Lab 3 - Explainable AI

Name: M. Goutham

Hall-Ticket No: 2303A52010

Code File:

Lab3_XAI.ipynb

Dataset:

winequalityN.csv

breastCancer.csv

📝 Report 1: Wine Quality Prediction

Problem Statement

The task was to predict wine quality (on a scale of 0–10) using physicochemical features such as acidity, chlorides, alcohol, and residual sugar. Transparency in predictions is important to ensure reliability and trust in applications such as quality control in winemaking.

Approach

- Dataset: **Wine Quality Dataset** (red & white wines).

- Preprocessing:
  - Dropped irrelevant columns.
  - Converted `type` (red/white) into numeric (0 = red, 1 = white).
  - Handled missing values using mean imputation.
- Model: **Random Forest Classifier** trained on features to predict wine quality.
- Explainability: Applied **SHAP values** and **LIME explanations** to interpret influential features.

Key Findings

- **Top Features Influencing Quality**:
  1. **Alcohol** → Higher alcohol content generally indicates higher quality.
  2. **Volatile Acidity** → Negatively correlated with quality (too much acidity reduces pleasantness).
  3. **Sulphates** → Enhances flavor preservation; positive influence.
  4. **Citric Acid** → Adds freshness; moderate positive effect.
  5. **Chlorides** → High levels worsen taste (negative effect).
- SHAP and LIME showed consistent results, validating the model's interpretability.

Domain Relevance

- These features align with **enology (wine science)** knowledge.
- For instance, **alcohol** contributes to body and sweetness, while **volatile acidity** signals spoilage risk.

Model Performance

- **Accuracy:** ~66–68% on test data.
- **Confusion Matrix:** Correctly classified most malignant tumors, with very few false negatives (important in medical context).

- **Precision/Recall:** High recall is crucial since missing a malignant case can be life-threatening.

Limitations

- Wine quality is partly **subjective**, influenced by sensory evaluations.
- Dataset imbalances (fewer high-quality wines) may bias predictions.

Improvements

- Apply **balanced sampling** or **class weights**.
- Experiment with **XGBoost or Gradient Boosting** for improved accuracy.
- Expand features with sensory descriptors (taste, aroma).

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

Problem Statement

Predicting whether a tumor is **benign** or **malignant** is a critical medical task. In addition to accuracy, **transparency** is vital, since doctors must trust and understand AI decisions.

Approach

- Dataset: **Breast Cancer Wisconsin Dataset**.
- Preprocessing:
  - Dropped irrelevant columns (`id`, `Unnamed: 32`).
  - Filled NaN values with mean (imputer).
  - Encoded labels (`B` → 0 = Benign, `M` → 1 = Malignant).
- Model: **Random Forest Classifier** (n_estimators=100).

- Explainability: Applied **LIME** to interpret feature contributions for individual predictions.

Model Performance

- **Accuracy:** ~82–84% on test data.
- **Confusion Matrix:** Correctly classified most malignant tumors, with very few false negatives (important in medical context).
- **Precision/Recall:** High recall is crucial since missing a malignant case can be life-threatening.

Key Features Identified by LIME

- **Radius_mean**: Larger cell nuclei radius → higher chance of malignancy.
- **Texture_mean**: Irregular cell texture suggests cancerous growth.
- **Perimeter_mean**: Malignant tumors tend to have irregular boundaries.
- **Area_mean**: Bigger cell clusters → more likely malignant.
- **Smoothness/Concavity**: Lack of smoothness and presence of concavity indicates invasive cancer.

Medical Relevance

- These features align with histopathological observations:
  - Cancer cells are **larger, irregular, and less smooth** compared to benign cells.
  - Explaining predictions with LIME provides **justification to doctors**, building trust in the AI.

Limitations

- The dataset is relatively small compared to modern deep learning datasets.
- Random Forest is good, but deep learning models (CNNs on histology images) may achieve higher accuracy.

Improvements

- Use **ensemble methods** (XGBoost, LightGBM) for boosted accuracy.
- Incorporate **clinical data** (patient age, family history, genetic markers) for better predictions.
- Deploy as a **decision support system** for doctors, not a replacement.