HR Analytics – Predicting Employee Attrition

---

1. Introduction

Employee attrition is a critical challenge for organizations, leading to increased recruitment costs, loss of knowledge, and reduced productivity. By leveraging HR data, companies can identify the key factors that influence employee turnover and take proactive measures to retain valuable talent. This project applies data analytics and machine learning techniques to understand attrition drivers and build predictive models for employee retention.

---

2. Abstract

The objective of this project is to analyze HR datasets to uncover patterns in employee attrition and develop a predictive model that identifies employees at risk of leaving. Exploratory Data Analysis (EDA) was performed to study attrition across departments, income levels, and tenure. Machine learning models were trained (Logistic Regression and Decision Tree), and model explainability was applied using SHAP values to identify critical features. An interactive Power BI dashboard was created to visualize insights and provide HR managers with actionable intelligence.

---

3. Tools Used

- **Python**: Data preprocessing, EDA, machine learning (Pandas, Seaborn, Scikit-learn).
- **SHAP**: Model explainability (feature importance).
- **Power BI**: Interactive dashboard and data visualization.
- **GitHub**: Version control and project repository.

---

4. Steps Involved in Building the Project

1. **Data Preparation**: Loaded IBM HR Analytics dataset (1,470 records), removed duplicates, cleaned ID-like columns, and handled missing values.
2. **EDA**:
    - Attrition rate overall: ~**16%** (Yes = 237, No = 1,233).
    - Higher attrition observed in **Sales** and **Research & Development** departments.
    - Younger employees (≤30 years) and low-income groups (<3k) showed higher turnover.
    - Overtime strongly correlated with attrition.
3. **Modeling**:

- Trained Logistic Regression and Decision Tree.
- Decision Tree achieved **~82% accuracy** with better recall for attrition cases.

4. **Explainability**:
    - SHAP summary highlighted **OverTime, MonthlyIncome, JobRole, and YearsAtCompany** as top predictors of attrition.
5. **Dashboard**: Built Power BI dashboard with KPIs, attrition by department, age group, income bin, job role, and filters for Gender, OverTime, Marital Status, and Education.

---

5. Key Insights & Recommendations

- **Top Drivers**: Overtime, Monthly Income, Tenure (YearsAtCompany), and Job Role.
- **At-Risk Segments**:
    - Employees aged ≤30 years.
    - Staff with <3k monthly income.
    - Sales Executives and Laboratory Technicians.
    - Employees with frequent overtime.
- **Recommendations**:
    - Implement workload balancing to reduce overtime.
    - Introduce career growth/mentorship programs for early-tenure employees.
    - Re-evaluate compensation bands for critical roles.
    - Strengthen retention strategies in Sales and R&D departments.

---

6. Conclusion

The HR Analytics project successfully demonstrated how data-driven methods can identify employees most at risk of attrition and provide actionable insights for retention. The combination of EDA, predictive modeling, and Power BI dashboards equips HR managers with an effective decision-support system. This solution can be extended by incorporating real-time HR data for proactive workforce planning.