Real World Data Science - Performance Measures

The world of academia or Kaggle competitions can be very different when it comes to deciding what makes a good model. Measures such as AUC, precision, recall and F1 score are used to measure the accuracy of classification models, especially those where there is an imbalance of classes in the training set.

Table

Description automatically generated with medium confidence*Confusion Matrix (*[*https://en.wikipedia.org/wiki/Confusion\_matrix*](https://en.wikipedia.org/wiki/Confusion_matrix)*)*

One of the best overall metrics for judging the performance of a classification model is the F1 score as it considers both precision and recall equally while handling imbalance better than accuracy. However, there are some instances where precision is more important than recall and vice versa.

Consider the following scenario from a marketing perspective. In a situation where you may be trying to attract new consumers over to your product who haven't historically purchased your product in the past, then there is some allowance for false positives in a model which determines product purchasers. Even though they might not be a purchaser now, they could become one in the future so it's still worthwhile to use meaningful campaigns towards them. Whereas with false negatives, the consumers may have purchased historically but because you are now ignoring them from a marketing perspective, you may lose some of that audience.

This is why another, potentially more useful metric than F1 score is to use cost matrix gain. That is, what is the financial impact of classifying a data point as a True Postive, True Negative, False Positive, or False Negative?