
BIRZEIT UNIVERSITY

Faculty of Engineering and Technology

Department of Electrical and Computer Engineering

# ARTIFICIAL INTELLIGENCE

# ENCS 3340

---

Project II

Raisin Dataset

---

Prepared by:

**Mohammed Abed Alkareem** **ID: 1210708**

**Mosa Sbeih** **ID: 1211250**

Instructor: **Dr. Yazan Abu Farha**

Section: **4**

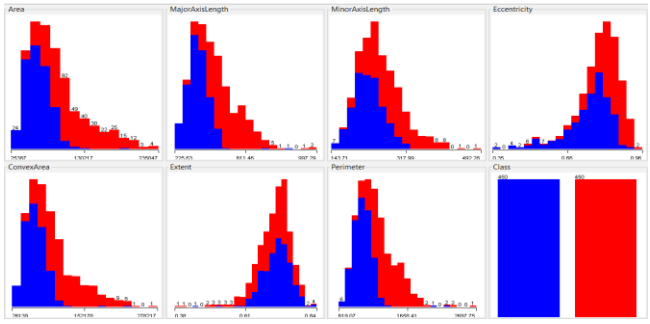
Date: **Saturday, June 22, 2024**

# 1. Introduction

This report evaluates three machine learning algorithms—Decision Tree (J48), Naïve Bayes, and Multilayer Perceptron (MLP)—using the Raisin Dataset. The performance of each model was analyzed through various preprocessing techniques and hyper-parameter adjustments.

# 2. Dataset and Attributes

- **Name**: Raisin Dataset
- **Instances**: 900
- **Attributes**: 8 (Area, MajorAxisLength, MinorAxisLength, Eccentricity, ConvexArea, Extent, Perimeter, Class)




# 3. Experiments and Results

## 3.1 Decision Tree (J48)

**Preprocessing**: Discretized of 5 bins to all attributes using the Discretize filter.

Hyper-Parameters

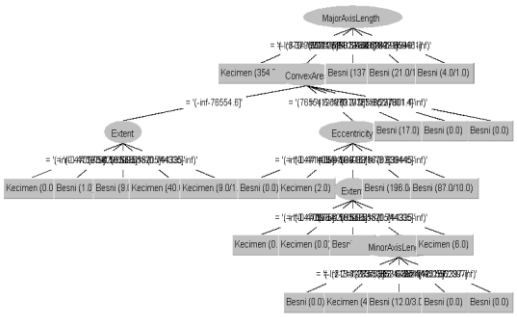
Initial

| Hyper-Parameters | Settings |
|---|---|
| confidenceFactor | 0.25 |
| minNumObj | 2 |
| useLaplace | FALSE |
| binarySplits | FALSE |
| Seed | 1 |
| f=US | 3 |

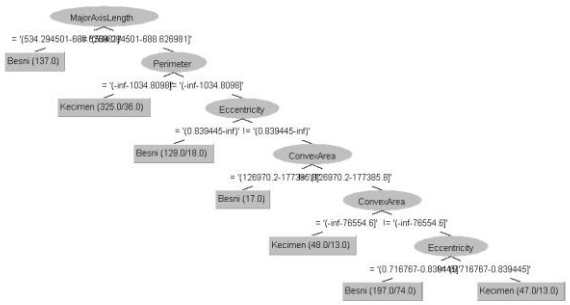
Updated

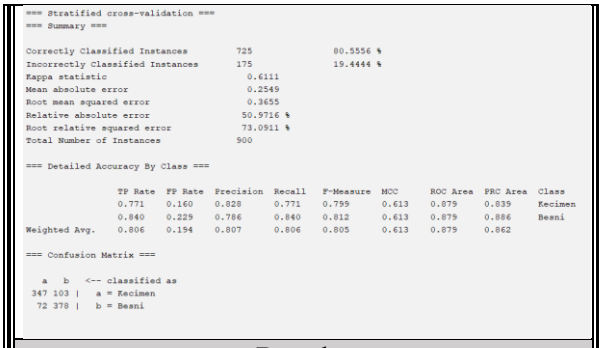
| Hyper-Parameters | Settings |
|---|---|
| confidenceFactor | 0.15 |
| minNumObj | 5 |
| useLaplace | TRUE |
| binarySplits | TRUE |
| Seed | 1 |
| f=US | 5 |

Decision Tree

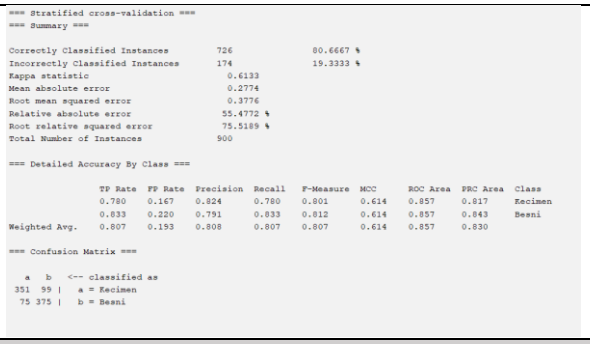



Decision Tree





```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         725               80.5556 %
Incorrectly Classified Instances       175               19.4444 %
Kappa statistic                          0.6111
Mean absolute error                      0.2549
Root mean squared error                  0.3655
Relative absolute error                 50.9716 %
Root relative squared error             73.0911 %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.771    0.160    0.828      0.771    0.799      0.613    0.879     0.839     Kecimen
                 0.840    0.229    0.786      0.840    0.812      0.613    0.879     0.886     Besni
Weighted Avg.    0.806    0.194    0.807      0.806    0.805      0.613    0.879     0.862

=== Confusion Matrix ===

   a   b   <-- classified as
 347 103 |   a = Kecimen
  72 378 |   b = Besni
```



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         726               80.6667 %
Incorrectly Classified Instances       174               19.3333 %
Kappa statistic                          0.6133
Mean absolute error                      0.2774
Root mean squared error                  0.3776
Relative absolute error                 55.4772 %
Root relative squared error             75.5189 %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.780    0.167    0.824      0.780    0.801      0.614    0.857     0.817     Kecimen
                 0.833    0.220    0.791      0.833    0.812      0.614    0.857     0.843     Besni
Weighted Avg.    0.807    0.193    0.808      0.807    0.807      0.614    0.857     0.830

=== Confusion Matrix ===

   a   b   <-- classified as
 351  99 |   a = Kecimen
  75 375 |   b = Besni
```

| Results | | |
|---|---|---|
| Confusion Matrix | Predicted Kecimen | Predicted Besni |
| True Kecimen | 347 (TP) | 103 (FP) |
| True Besni | 72 (FN) | 378 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.805 | 80.56% | 0.807 | 0.806 |

| Results | | |
|---|---|---|
| Confusion Matrix | Predicted Kecimen | Predicted Besni |
| True Kecimen | 351 (TP) | 99 (FP) |
| True Besni | 75 (FN) | 375 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.807 | 80.67% | 0.808 | 0.807 |

The initial Decision Tree (J48) model achieved 80.56% accuracy. Updated hyper-parameters improved accuracy to 80.67% but increased error metrics, indicating potential overfitting. This highlights the need for careful tuning to balance accuracy and error minimization.

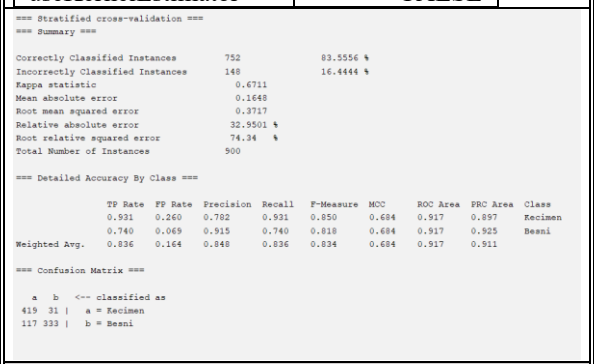
## 3.2 Naïve Bayes

**Preprocessing**: Applied Normalize filter to all numeric attributes.

Hyper-Parameters

Initial

| Hyper-Parameters | Initial Settings |
|---|---|
| useKernelEstimator | FALSE |



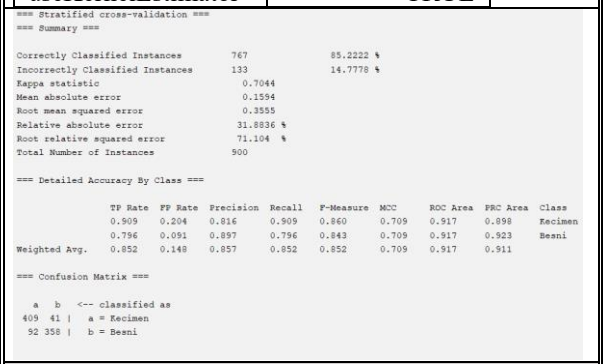
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         752               83.5556 %
Incorrectly Classified Instances       148               16.4444 %
Kappa statistic                          0.6711
Mean absolute error                      0.1648
Root mean squared error                  0.3717
Relative absolute error                 32.9501 %
Root relative squared error             74.34   %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.931    0.260    0.782      0.931    0.850      0.684    0.917     0.897     Kecimen
                 0.740    0.069    0.915      0.740    0.818      0.684    0.917     0.925     Besni
Weighted Avg.    0.836    0.164    0.848      0.836    0.834      0.684    0.917     0.911

=== Confusion Matrix ===

   a   b   <-- classified as
 419  31 |   a = Kecimen
 117 333 |   b = Besni
```

| Results | | |
|---|---|---|
| Confusion Matrix | Predicted Kecimen | Predicted Besni |
| True Kecimen | 419 (TP) | 31 (FP) |
| True Besni | 117 (FN) | 333 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.834 | 83.56% | 0.848 | 0.836 |

Updated

| Hyper-Parameters | Initial Settings |
|---|---|
| useKernelEstimator | TRUE |



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         767               85.2222 %
Incorrectly Classified Instances       133               14.7778 %
Kappa statistic                          0.7044
Mean absolute error                      0.1594
Root mean squared error                  0.3555
Relative absolute error                 31.8836 %
Root relative squared error             71.104  %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.909    0.204    0.816      0.909    0.860      0.709    0.917     0.898     Kecimen
                 0.796    0.091    0.897      0.796    0.843      0.709    0.917     0.923     Besni
Weighted Avg.    0.852    0.148    0.857      0.852    0.852      0.709    0.917     0.911

=== Confusion Matrix ===

   a   b   <-- classified as
 409  41 |   a = Kecimen
  92 358 |   b = Besni
```

| Results | | |
|---|---|---|
| Confusion Matrix | Predicted Kecimen | Predicted Besni |
| True Kecimen | 409 (TP) | 41 (FP) |
| True Besni | 92 (FN) | 358 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.852 | 85.22% | 0.857 | 0.852 |

The initial Naïve Bayes model achieved 83.56% accuracy. Updated hyper-parameters, including using a kernel estimator, improved accuracy to 85.22% and increased precision, recall, and F1-score, while reducing error metrics.

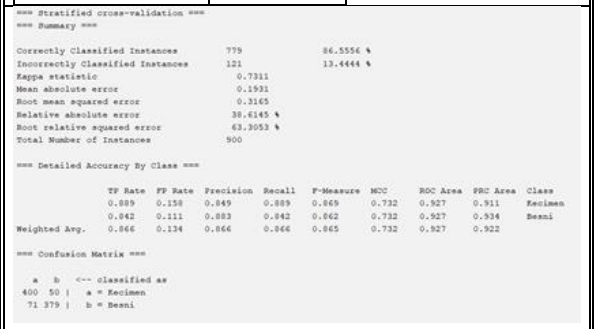
## 3.3 Multilayer Perceptron (MLP)

**Preprocessing**: Discretized of 5 bins to all attributes using the Discretize filter.

Hyper-Parameters

Initial

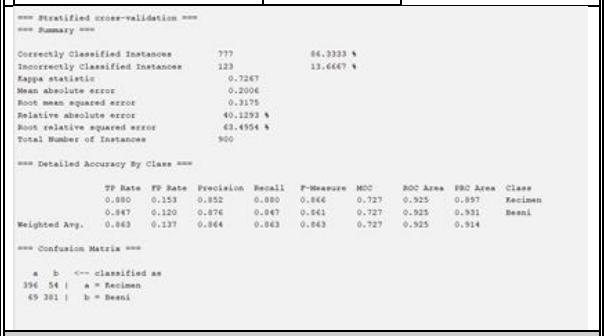
| Hyper-Parameters | Settings |
|---|---|
| hiddenLayer | a |
| trainingTime | 500 |
| learningRate | 0.3 |
| momentum | 0.2 |



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         779               86.5556 %
Incorrectly Classified Instances       121               13.4444 %
Kappa statistic                          0.7311
Mean absolute error                      0.1531
Root mean squared error                  0.3165
Relative absolute error                 38.6145 %
Root relative squared error             63.3053 %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.889     0.158     0.849     0.889     0.869     0.732     0.927     0.911     Kecimen
                 0.842     0.111     0.883     0.842     0.862     0.732     0.927     0.934     Besni
Weighted Avg.    0.866     0.134     0.866     0.866     0.865     0.732     0.927     0.922

=== Confusion Matrix ===

   a   b   <-- classified as
 400  50 |   a = Kecimen
  71 379 |   b = Besni
```

Updated

| Hyper-Parameters | Settings |
|---|---|
| hiddenLayer | 2,a,2 |
| trainingTime | 1000 |
| learningRate | 0.1 |
| momentum | 0.3 |



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         777               86.3333 %
Incorrectly Classified Instances       123               13.6667 %
Kappa statistic                          0.7267
Mean absolute error                      0.2006
Root mean squared error                  0.3175
Relative absolute error                 40.1293 %
Root relative squared error             63.4954 %
Total Number of Instances              900

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.880     0.153     0.852     0.880     0.866     0.727     0.925     0.897     Kecimen
                 0.847     0.120     0.876     0.847     0.861     0.727     0.925     0.931     Besni
Weighted Avg.    0.863     0.137     0.864     0.863     0.863     0.727     0.925     0.914

=== Confusion Matrix ===

   a   b   <-- classified as
 396  54 |   a = Kecimen
  69 301 |   b = Besni
```

Results

| Confusion Matrix | Predicted Kecimen | Predicted Besni |
|---|---|---|
| True Kecimen | 400 (TP) | 50 (FP) |
| True Besni | 71 (FN) | 379 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.865 | 86.56% | 0.866 | 0.866 |

Results

| Confusion Matrix | Predicted Kecimen | Predicted Besni |
|---|---|---|
| True Kecimen | 396 (TP) | 54 (FP) |
| True Besni | 69 (FN) | 301 (TN) |

| Metric | F measure | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Value | 0.863 | 86.33% | 0.864 | 0.863 |

The initial MLP model achieved 86.56% accuracy. Updated hyper-parameters accuracy to 86.33% no change, recall, and F1-score, while reducing error metrics.

# 4. Conclusion

In conclusion, tuning hyper-parameters improved the performance of Decision Tree, Naïve Bayes, but didn't improve MLP. The optimized settings led to higher accuracy and precision, particularly for Naïve Bayes. These results highlight the importance of hyper-parameter adjustments in enhancing machine learning models. Future work should focus on further fine-tuning and validation to ensure robust performance.