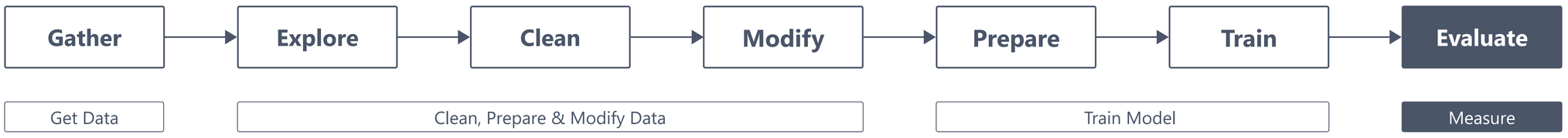
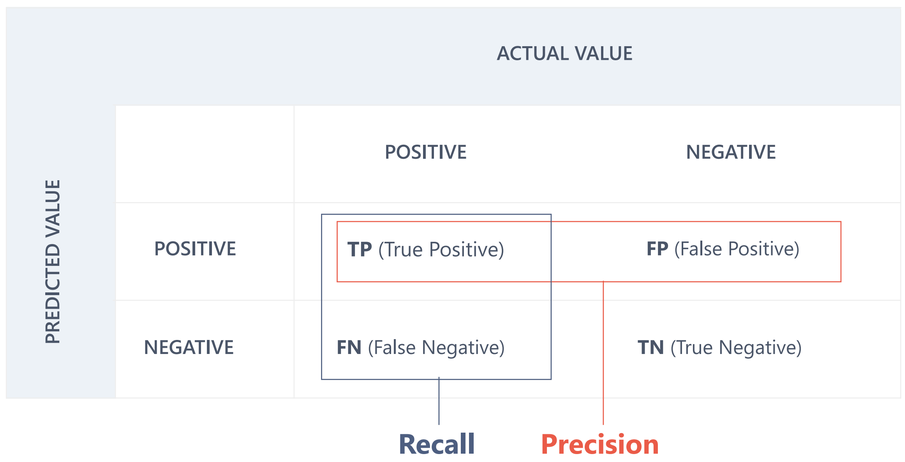
AI Workflow

Model Evaluation - Classification



CONFUSION MATRIX



TP (True Positive) & TN (True Negative):
The prediction matches the Truth correctly
E.g. we correctly predicted an object as a car

FP (False Positive) & FN (False Negative):
The prediction does not match the Truth
E.g. we incorrectly predicted an object as a car, while it was something else

METRICS

ACCURACY

Formula: $\frac{(TP + TN)}{(TP + TN + FP + FN)}$ OR $\frac{\#CORRECT\_PREDICTIONS}{\#TOTAL}$

Summary: How well does the model perform?

Example: Our model is 95% accurate

RECALL

Formula: $\frac{(TP)}{(TP + FN)}$ OR $\frac{\#CORRECT\_POSITIVE\_PREDICTIONS}{\#TRUE\_TRUTH\_VALUES}$

Summary: How often did we wrongly classify something as not true (= false?)

Example: 5% of the time we said it was not a car, while it was a car (we could've hit it)

PRECISION

Formula: $\frac{(TP)}{(TP + FP)}$ OR $\frac{\#CORRECT\_POSITIVE\_PREDICTIONS}{\#POSITIVE\_SAMPLES}$

Summary: How often are we correct in our positive prediction? (or how much are we being wrong?)

Example: 5% of the time we said an object was a car, while it actually was not. (wrong action will be taken - e.g. increasing speed)

F-SCORE (AND F1 SCORE)

Formula: $F_\beta = (1 + \beta^2) \frac{(PRECISION * RECALL)}{(\beta^2 * PRECISION) + RECALL}$

$F_1 = 2 * \frac{(PRECISION * RECALL)}{(PRECISION + RECALL)}$

Summary: Utilize the precision and recall to create a test's accuracy through the "harmonic mean". Also known as the Sørensen–Dice Coefficient

Example: Our model is 88% accurate based on high-business impact markers (#wrong detections and #false positives)

ROC CURVE

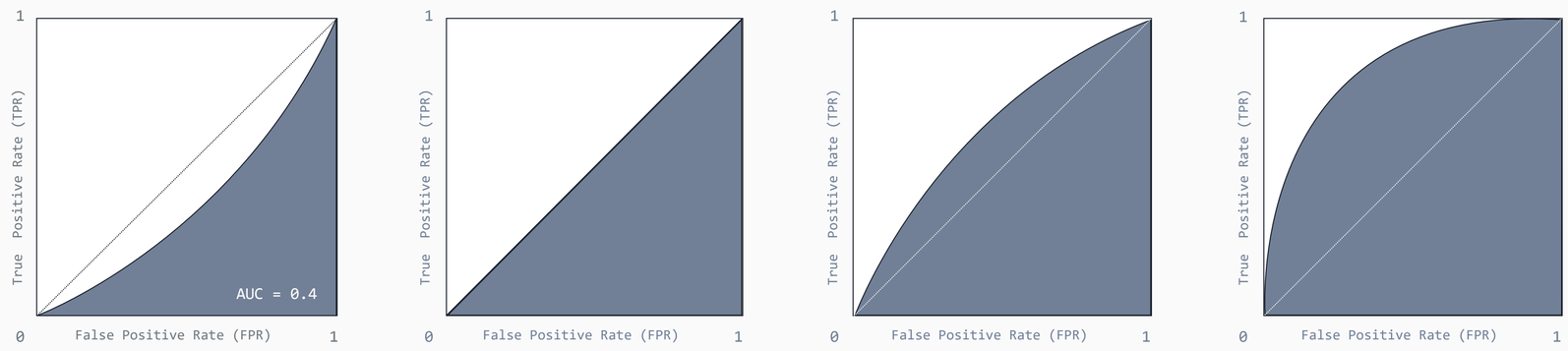
Formula: False Positive Rate (FPR) = X-Axis
True Positive Rate (TPR) = Y-Axis

$TPR = \frac{(TP)}{(TP + FN)}$ $FPR = \frac{(FP)}{(FP + TN)}$

Summary: The ROC curve allows us to select the optimal model and discard suboptimal ones.

Method:
1. Discretize the threshold for the confidence score (e.g. confidence score of [0, 1] becomes [0.0, ..., 0.9, 1.0]
2. Calculate the confusion matrix for the given threshold
3. Determine the TPR and FPR and plot them

Examples:



READ MORE

Performance Markers (https://xaviergeerinck.com/ai-performance-markers)
F1 Score (https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html#sklearn.metrics.f1_score)