# Classification metrics

- Accuracy = (TP + TN) / (TP + FP + FN + TN)
  - use this when the dataset classes are well balanced
- Precision = (TP) / (TP + FP)
  - use this when false positives are unacceptable
  - examples: credit default, crime prediction, etc.
  - If a model is optimized for precision, lots of true positives will fall through the cracks, but those that we do catch we can be more certain that it's not a false positive
- Recall / recall / hit rate / true positive rate / sensitivity
  - Use this when false negatives are unacceptable
  - probability of a positive test, conditioned on truly being positive
  - examples: medical screening / diagnosis
  - If a model is optimized for recall, you may get lots of false alarms, but you'll be very likely to capture all the real emergencies

$$\text{Recall} = \frac{\text{true positives}}{\text{total \# of positives}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Specificity
  - Probability of a negative test, conditioned on being truly negative

$$\text{specificity} = \frac{\text{true negatives}}{\text{total \# of negatives}} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

- F1 score
  - balance between precision and recall

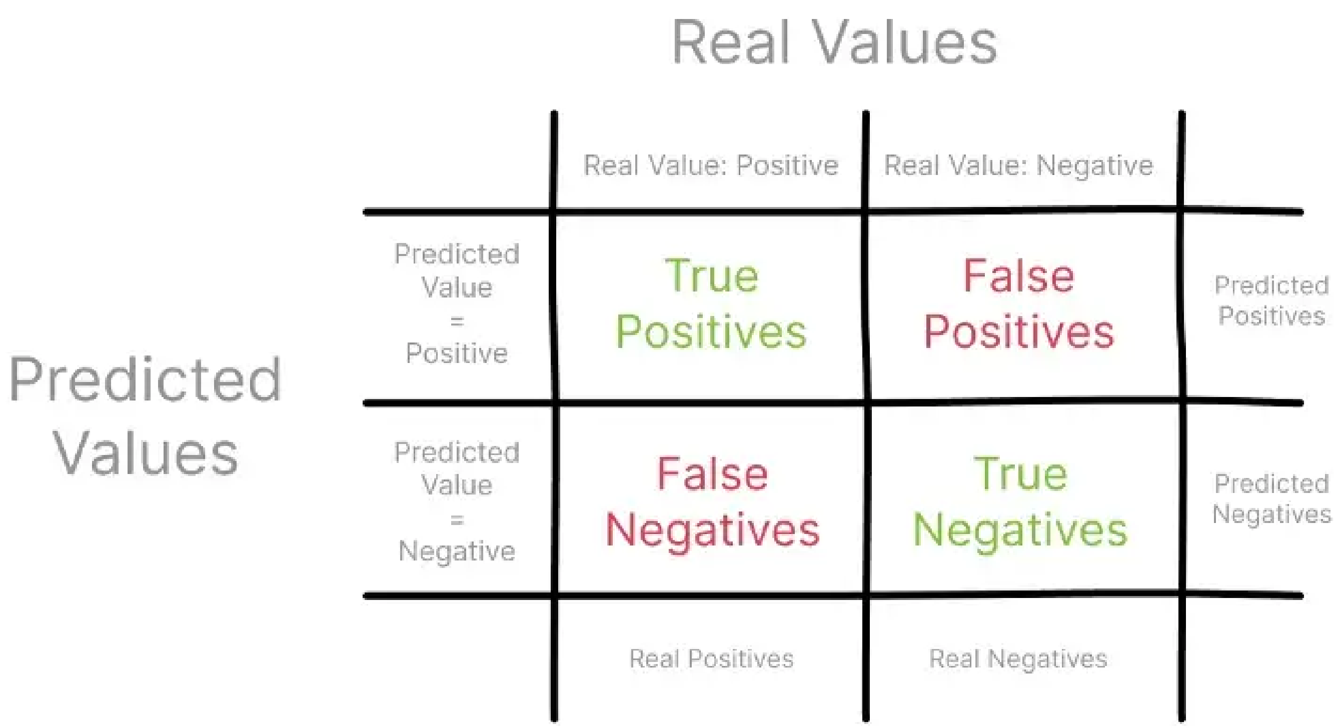
$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

- weighted F1 score
  - use $\beta$ paramerter to describe how much more importance to give to recall vs. precision

$$\text{F}_\beta = (1 + \beta^2) * \frac{\text{precision} * \text{recall}}{(\beta^2 * \text{precision}) + \text{recall}}$$

# ROC Analysis and the AUC - Area under the Curve

Back to the confusion matrix:




**Precision** = *Positive Predictive Value*

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

**Recall** = *True positive rate*

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

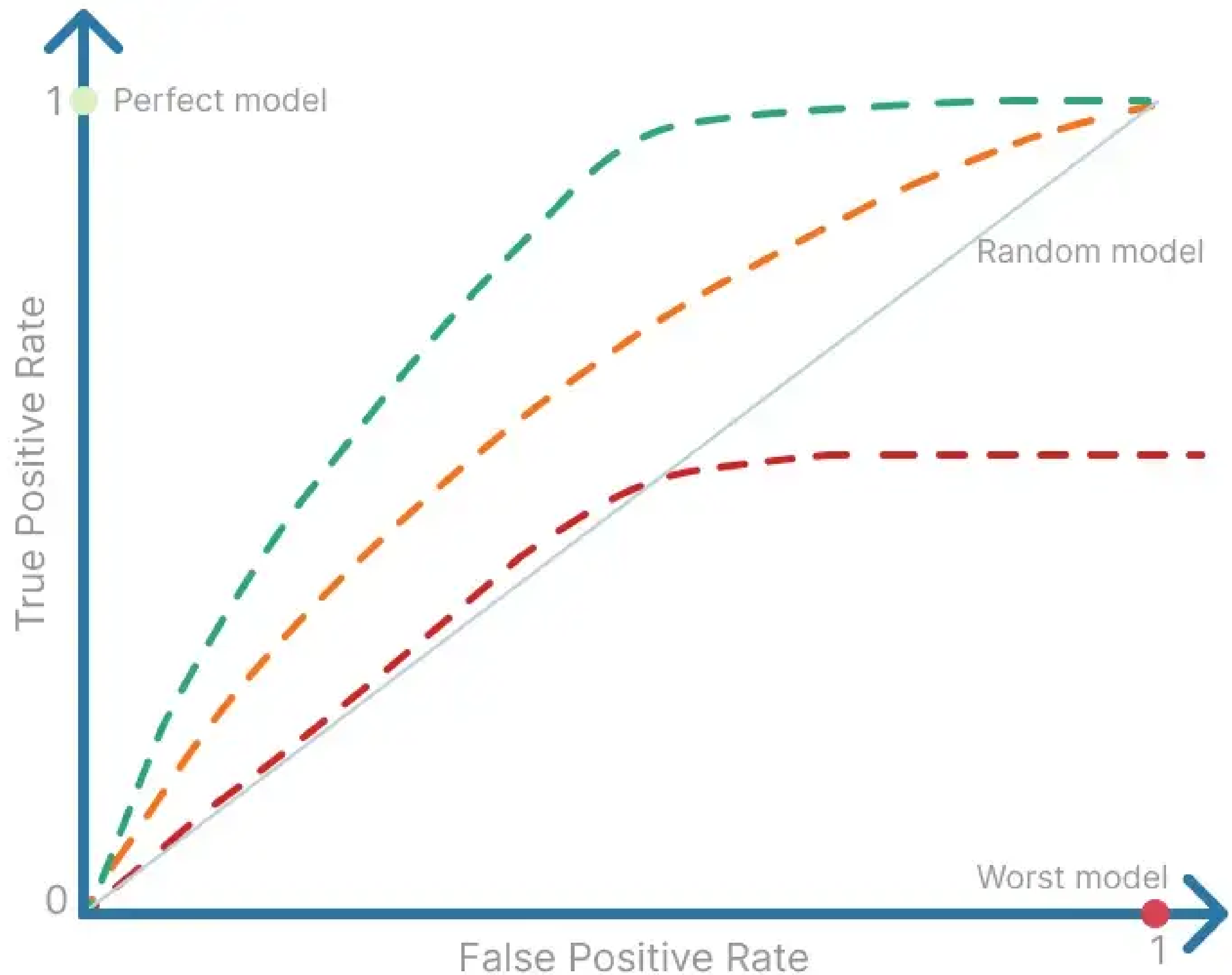
these generally provide a limited view on how the model is performing

A more robust alternative is ROC and AUC

**ROC**

- Take the probabilities that the model gives and calculate the FPR and TRP with many different thresholds for a prediction (e.g. 50%, 51%, 52%, ...)
- *False postive rate* on x-axis, *True positive rate* on y-axis

When you make the plot you get something like this:




Where green > orange > red

**AUC**

to sum up the results of the ROC analysis, use AUC (Area under the curve) (also called a **c-statistic**)

higher AUC generally means a better model

# References


- https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226
- https://en.wikipedia.org/wiki/Sensitivity_and_specificity
- https://towardsdatascience.com/roc-analysis-and-the-auc-area-under-the-curve-404803b694b9