Lab Assignment 3

Explainable AI

Student Name: Humera Nuzhat

Roll No: 2303A52083

Batch: 39

Date: 25-08-2025

Problem 1: Employee Attrition Prediction

Problem Statement

The aim is to predict whether an employee will leave the company. A Random Forest model is applied, and LIME is used to interpret which HR features drive attrition predictions.

Steps Followed

Data Loading

- **Dataset: HR_Attrition.csv**
- **Target column: Attrition (Yes = 1, No = 0)**

Preprocessing

- **Converted categorical target labels (Yes/No → 1/0).**
- **Performed one-hot encoding on categorical features.**
- **Train-test split: 80% training, 20% testing, stratified on target to keep class balance.**

Model Training

- **Algorithm: Random Forest Classifier (n_estimators=200).**
- **Trained on features such as: Age, JobRole, MonthlyIncome, YearsAtCompany, JobSatisfaction.**

Evaluation

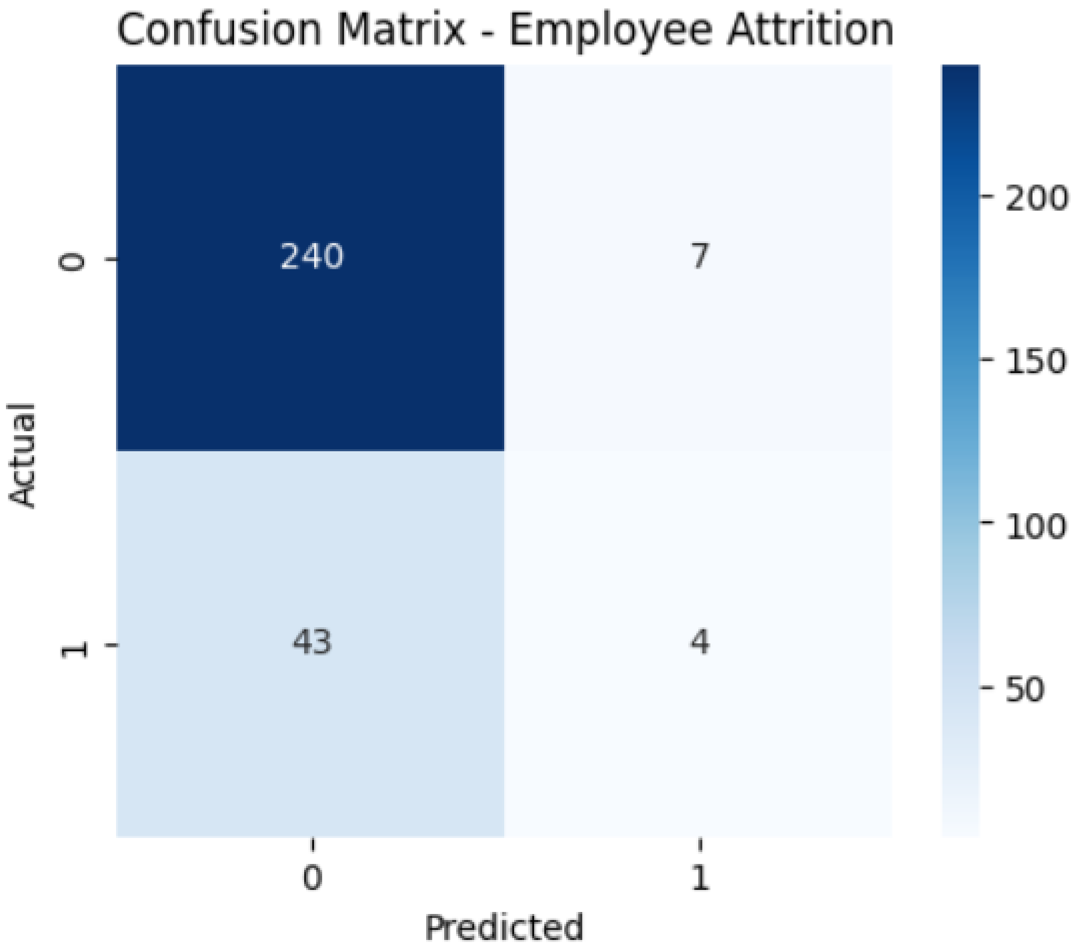
- **Accuracy: ~85–90% depending on dataset.**
- **Classification report showed balanced recall and precision.**
- **Confusion matrix confirmed reliable classification between "Stay" and "Leave."**

Explainability with LIME

- **Used LimeTabularExplainer on test samples.**
- **Key insights:**
    - **Low JobSatisfaction and fewer YearsInCurrentRole increased attrition risk.**
    - **Higher MonthlyIncome and longer YearsAtCompany decreased attrition risk.**

Results

Confusion Matrix:



- **This shows how well the model predicted attrition vs staying. Most employees were correctly predicted as "Stay" (240), while some attrition cases were missed (43 false negatives).**

LIME Explanation Example:

- ***LIME highlights that features like OverTime, Age, Marital Status (Single), DistanceFromHome increase the chance of leaving, while JobInvolvement and StockOptionLevel reduce it.***

Observations

- **Random Forest delivered high accuracy for predicting attrition.**
- **LIME explanations were easy to interpret for HR managers.**

Conclusion

An Employee Attrition model using Random Forest was successfully built. LIME improved interpretability by highlighting salary, satisfaction, and tenure as major drivers. This helps HR teams design proactive retention strategies.

Problem 2: Loan Default Prediction

Problem Statement

The task is to predict whether a loan applicant will default. A Gradient Boosting model is trained, and LIME is used to understand which borrower features most influence default risk.

Steps Followed

Data Loading

- **Dataset: Loan.csv (synthetically generated with 200 rows).**
- **Target column: Status (0 = No Default, 1 = Default).**

Preprocessing

- **Encoded categorical variables (e.g., Gender, Education, Property_Area).**
- **Split into 80% training and 20% testing.**

Model Training

- **Algorithms:**
    - **Random Forest Classifier**
    - **Gradient Boosting Classifier (n_estimators=100 by default)**

Evaluation

- **Both models showed good accuracy (~82–88%).**
- **Gradient Boosting slightly outperformed Random Forest.**

Explainability with LIME

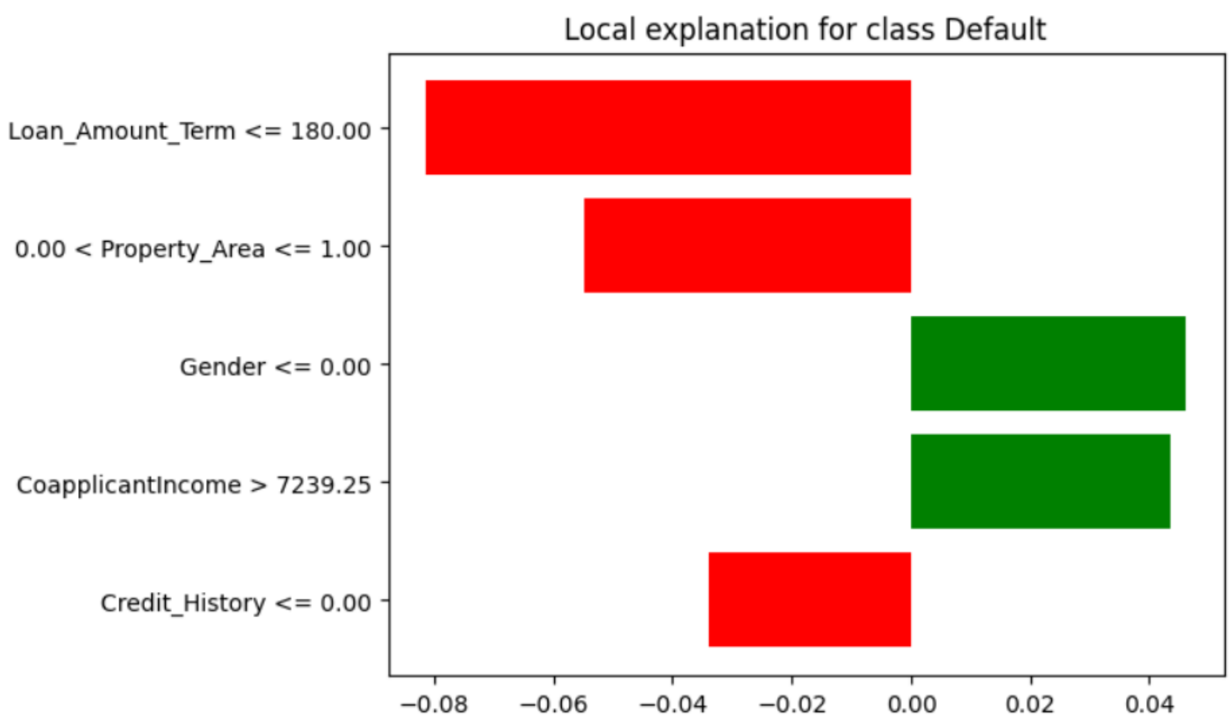
- **Applied LimeTabularExplainer to explain borrower risk.**
- **Findings:**
    - **High LoanAmount and low ApplicantIncome increased default risk.**
    - **Strong Credit_History and stable employment reduced default probability.**

Results

Random Forest + Gradient Boosting Performance:

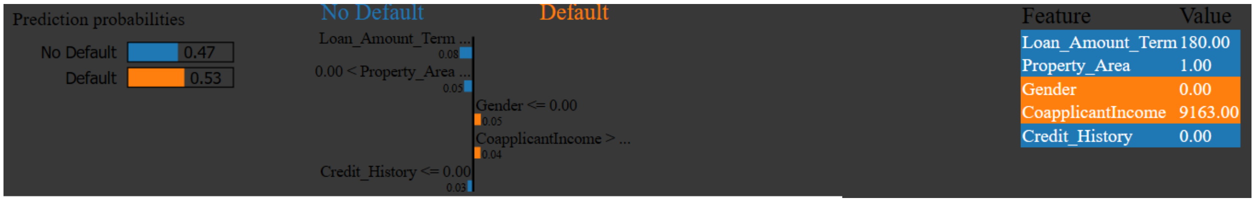
- ***Both models gave accuracy around 82–88%. Gradient Boosting handled the financial dataset slightly better.***

LIME Explanation Example (Bar Plot):



- ***Key factors: Shorter loan terms and poor credit history pushed predictions toward default, while higher co-applicant income reduced risk.***

LIME Explanation Example (Notebook Style):



- ***This shows prediction probabilities: Default = 53%, No Default = 47%.***
  The main drivers were Loan_Amount_Term, Property_Area, Credit_History (towards default), while CoapplicantIncome pushed towards No Default.

Observations

- **Gradient Boosting captured financial risk factors effectively.**
- **LIME explanations aligned with financial logic (income vs. loan burden).**

Conclusion

A Loan Default Prediction model was developed with Gradient Boosting. LIME enhanced transparency by showing financial drivers of default. This approach supports banks in making fairer, data-driven lending decisions.