HR Analytics – Predicting Employee Attrition

Internship Project Report

1. Introduction

Employee attrition is a major concern for organizations, affecting productivity and increasing hiring costs. In this project, we analyze HR data to understand the factors behind employee resignations and build a model to predict attrition risk.

2. Abstract

This project uses an IBM HR dataset to explore attrition patterns across departments, job roles, income levels, and work experience. We performed Exploratory Data Analysis (EDA) using Python and created an interactive dashboard in Power BI. A Logistic Regression model was trained to predict attrition, and SHAP values were used for model interpretation. This solution enables HR teams to identify and retain at-risk employees more effectively.

3. Tools Used

- Python (Pandas, Seaborn, Scikit-learn, SHAP)
- Power BI
- CSV Dataset: IBM HR Analytics (Kaggle)

4. Steps Involved

1. **Data Collection & Cleaning**
   a. Loaded IBM HR dataset in Python and Power BI
   b. Checked for missing values and encoded categorical data
2. **Exploratory Data Analysis (EDA)**
   a. Identified patterns in attrition by department, job role, income, and years at company
   b. Found Sales and Human Resources had relatively higher attrition
3. **Model Building**
   a. Built a Logistic Regression model using Scikit-learn
   b. Evaluated with Accuracy Score and Confusion Matrix
   c. Used SHAP for feature importance (top drivers: MonthlyIncome, JobRole, YearsAtCompany)

4. **Dashboard Creation**
    a. Created a Power BI dashboard showing:
        i.   Attrition by department, income, job role, and tenure
        ii. Slicers to filter by gender, marital status, education field, etc.

5. Conclusion

The model provides an accurate prediction of employee attrition and reveals key contributing factors. This insight can help HR departments take preventive measures, such as better salary planning, focused engagement for certain job roles, and improved retention strategies for early-tenure employees.

This project showcases the power of data in making strategic HR decisions.