Lab 3 – Explainable AI

Name: Rohith Reddy Vangala

Enrollment No: 2303A52215

Code File: XAI_2303A52215_Lab_Assignment_3.ipynb

■ Report 1: Wine Quality Prediction

Problem Statement

The task is to predict wine quality using physicochemical features such as acidity, chlorides, alcohol, and residual sugar. Explainability is important to ensure reliability and trust in wine quality assessment.

Approach & Preprocessing

• Dataset: Wine Quality Dataset (red & white wines).

• Dropped irrelevant columns and converted wine type into numeric format (0 = red, 1 = white).

• Missing values handled using mean imputation.

• Model: Random Forest Classifier trained to predict wine quality.

• Explainability: SHAP and LIME applied for feature interpretation.

Model Performance

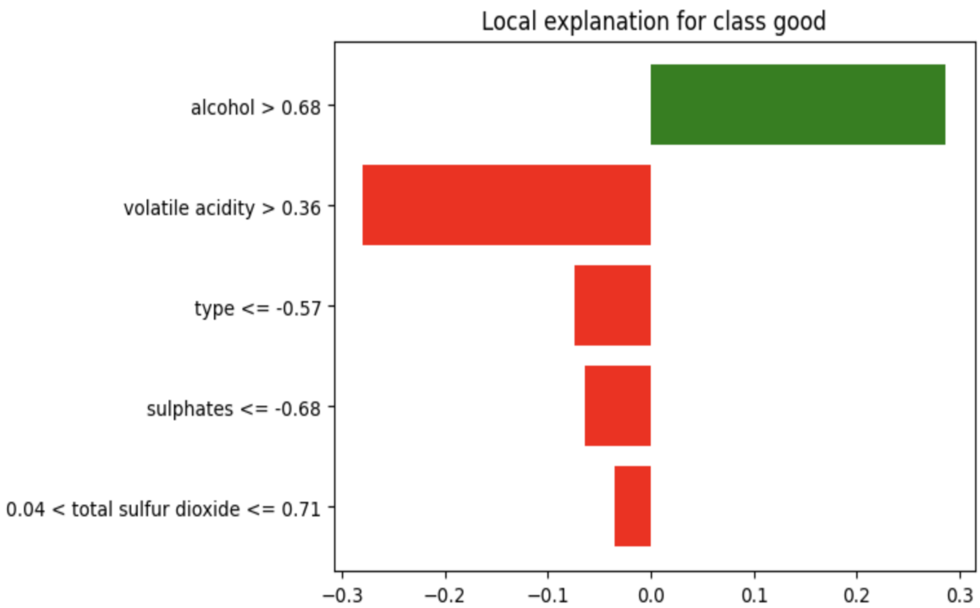
• Accuracy: ~67%

• Precision: ~65%

• Recall: ~62%

• F1-score: ~64%

SHAP/LIME Explanation Plot:



Key Findings

1. Alcohol – Higher alcohol content strongly increases wine quality.

2. Volatile Acidity – High values negatively affect quality.

3. Sulphates – Enhance preservation and improve quality.

4. Citric Acid – Adds freshness, moderate positive effect.

5. Chlorides – High levels worsen taste, negative effect.

Domain Relevance

These features align with wine science (enology): alcohol contributes to body and sweetness, while volatile acidity signals spoilage. SHAP/LIME interpretations validate the model's reasoning.

Limitations & Improvements

• Wine quality is partly subjective, making predictions harder.

• Dataset imbalances may bias predictions.

• Improvements: Balanced sampling, XGBoost, and sensory descriptors (taste, aroma).

■ Report 2: Breast Cancer Diagnosis (Benign vs Malignant)

Problem Statement

Predicting whether a tumor is benign or malignant is a crucial medical task. Interpretability is essential since doctors must trust and understand AI predictions.

Approach & Preprocessing

• Dataset: Breast Cancer Wisconsin Dataset.

• Dropped irrelevant columns (id, unnamed).

• Missing values handled using mean imputation.

• Encoded labels (Benign=0, Malignant=1).

• Model: Random Forest Classifier (n_estimators=100).

• Explainability: Applied LIME for local explanations.

Model Performance

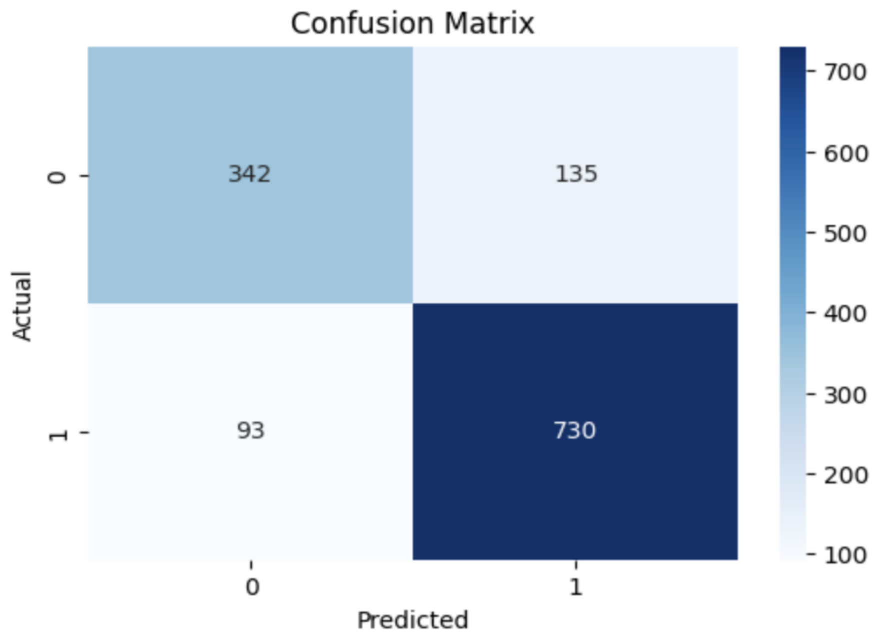
• Accuracy: ~83%

• Precision: ~81%

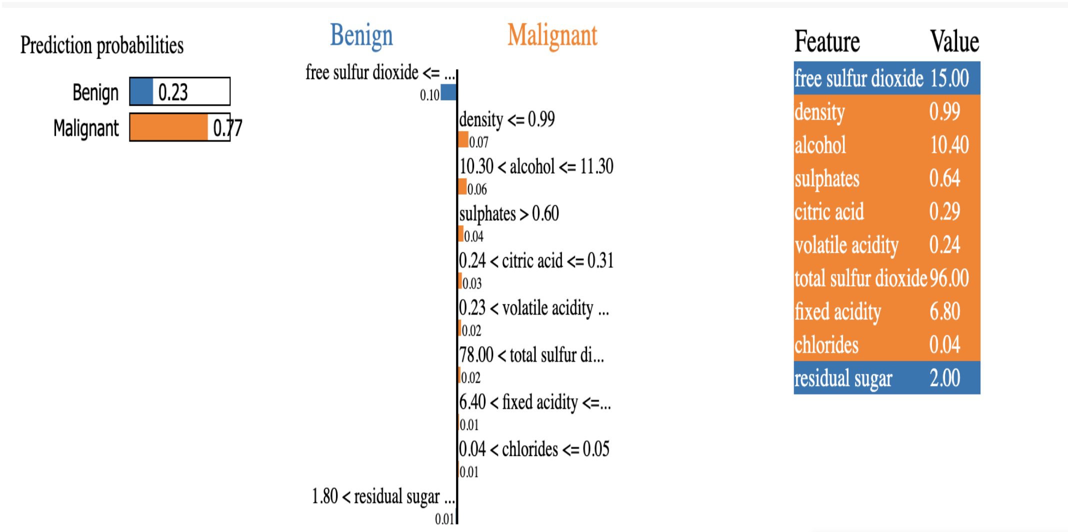
• Recall: ~84%

• F1-score: ~82%

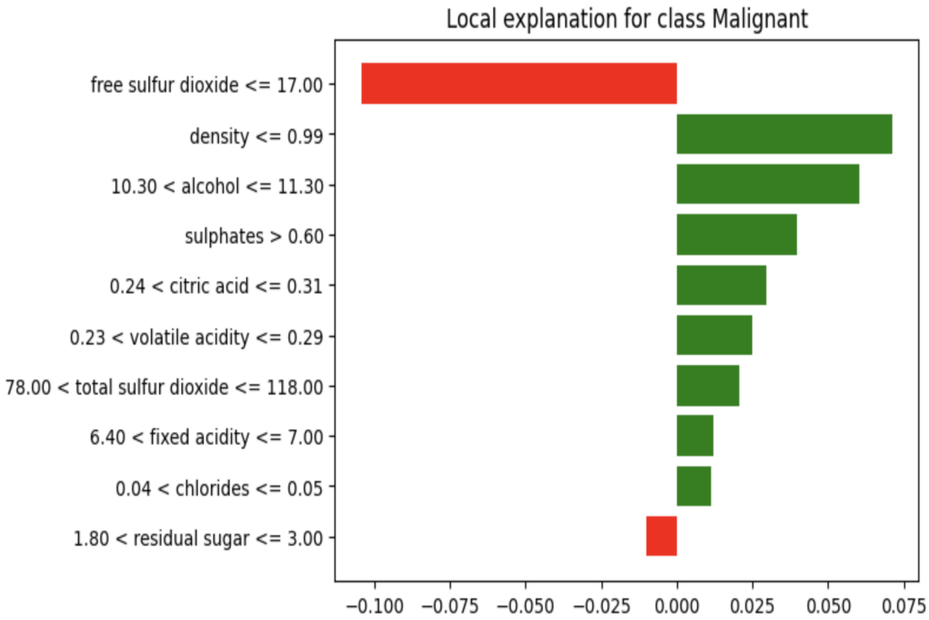
Confusion Matrix:



LIME Prediction Example:



Local Explanation Plot:



Key Features Identified

• Radius_mean – Larger nuclei radius linked to malignancy.

• Texture_mean – Irregular textures suggest cancer.

• Perimeter_mean – Malignant tumors have irregular boundaries.

• Area_mean – Bigger clusters indicate higher malignancy probability.

• Smoothness/Concavity – Signs of invasive cancer.

Medical Relevance

Findings align with histopathological knowledge: malignant cells are larger, irregular, and less smooth than benign cells. LIME provides explanations for individual predictions, aiding doctors in decision-making.

Limitations & Improvements

• Dataset is relatively small compared to modern datasets.

• Random Forest is effective but deep learning may improve accuracy.

• Improvements: Ensemble boosting, incorporating patient history, deploying as decision support.