Breast Cancer Classification using Logistic Regression

This project performs binary classification to detect whether a tumor is malignant or benign using logistic regression. The Breast Cancer dataset used here is provided as a CSV file (data.csv), and the task includes data preprocessing, model training, evaluation, and visualization.

# Dataset

- Source: Manually uploaded CSV file (data.csv)  
- Target Variable: diagnosis  
 - M: Malignant (1)  
 - B: Benign (0)  
- Features: 30 numerical columns related to tumor characteristics (mean, worst, standard error)

# Steps Performed

1. Data Cleaning:  
 - Removed irrelevant columns (id, Unnamed: 32)  
 - Converted diagnosis labels from M/B to 1/0  
  
2. Feature Scaling:  
 - Standardized the dataset using StandardScaler  
  
3. Train/Test Split:  
 - 80% Training / 20% Testing using train\_test\_split  
  
4. Model Building:  
 - Used LogisticRegression from scikit-learn  
  
5. Evaluation Metrics:  
 - Confusion Matrix  
 - Classification Report (Precision, Recall, F1-score)  
 - ROC-AUC Score  
 - ROC Curve Visualization

# Sample Results

Confusion Matrix:  
[[71 1]  
 [ 3 39]]  
  
Classification Report:  
 precision recall f1-score support  
 0.0 0.96 0.99 0.97 72  
 1.0 0.97 0.93 0.95 42  
  
Accuracy: 0.9649  
ROC-AUC Score: 0.99

# Concepts Covered

- Logistic Regression  
- Sigmoid Function  
- Binary Classification  
- Precision vs Recall  
- ROC and AUC Metrics  
- Threshold Tuning

# How to Run

# Clone this repository  
git clone https://github.com/yourusername/breast-cancer-logistic.git  
cd breast-cancer-logistic  
  
# Install required packages  
pip install pandas scikit-learn matplotlib  
  
# Run the notebook or Python file

# Notes

- Dataset should be placed in the same folder as the notebook (data.csv)  
- Make sure to standardize data before training the model

# References

- Scikit-learn Documentation: https://scikit-learn.org/  
- Dataset Source: UCI Breast Cancer Wisconsin Dataset

# Author

Pranitha Bokketi  
Learning Data Science | Building ML Projects