HR Analytics – Predict Employee Attrition

Introduction

Employee attrition is a major challenge for organizations as it leads to increased recruitment costs, loss of experienced talent, and reduced productivity. HR Analytics helps organizations use data-driven approaches to understand employee behavior and identify the factors that contribute to attrition. In this project, employee data such as age, salary, job role, department, and years of experience were analyzed to identify patterns and trends associated with employee attrition. The goal is to help organizations take proactive measures to retain valuable employees.

Abstract

This project focuses on analyzing HR employee data to predict employee attrition using data analytics and machine learning techniques. Exploratory Data Analysis (EDA) was performed to identify key patterns, such as the relationship between salary, experience, job role, and attrition. A classification model (Logistic Regression) was built using Scikit-learn to predict whether an employee is likely to leave the organization. The results were visualized using Power BI dashboards to provide clear insights. This project demonstrates how data analytics can help HR departments make informed decisions to improve employee retention.

Tools Used

• Python – Used for data analysis, preprocessing, and machine learning.
• Pandas – Used for data cleaning and manipulation.
• Seaborn and Matplotlib – Used for data visualization.
• Scikit-learn – Used to build and evaluate the machine learning model.
• Power BI – Used to create interactive dashboards and visualize attrition trends.

Steps Involved in Building the Project

1. Data Collection: The HR employee dataset was collected and imported into Python for analysis.
2. Data Cleaning: Missing values, duplicates, and inconsistent data were identified and handled.
3. Exploratory Data Analysis (EDA): Visualizations were created to understand attrition patterns by salary, age, department, and experience.
4. Data Preprocessing: Categorical data was converted into numerical format using encoding techniques.
5. Model Building: A Logistic Regression classification model was trained to predict

employee attrition.
6. Model Evaluation: The model was evaluated using accuracy score and confusion matrix.
7. Visualization: Power BI dashboards were created to visualize attrition trends and key influencing factors.
8. Insights and Recommendations: Based on analysis, recommendations were provided to reduce attrition.

Conclusion

The HR Analytics – Predict Employee Attrition project successfully identified key factors contributing to employee attrition using data analysis and machine learning techniques. The Logistic Regression model was able to predict employee attrition with good accuracy. The Power BI dashboard provided meaningful insights for HR departments to identify high-risk employees and take preventive actions. This project demonstrates the importance of data-driven decision-making in improving employee retention, reducing turnover costs, and enhancing organizational performance.