Employee Attrition Prediction Report

# 1. Dataset Description and Preprocessing

- Dataset: IBM HR Analytics Dataset  
- Shape: 1470 rows × 35 columns  
- Target Variable: Attrition (Yes/No)  
- Dropped Columns: EmployeeNumber, Over18, StandardHours, EmployeeCount  
- Encoding: Used LabelEncoder for categorical variables  
- Scaling: Applied StandardScaler for numerical features  
- Split: 80% training, 20% testing (stratified)

# 2. Exploratory Data Analysis (EDA)

- Attrition Distribution: ~16% of employees left (class imbalance)  
- High Attrition Observed In:  
 • Employees working OverTime  
 • Those with low JobSatisfaction and EnvironmentSatisfaction  
 • Employees with low MonthlyIncome  
 • Employees with < 2 years at the company  
- Visualizations Included:  
 • Countplots (Attrition vs. OverTime)  
 • Boxplots (Job Satisfaction vs. Attrition)  
 • Heatmap (Feature correlation with Attrition)

# 3. Model Used and Rationale

- Model: Logistic Regression  
- Reason:  
 • Simple, interpretable, effective for binary classification  
 • Fast training time and good baseline performance  
- Evaluation Metrics:  
 • Precision, Recall, F1-score  
 • Confusion Matrix (To evaluate false positives/negatives)

# 4. Model Explainability with LIME

- LIME (Local Interpretable Model-agnostic Explanations) was used to interpret predictions.  
- Top Features influencing individual predictions:  
 • OverTime  
 • JobSatisfaction  
 • MonthlyIncome  
 • YearsAtCompany

# 5. Actionable Insights to Reduce Attrition

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Key Factor | Insight from Data/Model | Actionable Strategy | | OverTime | Strongly linked with attrition | Monitor workloads, enforce fair overtime policies | | JobSatisfaction | Low job satisfaction leads to higher attrition | Conduct regular surveys, act on feedback | | EnvironmentSatisfaction | Poor satisfaction increases attrition | Improve workplace culture and facilities | | MonthlyIncome | Employees with lower salaries tend to leave | Evaluate and adjust compensation benchmarks | | YearsAtCompany | Newer employees are more likely to leave | Improve onboarding, mentorship, and career pathing | |

# 6. Challenges and Solutions

|  |  |
| --- | --- |
| Challenge | Solution |
| Imbalanced dataset | Used stratify during train-test split |
| Many categorical features | Encoded using LabelEncoder |
| Model interpretability | Used LIME for feature-level explanation |