How useful was your classifier?

A lightning tour of some basic metrics

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Example: 100 tumors, either malignant or benign, accuracy 91%

True Positive (TP): • Reality: Malignant • ML model predicted: Malignant • **Number of TP results: 1**	**False Positive (FP):** • Reality: Benign • ML model predicted: Malignant • **Number of FP results: 1**
False Negative (FN): • Reality: Malignant • ML model predicted: Benign • **Number of FN results: 8**	**True Negative (TN):** • Reality: Benign • ML model predicted: Benign • **Number of TN results: 90**

Always predicting benign tumors would achieve the exact same accuracy!

https://developers.google.com/machine-learning/crash-course/classification/accuracy

True Positives (TPs): 1	False Positives (FPs): 1
False Negatives (FNs): 8	True Negatives (TNs): 90

What proportion of positive identifications was actually correct?

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} = 0.5$$

When it predicts a tumor is malignant, it is correct 50% of the time

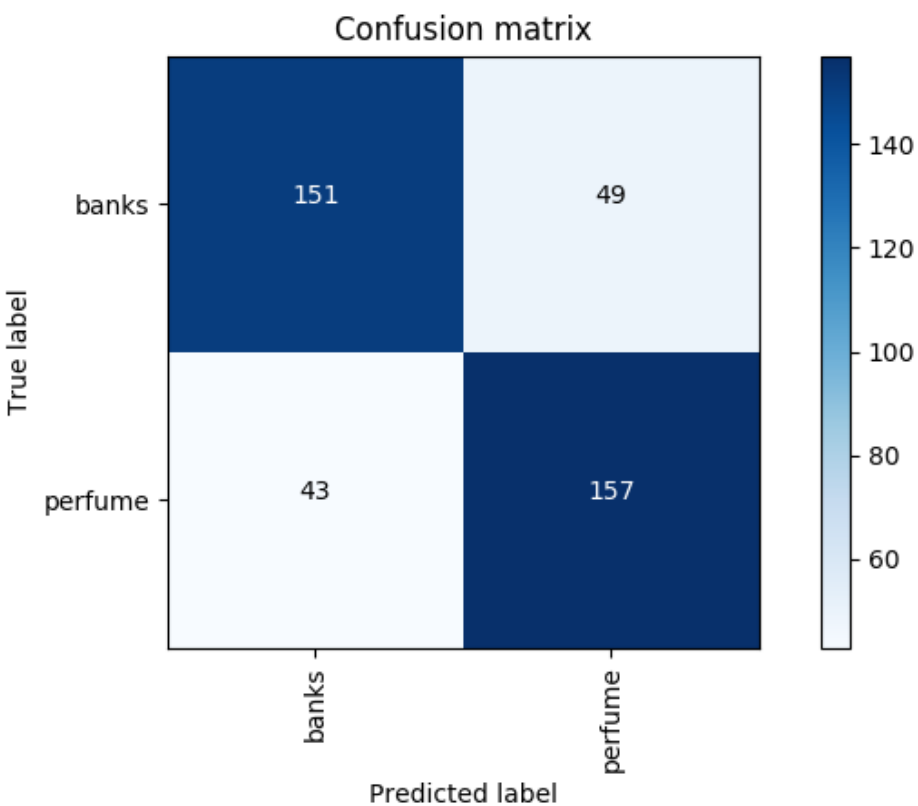
What proportion of actual positives was identified correctly?

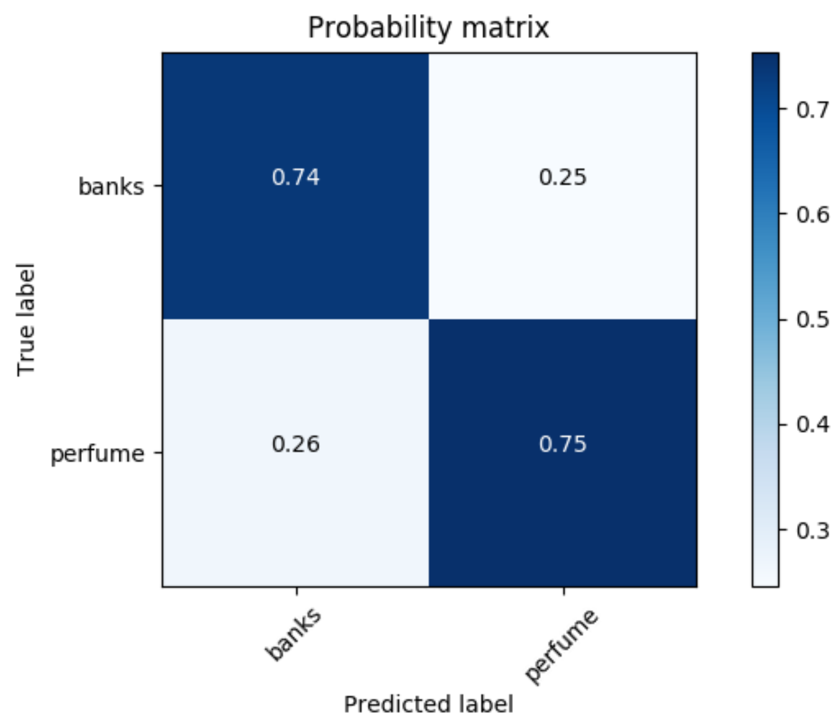
$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = 0.11$$

It correctly identifies 11% of all malignant tumors

https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

A Confusion Matrix is an easy way to visualise a classifiers strengths and weaknesses



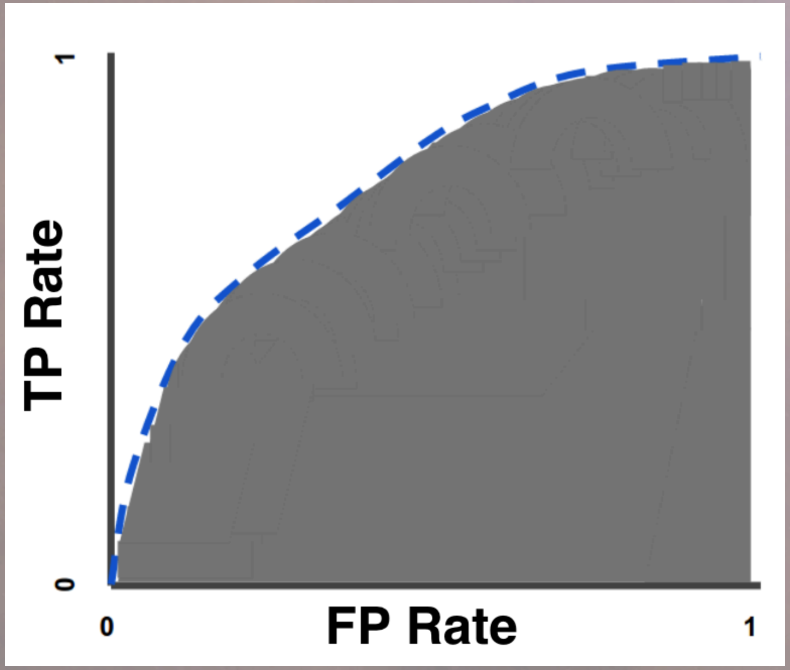


ROC curve (receiver operating characteristic curve)

Shows the performance of a classification model at all classification thresholds. This curve plots two parameters:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$



Area under ROC curve provides an aggregate measure of performance across all possible classification thresholds.

https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc