

### Exercise 1:

In this session, you will learn to:

- Construct and interpret ROC curves.
- Calculate the Area Under the Curve (AUC).
- Analyze the impact of data imbalance and decision thresholds on model performance.
- Collaborate with peers to compare results and discuss findings.

Below a table that consist of true labels ( $y$ ) and predicted probabilities of four different classifiers ( $\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3, \hat{\pi}_4$ ) generated from hypothetical models.

$y$	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$
1	0.99	0.10	0.01	0.7
1	0.60	0.05	0.40	0.9
1	0.95	0.07	0.05	0.2
1	0.70	0.15	0.30	0.8
0	0.80	0.01	0.20	0.5
0	0.10	0.08	0.90	0.1
0	0.30	0.02	0.70	0.3

### Tasks

- Step 1: Watch as the instructor demonstrates how to plot the ROC curve using  $\hat{\pi}_1$  and explains the steps.
- Step 2: Form groups of 4-6 people and
  - Complete the ROC curve for  $\hat{\pi}_1$ .
  - Plot the ROC curves for  $\hat{\pi}_2$ ,  $\hat{\pi}_3$ , and  $\hat{\pi}_4$ .
  - Manually calculate the AUC for each classifier and compare the results.
  - Compute the prevalence and the average of the predicted probability of each classifier across all 7 observations.
- Step 3: Within your group, discuss:
  - How the differences in predictions affect the ROC curves and AUC values.
  - The differences between average predicted probability and the prevalence.
  - Group A students: Assume you want to obtain a high partial AUC (pAUC) for low FPR values (e.g., using the constraint:  $\text{FPR} < 0.2$ ). Compare the pAUC of the four classifiers.
  - Key takeaways from comparing the four classifiers.
- Step 4: Formulate 1-2 challenging TRUE-FALSE questions about ROC curves and post them into the Etherpad in Moodle. Nominate a group leader to present one question to the class and explain its relevance.