Metrics for classification models evaluation

Confusion Matrix: a table showing correct predictions and types of incorrect predictions.

Precision: the number of true positives divided by all positive predictions. It is a measure of a classifier's exactness. Low precision indicates a high number of false positives.

Recall (or sensitivity): the number of true positives divided by the number of positive values in the test data. It is a measure of a classifier's completeness. Low recall indicates a high number of false negatives.

F1 score: the weighted average of precision and recall.

Area Under ROC Curve (AUROC): AUROC represents the likelihood of your model distinguishing observations from two classes. In other words, if you randomly select one observation from each class, what's the probability that your model will be able to rank them correctly ?