Week 7 - Evaluating Classification Models

Dataset: Suppose you have a binary classification problem with the following actual classes and predicted classes for a sample of 40 observations:

Actual Classes: 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0

Predicted Classes: 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0

Steps:

```
Create the Confusion Matrix:
    Manually create a 2x2 confusion matrix using the actual classes and predicted classes.

Calculate Evaluation Metrics:
    Calculate the following evaluation metrics:
        Accuracy
        Precision
        Recall (Sensitivity)
        Specificity
        F1 Score

Interpretation:
    Interpret the calculated evaluation metrics to assess the performance of the classificat
```

Example Solution:

Step 1: Create the Confusion Matrix: Predicted Negative (0) Predicted Positive (1) Actual Negative (0) TN FP Actual Positive (1) FN TP

Using the provided actual and predicted classes, we can construct the confusion matrix: Predicted Negative (0) Predicted Positive (1) Actual Negative (0) 15 2 Actual Positive (1) 1 22

Step 2: Calculate Evaluation Metrics:

```
Accuracy = (TP + TN) / Total = (15 + 22) / 40 = 37 / 40 = 0.925
Precision = TP / (TP + FP) = 22 / (22 + 2) = 22 / 24   0.917
Recall (Sensitivity) = TP / (TP + FN) = 22 / (22 + 1) = 22 / 23   0.957
Specificity = TN / (TN + FP) = 15 / (15 + 2) = 15 / 17   0.882
F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.917 * 0.957) / (0.917 + (
```

Step 3: Interpretation:

```
The accuracy of the model is 92.5%, indicating that 92.5% of the predictions are correct.
The precision of the model is approximately 91.7%, implying that 91.7% of the samples predict
The recall (sensitivity) of the model is approximately 95.7%, meaning that 95.7% of the actua
The specificity of the model is approximately 88.2%, indicating that 88.2% of the actual nega
The F1 score, which combines precision and recall, is approximately 93.7%, suggesting overall
```

Comparison:

Compare the model above to three other models.

Model 2

	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	16	1
Actual Positive (1)	5	18

Calculating Evaluation Metrics for the Model.

Accuracy = (TP + TN) / Total = (16 + 18) / 40 = 34 / 40 = 0.85 Precision = TP / (TP + FP) = 18 / (18 + 1) = 18 / 19 0.947 Recall (Sensitivity) = TP / (TP + FN) = 18 / (18 + 5) = 18 / 23 0.783 Specificity = TN / (TN + FP) = 16 / (16 + 1) = 16 / 17 0.941 F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.947 * 0.783) / (0.947 + 0.783) 0.857

Interpretation of the Model:

The accuracy of the model is 85%, indicating that 85% of the predictions are correct. The precision of the model is approximately 94.7%, implying that 94.7% of the samples predicted as positive are truly positive. The recall (sensitivity) of the model is approximately 78.3%, meaning that 78.3% of the actual positive samples are correctly identified. The specificity of the model is approximately 94.1%, indicating that 94.1% of the actual negative samples are correctly identified. The F1 score, which combines precision and recall, is approximately 85.7%, suggesting good overall performance of the model.

Model 3

	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	10	7
Actual Positive (1)	12	11

Calculating Evaluation Metrics for the Model:

Accuracy = (TP + TN) / Total = (10 + 11) / 40 = 21 / 40 = 0.525 Precision = TP / (TP + FP) = 11 / (11 + 7) = 11 / 18 0.611 Recall (Sensitivity) = TP / (TP + FN) = 11 / (11 + 12) = 11 / 23 0.478 Specificity = TN / (TN + FP) = 10 / (10 + 7) = 10 / 17 0.588 F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.611 * 0.478) / (0.611 + 0.478) 0.538

Interpretation of the Model:

The accuracy of the worse-performing model is 52.5%, indicating that only 52.5% of the predictions are correct. The precision of the model is approximately 61.1%, implying that 61.1% of the samples predicted as positive are truly positive. The recall (sensitivity) of the model is approximately 47.8%, meaning that only 47.8% of the actual positive samples are correctly identified. The specificity of the model is approximately 58.8%, indicating that 58.8% of the actual negative samples are correctly identified. The F1 score, which combines precision and recall, is approximately 53.8%, suggesting poor overall performance of the model.

Model 4

	Predicted Negative (0)	Predicted Positive (1)
Actual Negative (0)	16	1
Actual Positive (1)	0	23

Calculating Evaluation Metrics for the Model:

Accuracy = (TP + TN) / Total = (16 + 23) / 40 = 39 / 40 = 0.975 Precision = TP / (TP + FP) = 23 / (23 + 1) = 23 / 24 0.958 Recall (Sensitivity) = TP / (TP + FN) = 23 / (23 + 0) = 23 / 23 = 1 Specificity = TN / (TN + FP) = 16 / (16 + 1) = 16 / 17 0.941 F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.958 * 1) / (0.958 + 1) 0.978

Interpretation of the Model:

The accuracy of the better-performing model is 97.5%, indicating that 97.5% of the predictions are correct. The precision of the model is approximately 95.8%, implying that 95.8% of the samples predicted as positive are truly positive. The recall (sensitivity) of the model is 100%,

meaning that 100% of the actual positive samples are correctly identified. The specificity of the model is approximately 94.1%, indicating that 94.1% of the actual negative samples are correctly identified. The F1 score, which combines precision and recall, is approximately 97.8%, suggesting excellent overall performance of the model.