HR ANALYTICS – PREDICT EMPLOYEE ATTRITION

Introduction

Employee attrition has a major impact on an organization's stability, productivity, and overall costs. To maintain a strong and consistent workforce, HR teams must understand the causes of employee turnover and design effective retention strategies. This project applies Python and machine learning techniques to analyze employee data, identify the main drivers of attrition, and predict which employees are most likely to resign. The insights are presented through interactive Power BI dashboards, allowing HR professionals to make data-driven decisions that reduce turnover and strengthen workforce stability.

Abstract

This project aims to identify the key reasons behind employee resignations and predict future attrition using machine learning. Data cleaning, analysis, and model building were performed in Python (Jupyter Notebook), while Power BI was used to create interactive dashboards for HR data visualization and decision-making.

Tools Used

• Jupyter Notebook (Python libraries: Pandas, Seaborn, Scikit-learn, SHAP) • Power BI

Project Development Steps

1. Data Exploration and Cleaning - Loaded and examined the dataset in Jupyter Notebook. - Addressed missing values, removed duplicates, and dropped irrelevant columns. - Prepared a clean dataset for analysis and model training. 2. Exploratory Data Analysis (EDA) - Analyzed attrition distribution across departments and salary ranges. - Used boxplots to examine how "years since last promotion" impacts attrition. - Observed higher turnover rates in specific departments and among employees with longer promotion gaps. 3. Data Preprocessing - Encoded the target variable Attrition using Label Encoding. - Applied One-Hot Encoding to categorical features. - Conducted a stratified train-test split and standardized the feature values. 4. Model Building and Evaluation - Developed a Logistic Regression model with class balancing to address data imbalance. - Evaluated the model using accuracy (~76%), confusion matrix, classification report, and ROC-AUC metrics. - Achieved 62% recall, offering practical insights for HR policy and retention strategies. 5. Model Interpretation using SHAP - Used SHAP (SHapley Additive exPlanations) for interpretability of model predictions. - Identified YearsSinceLastPromotion, MonthlyIncome, and JobRole as the most significant factors influencing attrition. 6. Power BI Dashboard - Exported the cleaned dataset to Power BI for visualization. - Created interactive components including: • KPI Cards: Total Employees, Attrition Count, Attrition Percentage • Pie Chart: Attrition Yes vs No • Bar/Column Charts: Department and salary-wise attrition trends • Line Chart: Attrition % vs Years Since Last

Promotion • Slicers: Filters for department, gender, and marital status - Added screenshots of the dashboard for project presentation and submission.

Conclusion

The project successfully identified key factors driving employee attrition and developed a predictive model using systematic data analysis, machine learning, and visualization techniques. These insights empower HR departments to implement targeted retention initiatives, lower attrition rates, and improve organizational performance.