**AI Assignment**

**Breast canser database**

**Lecturer:** Sakina Juzar Neemuchwala

**Name:** feyintola lasekan

Sebastian Aguilera

**Analyzing the Dataset**

**Classification or Regression:**

This task involves categorizing tumors as either malignant (cancerous) or benign (non-cancerous), making it a classification problem.

**Dependent and Independent Variables:**

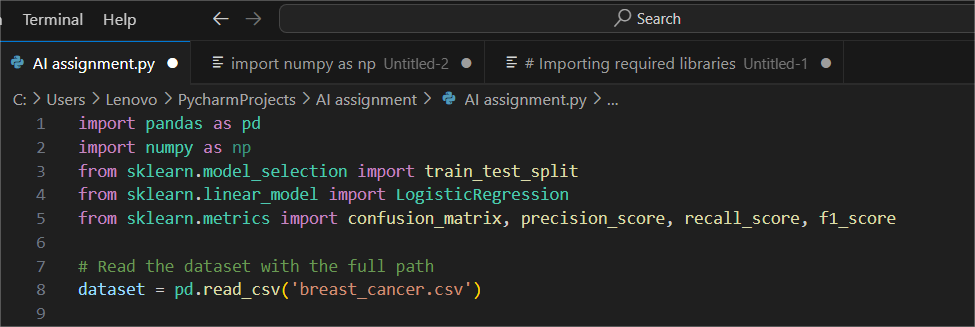
**Dependent Variable**: This is what we are trying to predict, which is the 'diagnosis' of the tumor (malignant or benign).

**Independent Variables**: These are the features used to make the prediction, such as measurements of the tumor (e.g., radius, texture, perimeter, area, etc.).

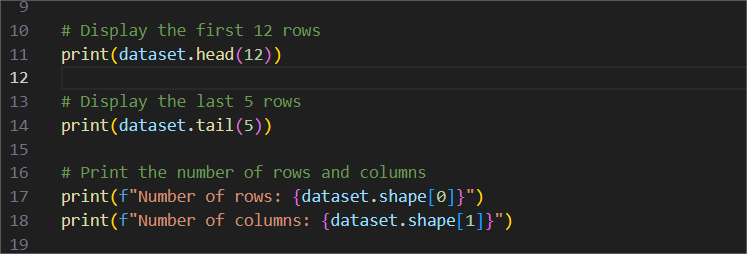
We will use the **Logistic Regression algorithm** for this classification task as it is well-suited for binary outcomes.

**2. Implementing the Solution**

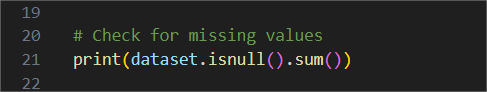
(a) Importing Libraries and Reading the Dataset



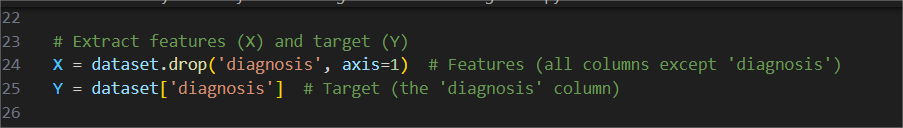
**(b) Displaying Rows and Dataset Shape**

****

**(c) Checking for the missing values**

****

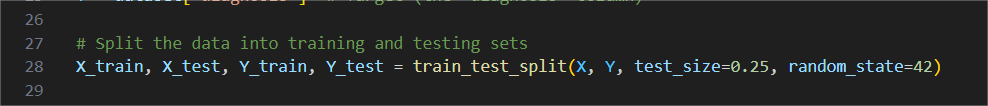
**(d) Seperating Features (X) and Target (Y)**

****

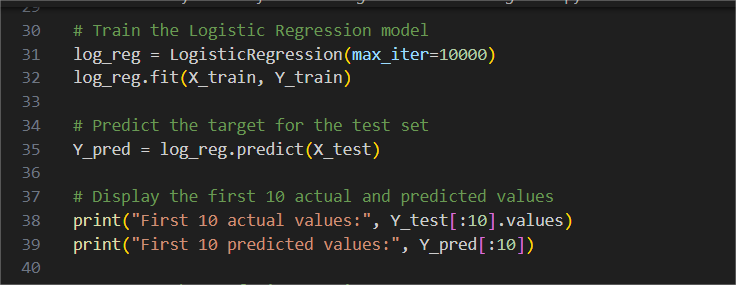
**X (features):** All coumns excepts the ‘diagnosis’ column

**Y (target):** The ‘diagnosis’ column which indicates if a tumor is malignant or benign

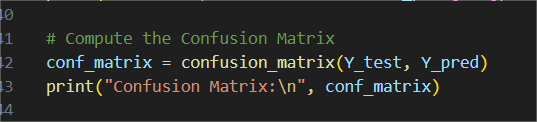
**Splitting the Data into Training and Testing sets**

****

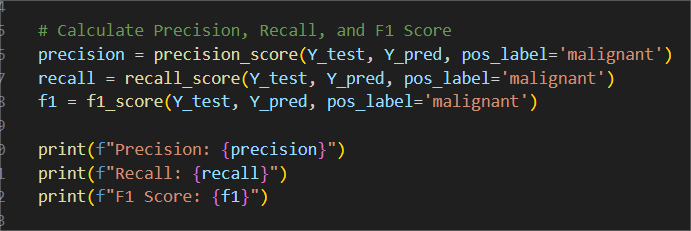
**Training the model**

****

**Evaluating the confusion matrix**

****

**Calculate Precision, Recall, and F1 Score**

****

**Explanation:**

The confusion matrix provides a summary of prediction results:

True Positives (TP): Correctly identified malignant cases.

True Negatives (TN): Correctly identified benign cases.

False Positives (FP): Incorrectly identified benign cases as malignant.

False Negatives (FN): Incorrectly identified malignant cases as benign.

**Precision:** The ratio of correctly predicted positive observations to the total predicted positives.

**Recall**: The ratio of correctly predicted positive observations to all observations in the actual class.

**F1 Score**: The weighted average of precision and recall, providing a balance between the two.

**Answers that the progam gave:**

id diagnosis radius\_mean texture\_mean perimeter\_mean ... compactness\_worst concavity\_worst concave points\_worst symmetry\_worst fractal\_dimension\_worst

0 842302 M 17.99 10.38 122.80 ... 0.6656 0.7119 0.26540 0.4601 0.11890

1 842517 M 20.57 17.77 132.90 ... 0.1866 0.2416 0.18600 0.2750 0.08902

2 84300903 M 19.69 21.25 130.00 ... 0.4245 0.4504 0.24300 0.3613 0.08758

3 84348301 M 11.42 20.38 77.58 ... 0.8663 0.6869 0.25750 0.6638 0.17300

4 84358402 M 20.29 14.34 135.10 ... 0.2050 0.4000 0.16250 0.2364 0.07678

5 843786 M 12.45 15.70 82.57 ... 0.5249 0.5355 0.17410 0.3985 0.12440

6 844359 M 18.25 19.98 119.60 ... 0.2576 0.3784 0.19320 0.3063 0.08368

7 84458202 M 13.71 20.83 90.20 ... 0.3682 0.2678 0.15560 0.3196 0.11510

8 844981 M 13.00 21.82 87.50 ... 0.5401 0.5390 0.20600 0.4378 0.10720

9 84501001 M 12.46 24.04 83.97 ... 1.0580 1.1050 0.22100 0.4366 0.20750

10 845636 M 16.02 23.24 102.70 ... 0.1551 0.1459 0.09975 0.2948 0.08452

11 84610002 M 15.78 17.89 103.60 ... 0.5609 0.3965 0.18100 0.3792 0.10480

[12 rows x 32 columns]

         id diagnosis radius\_mean texture\_mean perimeter\_mean ... compactness\_worst concavity\_worst concave points\_worst symmetry\_worst fractal\_dimension\_worst

564 926424 M 21.56 22.39 142.00 ... 0.21130 0.4107 0.2216 0.2060 0.07115

565 926682 M 20.13 28.25 131.20 ... 0.19220 0.3215 0.1628 0.2572 0.06637

566 926954 M 16.60 28.08 108.30 ... 0.30940 0.3403 0.1418 0.2218 0.07820

567 927241 M 20.60 29.33 140.10 ... 0.86810 0.9387 0.2650 0.4087 0.12400

568 92751 B 7.76 24.54 47.92 ... 0.06444 0.0000 0.0000 0.2871 0.07039

[5 rows x 32 columns]

Number of rows: 569

Number of columns: 32

id 0

diagnosis 0

radius\_mean 0

texture\_mean 0

perimeter\_mean 0

area\_mean 0

smoothness\_mean 0

compactness\_mean 0

concavity\_mean 0

concave points\_mean 0

symmetry\_mean 0

fractal\_dimension\_mean 0

radius\_se 0

texture\_se 0

perimeter\_se 0

area\_se 0

smoothness\_se 0

compactness\_se 0

concavity\_se 0

concave points\_se 0

symmetry\_se 0

fractal\_dimension\_se 0

radius\_worst 0

texture\_worst 0

perimeter\_worst 0

area\_worst 0

smoothness\_worst 0

compactness\_worst 0

concavity\_worst 0

concave points\_worst 0

symmetry\_worst 0

fractal\_dimension\_worst 0

dtype: int64

First 10 actual values: ['B' 'M' 'M' 'B' 'B' 'M' 'M' 'M' 'B' 'B']

First 10 predicted values: ['B' 'M' 'M' 'B' 'B' 'M' 'M' 'M' 'B' 'B']

Confusion Matrix:

 [[87 2]

 [ 5 49]]

Precision: 0.9607843137254902

Recall: 0.9074074074074074

F1 Score: 0.9333333333333333