confusion matrix cheat sheet chrissirico.com



classification metrics

Accuracy

- the proportion of all predictions that are correct
- appears high when dataset is imbalanced, even if model is no better than naïve (always predicts the majority class)

Precision / Positive Predictive Value

- accuracy of positive predictions
- proportion of searched w/ contraband

Recall / Sensitivity / True Positive Rate

- proportion of actual occurrences correctly predicted positive
- proportion of smugglers caught

Specificity / True Negative Rate

- proportion of non-occurrences correctly predicted negative
- proportion of innocent travelers passed

False Positive Rate / False Alarm Rate

- proportion of non-occurrences falsely predicted positive
- proportion of innocent travelers searched

False Negative Rate / Miss Rate

- proportion of actual occurrences falsely predicted negative
- proportion of smugglers passed

F1 Score

- harmonic mean of precision & recall
- balances tradeoff between multiple metrics
- assumes equal value/cost of TP, FP, TN, FN

	TP + TN TP + FN + FP + TN
	TP TP + FN
	TN
0	FP + TN
0	FP FP + TN
	FN
	TP + FN
2	precision • recall
	precision + recall

threshold-agnostic metrics vs. binary classification metrics

algorithms, features and hyperparameters. (E.g., log loss, Gini norm, AUC and P-R AUC.)

resulting from various thresholds.



More machine learning resources at **<u>chrissirico.com</u>**:

recall

- algorithm, hyperparameter, feature and threshold selection
- training data setup: group/stratified cross validation and time-based targets • model bias detection, mitigation and fairness metrics

© 2021 Chris Sirico.

- Each cell in the grid at far left represents an observation. A hypothetical classification model "circles" observations it thinks are positives. Actual positive occurrences show up as red dots and non-occurrences
- The confusion matrix tallies the model's correct and incorrect positive and negative predictions (true positives, false negatives, true positives

airport security analogy

- Your goal is to stop smugglers. Search travelers predicted to be smugglers and pass the rest.
- **• TP**: searched smugglers (nice work!) • FN: passed smugglers (oops!) **O FP**: searched, innocent travelers (oops!) **TN**: passed, innocent travelers (nice work!)
- Threshold-agnostic metrics are based on predicted probabilities. They are useful for describing the strength of model signal across all thresholds and for selecting
- **Classification metrics** defined to the left are based on binary (1, 0) predictions. They are useful for selecting a prediction threshold by comparing score tradeoffs

- **P-R AUC** (area under precision-recall curve)
- better for imbalanced datasets with