Experiment 9

K-Means Clustering

Aim: To demonstrate Clustering process on iris.arff dataset using K-means algorithm.

Tasks:

1. Load iris.arff dataset and remove target attribute.
2. Cluster the dataset using k-means algorithm.
3. Visualize the results.

Task 1: Load iris.arff dataset and remove target attribute.

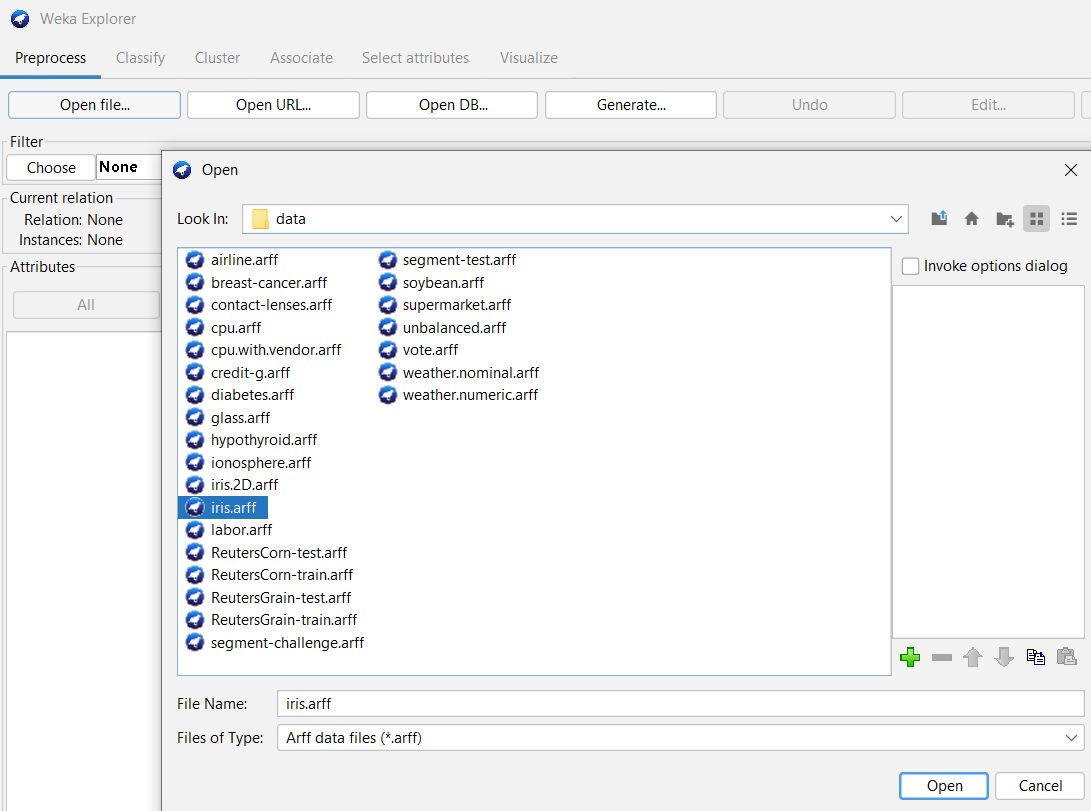
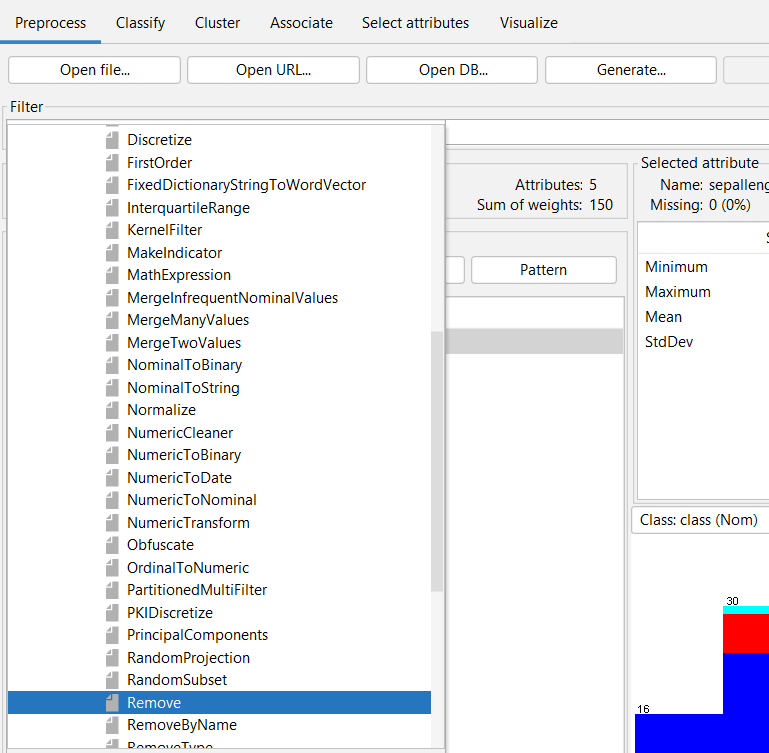
Load iris.arff from the Weka’s data folder.

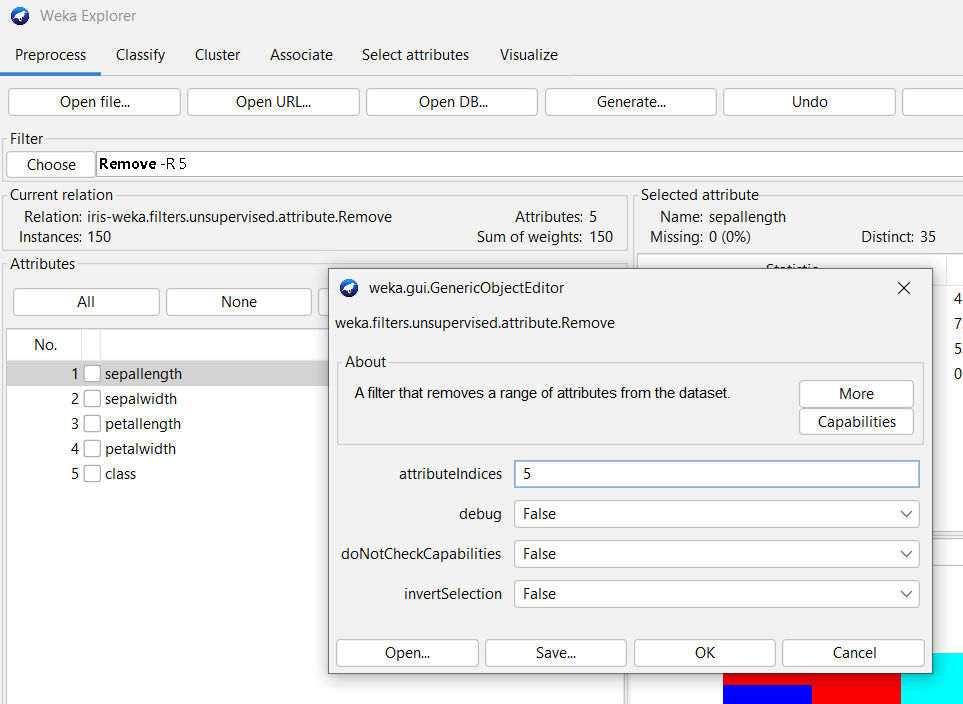
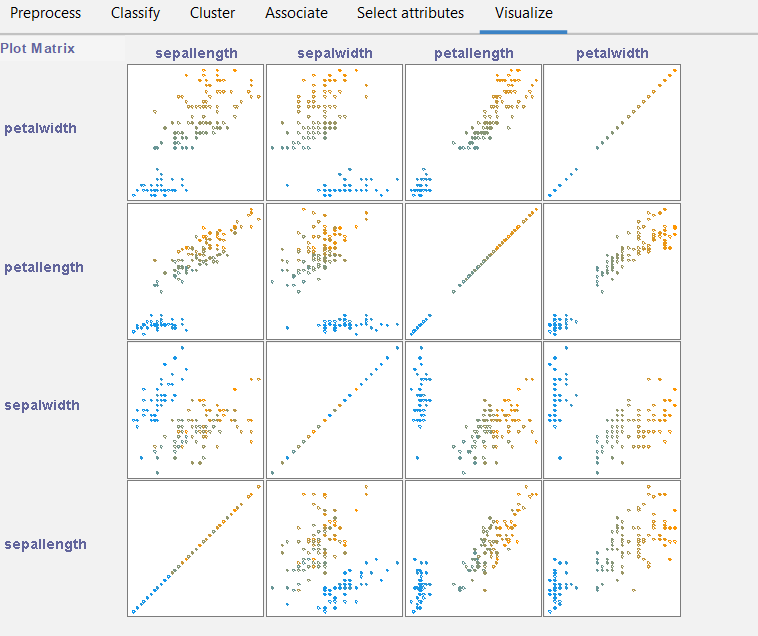
Select Choose 🡪 filter 🡪 unsupervised 🡪 attribute 🡪 Remove.

Click on filter and enter target attribute index.

Click on Apply.

Click on Visualize to observe the number of clusters that dataset may have.

Task 2: Cluster the dataset using k-means algorithm.

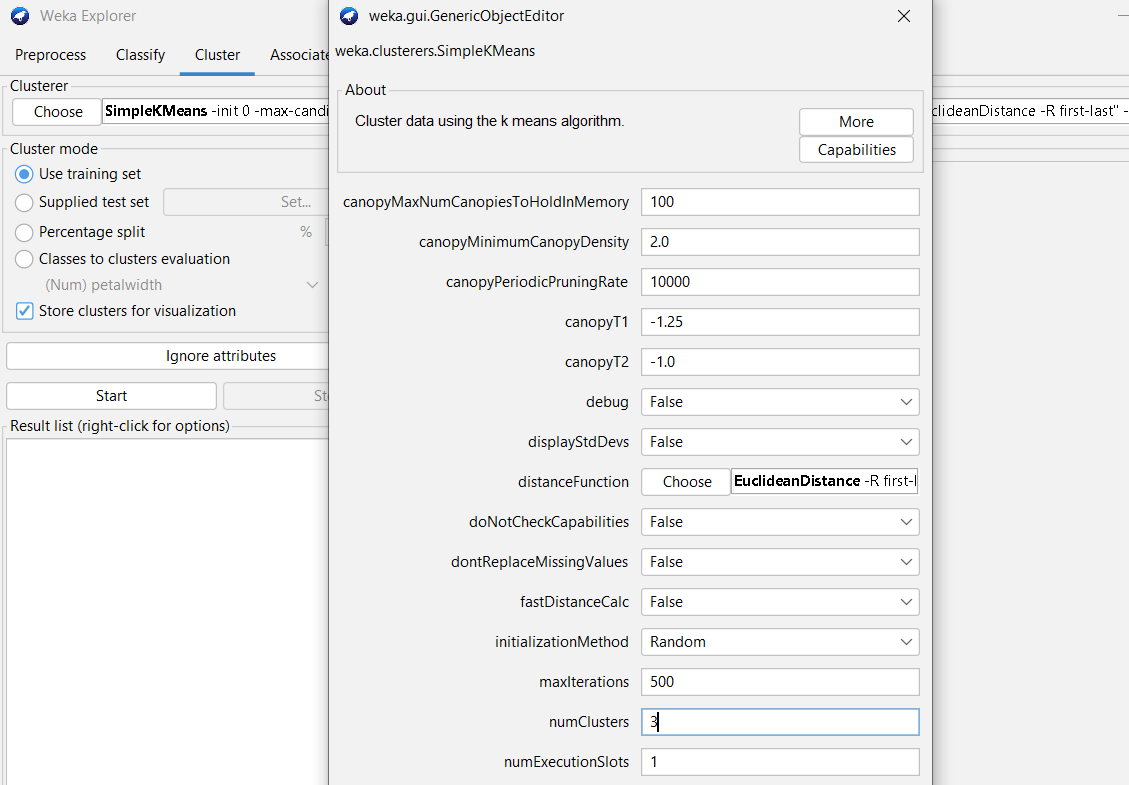
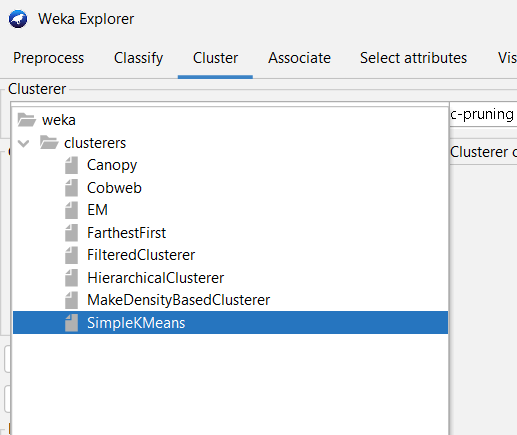
A clustering is an unsupervised learning technique to find groups of similar instances in the entire dataset.

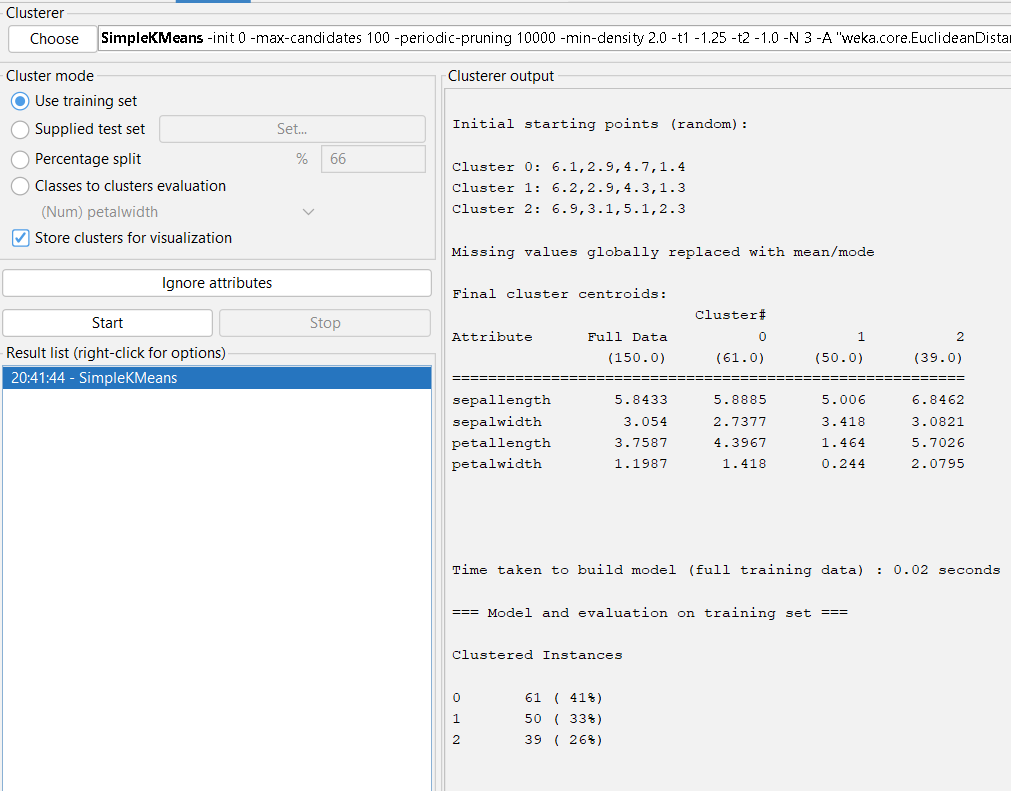
K-means clustering:

* + 1. Randomly K initial cluster centers are chosen.
    2. All instances are assigned to closest cluster centers according to the Euclidean distance measure.
    3. Cluster centers are updated based on the mean values of all instances in their respective clusters.
    4. Steps ii and iii are repeated until cluster centers are stabilized.

Steps:

* Open Explorer 🡪 Cluster window.
* Select the *Use training set* option in *Cluster mode* pane.
* Check the *Store clusters for visualization* option.
* Select Choose 🡪 Clusters 🡪 SimpleKMeans.
* Click on SimpleKMeans and set the number of clusters.
* Click on Start.





Observations:

Number of Iterations:

Starting random cluster centroids:

Final cluster centroids:

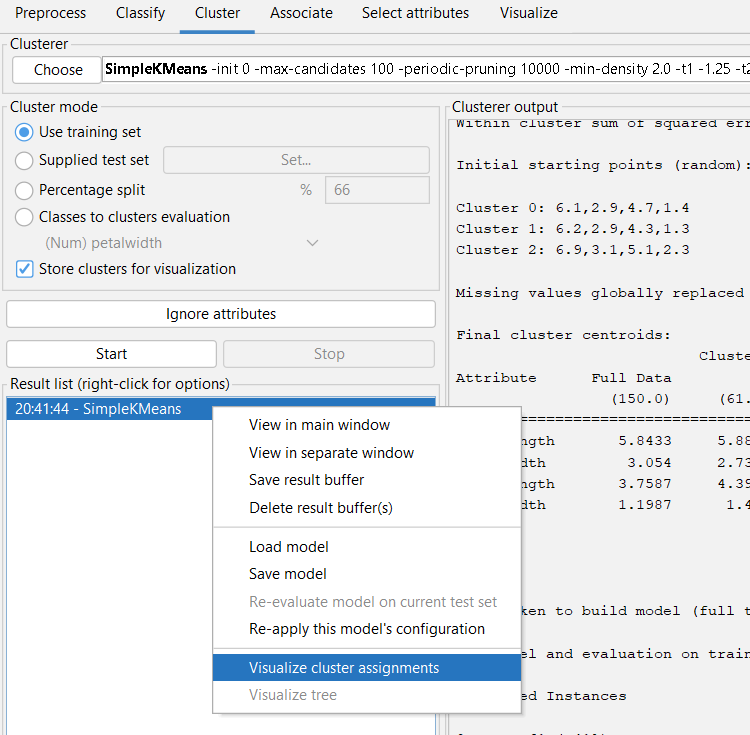
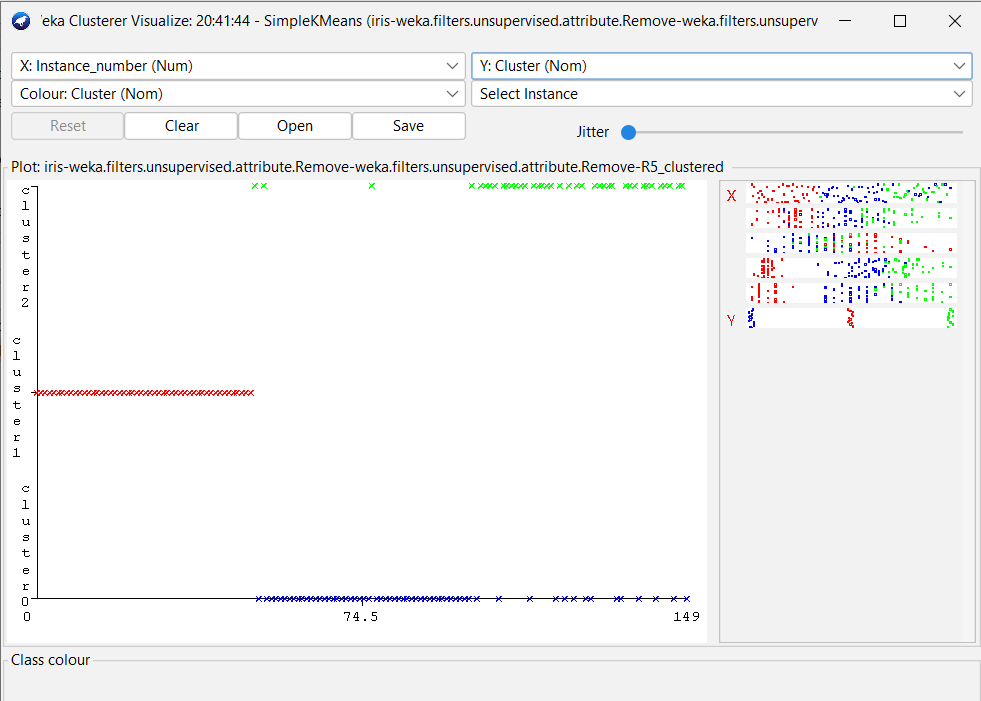
Number of instances in each cluster:

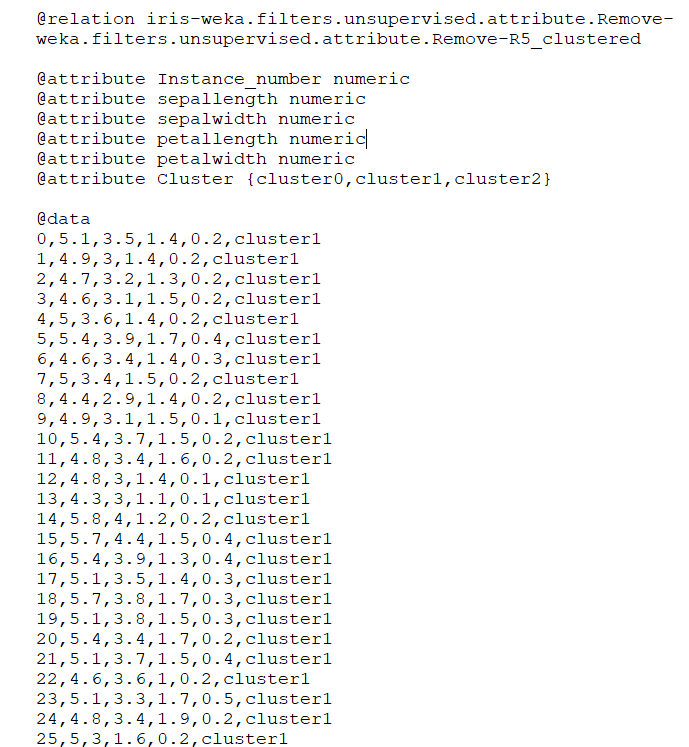
Task 3: Visualizing the clusters

Right click on the clustering model in Result list.

Click on Visualize cluster assignment.

Save the result in arff format.



Conclusion: