1. Project Objective

Analyze employee attrition and identify key factors influencing it. Predict at-risk employees to help HR take proactive actions and improve retention strategies.

2. Dataset Overview

The dataset contains 1470 employees and 35 features, including demographics, education, satisfaction level, job metrics, and salary information. The target variable is Attrition (Yes/No).

3. Data Cleaning & Preprocessing

- Checked for null values
- Encoded categorical variables into numeric form using Label Encoding
- Dropped redundant columns like EmployeeNumber, EmployeeCount, StandardHours, Over18
- Verified no empty or duplicate rows after preprocessing

4. Exploratory Data Analysis (EDA)

- Attrition Distribution: No = 1233, Yes = 237
- Job Role Distribution visualized using horizontal bar charts
- Correlation analysis revealed OverTime, WorkLifeBalance, and JobSatisfaction as key influencers
- Age and YearsAtCompany also show moderate correlation with attrition

5. Modeling Approach

- Split dataset into train-test sets (80%-20%)
- Models used: Logistic Regression and Random Forest Classifier
- Handled class imbalance using class_weight='balanced'
- Evaluated model performance using Accuracy, Precision, Recall, and F1-Score

6. Model Evaluation

Logistic Regression:
Accuracy: 67%
Recall for Attrition: 72%

Random Forest:
Accuracy: 84%
Recall for Attrition: 9%

Conclusion: Logistic regression performs better in identifying employees at risk of leaving, even though overall accuracy is lower.

7. Top Features Influencing Attrition

1. OverTime
2. JobSatisfaction
3. WorkLifeBalance
4. YearsAtCompany
5. EnvironmentSatisfaction
6. DistanceFromHome
7. JobInvolvement
8. MonthlyIncome
9. NumCompaniesWorked
10. JobRole

8. Insights

- Employees working overtime show higher attrition
- Poor work-life balance increases attrition risk
- Job satisfaction and environment satisfaction are strong retention factors
- Employees with low tenure are more likely to leave
- Distance from home contributes to attrition in some job roles

9. Recommendations for the Company

- Limit mandatory overtime to reduce burnout
- Implement flexible working policies for better work-life balance
- Regular satisfaction surveys to monitor morale
- Provide clear career growth paths and training opportunities
- Focus retention strategies on early-tenure employees
- Use predictive insights to identify high-risk employees early

10. Conclusion

The IBM HR Analytics Project successfully identified key factors contributing to employee attrition. By leveraging predictive analytics and data-driven insights, HR can make informed decisions to enhance employee satisfaction, improve retention, and maintain organizational productivity.