

Department: ENCS - Computer Systems Engineering - هندسة أنظمة الحاسوب

ENCS3340 – Artificial Intelligence

**Search Algorithms for Route Navigation**

1191590 – Tala Dweikat

1190102 – Mazen Batrawi

Instructor: Dr. Yazan Abu Farha

Section: 4

Date: 06 – 06 – 2022

# Abstract

We will learn how to use machine learning tools to test several algorithms for categorization tasks in this project. Our data set was Raisin data set. The data consists of several attributes;

Area, MajorAxisLength, Perimeter, MinorAxisLength, Eccentricity, ConvexArea, Extent and Class.

**Table of Contents**

[**Abstract** I](#_Toc105888975)

[**Algorithms** 1](#_Toc105888976)

[Decision Tree 1](#_Toc105888977)

[Naive Bayes 8](#_Toc105888978)

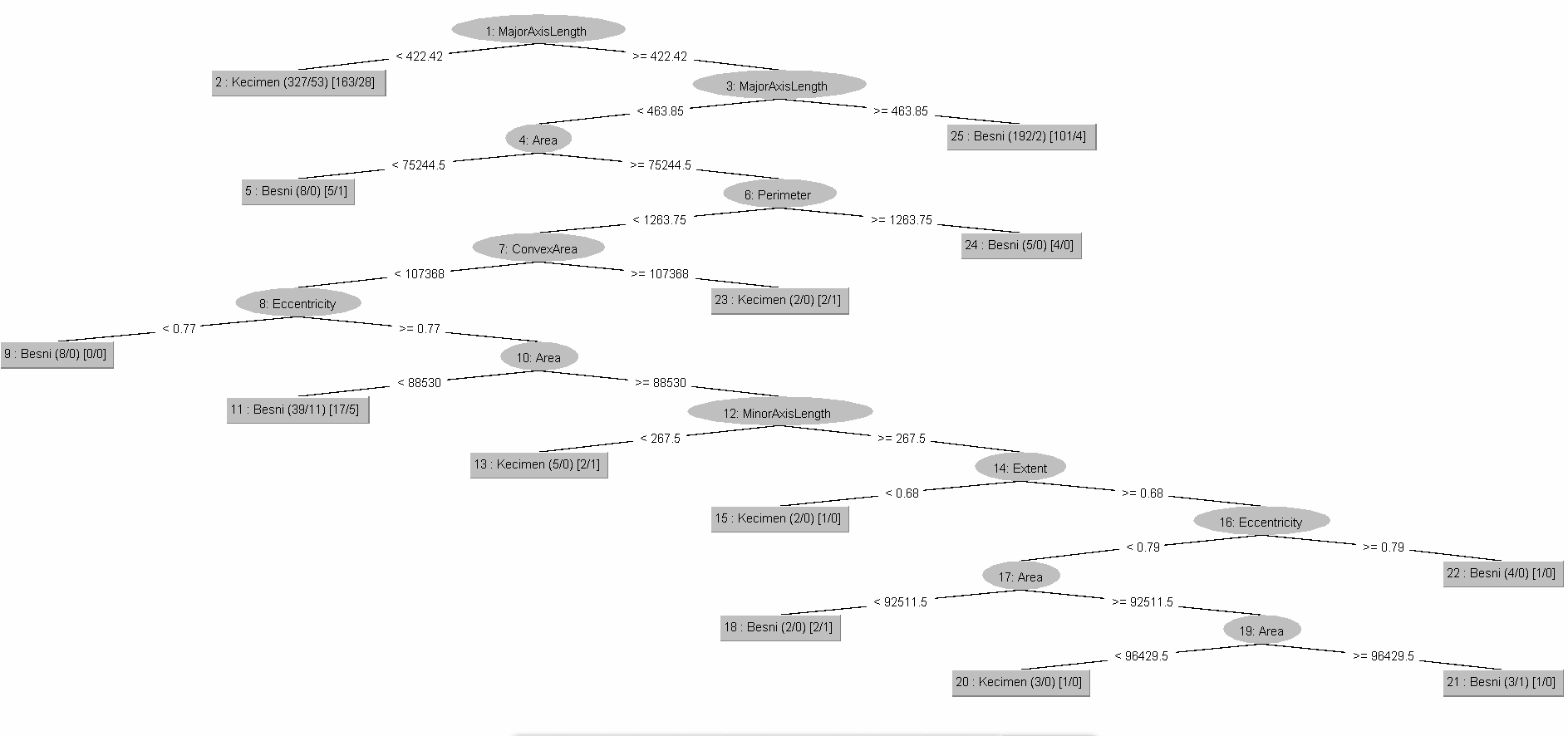
[Random Tree 12](#_Toc105888979)

[**Conclusion** 16](#_Toc105888980)

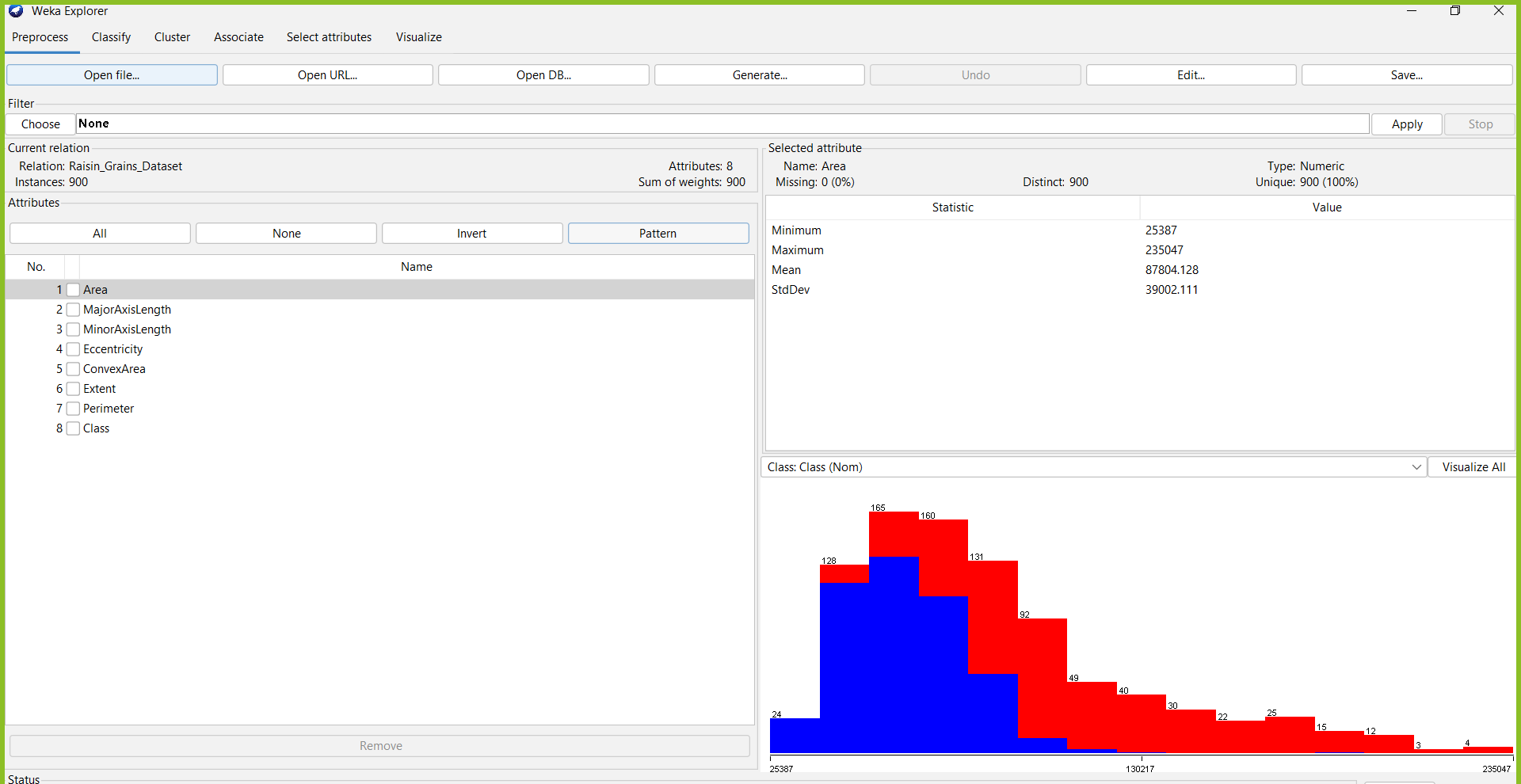
# 

# Algorithms

## **Decision Tree**



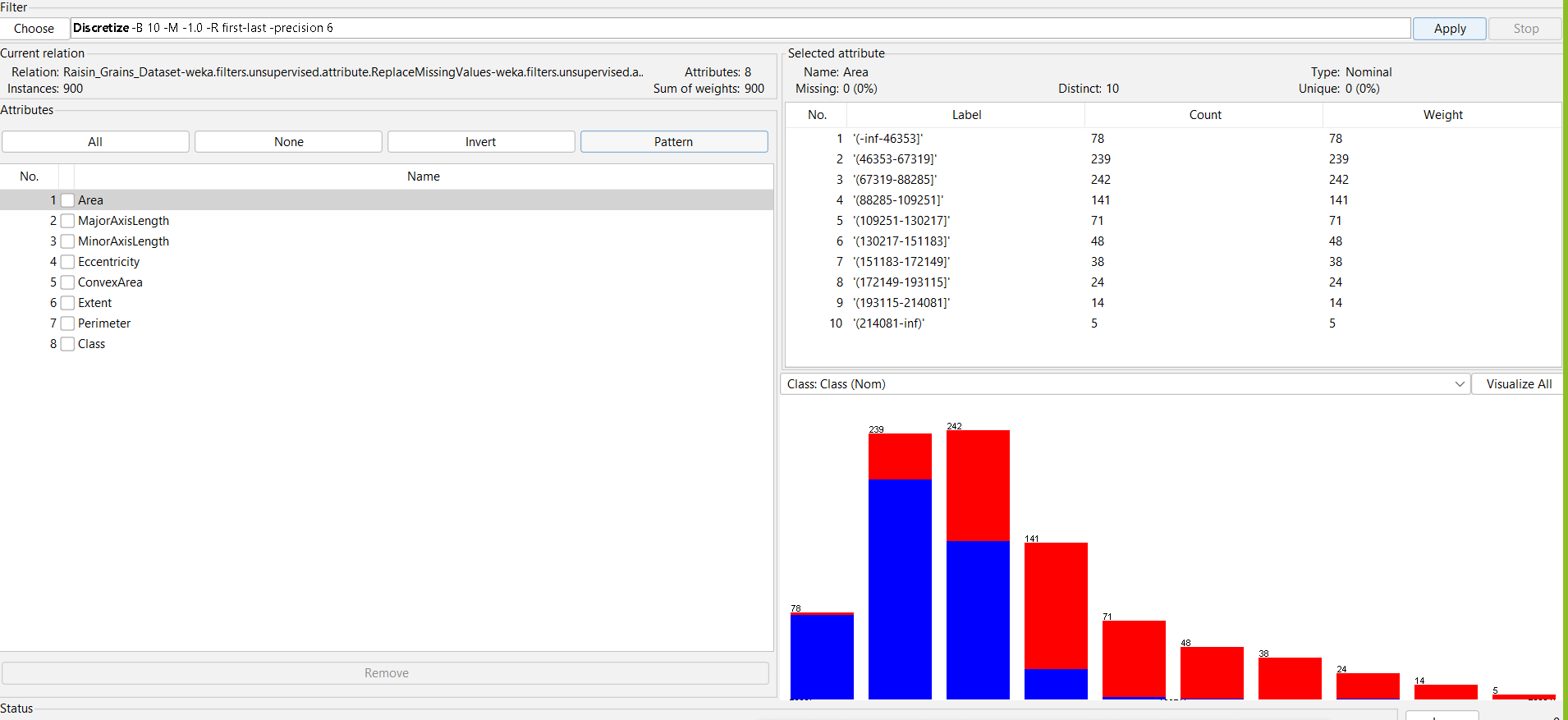
- This is the data we read from the file before preprocessing



- Then we applied the filter Discretize on the data:



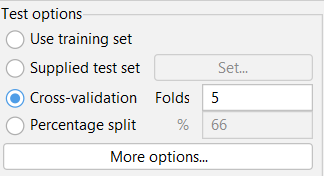
- The updated data after applying the filter is as follows:



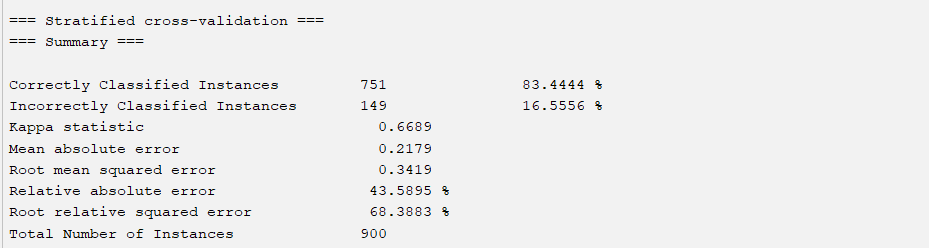
- The decision tree was chosen from the classifier shown below.



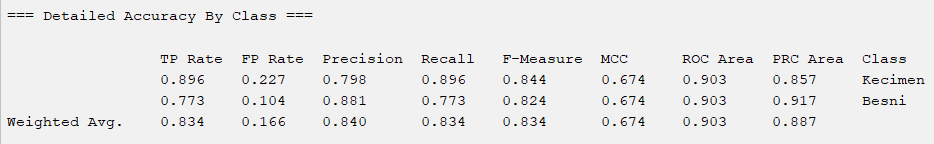
- The cross-validation folds was set to 5 from the test operations as required.



- We obtained an accuracy of 83.4% as shown in the figure below:



- We obtained the average weights of precision, recall and F-Measure as shown in the figure below:



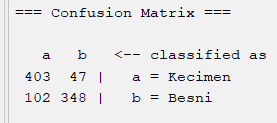
- We obtained the following values in the confusion matrix:

TP = 403

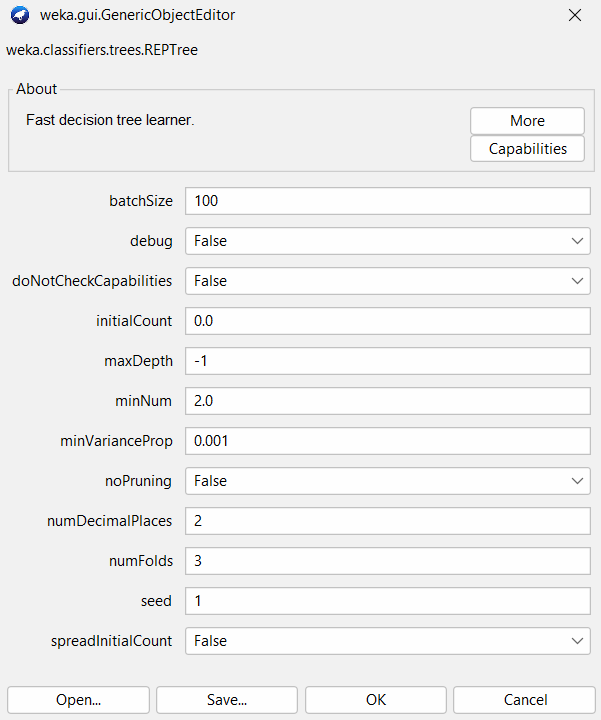
TN = 348

FP = 47

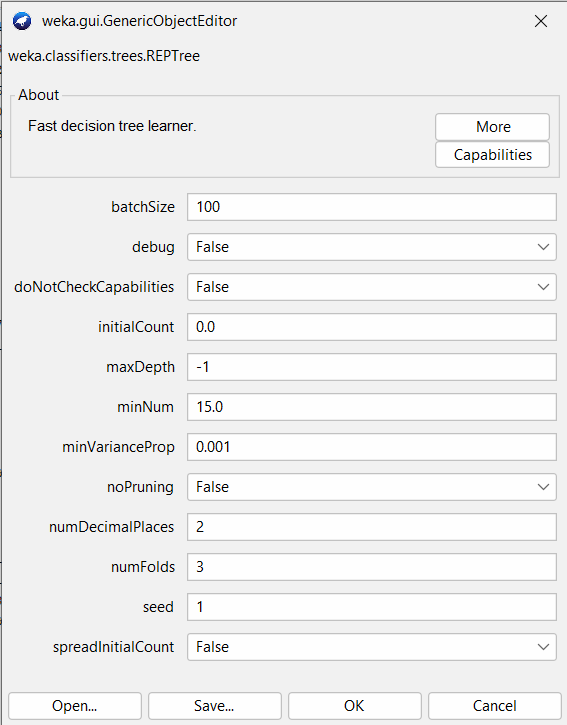
FN = 102



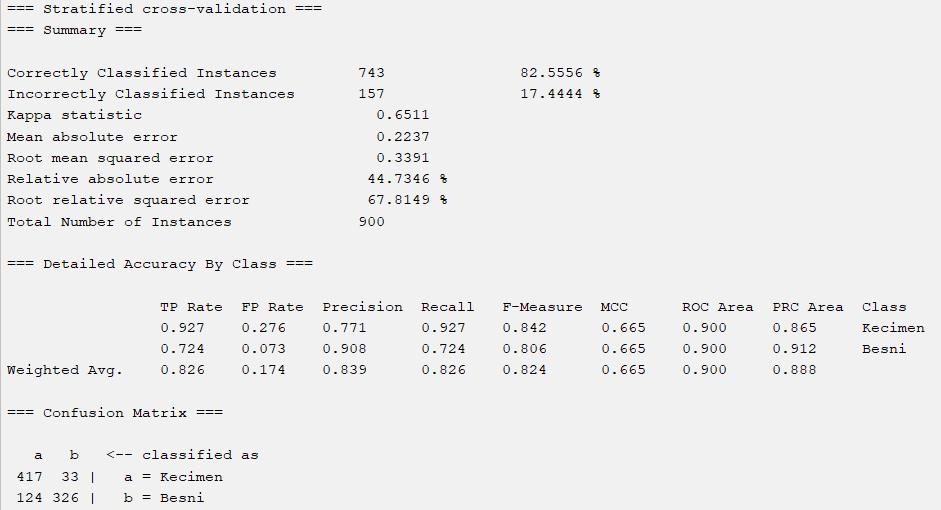
- Here are the initial hyper parameters:



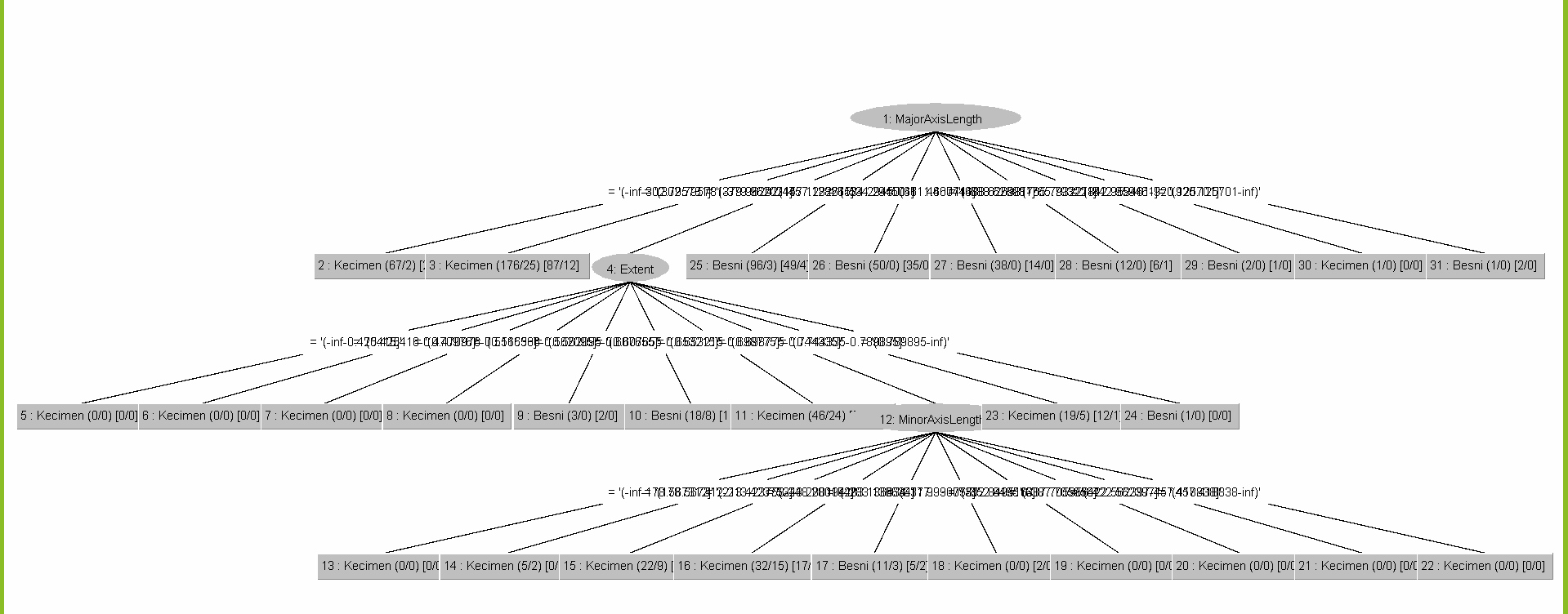
- Here are the hyper parameters after updating the minNum from 2 to 15:



- We can notice from the figure below that the accuracy, precision, recall, F-Measure and the confusion matrix changed:

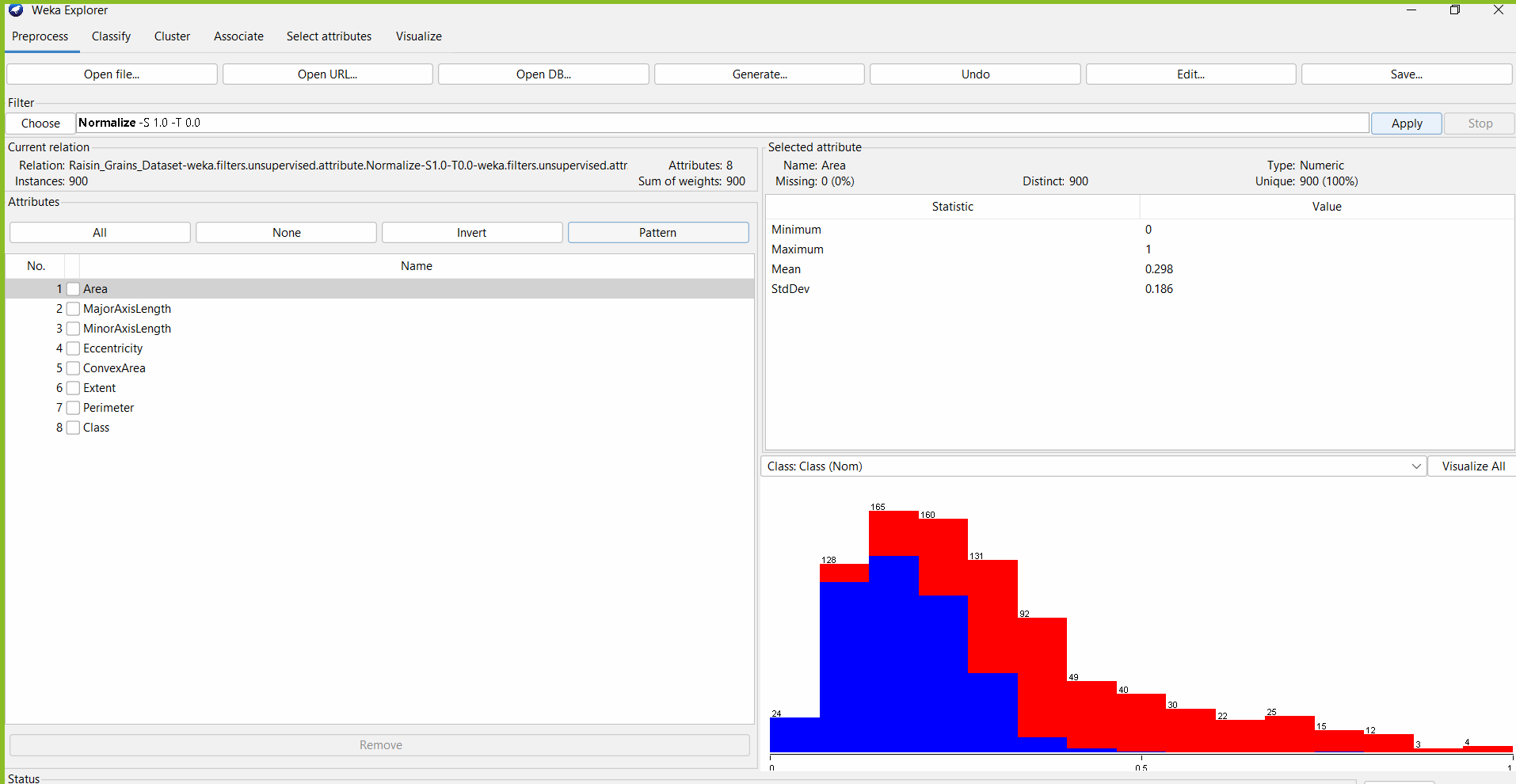


- Also, we can see that the decision tree has changed:

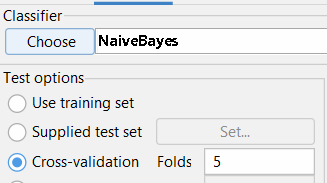


## **Naive Bayes**

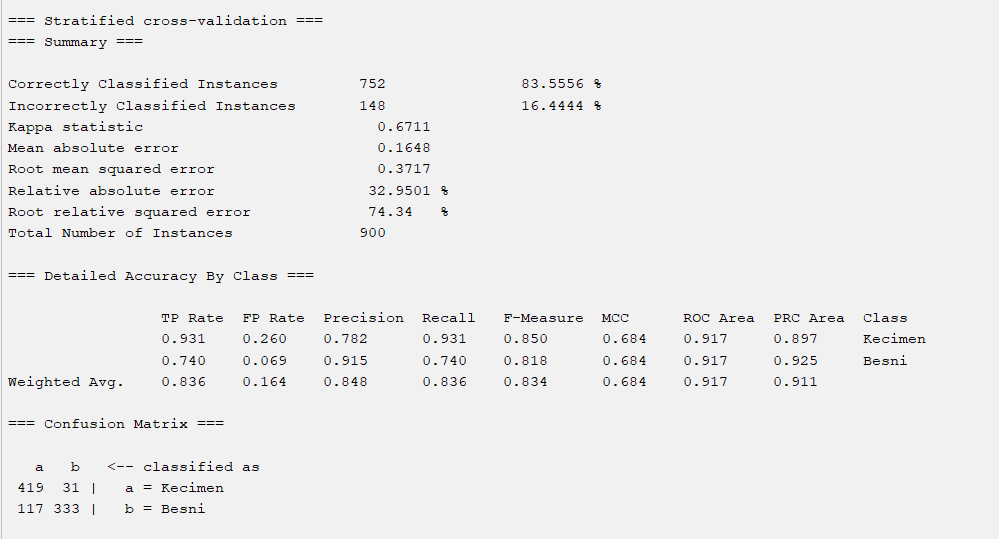
- We applied normalization for the data and got the following results:



- We chose the Naïve Bayes classifier and 5 cross-validation folds.



- After running the algorithm we got the following results:



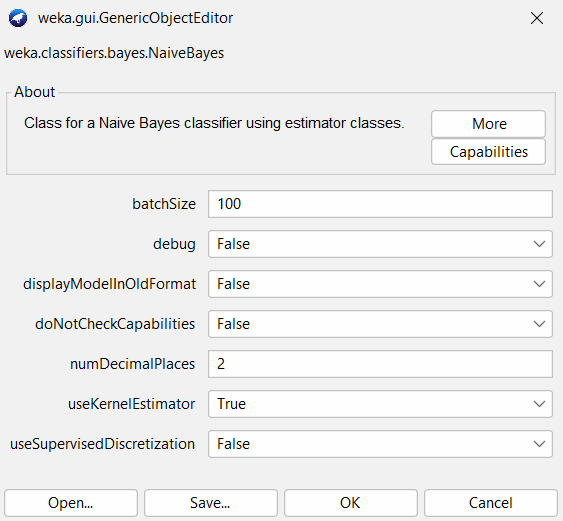
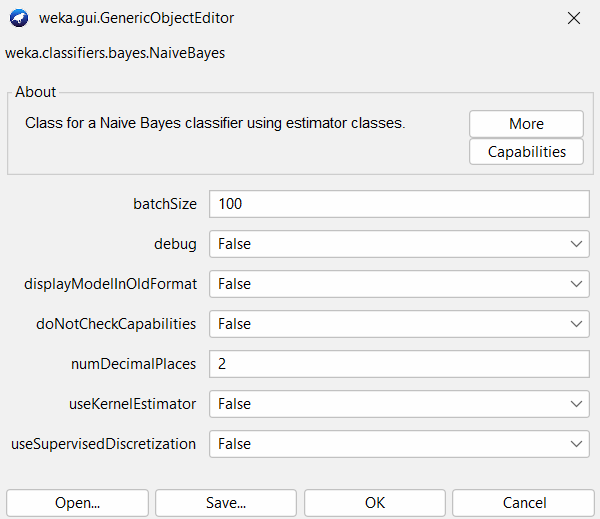
The accuracy is: 83.5556% (which is slightly higher than the decision tree).

The precision, recall and F-Measure are 0.848, 0.836 and 0.834 respectively.

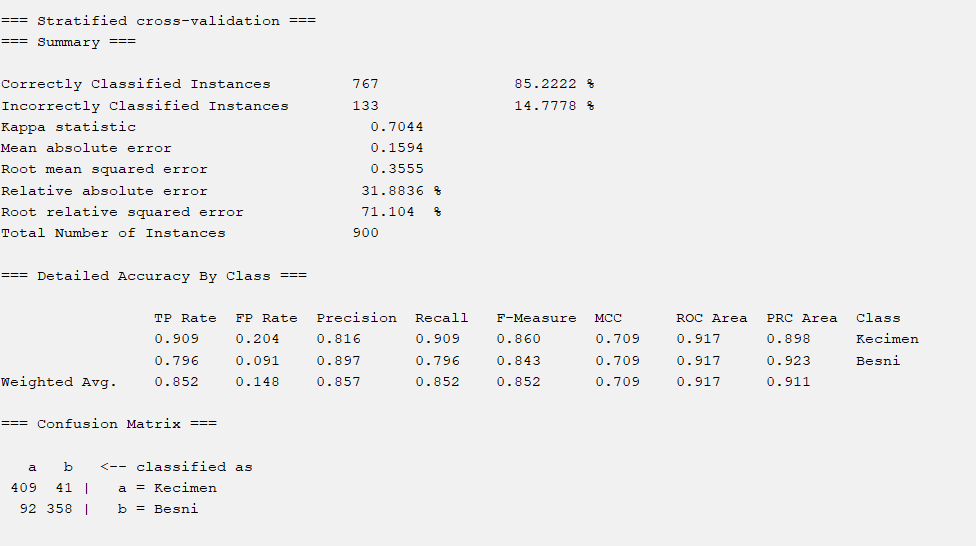
The confusion matrix has the following values:

TP = 419, FP = 31, FN = 117, TN = 333

- We changed the kernel estimator so that the algorithm uses it:



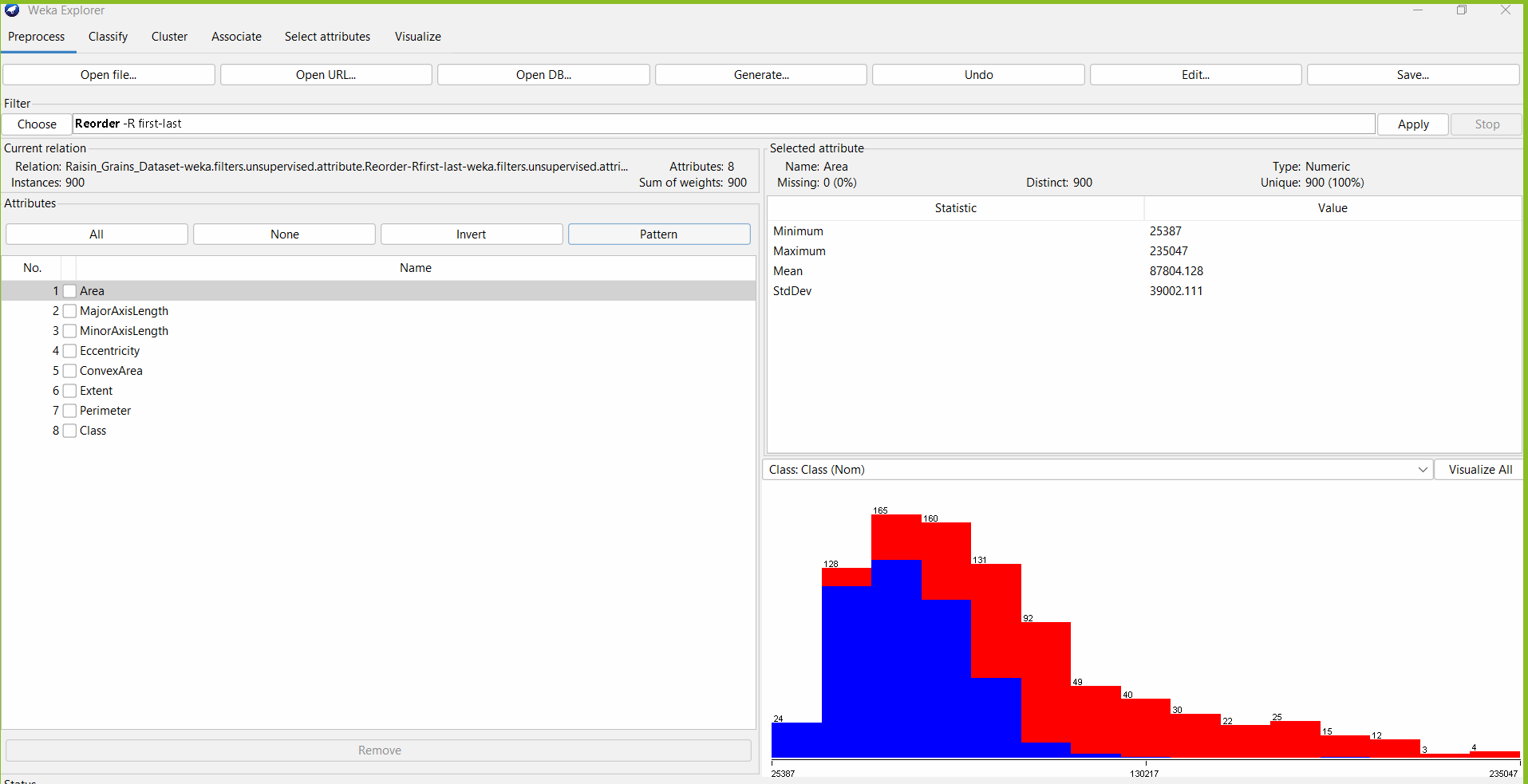
- We got the following results:



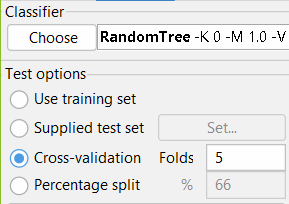
As we can see, the accuracy increased and the other values changed.

## **Random Tree**

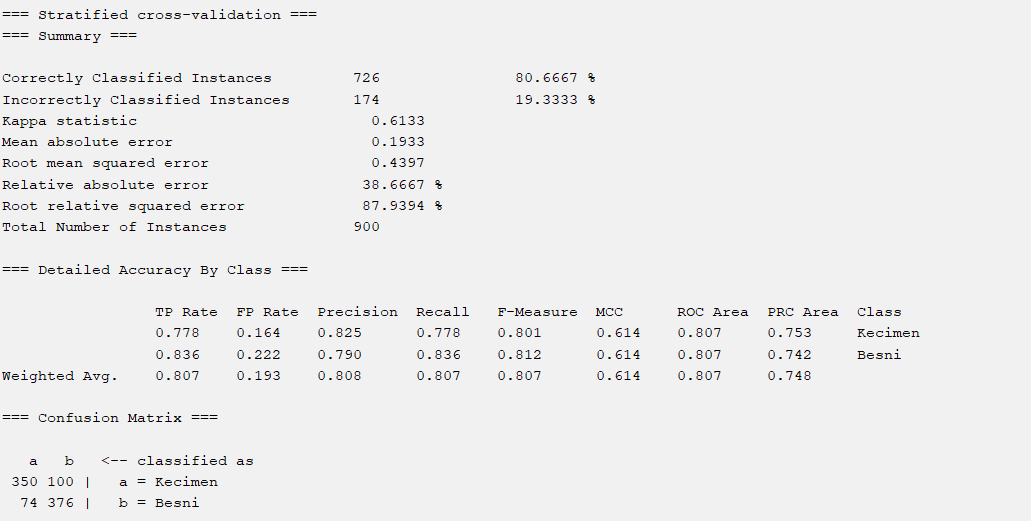
- We preprocessed the data with the reorder filter:



- We choose the random tree algorithm and set the cross-validation folds to 5:

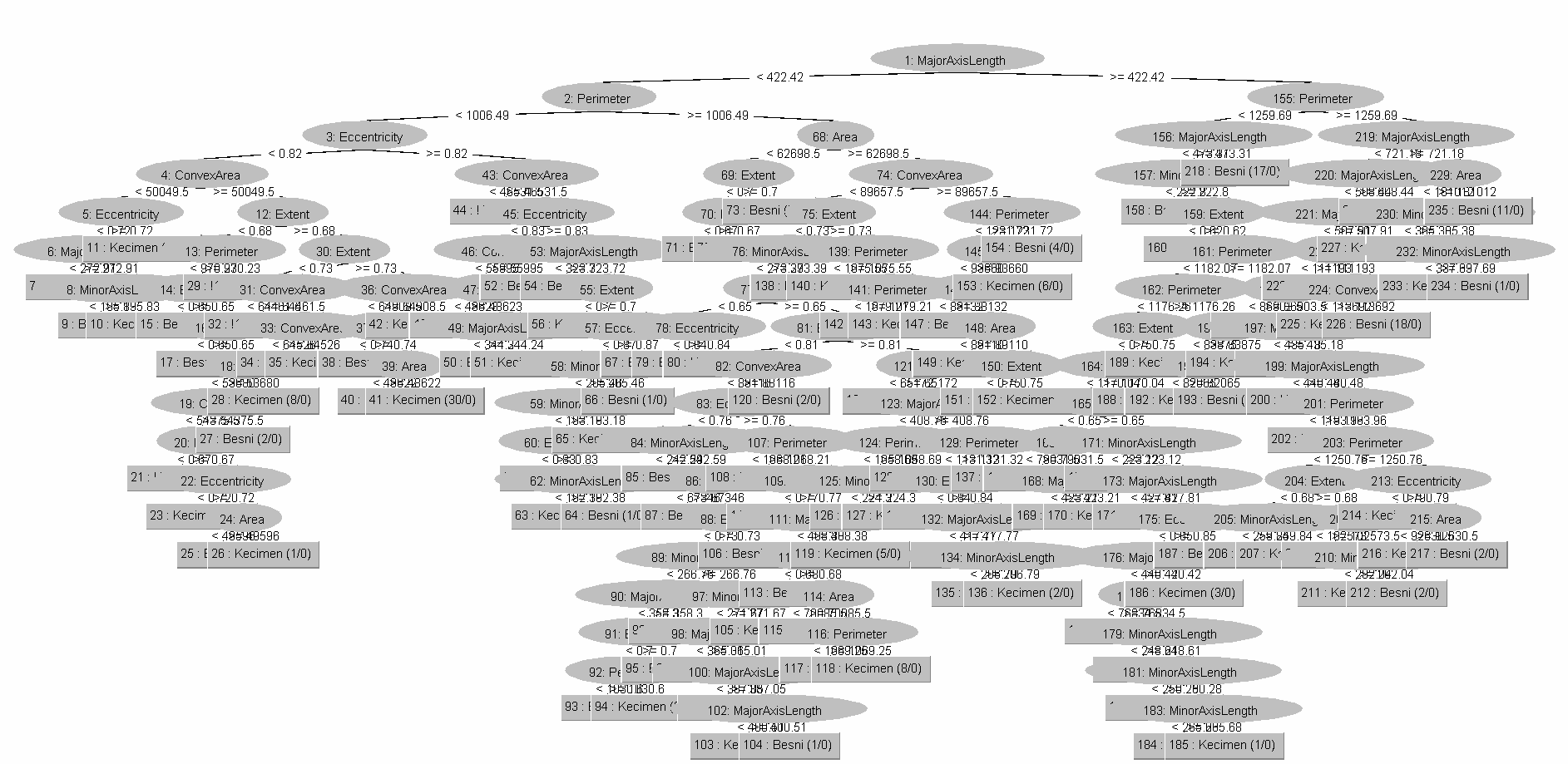


- We got the following results:

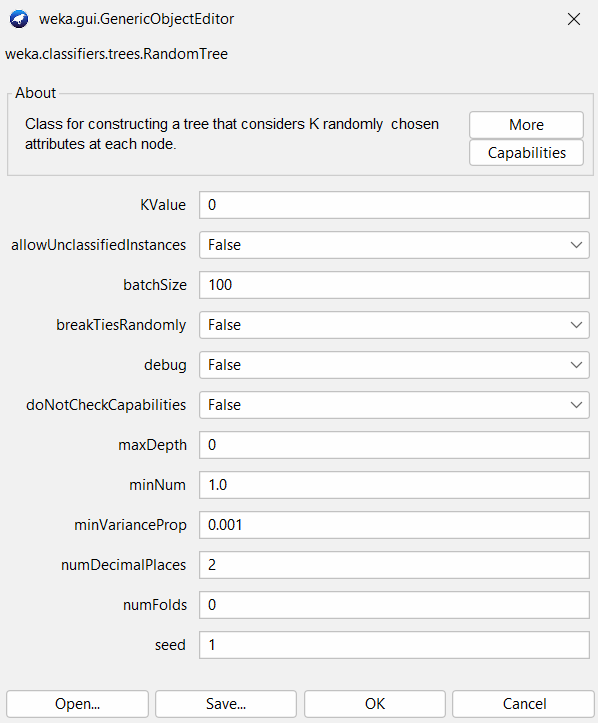
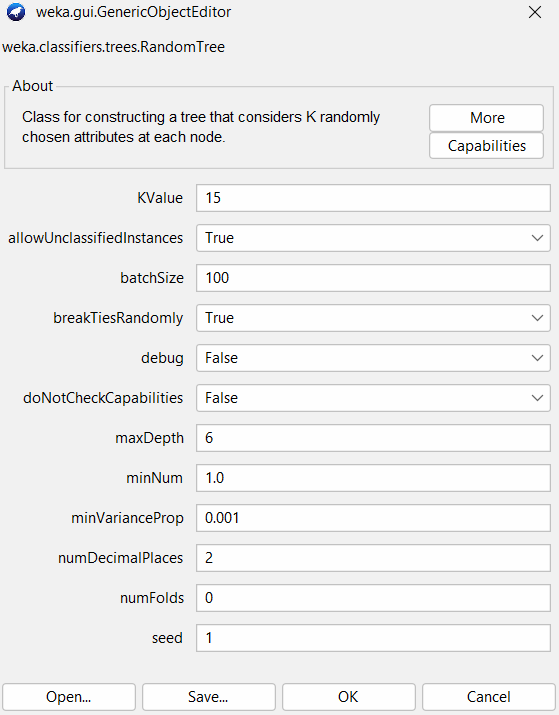


We can see that this algorithm has the least accuracy among the 3 applied algorithm to the data set with 80.6667% accuracy. The average precision, recall and F-Measure were 0.808, 0.807, and 0.807 respectively. The confusion matrix has the following values: TP = 350, FP = 100, FN = 74, TN = 376.

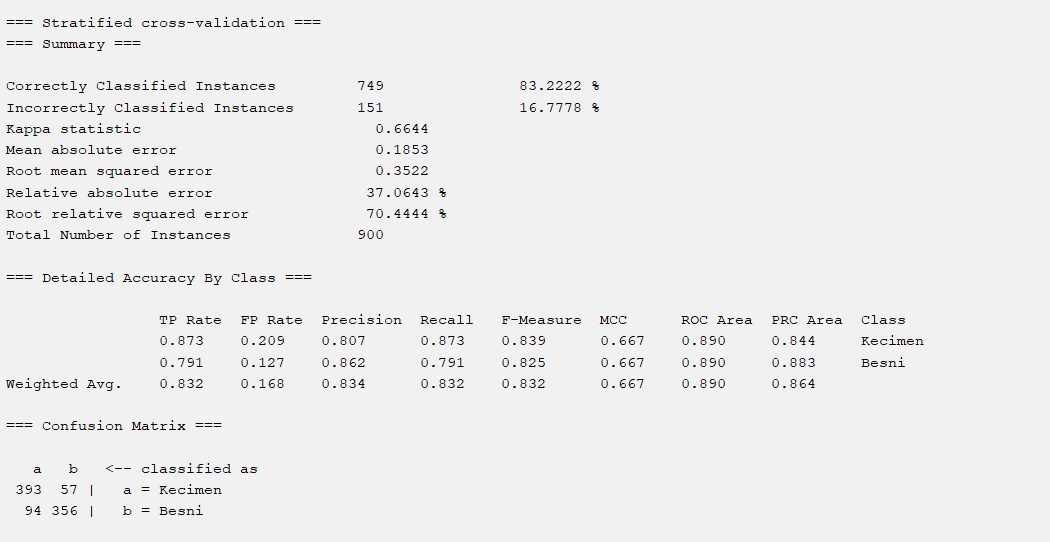
- This is the visualization of the data using random tree:



- We updated some hyper parameters as the following:

- We got the following reuslts:



We can see that the accuracy increased from 80.6667% to 83.2222%. The average precision, recall and F-Measure were 0.834, 0.832, and 0.832 respectively. The confusion matrix has the following values: TP = 393, FP = 56, FN = 94, TN = 356.

# Conclusion

Machine learning is very important to ease the job of humans by teaching the machines how to test the data and make correct decisions. In this project, we used WEKA program to test the raisin data set, and we used 3 different algorithms which are: Decision tree, Naïve Bayes and Random tree. Each algorithm displayed different success rate (accuracy), all of them had a range of accuracy between 80% and 85%. Each algorithm performs using a certain procedure and displays its results. Applying filters, changing the fold cross-validation and changing the hyper parameters will affect the results we obtain, and that was shown for each algorithm.