life balance were consistent risk signals.
- Age group 18–35 exhibited the highest attrition rates.

🔹 4. Predictive Modeling

- Used Decision Tree Classifier with depth control to avoid overfitting.
- Split data into training and test sets (80/20).
- Evaluated performance using a **confusion matrix** and **classification report**, achieving satisfactory accuracy and recall.
- Applied SHAP (SHapley Additive exPlanations) values to explain model predictions at a feature level:
  - OverTime, EnvironmentSatisfaction, and YearsSinceLastPromotion emerged as top drivers.

🔹 5. Dashboard Development (Power BI – 3 Pages)

Page 1: Executive Summary

- KPIs: Total Employees, Attrition Rate %, Avg Monthly Income, Avg Satisfaction.
- Donut Chart for attrition distribution.
- Bar chart: Department vs Attrition.
- Matrix: Gender × Job Role × Avg Tenure.
- Slicers for Department, Gender, Income, and Tenure.

Page 2: Attrition Risk Drivers

- Line chart: YearsSinceLastPromotion vs Attrition Rate.
- 100% stacked bar: BusinessTravel vs Attrition.
- Stacked bar: JobRole vs Attrition.

- Bar chart of top 5 manually ranked SHAP drivers.
- KPIs: Avg Promotion Gap (Attrited), OverTime %, High Risk Job Roles.

Page 3: Retention Strategy & Insights

- Line chart: Attrition Rate vs DistanceFromHome.
- Clustered chart: OverTime vs Attrition.
- Stacked bar: WorkLifeBalance vs Attrition.
- Bar: EducationField vs Attrition Rate.
- Table: High-Risk Employee Segments.
- Slicers: WorkLifeBalance, Distance, OverTime, Job Satisfaction.

Conclusion

This project provided a comprehensive analysis of employee attrition by integrating statistical and machine learning methods with visual storytelling. Insights such as high attrition among overworked and unpromoted staff members offer direct intervention opportunities for HR. The decision tree model, supported by SHAP explainability, gives transparent predictions, while the dashboard supports ongoing monitoring and strategic planning. This foundation empowers the organization to proactively reduce attrition and improve employee retention through data-backed HR policies.