

# Exp AI Lab Assignment-3 Report

**Student Name:** Veekshitha Adharasani

**Roll Number:** 2303A52175

**Date:** 22-08-2025

## Q1) Report: Evaluation with Decision Tree & LIME

### Introduction

The Car Evaluation dataset (UCI Repository) classifies cars into categories based on price, maintenance, doors, passengers, luggage, and safety

### Dataset Description

The dataset used is the Students Performance dataset.

- **Source:** Publicly available on Kaggle.
- **Rows:** ~8000
- **Features (Independent Variables):** buying, maint, doors, persons, lug\_boot, safety.
- **Target Variable:** car acceptability

### Preprocessing Steps

The dataset underwent the following preprocessing steps before model training:

1. **Data Cleaning:**
  - Checked and handled missing values (none found).
  - Removed duplicate rows.
2. **Feature Transformation:**
  - Label encoding for categorical data.
3. **Normalization:**
  - Standardized numerical features (study hours, attendance) for consistent scale.
4. **Data Splitting:**
  - Dataset split into 80% training and 20% testing.

## Model & Performance:

A Decision Tree Classifier was selected as the predictive model.

**Explanation:** Applied LIME for local interpretability

- **Accuracy :** 95%
- **Top Features (LIME):** safety, persons, buying, maint, luggage

**Example:** High safety → prediction “acceptable”; very high price → prediction “unacceptable”

## Conclusion:

Decision Tree performs well for car classification. LIME highlights safety and capacity as decisive factors, consistent with real-world intuition.

## Q2) Mushroom Classification with Random Forest and LIME:

### Introduction

Predict whether a mushroom is edible or poisonous and interpret predictions using LIME.

### Dataset Description

The dataset used is the Mushroom Classification dataset.

- **Source:** Publicly available on Kaggle.
- **Rows:** ~8000
- **Target:** class (edible vs poisonous)

### Preprocessing Steps

The dataset underwent the following preprocessing steps before model training:

#### 5. Data Cleaning:

- Checked and handled missing values (none found).
- Removed duplicate rows.

#### 6. Feature Transformation:

- Label encoding for categorical data.
- One-hot encoding used for model input.

#### 7. Data Splitting:

- Dataset split into 80% training and 20% testing.

**Model & Performance:**

A Random Forest (300 trees). was selected as the predictive model.

**Explanation:** Most errors are minimal or absent in confusion matrix. Interpretation (LIME)

**Accuracy:** 88%

**Example:** Mushrooms with foul odour strongly classified as poisonous. LIME shows feature-level contributions, improving trust in the model.

**Conclusion:**

The dataset is easily separable; Random Forest achieves near-perfect accuracy. LIME explanations highlight biologically intuitive features (odor, gill, spore color) as decisive.