HR Analytics: Employee Attrition Prediction – Project Report

1. Introduction

This project aims to understand and predict employee attrition to inform strategic HR initiatives. By analyzing factors contributing to turnover, we seek to develop actionable insights for enhancing employee retention and fostering a stable workforce.

2. Abstract

This report details an HR analytics project utilizing the WA_Fn-UseC_-HR-Employee-Attrition.csv dataset. It involved Exploratory Data Analysis (EDA) on key attrition factors (department, salary, promotions). A Decision Tree Classifier was built, achieving 81% accuracy, though facing challenges in predicting the minority 'attrition' class (Precision: 0.37, Recall: 0.25). SHAP value analysis was performed to interpret model predictions. Deliverables include a Power BI dashboard, model accuracy report, and attrition prevention suggestions. This project provides data-driven insights for proactive talent management.

3. Tools Used

- Python: Pandas (data manipulation), Seaborn (EDA visualization), Scikit-learn (Decision Tree model), SHAP (model interpretability).
- Power BI: Interactive dashboard development for attrition factor visualization.

4. Steps Involved in Building the Project

- Step 1: Data Acquisition & EDA (eda.ipynb)
- Loaded WA_Fn-UseC_-HR-Employee-Attrition.csv.
- Performed comprehensive EDA to analyze attrition trends across various dimensions (e.g., Department, MonthlyIncome, YearsSinceLastPromotion, JobRole, OverTime, Age, Job/Environment Satisfaction).

- Step 2: Data Preprocessing & Feature Engineering
- Converted categorical variables to numerical using pd.get_dummies.
- Dropped irrelevant columns (EmployeeNumber, EmployeeCount, Over18, StandardHours).
- Mapped Attrition to numerical (Yes: 1, No: 0).
- Step 3: Classification Model Building (classification_model.ipynb)

- o Split data (70% train, 30% test) with stratification for `Attrition` class.
- o Trained a `DecisionTreeClassifier` (`random_state=42`, `max_depth=5`).
- o Made predictions on the test set.

- Step 4: Model Evaluation & Reporting (`classification_model.ipynb`)
- o Confusion Matrix: [[339 31], [53 18]] (339 True Negatives, 18 True Positives; 31 False Positives, 53 False Negatives).
- o Classification Report: Precision: 0.86 (Class 0), 0.37 (Class 1); Recall: 0.92 (Class 0), 0.25 (Class 1).
- o Accuracy Score: 0.81 (81%).
- Step 5: SHAP Value Analysis (`classification_model.ipynb`)
- o Computed SHAP values using `shap.TreeExplainer` for model interpretability.
- o Code is included to generate `shap_attrition_long.csv` for Power BI visualization. Ensure the relevant cell in the notebook is executed to generate this file.
- Step 6: Power BI Dashboard Development (`hr_analytics.pbix`)
- o Designed an interactive dashboard to visualize attrition factors, including overall rate, departmental trends, job role impact, overtime influence, age groups, income levels, and employee tenure. Features interactive slicers.

- Step 7: Attrition Prevention Suggestions
- o Formulated data-driven recommendations based on EDA, model insights, and SHAP analysis to address identified attrition drivers.

5. Conclusion

This project successfully established a framework for understanding and predicting employee attrition. EDA provided initial insights, and the Decision Tree model achieved 81% accuracy, though with recognized limitations in predicting actual attrition cases (low recall for 'Yes' class). SHAP analysis offered crucial interpretability. The Power BI dashboard provides an interactive tool for stakeholders. Future efforts will focus on improving the predictive model's performance for the attrition class and refining prevention strategies through continuous monitoring. This project enables a proactive and data-informed approach to talent retention.