**Team Number:**

**Team Captain:**

**Team Members:**

**Activity on September 19, 2022**

**Problem 1 (5 Points)**

|  |  |  |
| --- | --- | --- |
| **Observation #** | **True Status** | **Posterior Probability** |
| **1** | **1** | **0.95** |
| **2** | **1** | **0.85** |
| **3** | **1** | **0.75** |
| **4** | **1** | **0.45** |
| **5** | **1** | **0.35** |
| **6** | **1** | **0.25** |
| **7** | **0** | **0.15** |
| **8** | **0** | **0.05** |
| **9** | **0** | **0.65** |
| **10** | **0** | **0.55** |
| **11** | **0** | **0.5** |
| **12** | **0** | **0.7** |

1. Calculate the following statistics: “True Positive”, “False Positive”, “True Negative”, “False Negative”, “Sensitivity”, “Specificity”, “Accuracy”, “Precision”, and F1 Score at cut-off probability at 0.72. (2 Points)
2. Calculate the model AUC (1 Points)
3. Calculate the model Gini index (1 Point)

**Problem 2 (5 Points)**

Calculate KS statistics for the following data:

|  |  |  |
| --- | --- | --- |
| **Decile** | **"Positive"** | **"Negative"** |
| 1 | 100 | 0 |
| 2 | 98 | 2 |
| 3 | 96 | 4 |
| 4 | 90 | 10 |
| 5 | 85 | 15 |
| 6 | 80 | 20 |
| 7 | 75 | 25 |
| 8 | 66 | 24 |
| 9 | 51 | 49 |
| 10 | 41 | 59 |

**Problem 3 (15 Points)**

**Data Used:** “Microsoft\_Results.CSV” with three variables: ID, HasDetect (observed value), and P\_ HasDetect (Model Predicted Probability).

**Problem 2.1 (0 Points)** Read the CSV file “Microsoft\_Results.CSV”

**Problem 2.2 (3 Points)** Write a program to calculate the following statistics: “True Positive”, “False Positive”, “True Negative”, “False Negative”, “Sensitivity”, “Specificity”, “Accuracy”, and “Precision” for any given cut-off probability (i.e., “P\_HasDetect”). The input data set has three variables and the output data set has nine variables including “Cut off Probability”, “True Positive”, “False Positive”, “True Negative”, “False Negative”, “Sensitivity”, “Specificity”, “Accuracy”, and “Precision”.

**Problem 2.3 (3 Points)** Write a program to calculate the AUC and Gini of the model.

**Problem 2.4** **(3 Points)** Write a program to produce the ROC curve of this model for each point on this curve being calculated at ten-percentile level (i.e., to calculate the (Sensitivity, 1-specificity) pair at cut-off probability of every ten percent).

**Problem 2.5** **(3 Points)** Write a program to produce the ROC curve of this model for each point on this curve being calculated at five-percentile level (i.e., to calculate the (Sensitivity, 1-specificity) pair at cut-off probability of every five percent).

**Problem 2.6 (3 Points)** Write a program to produce a lift chart of the model at decile level (i.e., every ten percent one point).

**Chapter IV Exercise (5 Points)** Work on Problem 1 and Problem 9 in the Textbook (Chapter 4, Page 189).