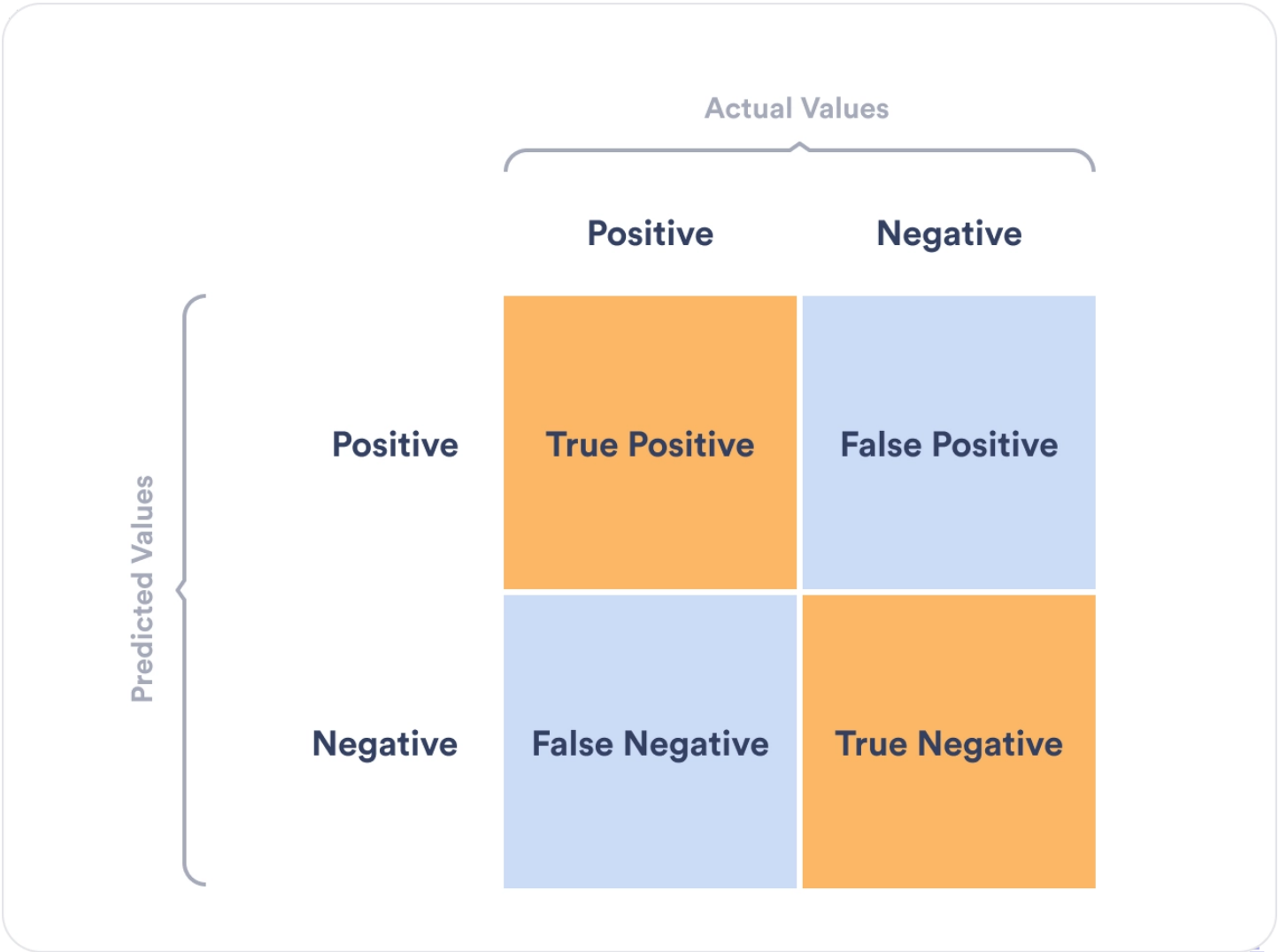
Confusion Matrix and Metrics

Md Abdullah-Al-Kafi

Lecturer, Department of CSE
Daffodil International University

What is a Confusion Matrix?

- A confusion matrix is a tool used to assess the performance of a classification model.
- It compares the actual and predicted classifications.



Confusion Matrix Structure

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- **TP**: Correct positive predictions
- **TN**: Correct negative predictions
- **FP**: Incorrect positive predictions (Type I Error)
- **FN**: Incorrect negative predictions (Type II Error)

Accuracy

Formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Example:

- Suppose we have a confusion matrix:

	Predicted Positive	Predicted Negative
Actual Positive	50	10
Actual Negative	5	35

Solution:

$$\text{Accuracy} = \frac{50 + 35}{50 + 35 + 5 + 10} = \frac{85}{100} = 0.85$$

Therefore, the accuracy is 85%.

Precision (Positive Predictive Value)

Formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Example:

- From the previous confusion matrix:

$$\text{Precision} = \frac{50}{50 + 5} = \frac{50}{55} = 0.91$$

Therefore, the precision is 91%.

Recall (Sensitivity or True Positive Rate)

Formula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Example:

- From the previous confusion matrix:

$$\text{Recall} = \frac{50}{50 + 10} = \frac{50}{60} = 0.83$$

Therefore, the recall is 83%.

F1 Score

Formula:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Example:

- Using precision $= 0.91$ and recall $= 0.83$:

$$\text{F1 Score} = 2 \times \frac{0.91 \times 0.83}{0.91 + 0.83} = 2 \times \frac{0.7553}{1.74} = 0.87$$

Therefore, the F1 Score is 87%.

Specificity (True Negative Rate)

Formula:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Example:

- From the confusion matrix:

$$\text{Specificity} = \frac{35}{35 + 5} = \frac{35}{40} = 0.875$$

Therefore, the specificity is 87.5%.

Matthews Correlation Coefficient (MCC)

Formula:

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Example:

$$\text{MCC} = \frac{(50 \times 35) - (5 \times 10)}{\sqrt{(50 + 5)(50 + 10)(35 + 5)(35 + 10)}}$$

Solving step-by-step:

$$\text{MCC} = \frac{1750 - 50}{\sqrt{(55)(60)(40)(45)}} = \frac{1700}{\sqrt{594000}} = \frac{1700}{771.67} = 0.88$$

Therefore, the MCC is 0.88.

Conclusion

- The confusion matrix provides a powerful way to evaluate classification models.
- Different metrics like accuracy, precision, recall, F1 score, and MCC offer insights into model performance, especially when dealing with imbalanced datasets.
- These metrics help in making decisions about trade-offs in model performance (e.g., precision vs recall).