Lab 3 - Explainable AI

Name: BELLAMPALLY KOWSHIK REDDY

Enrollment No: 2303A52002

Code File: XAI_2303A52002_Lab_Assignment_3.ipynb

Report 1: Wine Quality Prediction

Problem Statement

Predict wine quality (scale 0-10) using physicochemical properties like acidity, alcohol, residual sugar, and sulphates.

Approach

Dataset: Wine Quality Dataset (red & white wines)

Preprocessing:

Dropped irrelevant columns

Converted type (red/white) to numeric (0 = red, 1 = white)

Handled missing values with mean imputation

Model Used: Random Forest Classifier

Explainability: SHAP and LIME

Key Findings

Important Features:

Alcohol - Higher alcohol indicates better quality

Volatile Acidity - Negative impact on taste

Sulphates - Positive effect on flavor preservation

Citric Acid - Adds freshness

Chlorides - Higher values reduce quality

Model Performance:

Accuracy: ~67%

Precision: ~65%

Recall: ~62%

F1-score: ~64%

(Insert wine quality SHAP/LIME image here)

Report 2: Breast Cancer Diagnosis

Problem Statement

Predict benign vs malignant tumors with model interpretability for medical use.

Approach

Dataset: Breast Cancer Wisconsin Dataset

Preprocessing:

Dropped irrelevant columns (id, unnamed)

Imputed missing values with mean

Encoded labels (B = 0, M = 1)

Model Used: Random Forest Classifier (100 trees)

Explainability: LIME (local explanations)

Key Findings

Important Features:

Radius_mean: Larger nuclei radius -> higher malignancy risk

Texture_mean: Irregular texture suggests cancer

Perimeter_mean & Area_mean: Bigger, irregular clusters -> malignant

Smoothness/Concavity: Lack of smoothness -> invasive cancer

Model Performance:

Accuracy: ~83%

Precision: ~81%

Recall: ~84%

F1-score: ~82%

Conclusion (Lab 3)

Random Forest performed well for both Wine and Breast Cancer datasets

SHAP & LIME provided transparency into predictions

Limitations: Subjectivity in wine dataset, small size for medical dataset

Future Work: Explore XGBoost / Gradient Boosting, include more features