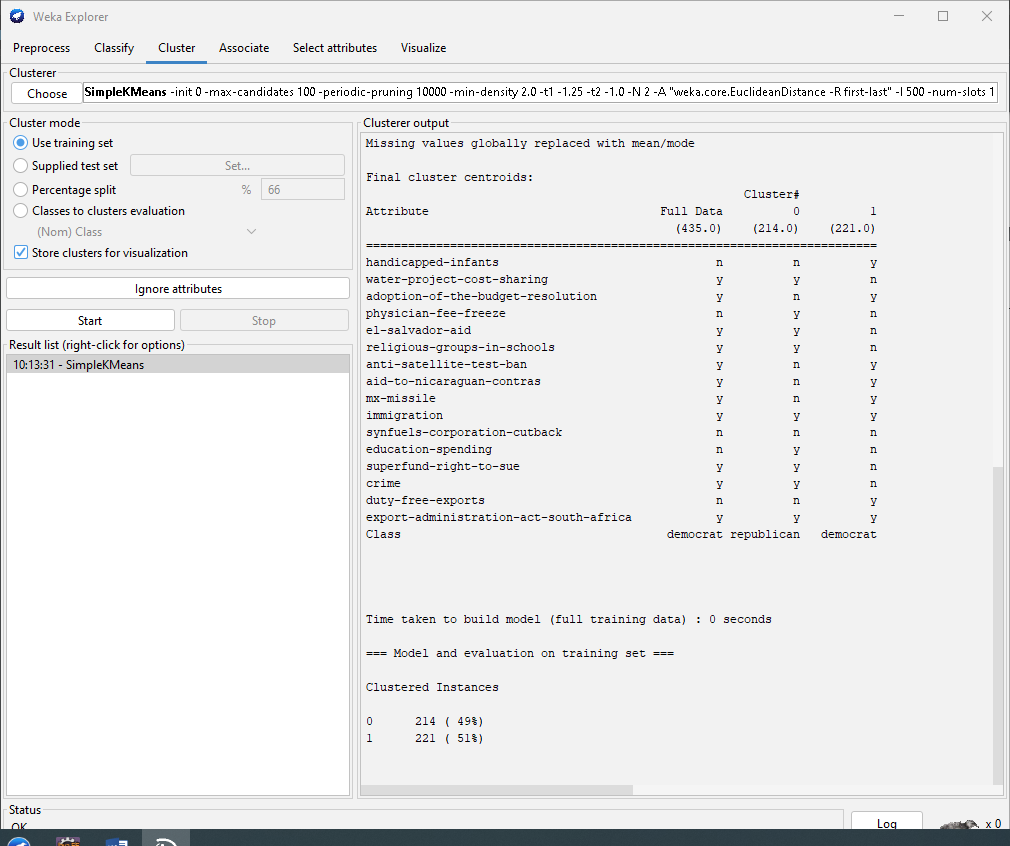
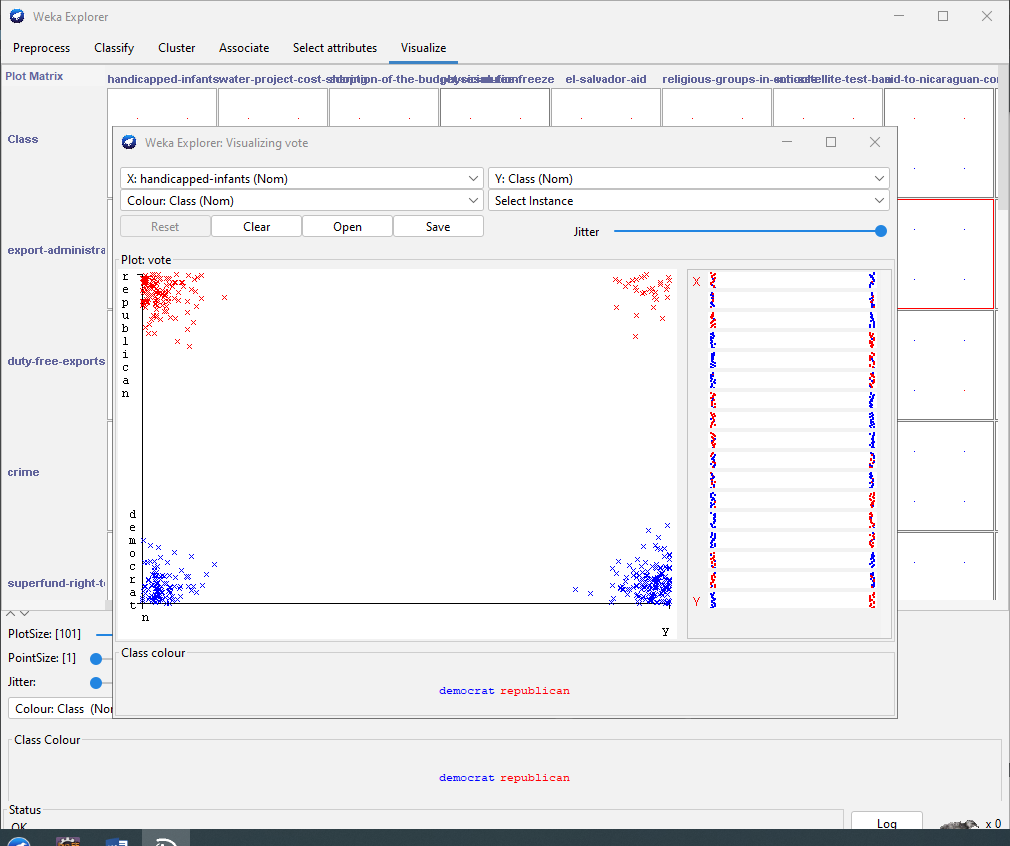
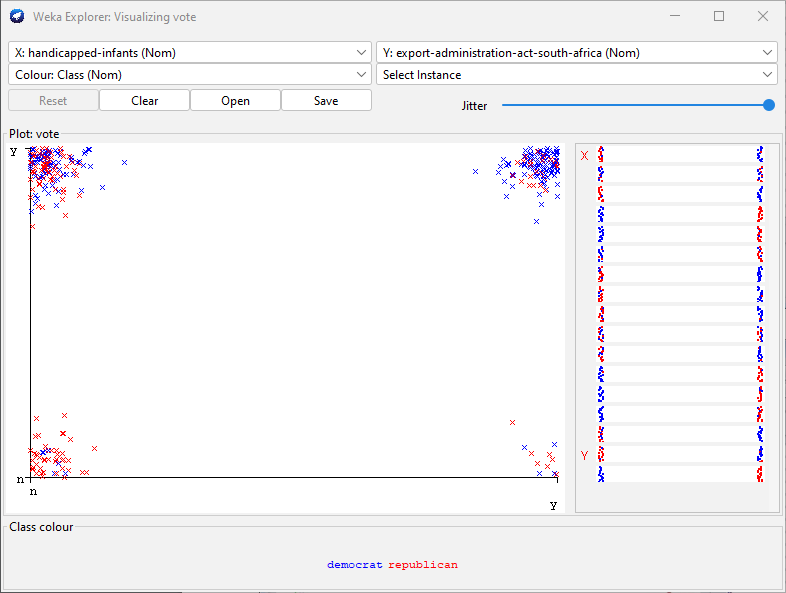
21083791 – Lê Hoàng Khang – Lab 8-9: : CLUSTERING DATA

**Exercise 1:** This experiment illustrates the use of simple k-mean clustering with Weka. The sample data set used for this example is based on the vote.arff data set. This document assumes that appropriate pre-processing has been performed.

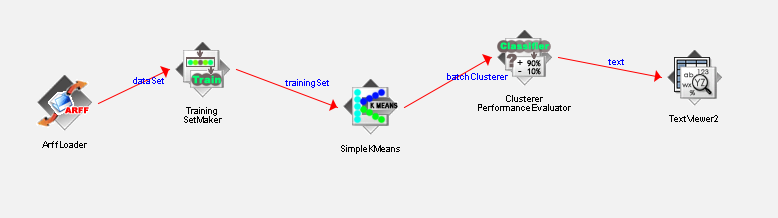
**Using Weka Explore**

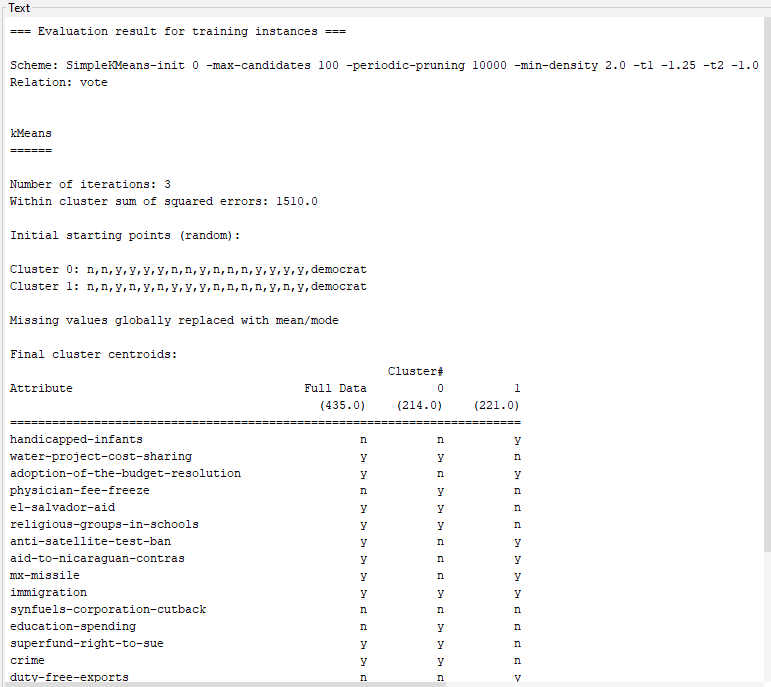






Using Weka KnowledgeFlow





Exercise 3: The goal of this data mining study is to find groups of animals in the zoo dataset, and to check whether these groups correspond to the real animal types in the dataset.

1. What types of variables are in this dataset?

- Có hai loại kiểu dữ liệu là Nominal và Numeric

b. How many rows / cases are there?

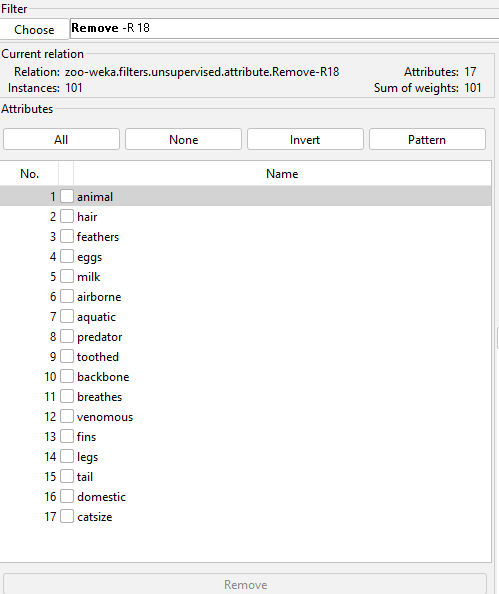
- Dữ liệu có 101 dòng

c. How many animal types are represented in this dataset? List them here.

- Có 100 loài động vật trong dataset

d. After removing the type attribute, go to the Cluster tab. How many clustering algorithms are available in Weka?

- Remove thuộc tính type



Có 8 thuộc tính phân cụm trong weka

e. List the clustering algorithms seen in class, and map these to the ones provided in Weka.

- SimpleKMeans, HierarchicalClusterer

Thuật toán HierarchicalClusterer

A screenshot of a computer

Description automatically generated

f. Start using the SimpleKMeans clusterer choosing 7 clusters. Do the clusters learnt and their centroids seem to match the animal types?

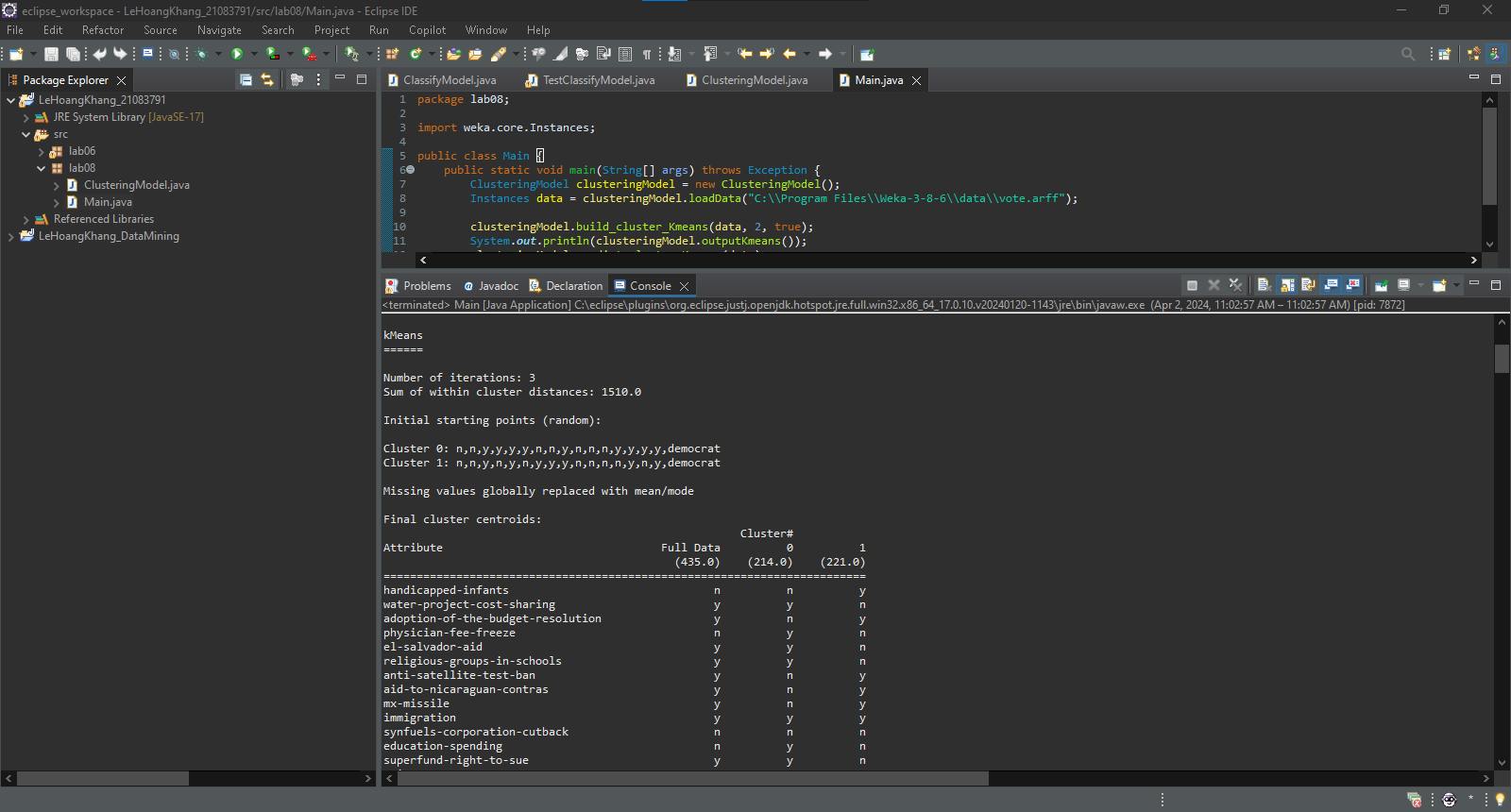
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Exercise 5.



A screenshot of a computer

Description automatically generated