HR Attrition – Final Project Report

🔹 Introduction

Employee attrition remains a critical challenge in Human Resource Management. Retaining skilled employees helps reduce hiring costs, improve productivity, and preserve organizational knowledge. This project aims to predict employees likely to leave the company using historical HR data and machine learning techniques.

🔹 Abstract

This project analyzes employee behavior and workplace-related factors to identify patterns that contribute to attrition. Key tasks included data cleaning, feature encoding, exploratory data analysis, and building multiple classification models including Logistic Regression, Decision Tree, Random Forest, and XGBoost. To address class imbalance, SMOTE was applied. The results were visualized through an interactive Power BI dashboard for effective stakeholder communication and decision-making.

🔹 Tools & Technologies Used

- **Python** – Data cleaning, encoding, SMOTE, modeling
- **Jupyter Notebook** – Code development and experimentation
- **Scikit-learn & XGBoost** – Classification algorithms
- **Pandas, Seaborn, Matplotlib** – EDA and visualization
- **Power BI** – Dashboard development for insights

🔹 Steps Involved in the Project

1. **Data Preprocessing**
    - Dropped non-informative columns (e.g., EmployeeCount, Over18)
    - Encoded categorical features using LabelEncoder and One-Hot Encoding
2. **Exploratory Data Analysis (EDA)**
    - Examined attrition by gender, department, job role, and salary band
    - Analyzed impact of promotions and overtime
3. **SMOTE for Class Imbalance**
    - Balanced the dataset using SMOTE (original attrition ratio: ~16%)
4. **Model Building & Evaluation**
    - Trained and compared Logistic Regression, Decision Tree, Random Forest, and XGBoost
    - Evaluated using accuracy, recall, F1-score, and confusion matrix

5. **Power BI Dashboard**
    - Built an interactive dashboard visualizing KPIs, department-level attrition, and employee segments at risk

---

🔹 Conclusion

Logistic Regression (after SMOTE) achieved the most balanced results, with an **85.7% test accuracy**, **0.40 recall**, and **0.47 F1-score** for the attrition class. Key influencing factors included **OverTime**, **Salary Band**, and **Years Since Last Promotion**. The Power BI dashboard enables HR teams to monitor and respond to attrition trends effectively using data-driven strategies.