Breast Cancer Prediction

Breast cancer is one of the leading causes of death among women worldwide. Early detection and accurate diagnosis play a critical role in increasing survival rates.

You are given a dataset containing various **medical features extracted from breast cancer biopsy images**. Your task is to **develop a machine learning model** that can accurately predict whether a tumor is **Malignant (harmful)** or **Benign (non-harmful)** based on the given features.

Dataset Details

- **Source:** Breast Cancer Wisconsin Dataset

- **Number of Records:** 569

- **Number of Features:** 30 (real-valued features)

Note : Use load_breast_cancer() method from sklearn to load the dataset.

Features:

- Mean Radius

- Mean Texture

- Mean Perimeter

- Mean Area

- Mean Smoothness

- Mean Compactness

- Mean Concavity

- Mean Symmetry

- Worst Radius, Worst Texture, … (and other statistical measurements)

Target Variable:

- 0 → Malignant

- 1 → Benign

Objectives

1. Load and explore the dataset.

2. Perform data preprocessing steps:

    ○ Handle missing values (if any)

    ○ Normalize or scale features

3. Perform **exploratory data analysis (EDA)**:

    ○ Summary statistics

    ○ Visualization of feature correlations

4. Split the dataset into **training and testing sets**.

5. Build a **machine learning classification model** to predict tumor type.

6. Evaluate the model using:

    ○ Accuracy

    ○ Confusion Matrix

    ○ Precision, Recall, F1-Score

7. Provide your **observations and conclusions**.

Expected Deliverables

Code File:

    ○ Data loading

    ○ Preprocessing

    ○ Model building

    ○ Evaluation