**Performance Measures**

There are three things to note in the performance summary for classification algorithms:

1. **Classification accuracy**. This the ratio of the number of correct predictions out of all predictions made, often presented as a percentage where 100% is the best an algorithm can achieve. If your data has unbalanced classes, you may want to look into the Kappa metric which presents the same information taking the class balance into account.

**Kappa Statistics:** The Kappa statistic (or value) is a metric that compares an **Observed Accuracy** with an **Expected Accuracy** (random chance). The kappa statistic is used not only to evaluate a single classifier, but also to evaluate classifiers amongst themselves. In addition, it takes into account random chance (agreement with a random classifier), which generally means it is less misleading than simply using accuracy as a metric. Computation of **Observed Accuracy** and **Expected Accuracy** is integral to comprehension of the kappa statistic, and is most easily illustrated through use of a confusion matrix.

1. **Accuracy by class**. Take note of the true-positive(TP) and false-positive(FP) rates for the predictions for each class which can be instructive of the class breakdown for the problem is uneven or there are more than two classes. This can help you interpret the results if predicting one class is more important than predicting another.
2. **Confusion matrix**. A table showing the number of predictions for each class compared to the number of instances that actually belong to each class. This is very useful to get an overview of the types of mistakes the algorithm made.

Training data – 1000 observations

Validation data – 1111 observations

119 variables/predictors

**Summary**

Models: KNN, SVM, RF, C50, CART

Number of resamples: 10

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

KNN 0.9800 0.9899 1.00 0.9933 1.0000 1 1

SVM 0.9802 0.9899 0.99 0.9920 0.9975 1 0

RF 1.0000 1.0000 1.00 1.0000 1.0000 1 2

C50 0.9802 1.0000 1.00 0.9980 1.0000 1 0

RPART 0.9900 0.9926 1.00 0.9970 1.0000 1 0

Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

KNN 0.9594 0.9796 1.0000 0.9865 1.0000 1 1

SVM 0.9602 0.9796 0.9798 0.9839 0.9951 1 0

RF 1.0000 1.0000 1.0000 1.0000 1.0000 1 2

C50 0.9608 1.0000 1.0000 0.9961 1.0000 1 0

RPART 0.9798 0.9851 1.0000 0.9940 1.0000 1 0

**KNN Model**

**Accuracy and Kappa Statistics**

|  |  |  |
| --- | --- | --- |
| K | Accuracy | Kappa |
| 5 | 0.9924874 | 0.9847971 |

**Confusion Matrix**

      Reference

Prediction 0-0 2-1 2-2 3-2

       0-0 0   0 0 0

       2-1   0 351   0   5

       2-2   0   0   0   0

       3-2   0   1   0 442

 Accuracy (average): 0.9925

**SVM Model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Cost** | **Loss** | **Accuracy** | **Kappa** |
| **2.00** | **L1** | **0.9929998** | **0.9858851** |

**Confusion Matrix**

Prediction 0-0 2-1 2-2 3-2

       0-0   0   0   0   0

       2-1   1 440   1   3

       2-2   0   0   0   0

       3-2   0   1   0 554

 Accuracy (average): 0.994

**Random Forest Model**

|  |  |  |
| --- | --- | --- |
| **Mtry** | **Accuracy** | **Kappa** |
| **2** | 1.0000000 | 1.0000000 |

**Confusion Matrix**

Reference

Prediction 0-0 2-1 2-2 3-2

0-0 0 0 0 0

2-1 0 352 0 0

2-2 0 0 0 0

3-2 0 0 0 445

Accuracy (average): 1

**C5.0 Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Winnow** | **Trials** | **Accuracy** | **Kappa** |
| **Rules** | **FALSE** | **10** | **0.9980198** | **0.9960792** |

**Confusion Matrix**

Reference

Prediction 0-0 1-0 2-0 3-0 0-1 1-1 2-1 3-1 0-2 1-2 2-2 3-2 4-2

0-0 0 0 0 0 0 0 0 0 0 0 0 0 0

1-0 0 0 0 0 0 0 0 0 0 0 0 0 0

2-0 0 0 0 0 0 0 0 0 0 0 0 0 0

3-0 0 0 0 0 0 0 0 0 0 0 0 0 0

0-1 0 0 0 0 0 0 0 0 0 0 0 0 0

1-1 0 0 0 0 0 0 0 0 0 0 0 0 0

2-1 0 0 0 0 0 0 1 0 6 42 38 23 17

3-1 0 0 0 0 0 0 0 0 0 0 0 0 0

0-2 0 0 0 0 0 0 0 0 0 0 0 0 0

1-2 0 0 0 0 0 0 0 0 0 0 0 0 0

2-2 0 0 0 0 0 0 0 0 0 0 0 0 0

3-2 78 208 165 85 30 143 86 47 18 69 16 17 22

4-2 0 0 0 0 0 0 0 0 0 0 0 0 0

Accuracy : 0.0162

**CART Model**

|  |  |  |
| --- | --- | --- |
| **Cp** | **Accuracy** | **Kappa** |
| **0.0000** | **0.9970099** | **0.9939875** |

Confusion Matrix

Reference

Prediction 0-0 1-0 2-0 3-0 0-1 1-1 2-1 3-1 0-2 1-2 2-2 3-2 4-2

0-0 0 0 0 0 0 0 0 0 0 0 0 0 0

1-0 0 0 0 0 0 0 0 0 0 0 0 0 0

2-0 0 0 0 0 0 0 0 0 0 0 0 0 0

3-0 0 0 0 0 0 0 0 0 0 0 0 0 0

0-1 0 0 0 0 0 0 0 0 0 0 0 0 0

1-1 0 0 0 0 0 0 0 0 0 0 0 0 0

2-1 0 0 0 0 0 0 0 0 0 8 11 12 3

3-1 0 0 0 0 0 0 0 0 0 0 0 0 0

0-2 0 0 0 0 0 0 0 0 0 0 0 0 0

1-2 0 0 0 0 0 0 0 0 0 0 0 0 0

2-2 0 0 0 0 0 0 0 0 0 0 0 0 0

3-2 78 208 165 85 30 143 87 47 24 103 43 28 36

4-2 0 0 0 0 0 0 0 0 0 0 0 0 0

Accuracy : 0.0252