

**WMCS009-05**  
**Information Systems**

## **Assignment 3 - Association Analysis**



**Dr. George Azzopardi**

# Task

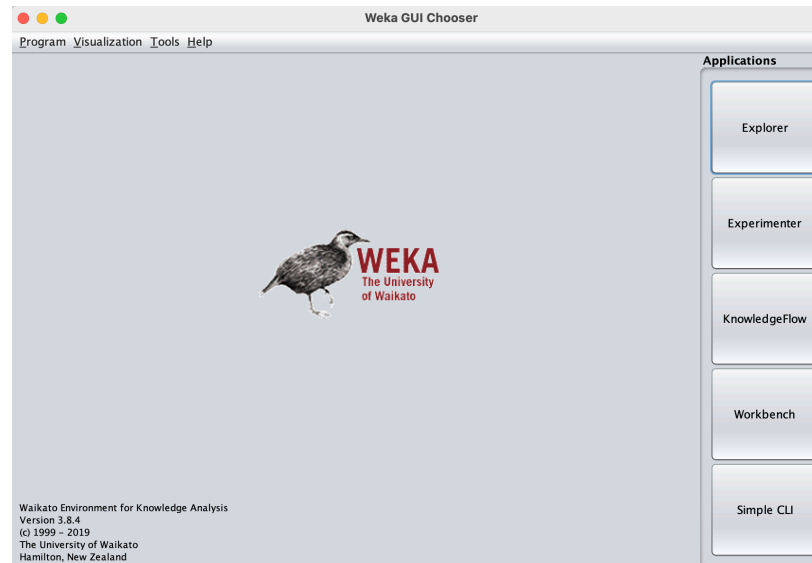
- Use Jupyter notebook to implement the Apriori algorithm in Python
- **Input**
  - CSV file: It must contain a binary representation of all concerned items. The first row must contain the header (i.e. the names of the items), and the remaining rows must contain only 1s and 0s separated by commas
  - Minimum support - the threshold used to select the frequent itemsets
  - Minimum confidence - the threshold used to determine the association rules
- **Output**
  - List of association rules that satisfy the given minimum support and confidence
- **Deliverable**
  - The python notebook and a short report describing how the main steps of the algorithm were implemented. These include the self-join and pruning (based on the Apriori principle) in determining the frequent itemsets, and the non-monotonicity property in determining the association rules.

# Rubric

- Use the provided myDataFile.csv as your test case. Test your algorithm with a minimum ***support of 0.005*** and ***confidence of 0.6***. List the number of frequent itemsets per layer and the association rules in the form “A -> B (confidence = ?” where A and B can be comma separated lists. **(3 points for correct output)**
- Correct implementation of the Apriori principle in determining the frequent itemsets **(3 points)**
- Correct implementation of the non-monotonicity property in the determination of the association rules **(3 points)**
- Readability of the algorithm and report **(1 point)**

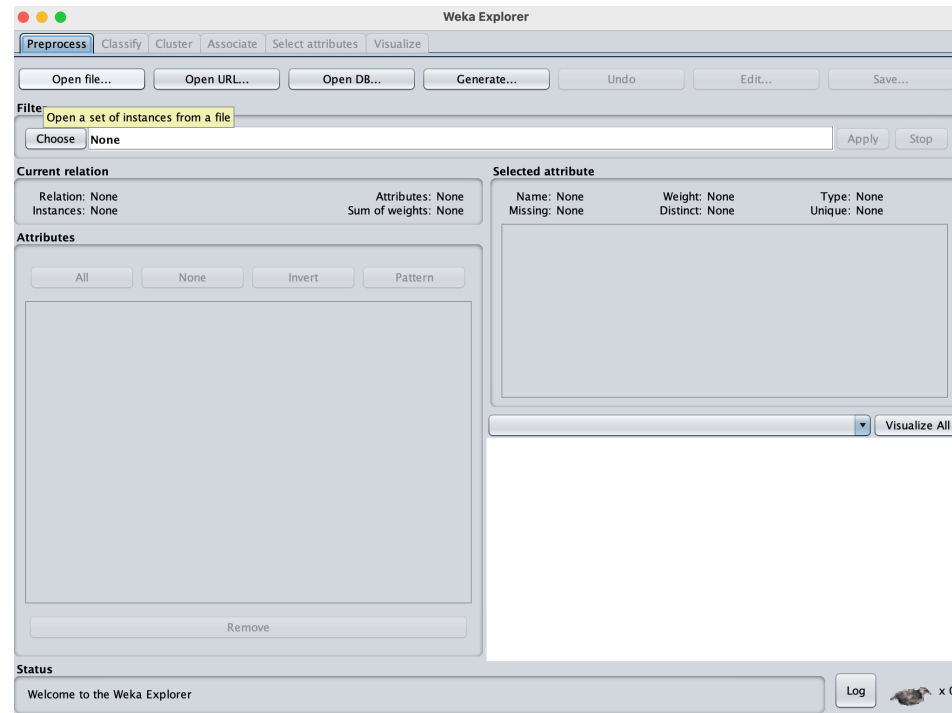
# Additional Information

- Use the Weka tool to check your results
  - Download it from here: [https://waikato.github.io/weka-wiki/downloading\\_weka/](https://waikato.github.io/weka-wiki/downloading_weka/)
- Click on Explorer



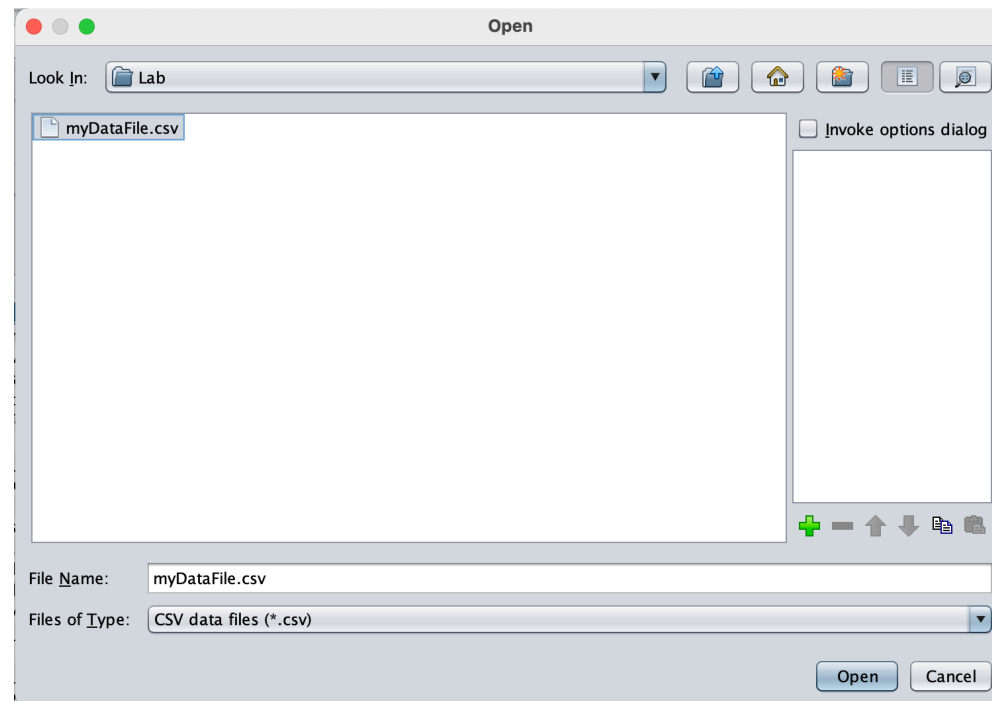
# Additional Information

- Click on open file



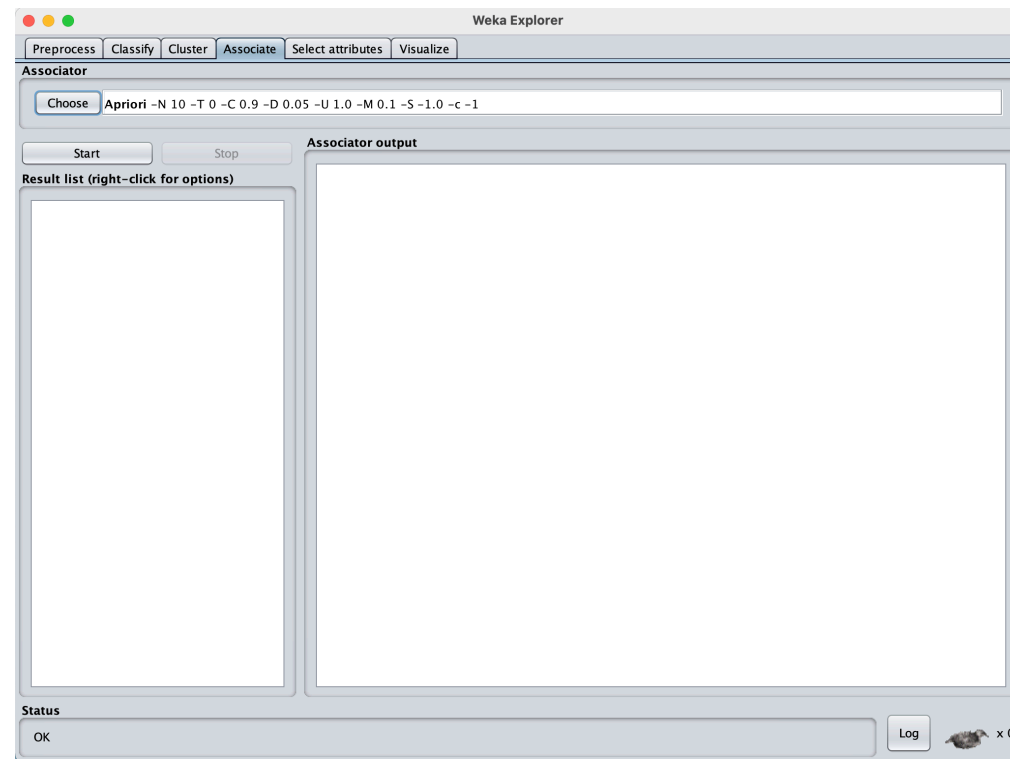
# Additional Information

- Locate myDataFile.csv



# Additional Information

- Click on the **Associate** tab and click in the field next to the **Choose** button



# Additional Information

- Set the minimum support and the minimum confidence.
  - Click OK followed by Start



The screenshot shows a Java Swing window titled "weka.gui.GenericObjectEditor" with a subtitle "weka.associations.Apriori". The window contains an "About" section with a text box stating "Class implementing an Apriori-type algorithm." and buttons for "More" and "Capabilities". Below this is a list of configuration parameters for the Apriori algorithm, each with a text input field or a dropdown menu. The parameters and their values are: "car" (False), "classIndex" (-1), "delta" (0.05), "doNotCheckCapabilities" (False), "lowerBoundMinSupport" (0.005), "metricType" (Confidence), "minMetric" (0.6), "numRules" (9999), "outputItemSets" (False), "removeAllMissingCols" (False), "significanceLevel" (-1.0), "treatZeroAsMissing" (False), "upperBoundMinSupport" (1.0), and "verbose" (False). At the bottom of the window are four buttons: "Open...", "Save...", "OK", and "Cancel".

| Parameter              | Value      |
|------------------------|------------|
| car                    | False      |
| classIndex             | -1         |
| delta                  | 0.05       |
| doNotCheckCapabilities | False      |
| lowerBoundMinSupport   | 0.005      |
| metricType             | Confidence |
| minMetric              | 0.6        |
| numRules               | 9999       |
| outputItemSets         | False      |
| removeAllMissingCols   | False      |
| significanceLevel      | -1.0       |
| treatZeroAsMissing     | False      |
| upperBoundMinSupport   | 1.0        |
| verbose                | False      |