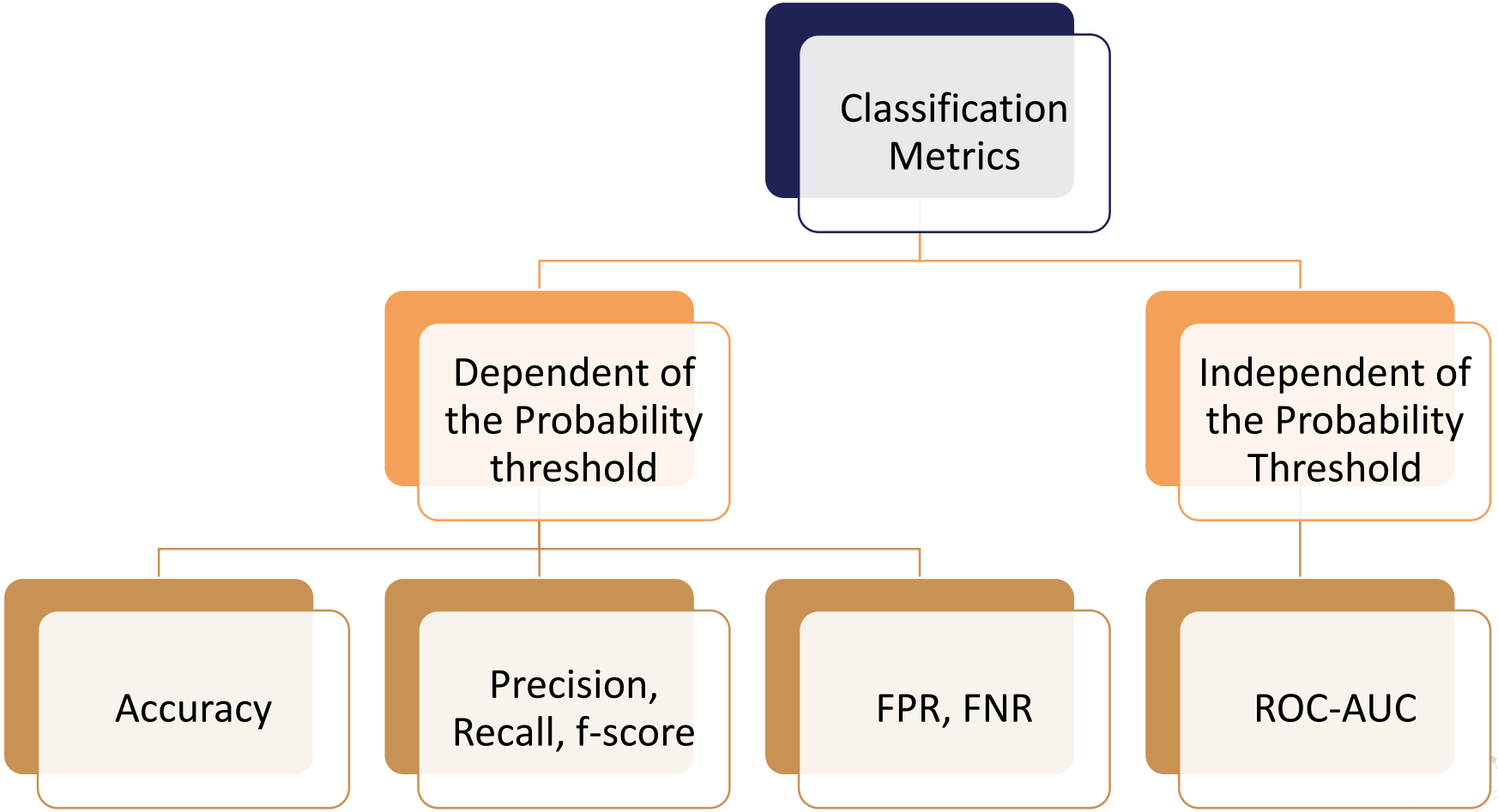
Train In Data

Classification Metrics

Overview

Classification Metrics



Accuracy

- Percentage or fraction of correct predictions
- Fraction of the predictions that the model got right.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

- **TN**: # of negative samples correctly classified - True Negatives
- **FP**: # of negative samples incorrectly classified as positive - False Positives
- **FN**: # of positive samples incorrectly classified as negative - False Negatives
- **TP**: # of positive samples correctly classified - True positives

Precision and Recall

Recall is True Positive divided by Total no. of actual Positive classes i,e True Positive + False Negative (Actually it was Positive but predicted Negative). E.g: Suppose your gf asks you to recall how many number of times you both went for a date. So, the total number of dates you were able to recall correctly will be your True Positive(TP) and Total no. of dates you recalled incorrectly will be your False Negative(FN) i.e they were positive but you recalled it incorrectly (negative).

- True Positive Rate (Recall or Sensitivity)

$$TP_{rate} = TP / (TP + FN)$$

- Positive Predictive Value (Precision)

$$PP_{value} = TP / (TP + FP)$$

Precision is True Positive divided by Total no. of Classes predicted as Positive by the Model (no matter it was actually positive or not) i.e True Positive + False Positive.

- F1 score - weighted harmonic mean of precision and recall.

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

FPR and FNR

- False Positive Rate, FPR = FP / (FP + TN)

  is out of all the Negative Class (TN + FP) , how many of them were incorrectly predicted as Positive though they were actually Negative i,e FP

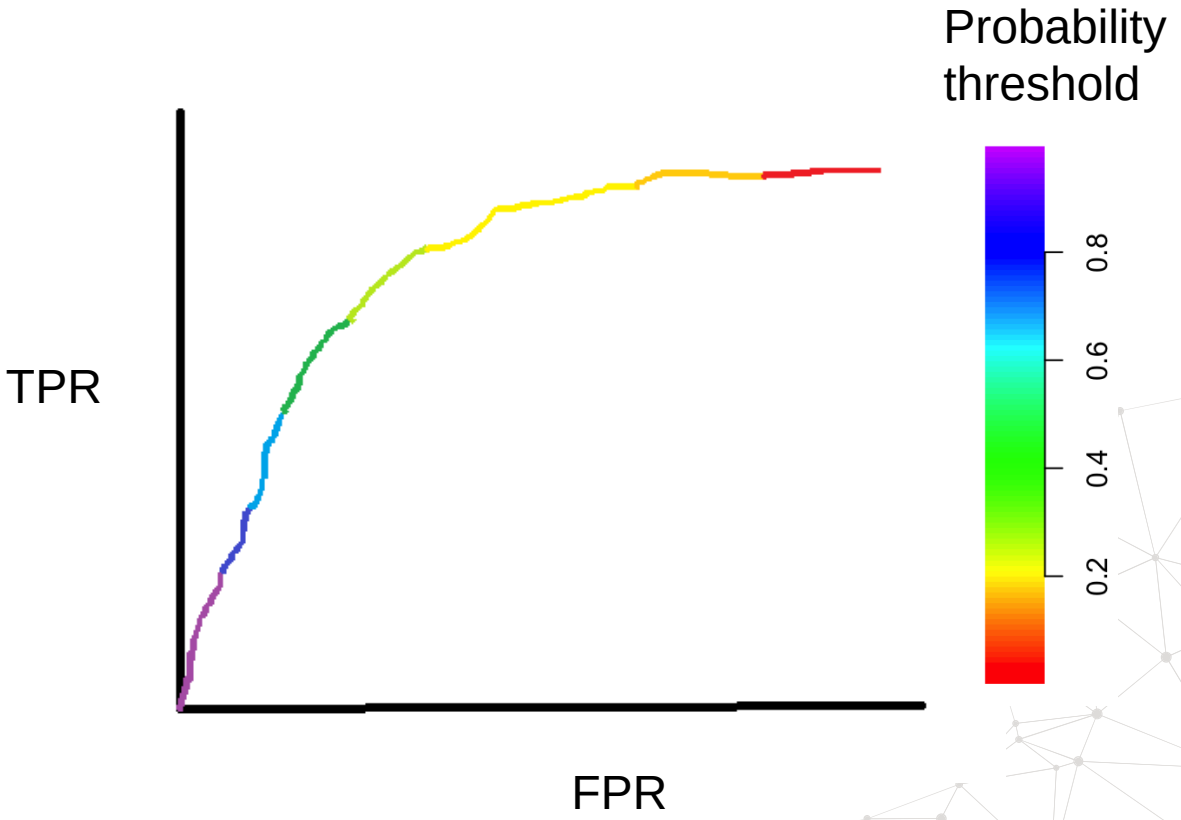
- False Negative Rate FNR = FN / (TP + FN)

  is out of all the Positive Class (TP + FN) , how many of them were incorrectly predicted as Negative though they were Positive i,e FN

Receiving Operating Characteristic Curve (ROC Curve)

- Plots benefits (TPR) vs costs (FPR) at different classification thresholds.

- ROC-AUC: area under the ROC curve



Loss Function

$$L_{\log}(y, p) = -(y \log(p) + (1 - y) \log(1 - p))$$

Train In Data

THANK YOU

www.trainindata.com