**Metrics for classification**

In Chapter 1, you evaluated the performance of your k-NN classifier based on its accuracy. However, as Andy discussed, accuracy is not always an informative metric. In this exercise, you will dive more deeply into evaluating the performance of binary classifiers by computing a confusion matrix and generating a classification report.

You may have noticed in the video that the classification report consisted of three rows, and an additional *support* column. The *support* gives the number of samples of the true response that lie in that class - so in the video example, the support was the number of Republicans or Democrats in the test set on which the classification report was computed. The *precision*, *recall*, and *f1-score* columns, then, gave the respective metrics for that particular class.

Here, you'll work with the [PIMA Indians](https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes) dataset obtained from the UCI Machine Learning Repository. The goal is to predict whether or not a given female patient will contract diabetes based on features such as BMI, age, and number of pregnancies. Therefore, it is a binary classification problem. A target value of 0 indicates that the patient does *not* have diabetes, while a value of 1 indicates that the patient *does* have diabetes. As in Chapters 1 and 2, the dataset has been preprocessed to deal with missing values.