Explainable Ai

Assignment -3

Name : K.SAI TEJA

HTNO : 2303A52325

Batch : 35

Instructor : Dr.Vairachilai Shenbagavel

1. Introduction :

In this assignment, we explored **Explainable Artificial Intelligence (XAI)** using the **LIME** (Local Interpretable Model-agnostic Explanations) technique. Two different problems were addressed:

- Predicting deposit subscription in a **bank marketing campaign**.
- Classifying **house prices** as expensive or cheap.

The aim was to build predictive models using **Gradient Boosting** and **Random Forest** and interpret their predictions using LIME.

2. Problem Statements:

Problem 1: Bank Marketing Campaign

- **Objective**: Predict whether a client subscribes to a deposit based on marketing data.
- **Dataset**: bank.csv
- **Algorithm Used**: **Gradient Boosting Classifier**
- **Explainability Tool**: LIME

Problem 2: House Price Classification

- **Objective**: Classify California houses as **expensive** or **cheap** based on their features.
- **Dataset**: housing.csv
- **Algorithm Used**: **Random Forest Classifier**
- **Explainability Tool**: LIME

3. Methodology

Step 1 — Data Preprocessing

- Encoded categorical variables using **Label Encoding**.
- Created binary classification targets:
    - **Bank dataset**: deposit column (Yes/No).
    - **Housing dataset**: Converted median_house_value into **cheap (0)** or **expensive (1)**.
- Split both datasets into **training (80%)** and **testing (20%)** sets.

Step 2 — Model Building

- **Gradient Boosting** for bank marketing data:
    - Built a predictive model for deposit subscription.
- **Random Forest** for house price classification:
    - Used an ensemble approach for better accuracy.

Step 3 — Model Evaluation

- Used **Accuracy** and **Classification Report** to evaluate models.
- Achieved **high accuracy** on both datasets.

Step 4 — Explainability with LIME

- Applied **LIME** to interpret predictions:
    - Selected a **single test instance** from each dataset.
    - Visualized **feature contributions** to the prediction.
    - Identified **top influencing factors** for each model.

4. Results

Bank Marketing Campaign

- **Model Accuracy**: 0.84639
- **Top 5 Influencing Features**:
    1. duration
    2. month
    3. contact
    4. pdays
    5. housing
- **LIME Visualization**: Showed which features pushed a client towards **Yes** or **No**.

House Price Classification

- **Model Accuracy**: 0.897286
- **Top 5 Influencing Features**:
    1. Median_income
    2. longitude
    3. latitude
    4. ocean_proximity
    5. population
- **LIME Visualization**: Explained which features contributed to predicting **expensive** or **cheap** houses.

---

6. Insights

From Bank Dataset

- Features like **balance**, **duration**, and **age** strongly influence deposit subscription.
- Clients with **higher balance** and **longer campaign calls** are more likely to subscribe.

From Housing Dataset

- **Median income** and **location coordinates** are the strongest predictors of house prices.
- Proximity to the **ocean** also significantly impacts house value.

7. Conclusion

- Successfully built predictive models for both datasets.
- **LIME** provided clear explanations of model predictions.
- Insights derived can assist:

- **Banks** in targeting potential customers effectively.
- **Real-estate stakeholders** in understanding factors influencing house prices.