Artificial Intelligence

Assignment 1

Assignment - 30 Marks

*Objective: To analyze the given dataset and apply **linear regression** **OR** **logistic regression** algorithms as required.*

Dataset:

Breast cancer is the most common cancer amongst women in the world. It accounts for 25% of all cancer cases. It starts when cells in the breast begin to grow out of control.
The key challenge against its detection is how to classify tumors into malignant (cancerous) or benign(non-cancerous). We ask you to complete the analysis of classifying these tumors using machine learning.

1. Analyze the dataset and explain whether this is a Regression or Classification and which algorithm will you apply for this problem.
   Explain the concept of dependent and independent variables and which is the **dependent variable in this dataset.**

   (Explanation should be included in a Word document – 5 Marks)

2. Using Jupyter Notebook or any suitable IDE:
   (a) import the necessary libraries
   (b) read the dataset

   (Code – 2 Marks)

3. Analyze the dataset by
   (a) display the first 12 rows
   (b) display the last 5 rows
   (c) print the number of rows and columns in the dataset

   (Code – 3 Marks)

4. Check if this dataset has any missing values and print the result

   (Code – 2 Marks)

University of Europe for Applied Sciences

Artificial Intelligence

Assignment 1

5. Implement the correct algorithm required for this task (Linear or Logistic)
   (a) Extract the **X feature** and **Y target**

(Code – 2 Marks)
(Include explanation in the Word document about X feature and Y target
(Explanation – 1 Mark)

   (b) Divide the data into training and testing sets to evaluate the model's performance on unseen data. Keep the **test size = 0.25** and **random state = 42**

(Code – 2 Marks)

   (c) Implement the algorithm and find the first 10 values for the actual data and the predicted values

(Code – 2 Marks)

6. Result
   (a) Show the result with the **Confusion Matrix**

(Code – 2 Marks)
(In your Word document, explain the values obtained in this matrix – 2 Marks)

   (b) Calculate and print the **Precision, Recall** and **F1 Score**

(Code – 3 Marks)
(Explanation – 1 Mark)

7. Conclusion
   List your learnings from this assignment and your overall understanding of the ML algorithm you implemented.

(Code – 3 Marks)