HR Analytics - Predict Employee Attrition

Abstract

This project aims to predict employee attrition using a Decision Tree Classifier by analyzing HR metrics like salary, job role, and promotions. SHAP (SHapley Additive exPlanations) is used to interpret model decisions, revealing key features that influence attrition. A Power BI dashboard complements the model by visually presenting attrition trends for better HR decision-making.

Introduction

Employee retention is vital for organizational success. Early identification of attrition risks helps HR take proactive actions. This project integrates machine learning and data visualization to analyze employee data, build a predictive model, interpret it using SHAP, and display insights through an interactive Power BI dashboard.

Tools Used

- **Python** – For data manipulation, analysis, and modeling
- **Pandas, NumPy** – For data preprocessing
- **Matplotlib, Seaborn** – For visualizing data patterns
- **Scikit-learn** – For model development and evaluation
- **SHAP** – For explainable AI visualizations
- **Power BI** – For building interactive dashboards

Steps Involved in Building the Project

1. Data Collection & Cleaning

- The HR dataset with 35 features was imported.
- Dropped irrelevant columns such as EmployeeNumber, Over18, StandardHours, and EmployeeCount.

2. Exploratory Data Analysis (EDA)

- Visualized attrition trends across departments, salary bands, gender, education fields, and promotion status.
- Used pd.qcut() to categorize MonthlyIncome into salary bands (Low, Medium, High, Very High).

3. Label Encoding

- Converted categorical variables such as Department, Gender, and EducationField into numerical values using LabelEncoder.

4. Train-Test Split

- Data was split into 80% training and 20% testing using train_test_split.

5. Model Building

- Trained a **Decision Tree Classifier** with max_depth=5.
- Achieved an accuracy of **82.65%**.
- Observed high precision for the "No Attrition" class, but lower recall for the "Yes" class, indicating a class imbalance.

6. Model Evaluation

- Used classification_report and confusion_matrix for evaluation.
- Found that while overall accuracy was good, the model struggled to correctly identify employees who would actually leave (low recall for "Yes").

7. Model Explainability with SHAP

- Implemented shap.TreeExplainer to calculate SHAP values on the test dataset.
- Generated SHAP summary_plot to interpret the impact of features such as:
    - MonthlyIncome ,JobRole ,YearsSinceLastPromotion ,OverTime

JobSatisfaction

Power BI Dashboard Insights

Page 1: Employee Profile and Attrition Summary

- Displays Total Employees (1470), Attrition Count (237), Active Employees (1233), Attrition Rate (16.12%), and Average Age (37).
- Visuals include:
    - **Pie Charts** for attrition by Department, Gender, and Marital Status
    - **Bar Charts** for attrition rate by Education Field and OverTime
    - **Combined Charts** showing Gender vs. Attrition and Department-wise counts

Page 2: Job Features and Work Conditions

- Analyzes attrition in relation to:

Promotion Status, Salary Bands, Job Satisfaction Rating by Job Role, Years at Company

Years Since Last Promotion

- Interactive filters for Department, Gender, and Monthly Income allow in-depth analysis and targeted HR insights.

Conclusion

The Decision Tree model achieved 82.65% accuracy and provided a solid baseline for predicting employee attrition. While class imbalance affected prediction of "Yes" attrition, SHAP helped interpret key drivers. The Power BI dashboard enhanced understanding through interactive visuals, supporting better HR decision-making.