Lab 4 Classification Models

In this lab the objective was to report various scores on three different machine learning classification models. We were then to describe the key differences between the three models and list the pros and cons of choosing one over the other.

**Naïve Bayes**

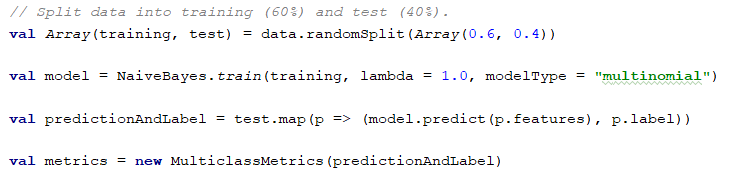
Confusion Matrix: 7.0 1.0

0.0 19.0

Accuracy: 96.2%

Precision: .95

Recall: .875



The code above is the code that setup the data into training and prediction datasets so that it could be used in the NaiveBayes function. The splitting method is used on the other classification methods as well.

Naïve Bayes is useful on large datasets. The dataset that we used was pretty large, so you see that it ends up with a pretty good accuracy score.

**Decision Tree**

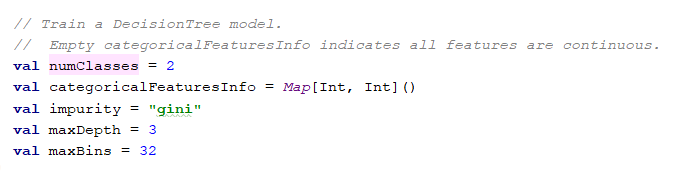
Confusion Matrix: 9.0 0.0

0.0 22.0

Accuracy: 95.2%

Precision: 1.0

Recall: 1.0



These are the settings that were used for the decision tree. These are pretty standard settings that produce a pretty reliable result.

**Random Forest**

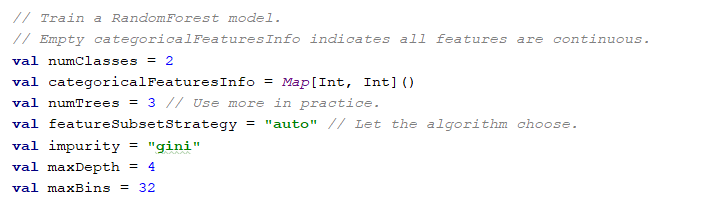
Confusion Matrix: 14.0 0.0

1.0 12.0

Accuracy: 96.3%

Precision:93%

Recall: 92.3%



These are the settings that were used for the Random Forest. They are also pretty standard settings that produce a pretty reliable result.

**Which to Use**

When determining which classification model to use you need to keep in mind these characteristics of your dataset. In this case, the dataset was rather large, which is why Naïve Bayes performed as well as it did. The other options can be very good if you are looking