Project Report: HR Analytics - Predicting Employee Attrition

Name: Rishabh Kumar

Internship: Elevate Labs & Ministry of MSME Project Phase

---

1. Introduction

Employee attrition, or turnover, is a significant challenge for organizations, leading to increased recruitment costs, loss of productivity, and decreased morale. Proactively identifying employees who are at a high risk of leaving is crucial for implementing targeted retention strategies. This project aims to analyze a comprehensive HR dataset (HR_Analytics.csv) to understand the primary factors driving employee resignation and to build a predictive model to identify future attrition.

2. Abstract

This project involves a detailed analysis of employee data to uncover the root causes of attrition. The process included data cleaning, exploratory data analysis (EDA), and the development of a machine learning classification model to predict the likelihood of an employee leaving. Key insights and factors influencing attrition were visualized using an interactive Power BI dashboard (HR Analytics Dashboard.pbix). The final model and dashboard serve as a tool for the HR department to make data-driven decisions and develop strategies to improve employee retention.

3. Tools Used

- **Python:** Used for data cleaning, analysis, and machine learning.

    o **Pandas:** For data manipulation and analysis. o **Scikit-learn (Sklearn):**

      For building and evaluating the classification model.

    o **Seaborn/Matplotlib:** For initial data visualization during the EDA phase.

- **Power BI:** Used to create the final interactive dashboard for visualizing attrition factors and key metrics.

- **CSV:** As the data source format (**HR_Analytics.csv**).

4. Steps Involved in Building the Project

1. **Data Loading and Cleaning:** The HR_Analytics.csv dataset was loaded into a Pandas Data Frame. The data was checked for missing values, duplicates, and inconsistencies, which were handled accordingly to ensure data quality.

2. **Exploratory Data Analysis (EDA):** A comprehensive EDA was performed to understand the data's structure and find initial patterns. This involved analyzing department-wise attrition rates, the impact of salary bands, promotion history, and employee satisfaction levels on attrition.

3. **Feature Engineering & Preprocessing:** Categorical features (like 'Department' or 'Salary') were converted into numerical formats using techniques like one-hot encoding so they could be used by the machine learning model.

4. **Model Building & Training:** The dataset was split into training and testing sets. A classification model, such as **Logistic Regression** or a **Decision Tree** (as suggested in the project guide), was trained on the training data to learn the patterns that predict attrition.

5. **Model Evaluation:** The model's performance was evaluated on the unseen test data. Key metrics, including **Accuracy** and a **Confusion Matrix**, were generated to understand how well the model predicts attrition.

6. **Dashboard Visualization:** The HR Analytics Dashboard.pbix file was created to visualize the key findings. This dashboard interactively displays the main factors contributing to attrition (e.g., satisfaction level, last evaluation, number of projects) and allows stakeholders to filter data by department, salary, and other dimensions.

5. Conclusion

This project successfully developed a predictive model for employee attrition and a supporting Power BI dashboard. The analysis identified low employee satisfaction, high average monthly working hours, and a lack of promotion in the last 5 years as the strongest predictors of an employee leaving.

The model can be used by the HR department to proactively identify "at-risk" employees, allowing for timely interventions. The dashboard provides a clear and accessible way for management to monitor attrition trends and understand its drivers, ultimately enabling the company to implement more effective retention strategies and reduce turnover costs.