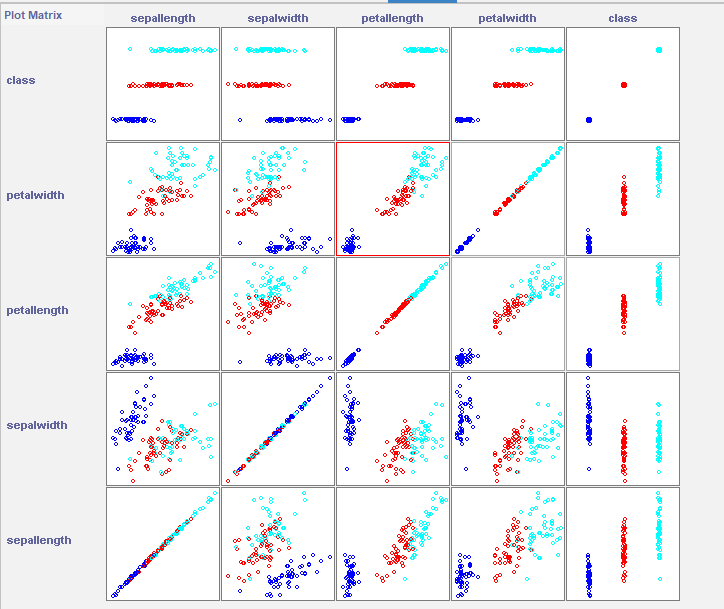
Experiment 1

**Aim:** Perform data Pre-Processing task using Weka data mining tool.

**Theory:** WEKA - an open source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems.

Everything except association rule mining is done on the Iris dataset.

## 1.Visualization



 **Sepal Length and Sepal Width**: Weak negative correlation.

 **Sepal Length and Petal Length**: Strong positive correlation.

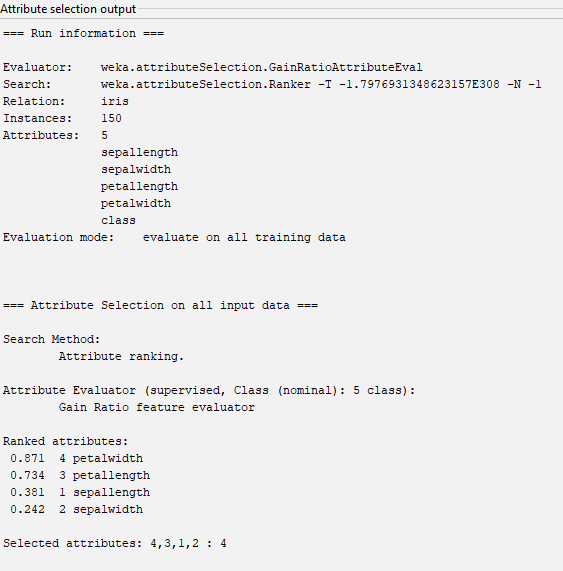
 **Sepal Length and Petal Width**: Strong positive correlation.

 **Sepal Width and Petal Length**: Weak negative correlation.

 **Sepal Width and Petal Width**: Weak negative correlation.

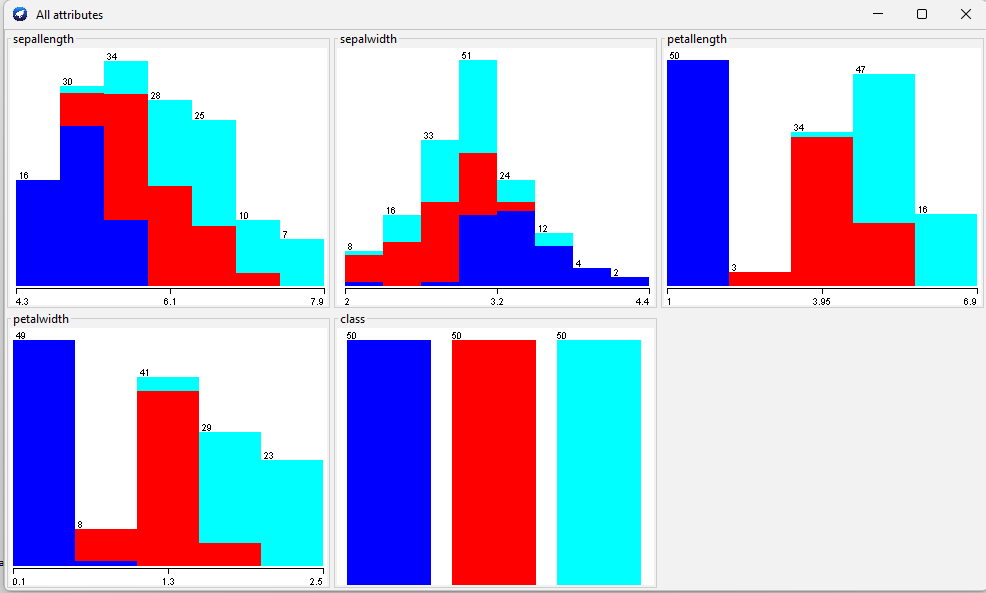
 **Petal Length and Petal Width**: Strong positive correlation.

## 2. Select Attributes

****

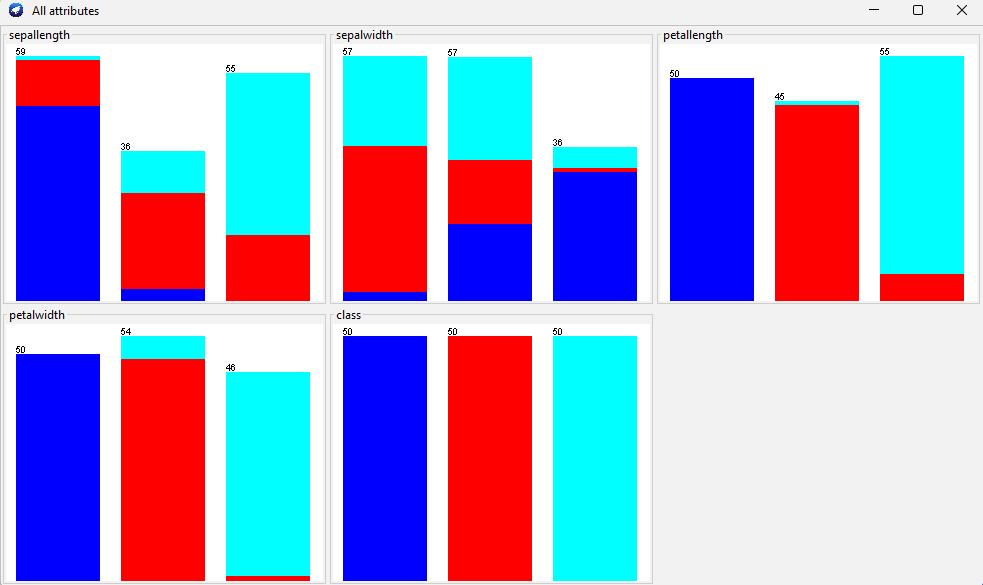
## 3. Pre-processing

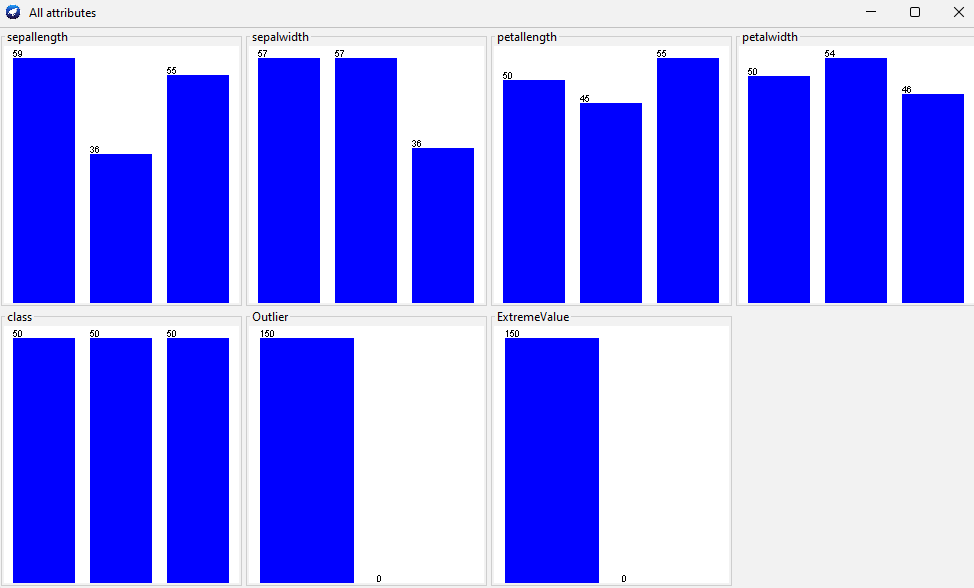
### Visualize All



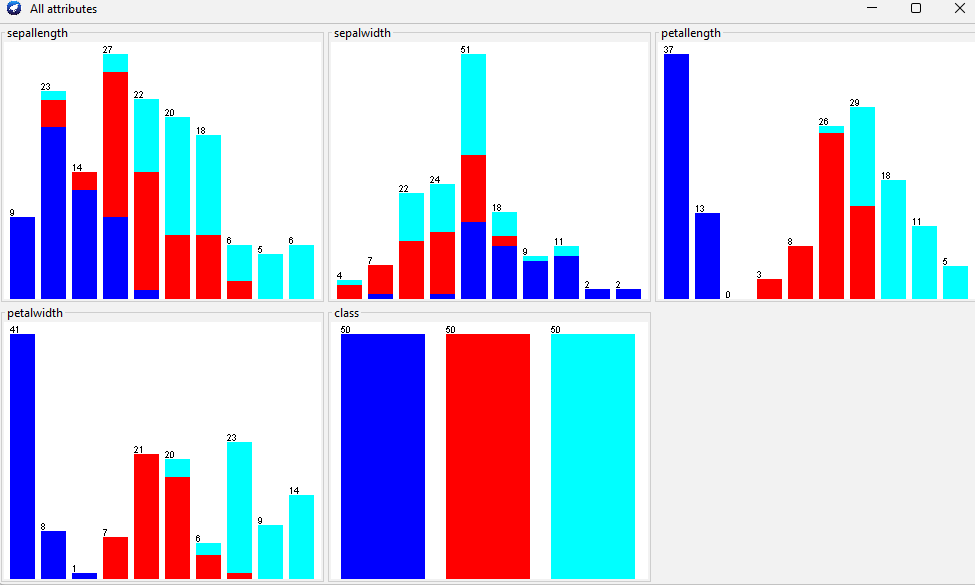
### Filter (Discretization)

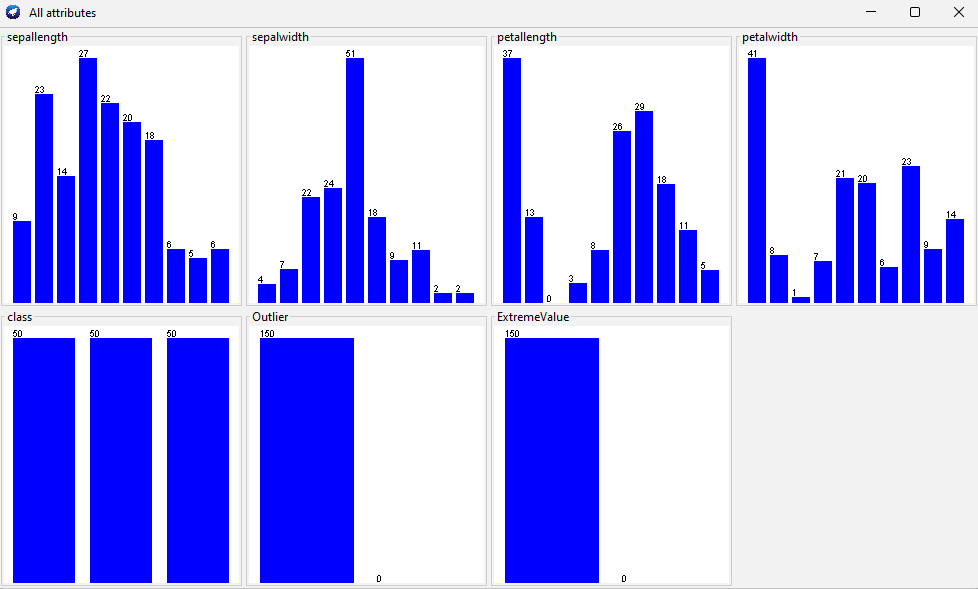
#### Supervised



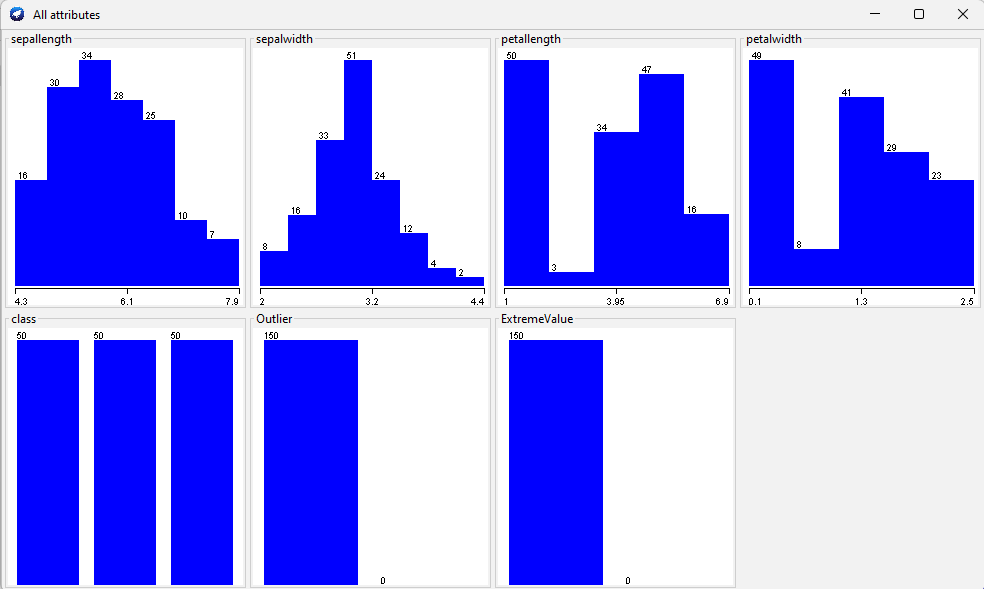


#### Unsupervised



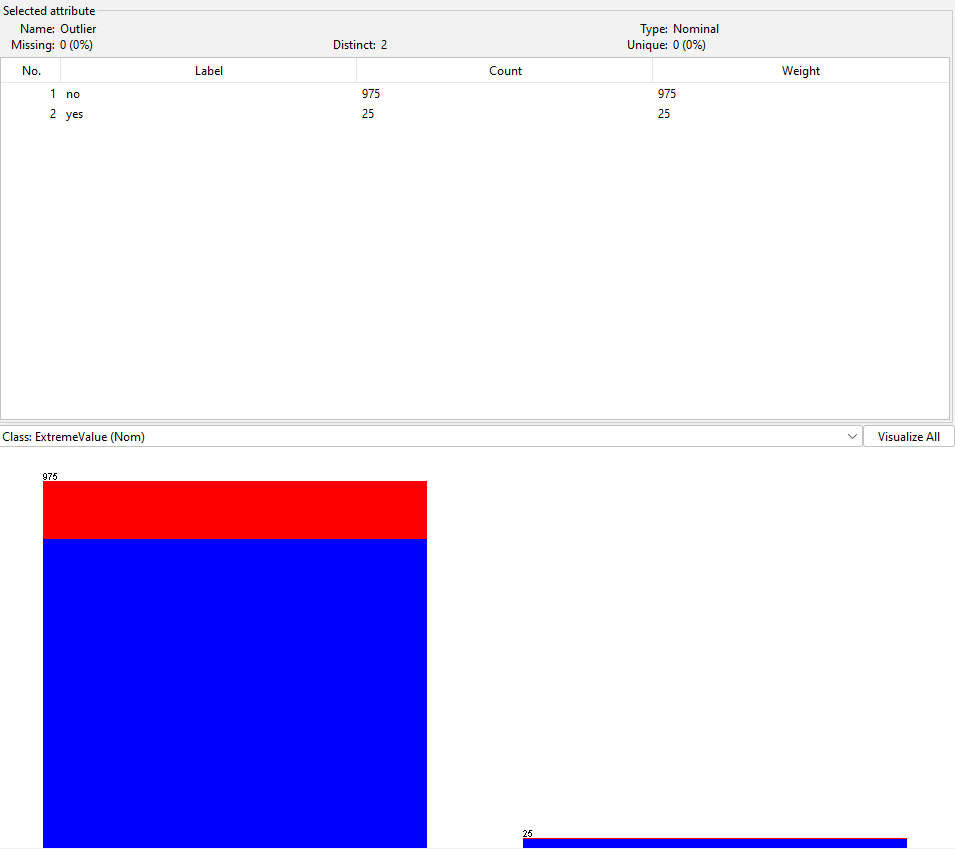


### IQR



### Remove values

Before Removing outliers

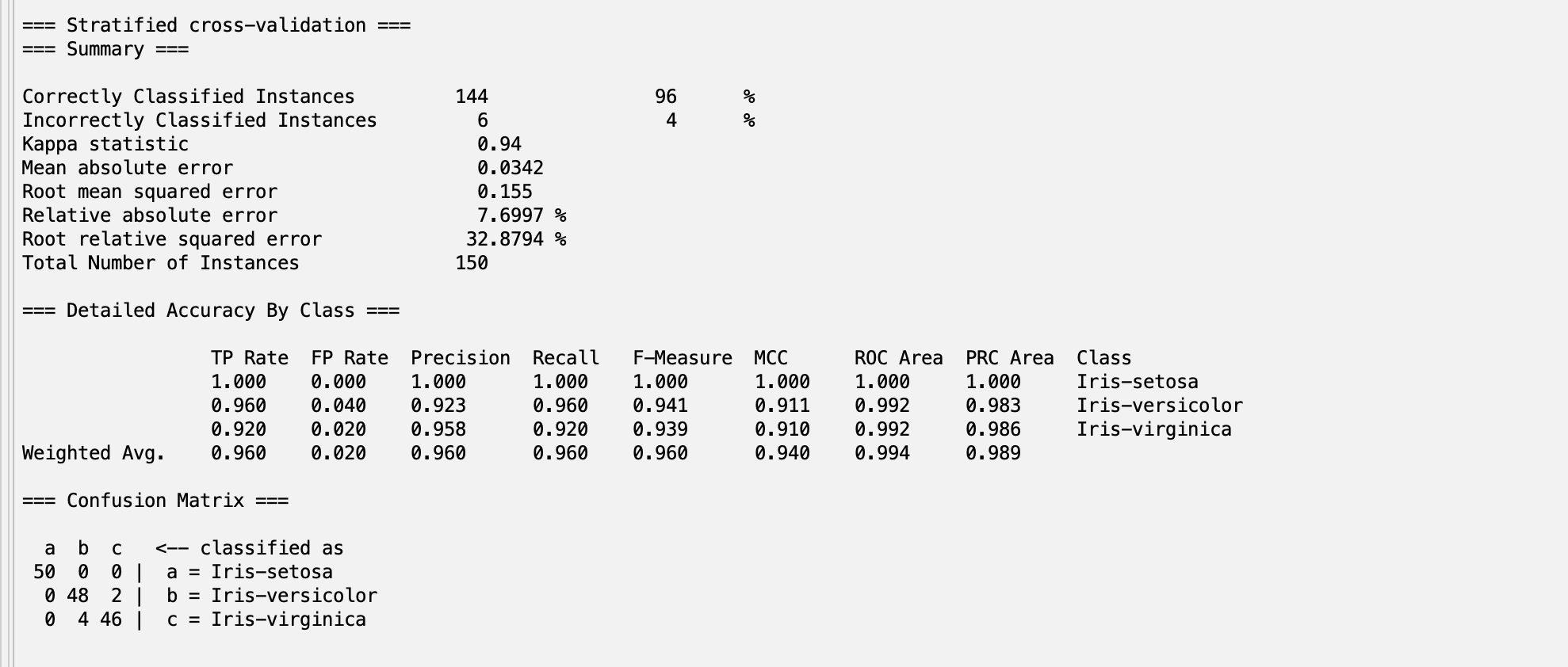
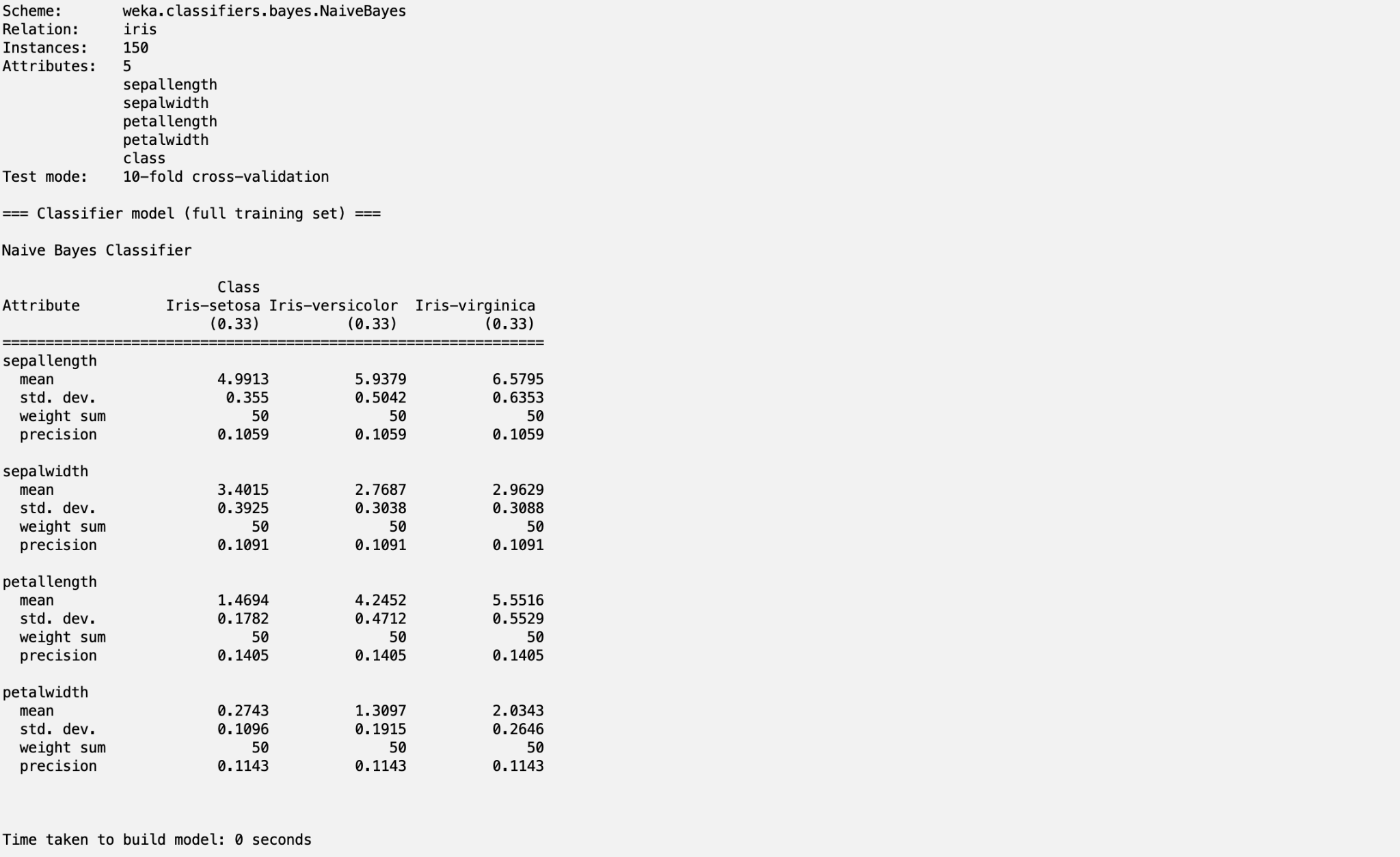


After removing outliers

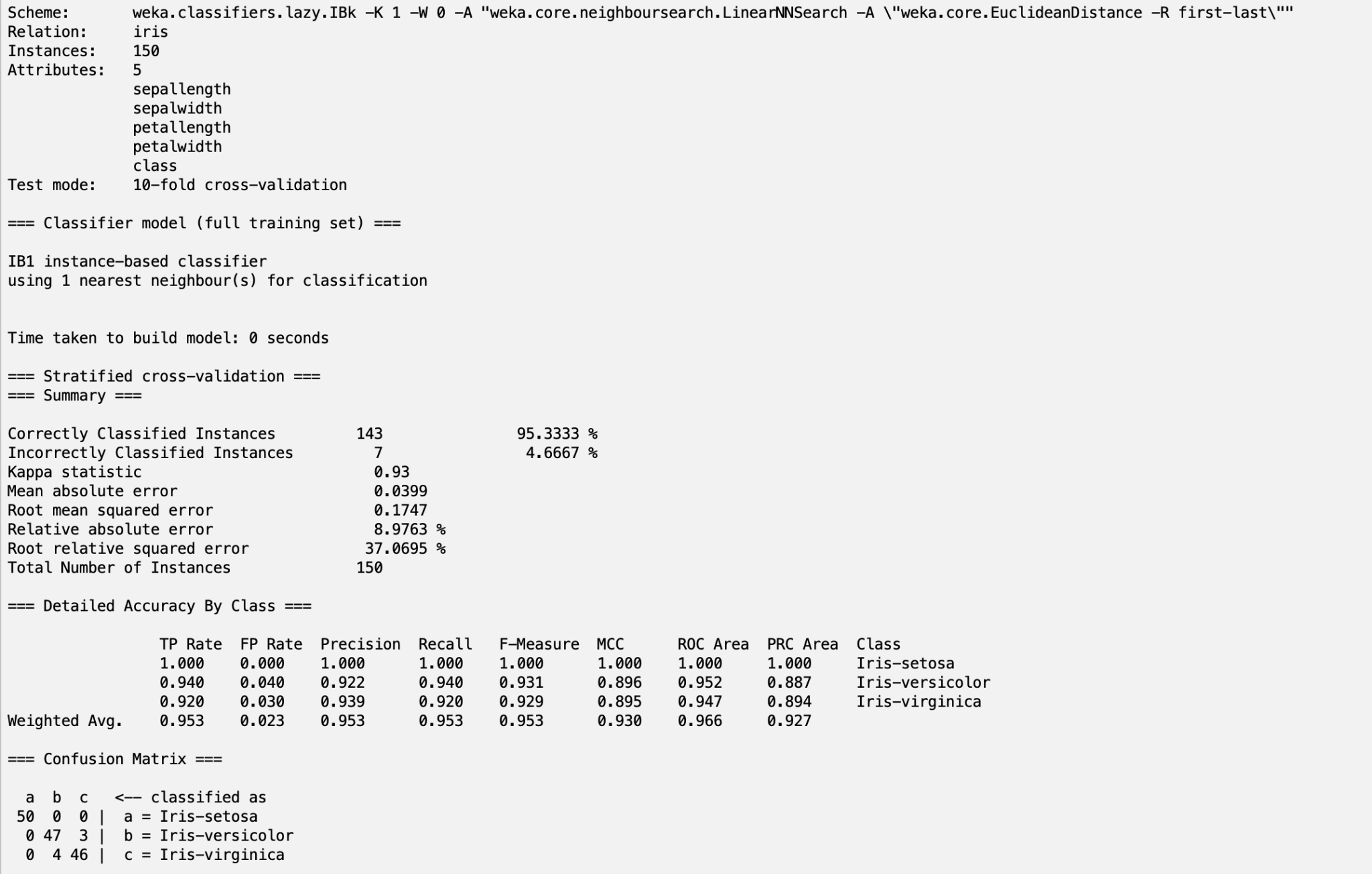


## 4. Classification

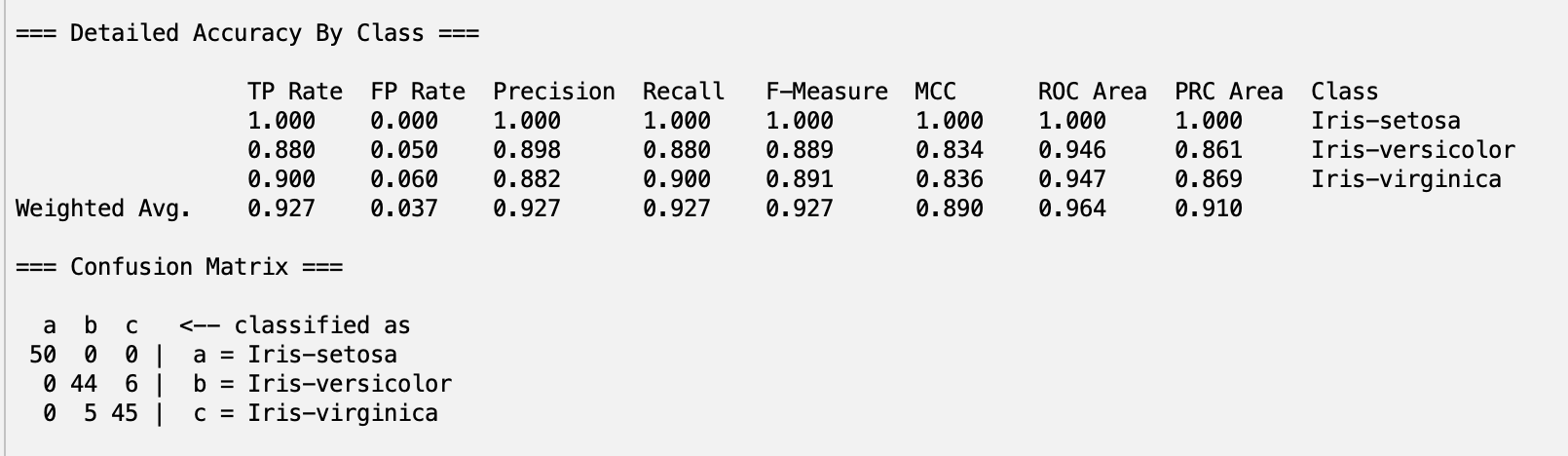
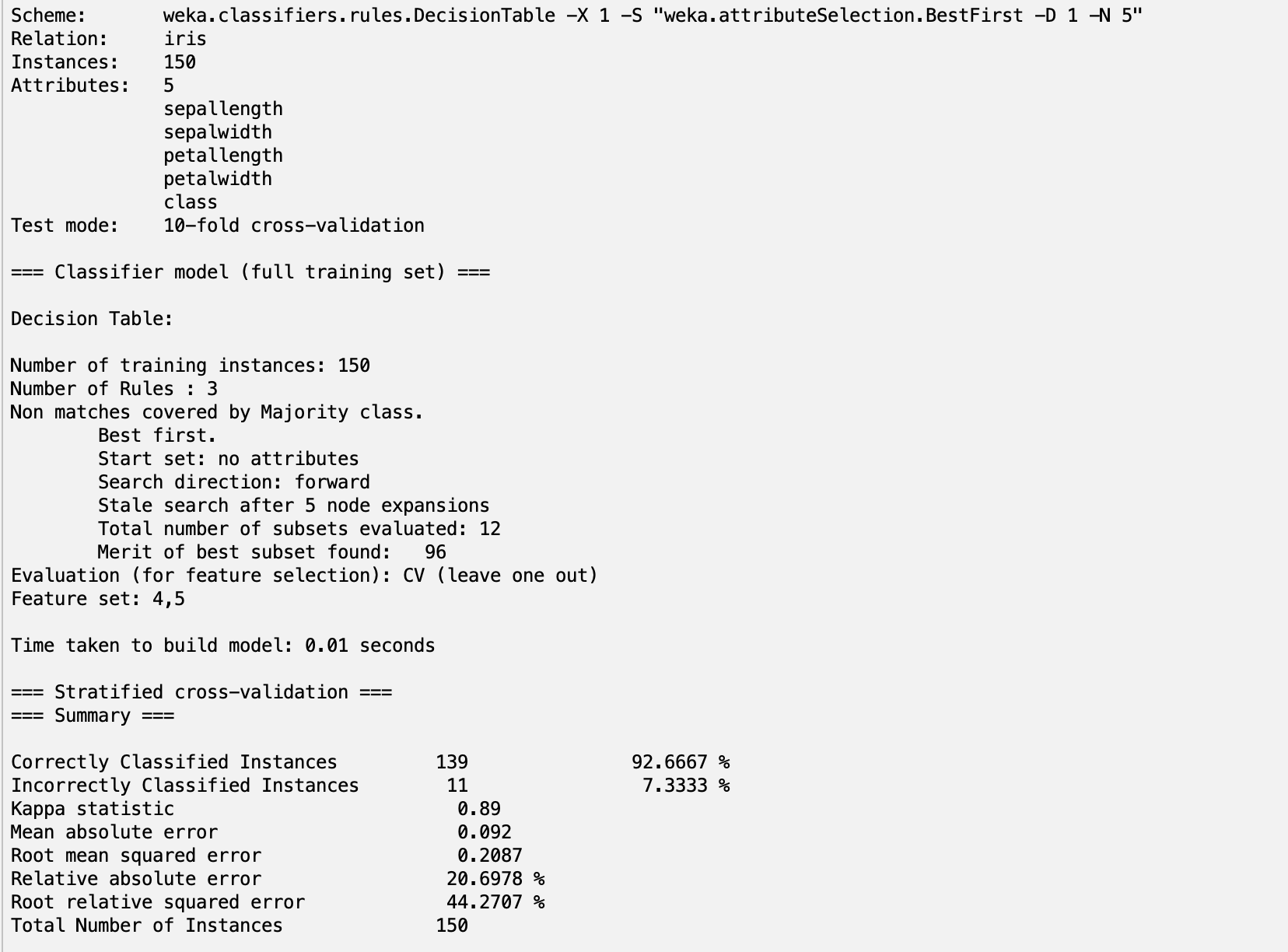
### Naive Bayes:



### kNN (k is 1)

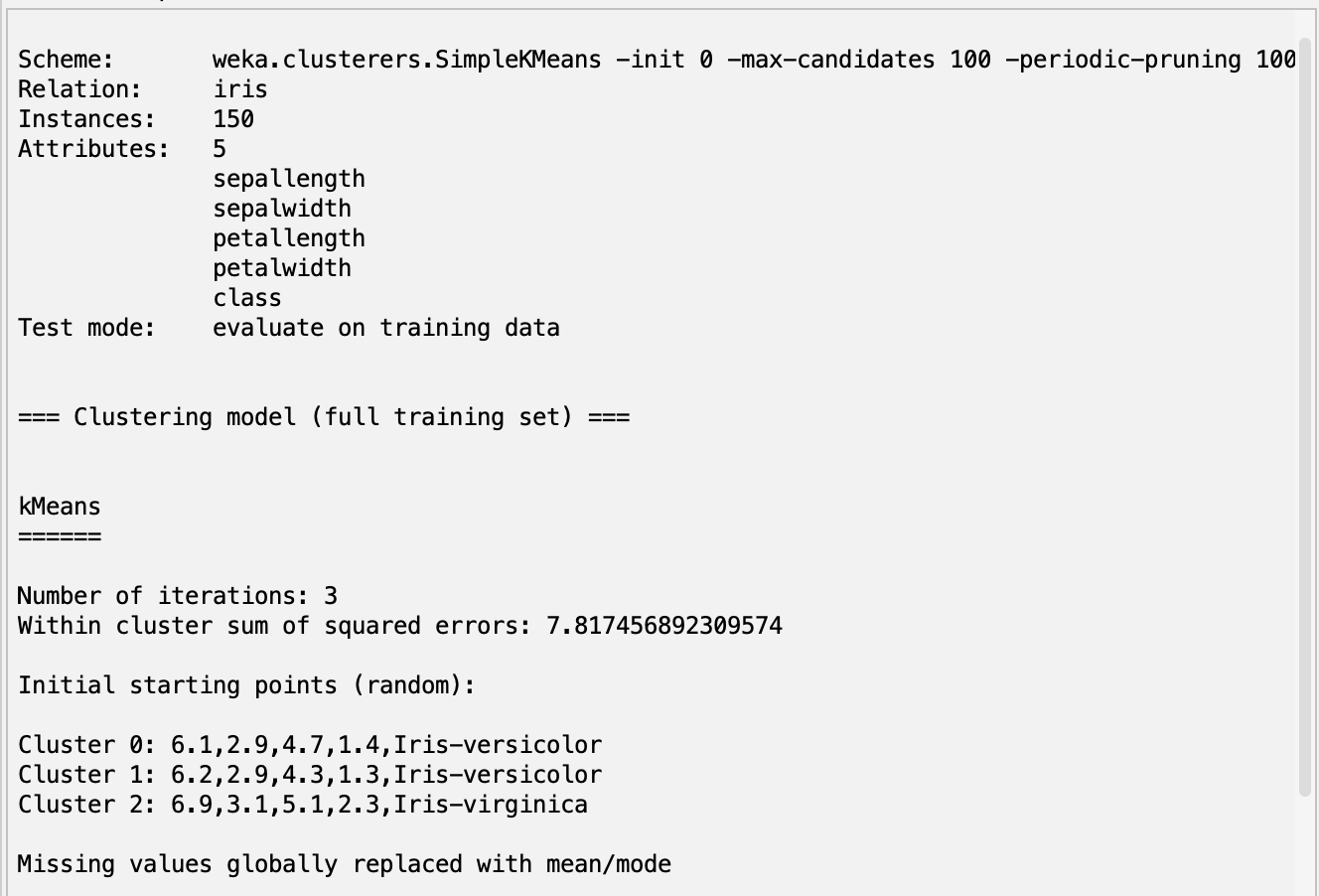


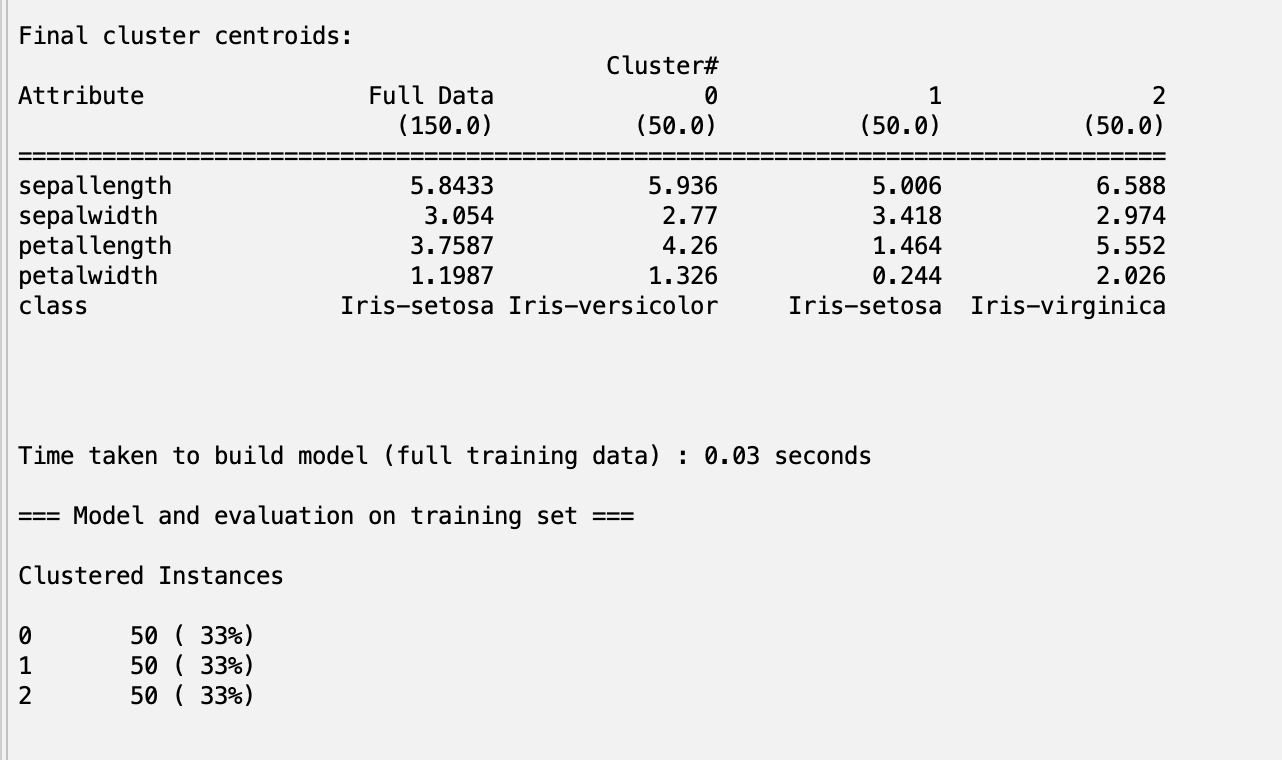
### Decision Table



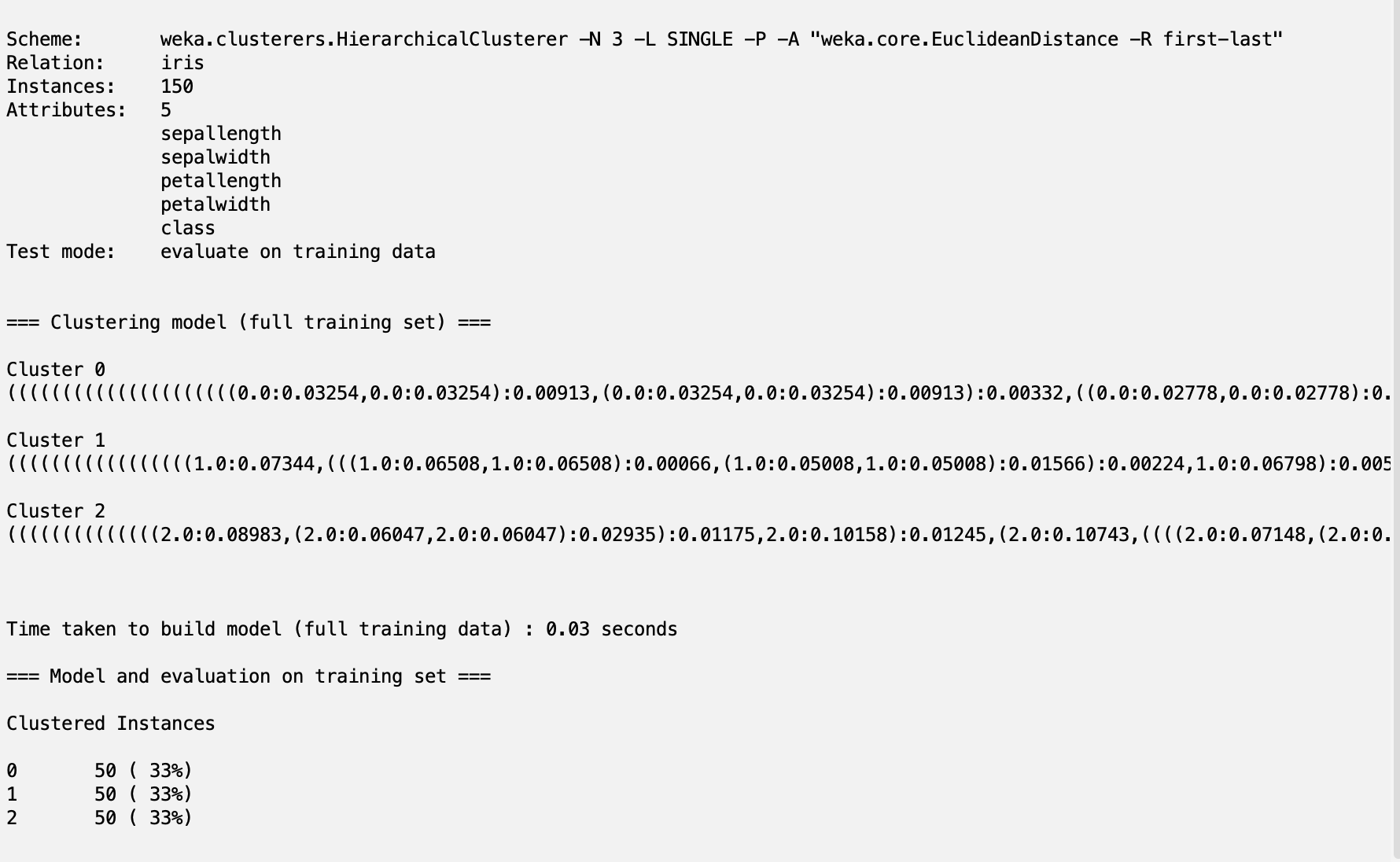
## 5. Clustering

### Simple kmeans:



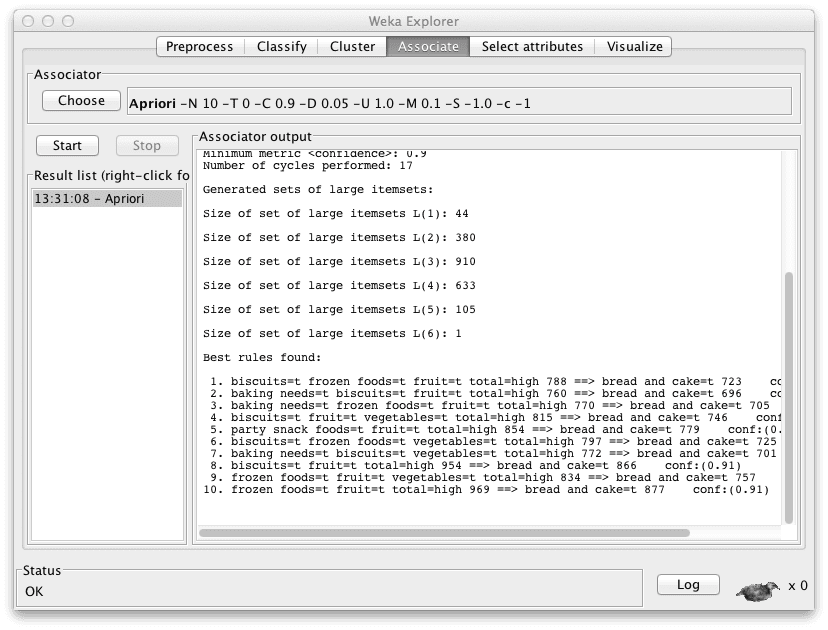


### Hierarchical clustering



## 6. Association Rule mining

This is done on the supermarket dataset.



**Conclusion:**

The analysis of the Iris dataset using various Weka tools provided valuable insights into the data's structure, patterns, and relationships. The combination of preprocessing, classification, clustering, and association rule mining techniques yielded a comprehensive understanding of the dataset, which can inform future research and applications. The Apriori algorithm used on the supermarket dataset deduced rules in the antecedent => consequent format. This is much more efficient than trying to deduce rules by hand.