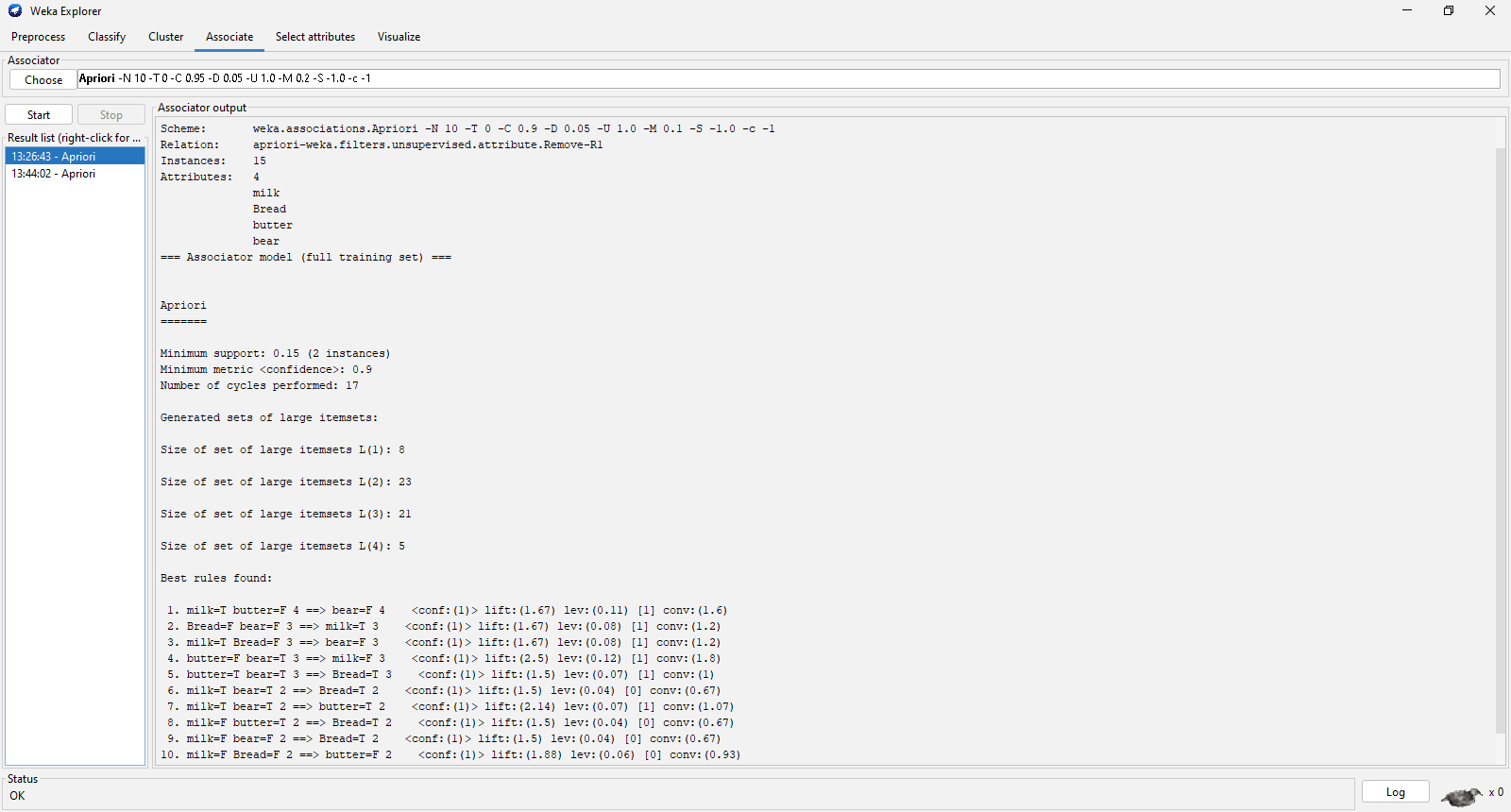
Lab \_ 3 \_4 \_ 5

Exercise\_01

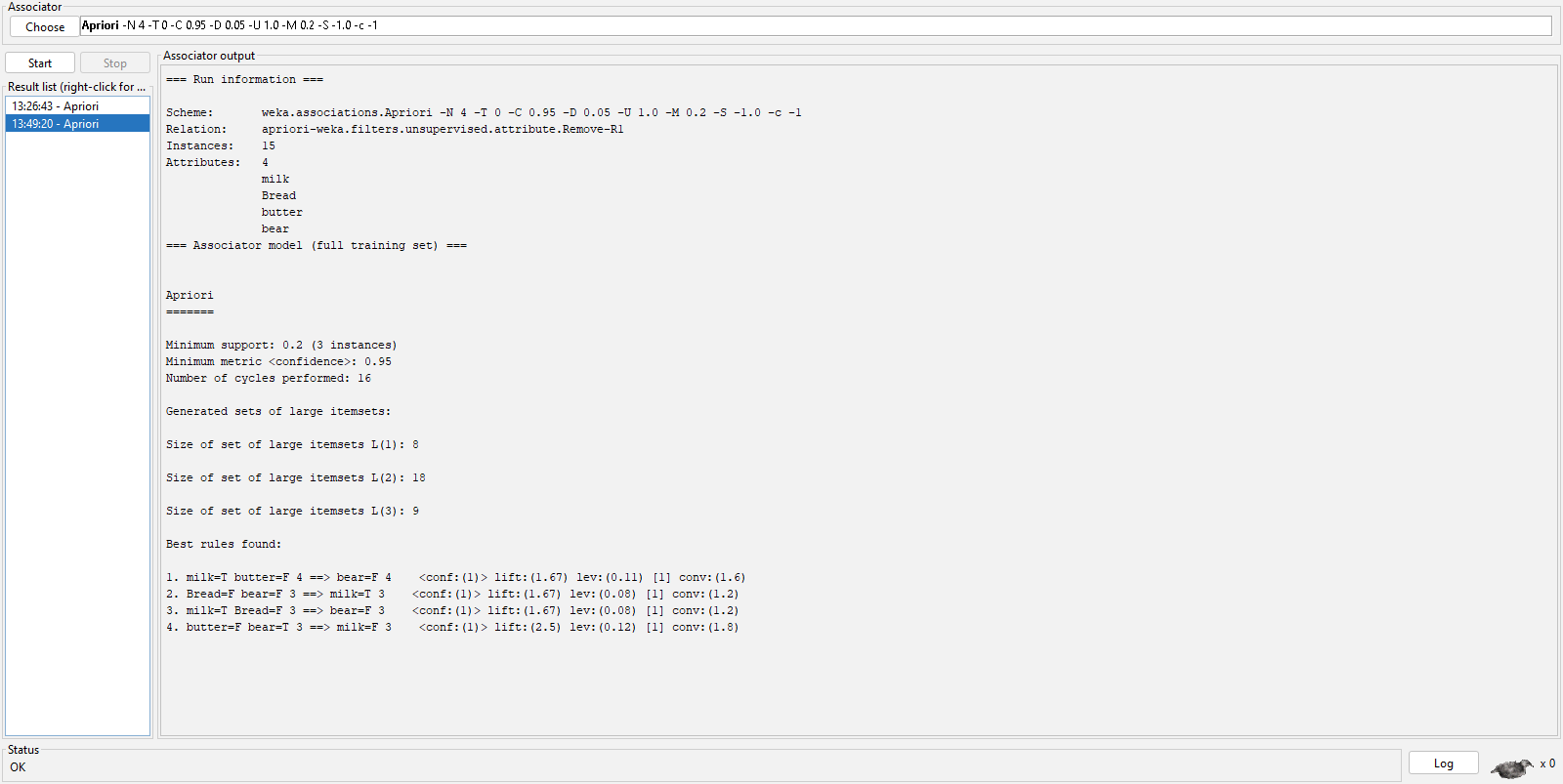
1. Analyse the association rules mined from the data set. What are their confidence and support values? Examine the number of large itemsets.

* Apriori Parameters:
  + Minimum Support: 0.15 (2 instances)
  + Minimum Confidence: 0.9
  + Number of cycles performed: 17
* Generated Sets of Large Itemsets:
  + Size of set of large itemsets L(1): 8
  + Size of set of large itemsets L(2): 23
  + Size of set of large itemsets L(3): 21
  + Size of set of large itemsets L(4): 5
* Best Rules Found:
  + The output provides the top 10 association rules based on confidence and lift.
  + Each rule is presented with its antecedent, consequent, confidence, lift, leverage, and conviction.
* Example Rules:
  + Rule 1: If milk is true and butter is false, then bear is false. (Confidence: 1, Lift: 1.67)
  + Rule 4: If butter is false and bear is true, then milk is false. (Confidence: 1, Lift: 2.5)
  + Rule 5: If butter is true and bear is true, then Bread is true. (Confidence: 1, Lift: 1.5)



1. Try changing different values for the following parameters \lowerBoundMinSupport" (min threshold for support), \minMetric" (min threshold for confidence). As you change these parameter values what do you notice about the rules that are found by the associator? Note that the parameter \numRules" limits the maximum number of rules that the associator looks for, you can try changing this value.

After change lowerBoundMinSupport from 0.1 into 0.2 and minMetric from 0.9 into 0.95 and numRules from 10 into 4



numRules now is only have 4

1. Apriori Parameters:
   * Minimum Support: 0.2 (3 instances)
   * Minimum Confidence: 0.95
   * Number of cycles performed: 16
2. Generated Sets of Large Itemsets:
   * Size of set of large itemsets L(1): 8
   * Size of set of large itemsets L(2): 18
   * Size of set of large itemsets L(3): 9
3. Best Rules Found:
   * Rule 1: If milk is true and butter is false, then bear is false. (Confidence: 1, Lift: 1.67)
   * Rule 2: If Bread is false and bear is false, then milk is true. (Confidence: 1, Lift: 1.67)
   * Rule 3: If milk is true and Bread is false, then bear is false. (Confidence: 1, Lift: 1.67)
   * Rule 4: If butter is false and bear is true, then milk is false. (Confidence: 1, Lift: 2.5)

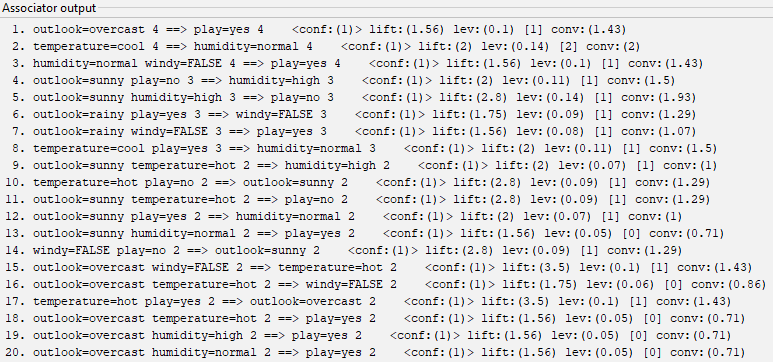
Exercise\_02

* 1. Load the data in Weka Explorer. Select the Associate tab. How many different associations rule mining algorithms are available?
* 12 + 47 + 39 + 6 = 104 rules

1. Choose Apriori algorithm with the following parameters (which you can select by clicking on the chosen algorithm: support threshold = 15% (lowerBoundMinSupport = 0.15), confidence threshold = 90% (metricType = confidence, minMetric = 0.9), number of rules = 50 (numRules = 50). After starting the algorithm, how many rules do you find? Could you use the regular weather dataset to get the results? Explain why.

* 12 + 47 + 39 + 6 = 104 rules
* I cannot use the regular weather dataset to obtain results because numeric values need to be preprocessed into nominal ones to transform the weather into a discrete structured format.

1. Paste a screenshot of the Explorer window showing at least the first 20 rules.



1. Based on the output, what is the support for this item set? outlook = rainy humidity = normal windy = FALSE play = yes

Itemsets = 6

{outlook = rainy, humidity = normal, windy = FALSE, play = yes} 2

* 2/6 = 1/3

1. Apriori algorithm generates association rules from frequent itemsets. How many itemsets of size 4 were found? Which rule(s) have been generated from itemset of size 4 (temperature=mild, windy=false, play=yes, outlook=rainy)? List their numbers in the list of rules.

* How many itemsets of size 4 were found: 6
* Rules 7
  + outlook=rainy temperature=mild play=yes (Size 3)
  + outlook=rainy temperature=mild windy=FALSE (Size 3)
  + outlook=rainy temperature=mild play=yes (Size 3)
  + outlook=rainy temperature=cool humidity=normal (Size 3)

1. What is the total number of possible rules for the weather data for each combination of values in the following table?

Number of Rules=3Number of Attributes−2Number of Attributes+1

Assuming you have 5 attributes in your dataset, the number of rules without any constraints would be 35−26+1=244.

* Number of rules = 244 \* Minimun confidence \* Minimum support

|  |  |  |  |
| --- | --- | --- | --- |
| Minimun confidence | Minimum support | Number of rules | Values |
| 0.9 | 0.3 | 65.88 |  |
| 0.9 | 0.2 | 43.92 |  |
| 0.9 | 0.1 | 21.96 |  |
| 0.8 | 0.3 | 58.56 |  |
| 0.8 | 0.2 | 39.04 |  |
| 0.8 | 0.1 | 19.52 |  |
| 0.7 | 0.3 | 51.24 |  |
| 0.7 | 0.2 | 34.16 |  |
| 0.7 | 0.1 | 17.08 |  |