ARCHITHA OJHA

2303A52170

BATCH 41

Q1)Report: Car 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.

Objective:

Train a Decision Tree model and use LIME to interpret predictions.

Dataset:

Rows: 1,728

Features: 6 (buying, maint, doors, persons, lug_boot, safety)

Target: car acceptability (unacc, acc, good, vgood)

Preprocessing:

Label encoding for categorical data.

Train-test split:(80–20).

Model: Decision Tree Classifier.

Explanation: Applied LIME for local interpretability.

Results:

Accuracy: ~95% on test set.

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:

Dataset: mushrooms.csv (local copy)

Problem:

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

Preprocessing:

Target: class (edible vs poisonous).

Features: all categorical (cap-shape, odor, gill-size, etc.).

Missing values filled with mode.

One-hot encoding used for model input.

Model. Random Forest (300 trees).

Train/test split: 80/20.

Results:

Accuracy: ~100% (dataset is known to be highly separable).

Most errors are minimal or absent in confusion matrix.

Interpretation (LIME).

Example:

Mushrooms with foul odor 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.