Additional Results for Task 3 – Course 3 – WiFi Fingerprinting

By Roberto Siegert – Nov. 18th 2019

Part II

Confusion Matrices and Performance Measurements

Analysis

When looking into the entire population of WAP signal levels, we obtained the following Confusion Matrix,

```
> knn_CM
Confusion Matrix and Statistics

          Reference
Prediction    0    1    2
         0 1312    0    0
         1    0 1299    0
         2    0    0 2373

Overall Statistics

               Accuracy : 1
                 95% CI : (0.9993, 1)
    No Information Rate : 0.4761
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Sensitivity            1.0000   1.0000   1.0000
Specificity            1.0000   1.0000   1.0000
Pos Pred Value         1.0000   1.0000   1.0000
Neg Pred Value         1.0000   1.0000   1.0000
Prevalence             0.2632   0.2606   0.4761
Detection Rate         0.2632   0.2606   0.4761
Detection Prevalence   0.2632   0.2606   0.4761
Balanced Accuracy      1.0000   1.0000   1.0000
```

With Accuracy and Kappa values equivalent to 1.

But we can continue evaluating at higher levels of granularity by sorting out the population of WAP signal levels, by individual building, and hence creating three(3) decoupled new datasets.

First we evaluate the “K- Nearest Neighbor” algorithm for the 3x new datasets, obtaining:

For Building 1, we obtained

```
> knn_CM1
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3
         0 264   3   2   0
         1   0 336   1   0
         2   0   0 356   0
         3   0   0   1 347

Overall Statistics

               Accuracy : 0.9947
                 95% CI : (0.989, 0.9978)
    No Information Rate : 0.2748
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9928
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   0.9912   0.9889   1.0000
Specificity            0.9952   0.9990   1.0000   0.9990
Pos Pred Value         0.9814   0.9970   1.0000   0.9971
Neg Pred Value         1.0000   0.9969   0.9958   1.0000
Prevalence             0.2015   0.2588   0.2748   0.2649
Detection Rate         0.2015   0.2565   0.2718   0.2649
Detection Prevalence   0.2053   0.2573   0.2718   0.2656
Balanced Accuracy      0.9976   0.9951   0.9944   0.9995
```

And similarly for DataSet2 and DataSet2, - building 2 and 3, respectively.

```
knn_CM2
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3
         0 342   0   0   0
         1   0 371   0   0
         2   0   0 348   0
         3   0   0   1 237

Overall Statistics

               Accuracy : 0.9992
                 95% CI : (0.9957, 1)
    No Information Rate : 0.2856
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.999
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   1.0000   0.9971   1.0000
Specificity            1.0000   1.0000   1.0000   0.9991
```

```
Pos Pred Value         1.0000   1.0000   1.0000   0.9958
Neg Pred Value         1.0000   1.0000   0.9989   1.0000
Prevalence             0.2633   0.2856   0.2687   0.1824
Detection Rate         0.2633   0.2856   0.2679   0.1824
Detection Prevalence   0.2633   0.2856   0.2679   0.1832
Balanced Accuracy      1.0000   1.0000   0.9986   0.9995
> knn_CM3
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3   4
         0 485   2   0   2   1
         1   0 538   3   0   0
         2   0   0 391   0   1
         3   0   0   0 675   2
         4   0   0   0   0 271

Overall Statistics

               Accuracy : 0.9954
                 95% CI : (0.9917, 0.9977)
    No Information Rate : 0.2855
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9941
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3 Class: 4
Sensitivity            1.0000   0.9963   0.9924   0.9970   0.9855
Specificity            0.9973   0.9984   0.9995   0.9988   1.0000
Pos Pred Value         0.9898   0.9945   0.9974   0.9970   1.0000
Neg Pred Value         1.0000   0.9989   0.9985   0.9988   0.9981
Prevalence             0.2046   0.2278   0.1662   0.2855   0.1160
Detection Rate         0.2046   0.2269   0.1649   0.2847   0.1143
Detection Prevalence   0.2067   0.2282   0.1653   0.2855   0.1143
Balanced Accuracy      0.9987   0.9973   0.9959   0.9979   0.9927

>
```

Now, we can evaluate other algorithms:

Evaluation of KKNN algorithm

```
> kknn_1
k-Nearest Neighbors

3939 samples
 200 predictor
   4 classes: '0', '1', '2', '3'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3544, 3546, 3547, 3545, 3544, 3544, ...
Resampling results across tuning parameters:

  kmax  Accuracy   Kappa
   5    0.9824846  0.9765502
   7    0.9845176  0.9792678
   9    0.9850266  0.9799492
  11    0.9850266  0.9799492
  13    0.9850266  0.9799492

Tuning parameter 'distance' was held constant at a value of 2
```

```
Tuning parameter 'kernel' was held constant at a value of optimal
Accuracy was used to select the optimal model using  the largest value.
The final values used for the model were kmax = 13, distance = 2 and kernel = optimal.
> kknn_2
k-Nearest Neighbors

3897 samples
 207 predictor
   4 classes: '0', '1', '2', '3'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3508, 3508, 3508, 3507, 3508, 3507, ...
Resampling results across tuning parameters:

  kmax  Accuracy   Kappa
   5    0.9935877  0.9913787
   7    0.9935877  0.9913787
   9    0.9935877  0.9913787
  11    0.9935877  0.9913787
  13    0.9935877  0.9913787

Tuning parameter 'distance' was held constant at a value of 2
Tuning parameter 'kernel' was held constant at a value of optimal
Accuracy was used to select the optimal model using  the largest value.
The final values used for the model were kmax = 13, distance = 2 and kernel = optimal.
> kknn_3
k-Nearest Neighbors

7121 samples
 203 predictor
   5 classes: '0', '1', '2', '3', '4'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 6410, 6408, 6409, 6409, 6409, 6409, ...
Resampling results across tuning parameters:

  kmax  Accuracy   Kappa
   5    0.9827279  0.9779693
   7    0.9834301  0.9788638
   9    0.9835707  0.9790432
  11    0.9835707  0.9790432
  13    0.9835707  0.9790432

Tuning parameter 'distance' was held constant at a value of 2
Tuning parameter 'kernel' was held constant at a value of optimal
Accuracy was used to select the optimal model using  the largest value.
The final values used for the model were kmax = 13, distance = 2 and kernel = optimal.
```

Now, we can evaluate the Random Forest algorithm:

```
> rf_CM1
Confusion Matrix and Statistics
```

```
          Reference
Prediction   0   1   2   3
         0 264   1   0   0
         1   0 337   1   0
         2   0   1 359   0
         3   0   0   0 347

Overall Statistics

               Accuracy : 0.9977
                 95% CI : (0.9933, 0.9995)
    No Information Rate : 0.2748
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9969
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   0.9941   0.9972   1.0000
Specificity            0.9990   0.9990   0.9989   1.0000
Pos Pred Value         0.9962   0.9970   0.9972   1.0000
Neg Pred Value         1.0000   0.9979   0.9989   1.0000
Prevalence             0.2015   0.2588   0.2748   0.2649
Detection Rate         0.2015   0.2573   0.2740   0.2649
Detection Prevalence   0.2023   0.2580   0.2748   0.2649
Balanced Accuracy      0.9995   0.9965   0.9981   1.0000
> rf_CM2
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3
         0 342   1   0   0
         1   0 370   0   0
         2   0   0 349   1
         3   0   0   0 236

Overall Statistics

               Accuracy : 0.9985
                 95% CI : (0.9944, 0.9998)
    No Information Rate : 0.2856
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9979
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   0.9973   1.0000   0.9958
Specificity            0.9990   1.0000   0.9989   1.0000
Pos Pred Value         0.9971   1.0000   0.9971   1.0000
Neg Pred Value         1.0000   0.9989   1.0000   0.9991
Prevalence             0.2633   0.2856   0.2687   0.1824
```

```
Detection Rate         0.2633   0.2848   0.2687   0.1817
Detection Prevalence   0.2640   0.2848   0.2694   0.1817
Balanced Accuracy      0.9995   0.9987   0.9995   0.9979
> rf_CM3
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3   4
         0 485   1   0   0   0
         1   0 539   1   0   0
         2   0   0 393   0   2
         3   0   0   0 677   1
         4   0   0   0   0 272

Overall Statistics

               Accuracy : 0.9979
                 95% CI : (0.9951, 0.9993)
    No Information Rate : 0.2855
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9973
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3 Class: 4
Sensitivity            1.0000   0.9981   0.9975   1.0000   0.9891
Specificity            0.9995   0.9995   0.9990   0.9994   1.0000
Pos Pred Value         0.9979   0.9981   0.9949   0.9985   1.0000
Neg Pred Value         1.0000   0.9995   0.9995   1.0000   0.9986
Prevalence             0.2046   0.2278   0.1662   0.2855   0.1160
Detection Rate         0.2046   0.2273   0.1658   0.2855   0.1147
Detection Prevalence   0.2050   0.2278   0.1666   0.2860   0.1147
Balanced Accuracy      0.9997   0.9988   0.9982   0.9997   0.9945
```

Now, we evaluate the C5.0 algorithm, also included in the Caret package:

```
C50_CM1
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3
         0 263   1   0   0
         1   1 338   2   0
         2   0   0 357   0
         3   0   0   1 347

Overall Statistics

               Accuracy : 0.9962
                 95% CI : (0.9911, 0.9988)
    No Information Rate : 0.2748
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9949
```

```
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            0.9962   0.9971   0.9917   1.0000
Specificity            0.9990   0.9969   1.0000   0.9990
Pos Pred Value         0.9962   0.9912   1.0000   0.9971
Neg Pred Value         0.9990   0.9990   0.9969   1.0000
Prevalence             0.2015   0.2588   0.2748   0.2649
Detection Rate         0.2008   0.2580   0.2725   0.2649
Detection Prevalence   0.2015   0.2603   0.2725   0.2656
Balanced Accuracy      0.9976   0.9970   0.9958   0.9995
> C50_CM2
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3
         0 342   3   0   0
         1   0 368   0   0
         2   0   0 346   0
         3   0   0   3 237

Overall Statistics

               Accuracy : 0.9954
                 95% CI : (0.99, 0.9983)
    No Information Rate : 0.2856
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9938
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            1.0000   0.9919   0.9914   1.0000
Specificity            0.9969   1.0000   1.0000   0.9972
Pos Pred Value         0.9913   1.0000   1.0000   0.9875
Neg Pred Value         1.0000   0.9968   0.9969   1.0000
Prevalence             0.2633   0.2856   0.2687   0.1824
Detection Rate         0.2633   0.2833   0.2664   0.1824
Detection Prevalence   0.2656   0.2833   0.2664   0.1848
Balanced Accuracy      0.9984   0.9960   0.9957   0.9986
> C50_CM3
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2   3   4
         0 485   1   0   0   0
         1   0 539   0   0   0
         2   0   0 394   0   1
         3   0   0   0 677   0
         4   0   0   0   0 274

Overall Statistics
```

```
               Accuracy : 0.9992
                 95% CI : (0.997, 0.9999)
    No Information Rate : 0.2855
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9989
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3 Class: 4
Sensitivity            1.0000   0.9981   1.0000   1.0000   0.9964
Specificity            0.9995   1.0000   0.9995   1.0000   1.0000
Pos Pred Value         0.9979   1.0000   0.9975   1.0000   1.0000
Neg Pred Value         1.0000   0.9995   1.0000   1.0000   0.9995
Prevalence             0.2046   0.2278   0.1662   0.2855   0.1160
Detection Rate         0.2046   0.2273   0.1662   0.2855   0.1156
Detection Prevalence   0.2050   0.2273   0.1666   0.2855   0.1156
Balanced Accuracy      0.9997   0.9991   0.9997   1.0000   0.9982
```