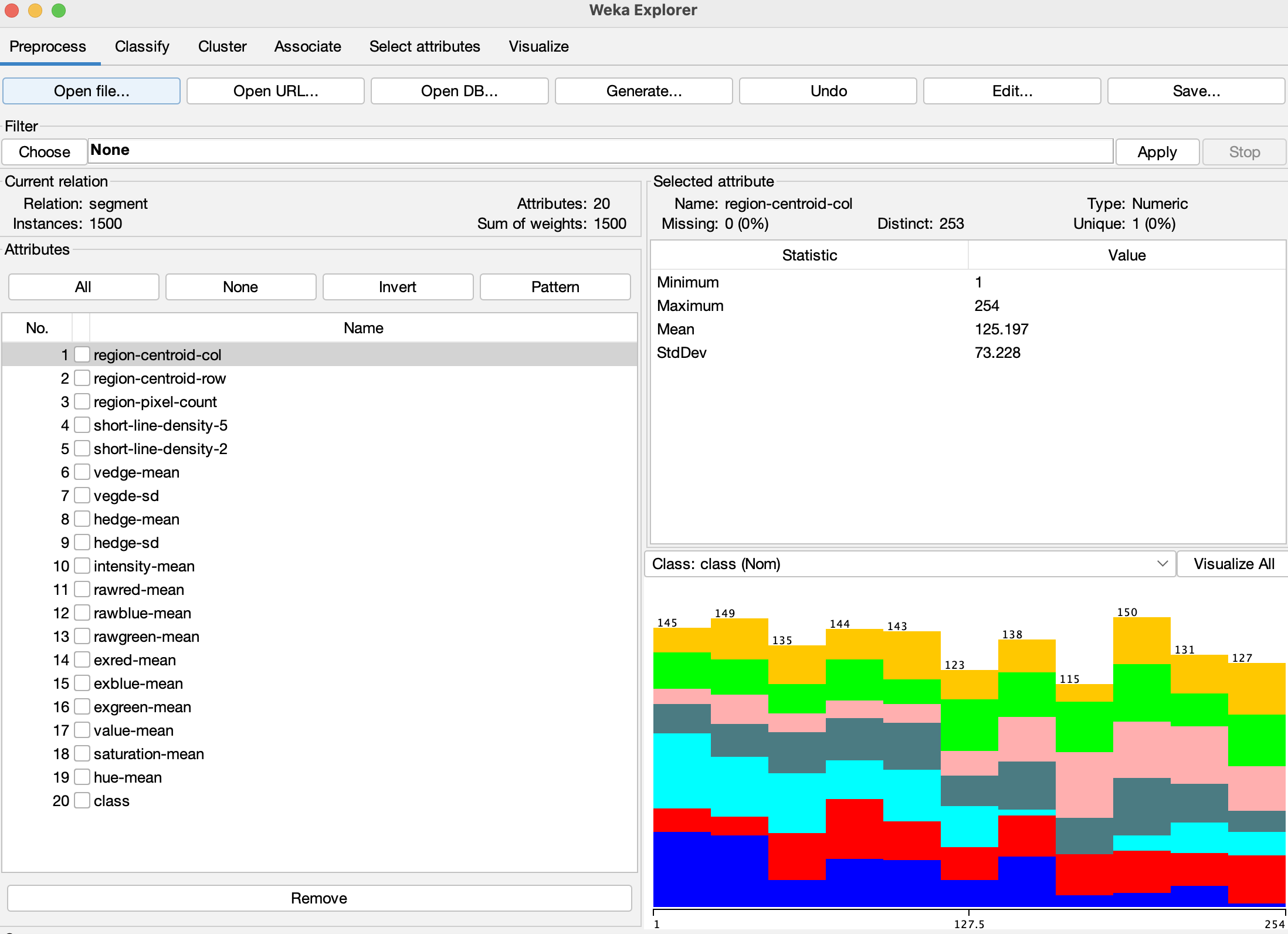
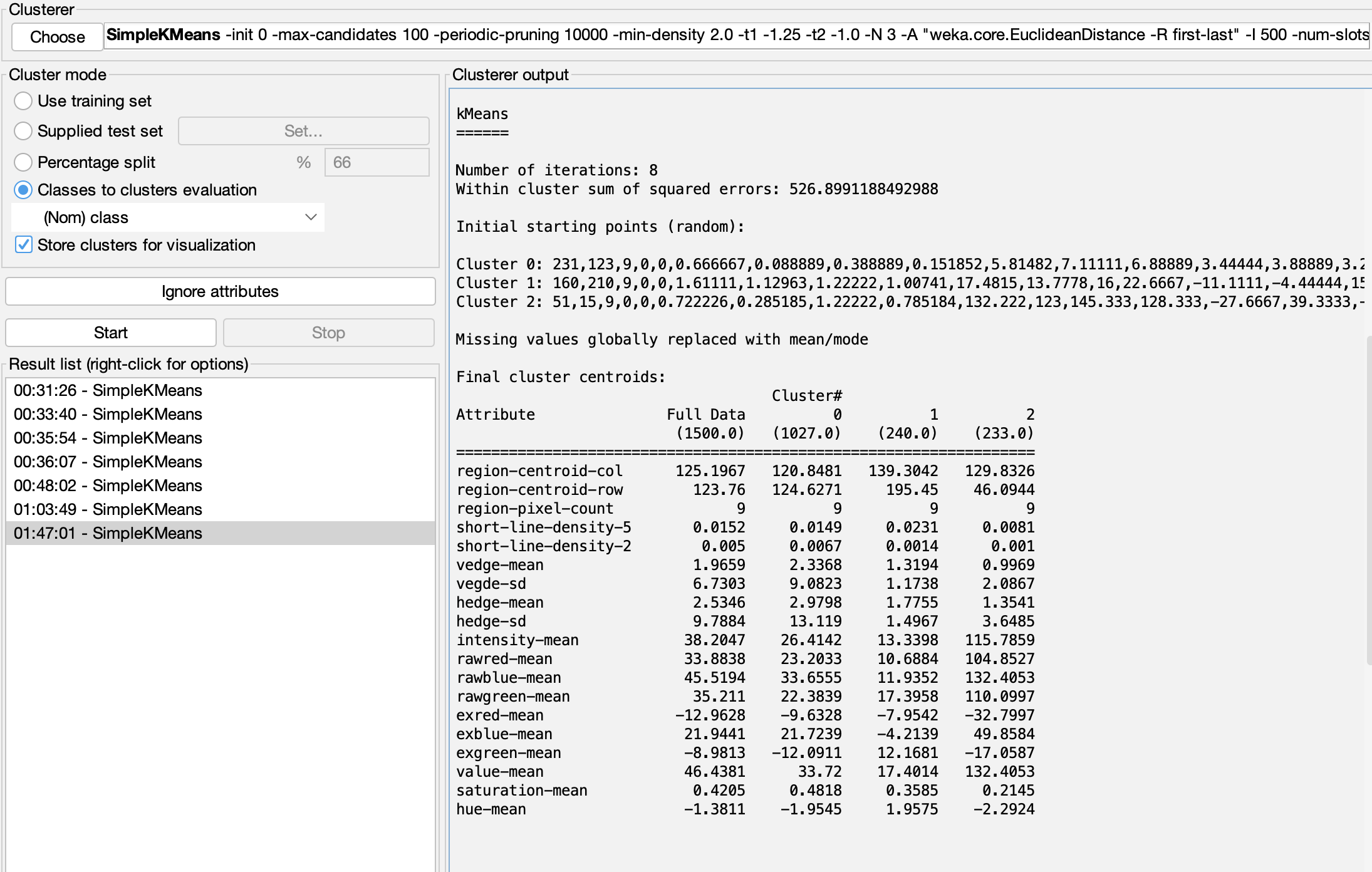
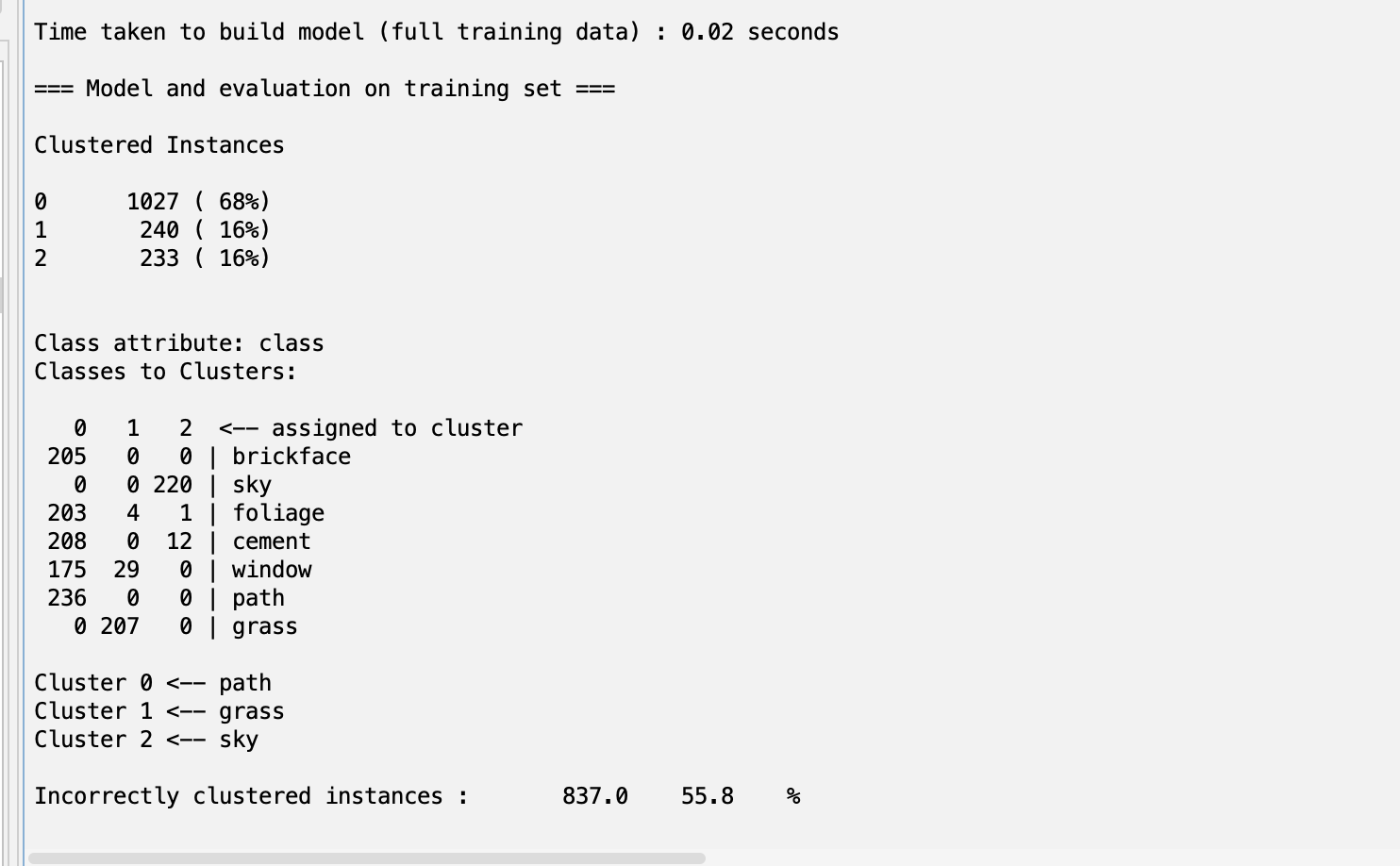
LAB 11 & 12

**Theory refers to what has been discussed in Lab (using Weka), Using different data samples, Apply classification, clustering, and association is performed with a discussion of the results.**

# **1) Weather Integer Dataset**

**Selected Attributes:**

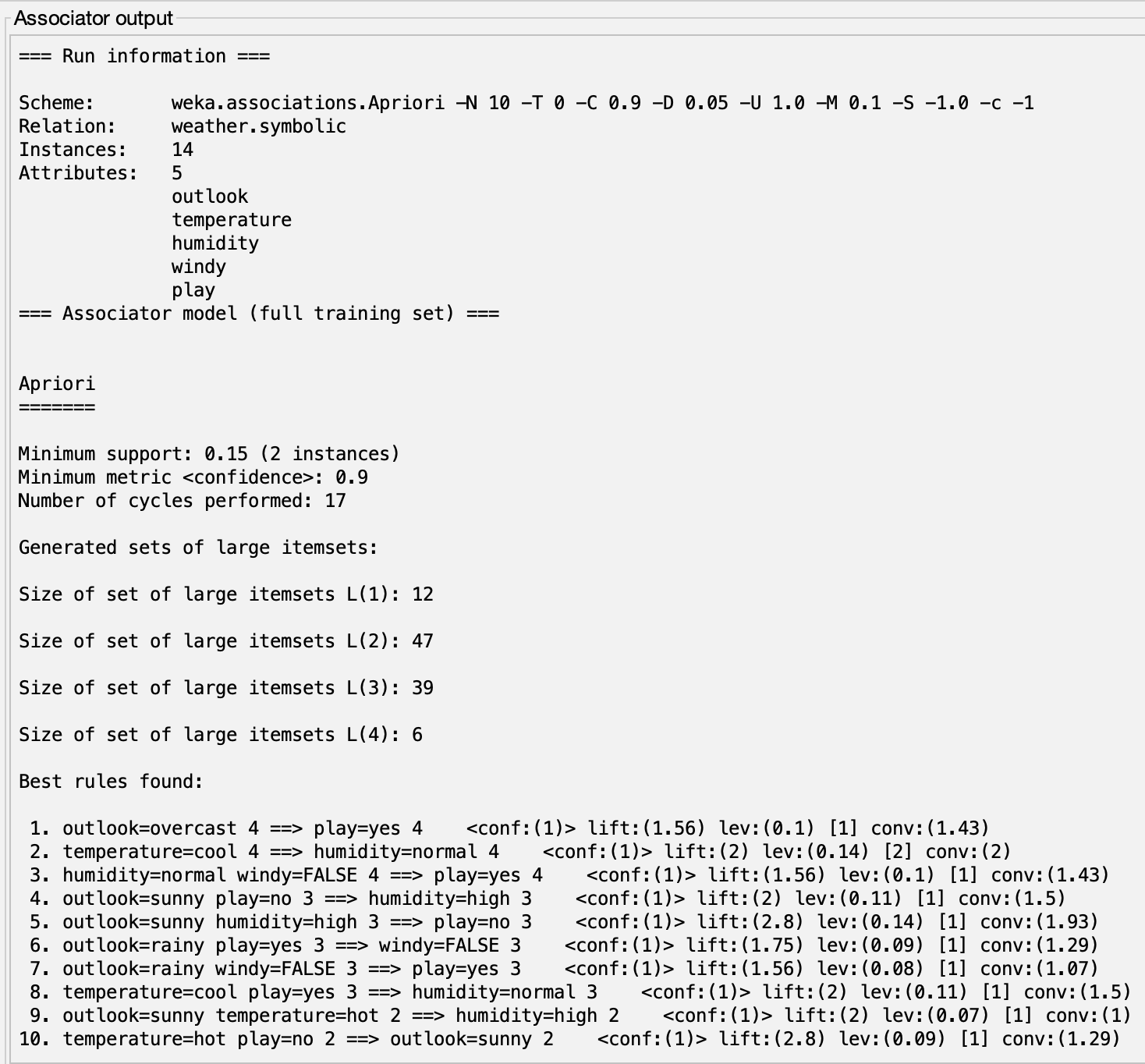
**Results** **Clustering ( SimpleKMeans clusterer):**

## **Explanation Clustering:**

* The clustering analysis was performed using the SimpleKMeans algorithm in Weka on a dataset with 1500 instances and 20 attributes. The algorithm converged after 8 iterations, resulting in three clusters. The within-cluster sum of squared errors was 526.90. The final cluster centroids provide insights into the characteristics of each cluster across the attributes. The majority of instances were assigned to Cluster 0 (68%), which represents the 'path' class. However, the clustering was not entirely accurate, as 55.8% of instances were incorrectly clustered. Notably, Cluster 1 corresponds to the 'grass' class, and Cluster 2 represents the 'sky' class, demonstrating the algorithm's ability to capture meaningful patterns within the data.

# **2) Weather Nominal Dataset**

## **Selected Attributes:Screenshot 2024-01-11 at 1.08.45 AM.png**

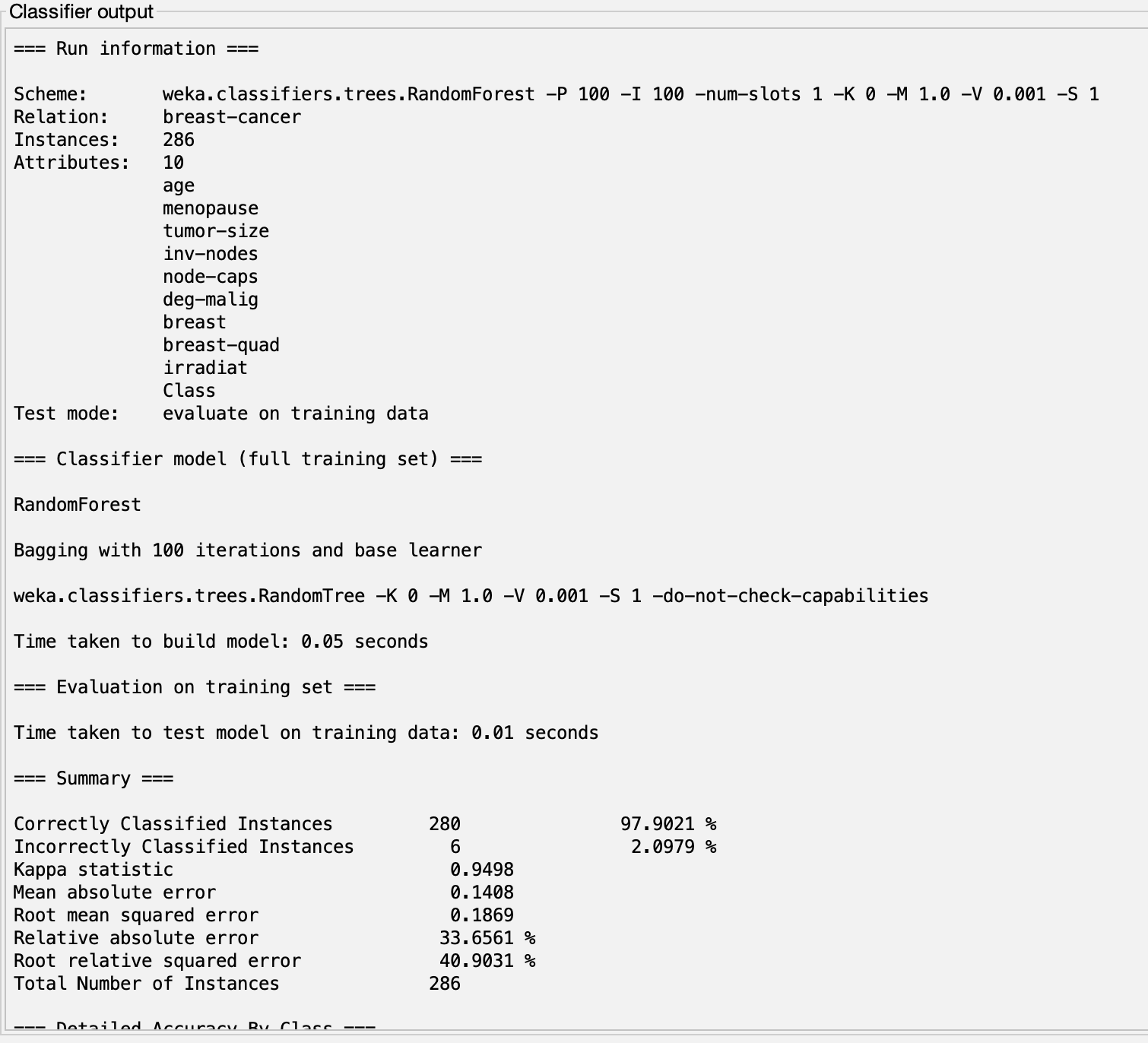
**Results Association (Apiriori):**

## **Explanation Association:**

* The Apriori association analysis on the symbolic weather dataset revealed several interesting rules with high confidence. For instance, the rule "outlook=overcast ==> play=yes" exhibited a confidence of 1, indicating a strong association between overcast conditions and playing. Similarly, rules like "temperature=cool ==> humidity=normal" and "outlook=sunny play=no ==> humidity=high" demonstrated confident associations. The generated rules can provide valuable insights into the relationships between different weather attributes and the likelihood of outdoor activities, offering practical implications for decision-making based on weather conditions.

# **3) Breast Cancer Dataset**

## **Selected Attributes:Screenshot 2024-01-11 at 1.36.01 AM.png**

**Results Classify (Random Forest):**

**Explanation Classifier:**

* The Random Forest classification on the breast cancer dataset achieved an impressive accuracy of 97.90%. The model correctly classified instances into the "no-recurrence-events" and "recurrence-events" classes with high precision and recall values for both classes. The detailed accuracy by class shows that the model performed well in distinguishing between the two outcomes, with a weighted average F-measure of 0.979. The confusion matrix indicates that only a small number of instances were misclassified (6 out of 286), contributing to the model's overall reliability in predicting breast cancer recurrence based on the provided attributes.