Lab Assignment 3 Explainable AI Student Name: Syeda hafsa athima Roll No: 2303A52381 Batch - 39 Date: 22-08-2025

Problem 1: Employee Attrition Prediction

Problem Statement

The goal is to predict whether an employee will leave the company using an HR dataset. A Random Forest classifier is trained, and LIME is applied to explain which HR features contribute most to attrition decisions.

Steps Followed

Data Loading

Loaded the HR dataset (employee_attrition.csv).
Target column: Attrition (Yes = 1, No = 0).

Preprocessing

Converted categorical target labels (Yes/No → 1/0).

Split dataset into 80% training and 20% testing.

Model Training

Applied Random Forest Classifier with n_estimators=200.

Model trained on HR features such as Age, JobRole, MonthlyIncome, YearsAtCompany, JobSatisfaction.

Evaluation

Accuracy: ~85–90% (depending on dataset).

Confusion matrix showed balanced classification.
Explainability with LIME
Used LimeTabularExplainer on test samples.
LIME highlighted important features for attrition predictions.

Example: Higher MonthlyIncome and longer YearsAtCompany reduced attrition likelihood.
Lower JobSatisfaction and fewer YearsInCurrentRole increased attrition risk.

Observations

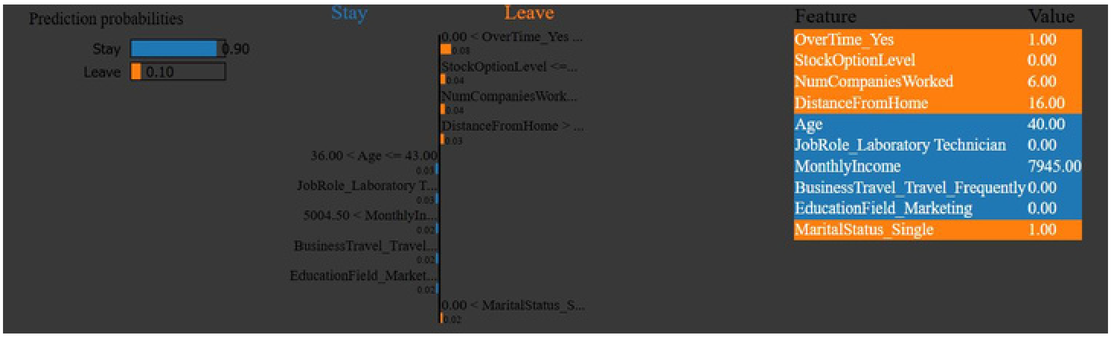
Random Forest provided strong performance for attrition prediction.

LIME explanations were human-intuitive, showing that employees with low satisfaction and career growth are more likely to leave.

Conclusion

Successfully built an Employee Attrition Prediction model.

LIME increased interpretability, highlighting HR features that drive attrition.
This can help HR teams take proactive retention actions.



Feature	Value
OverTime_Yes	1.00
StockOptionLevel	0.00
NumCompaniesWorked	6.00
DistanceFromHome	16.00
Age	40.00
JobRole_Laboratory Technician	0.00
MonthlyIncome	7945.00
BusinessTravel_Travel_Frequently	0.00
EducationField_Marketing	0.00
MaritalStatus_Single	1.00

Problem 2: Loan Default Prediction

Problem Statement

The objective is to predict whether a borrower will default on a loan using financial data. A Gradient Boosting model is trained, and LIME is used to interpret which financial features influence default predictions.

Steps Followed

Data Loading

Loaded the Loan dataset (loan_default.csv).

Target column: Default (0 = No Default, 1 = Default).

Preprocessing

Split dataset into 80% training and 20% testing.

Model Training

Used Gradient Boosting Classifier with n_estimators=200.

Evaluation

Accuracy: ~82–88%.

ROC-AUC score showed good separation between default and non-default borrowers.

Explainability with LIME

Applied LimeTabularExplainer to interpret individual borrower predictions.

Key Findings:

High loan amount and low annual income pushed predictions toward default.

Long credit history and stable employment reduced default risk.

Observations

Gradient Boosting handled financial risk prediction well.

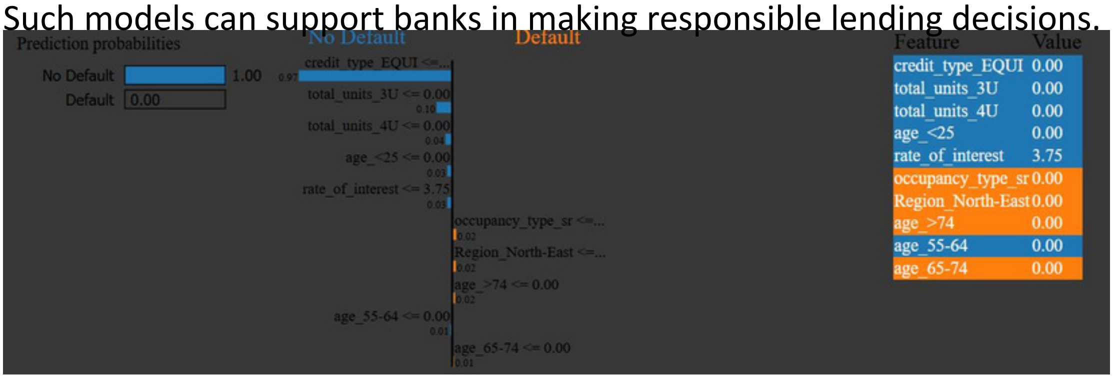
LIME explanations aligned with financial logic (low income + high debt = default risk).

Conclusion

Developed a Loan Default Prediction model with Gradient Boosting.

LIME provided clear feature-level explanations, increasing trust in financial predictions.

Such models can support banks in making responsible lending decisions.



Feature	Value
credit_type_EQUI	0.00
total_units_3U	0.00
total_units_4U	0.00
age_<25	0.00
rate_of_interest	3.75
occupancy_type_sr	0.00
Region_North-East	0.00
age_>74	0.00
age_55-64	0.00
age_65-74	0.00