LAB 3+4

Exercise 1: Illustrate the Apriori algorithm in Weka to find association rules from the transaction

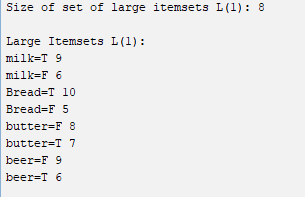
database.

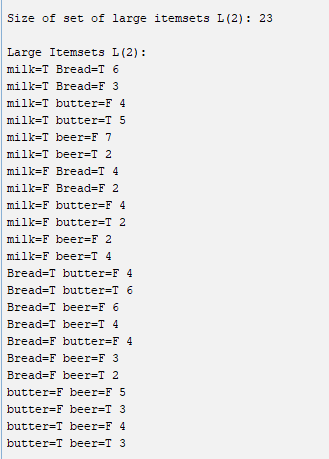
- Analyse the association rules mined from the data set. What are their confidence and support

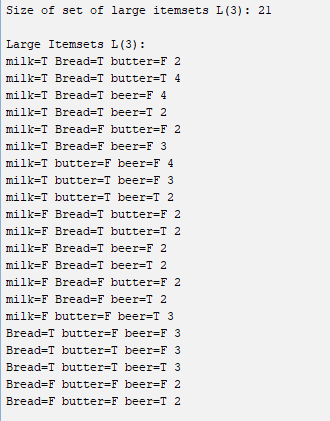
values? Examine the number of large itemsets.

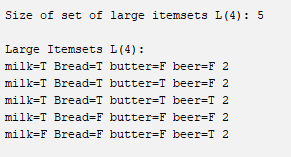
* In simple terms, the Apriori algorithm searches for all itemsets that have a support greater than or equal to a minimum support threshold. The support of an itemset is the percentage of transactions in the dataset that contain that itemset. Once all frequent itemsets are found, association rules can be generated from them.











- Try changing different values for the following parameters \lowerBoundMinSupport" (min

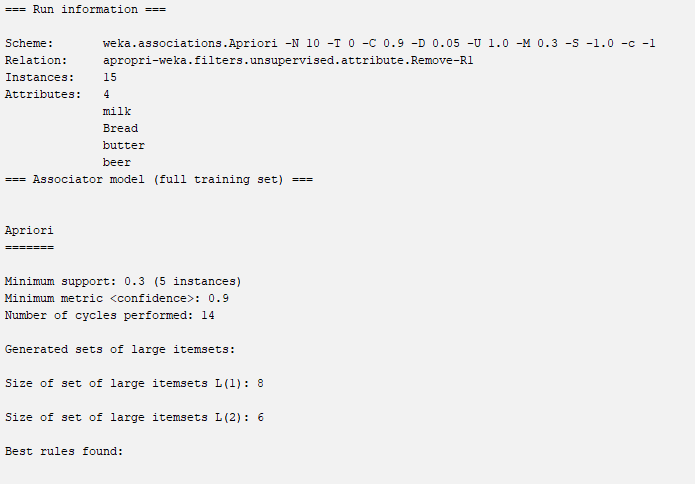
threshold for support), \minMetric" (min threshold for confidence). As you change these

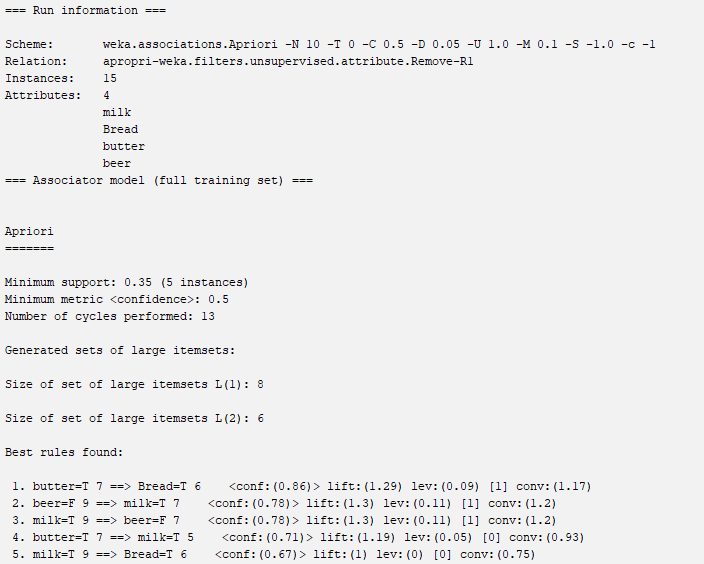
parameter values what do you notice about the rules that are found by the associator? Note

that the parameter \numRules" limits the maximum number of rules that the associator looks

for, you can try changing this value.

* Khi thử tăng lượng MinSupport hoặc giảm lượng minMetric thì số lượng luật tìm được cũng giảm theo 1 số lượng nhất định.





Exercise 2: Find association rules in the weather.nominal dataset by Weka explorer

a. Load the data in Weka Explorer. Select the Associate tab. How many different associations rule

mining algorithms are available?

-> 2 : Apriori and FilteredAssociator

b. 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

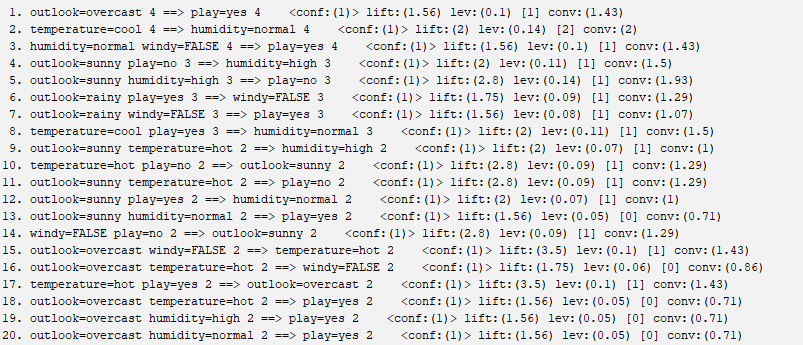
= 50). After starting the algorithm, how many rules do you find? Could you use the

regular weather dataset to get the results? Explain why.

- Tổng số luật tìm được: 12 + 47 + 39 + 6 = 104

- Không thể dùng regular weather dataset để lấy kết quả từ Apriori vì loại của các thành phần thuộc dạng rời rạc (nominal) còn các thành phần thuộc phần numeric bắt buộc phải tiền xử lý mới có thể tiến hành sử dụng công thức.

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



d. Based on the output, what is the support for this item set?

outlook = rainy humidity = normal windy = FALSE play = yes : 1/3

e. Apriori algorithm generates association rules from frequent itemsets. How many itemsets of size

4 were found?

* 6

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.

outlook=rainy temperature=mild windy=FALSE play=yes 2

outlook=rainy temperature=mild windy=FALSE 2

outlook=rainy temperature=mild play=yes 2

outlook=rainy windy=FALSE play=yes 3

outlook=rainy temperature=mild 3

outlook=rainy windy=FALSE 3

outlook=rainy play=yes 3

temperature=mild windy=FALSE 3

temperature=mild play=yes 4

windy=FALSE play=yes 6

outlook=rainy 5

temperature=mild 6

windy=FALSE 8

play=yes 9

f. What is the total number of possible rules for the weather data for each combination of values in

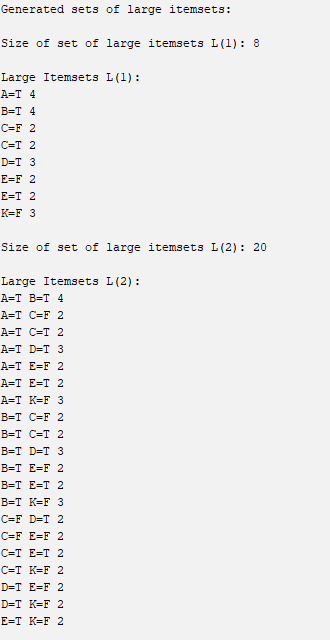
the following table?

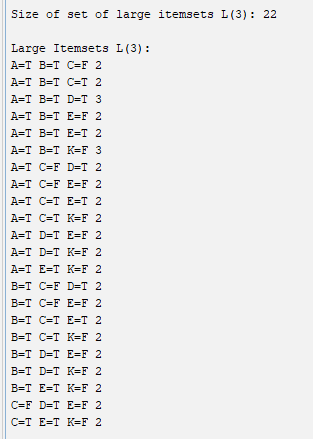
|  |  |  |
| --- | --- | --- |
| Minimum confidence | Minimum support | Numbers of rules |
| 0.9 | 0.3 | 22 |
| 0.9 | 0.2 | 42 |
| 0.9 | 0.1 | 104 |
| 0.8 | 0.3 | 22 |
| 0.8 | 0.2 | 42 |
| 0.8 | 0.1 | 104 |
| 0.7 | 0.3 | 22 |
| 0.7 | 0.2 | 42 |
| 0.7 | 0.1 | 104 |

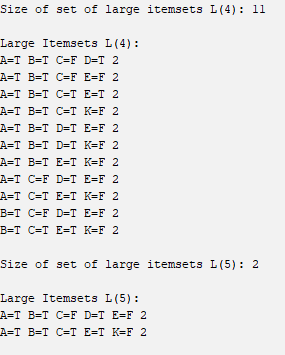
Exercise 3: The 'database' below has four transactions:

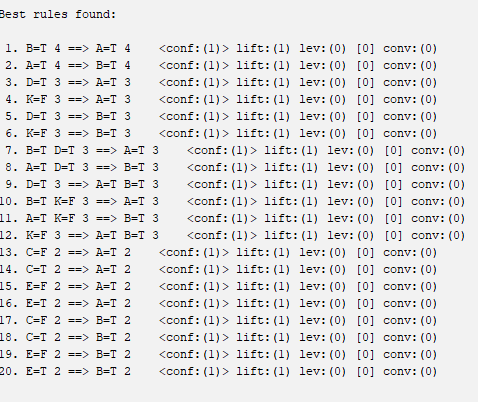
a. Use manually the Apriori algorithm to find all frequent itemsets and all association rules with

minimum support is 60% and the minimum confidence is 80%.



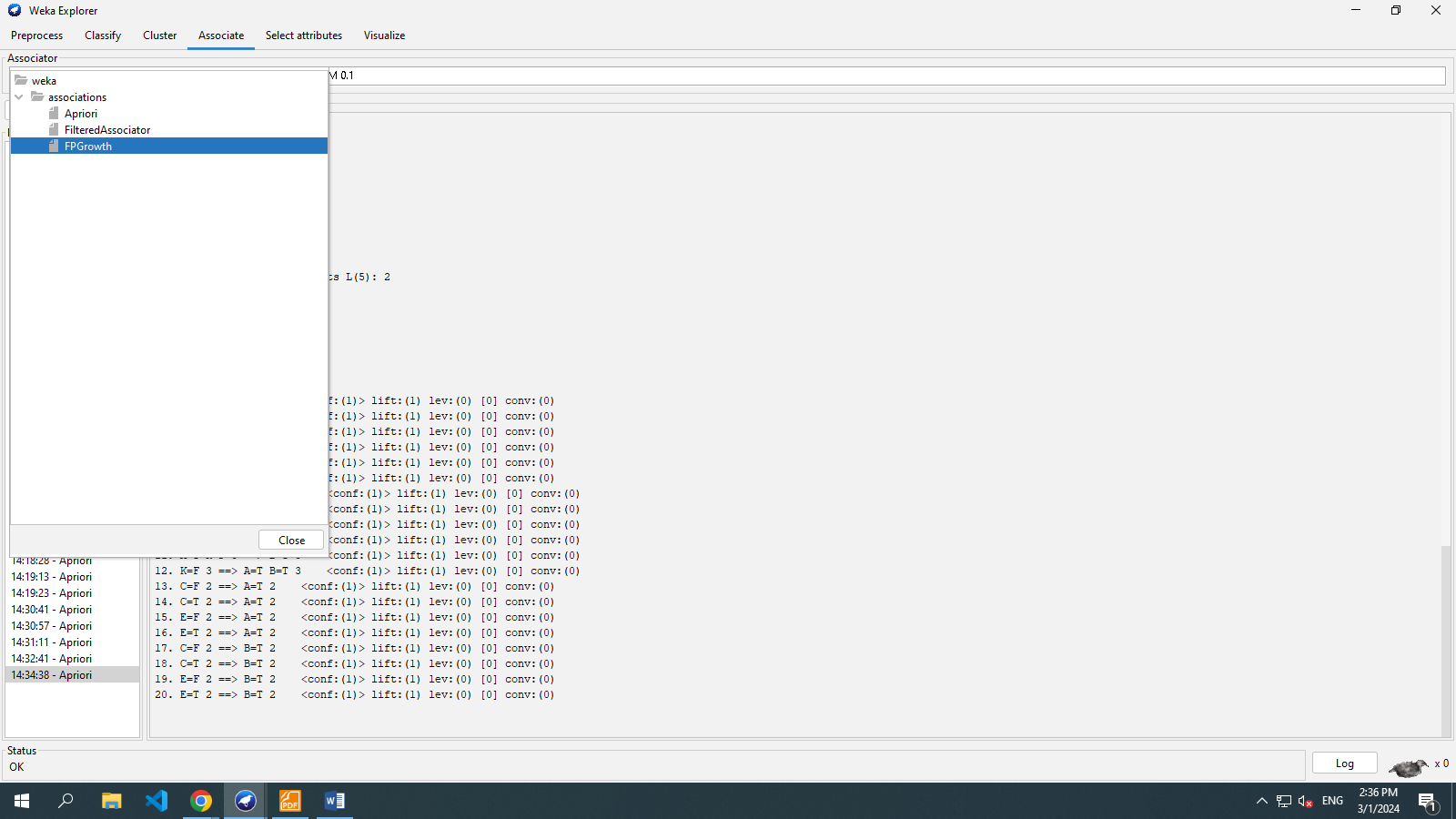






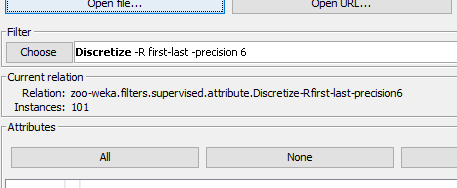
b. What association rules can be found in this set using WEKA, if the minimum support is 60%

and the minimum confidence is 80%.



Exercise 4:

- Download the zoo dataset and load it into Weka.

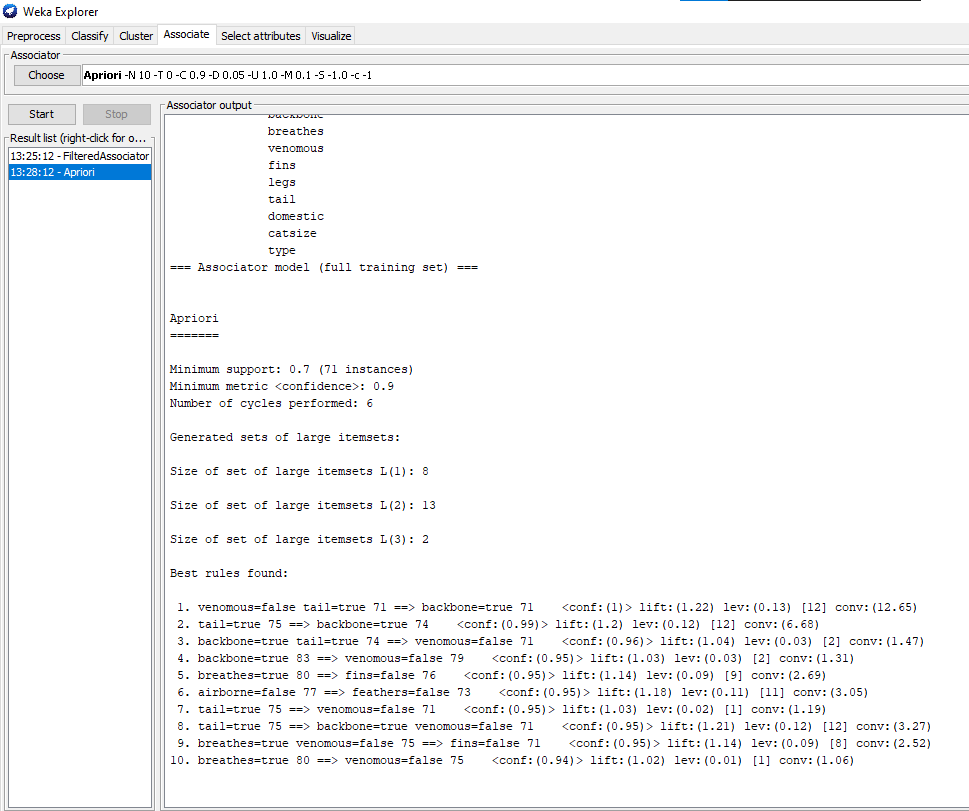


- Examine the attributes. Are all attributes nominal?

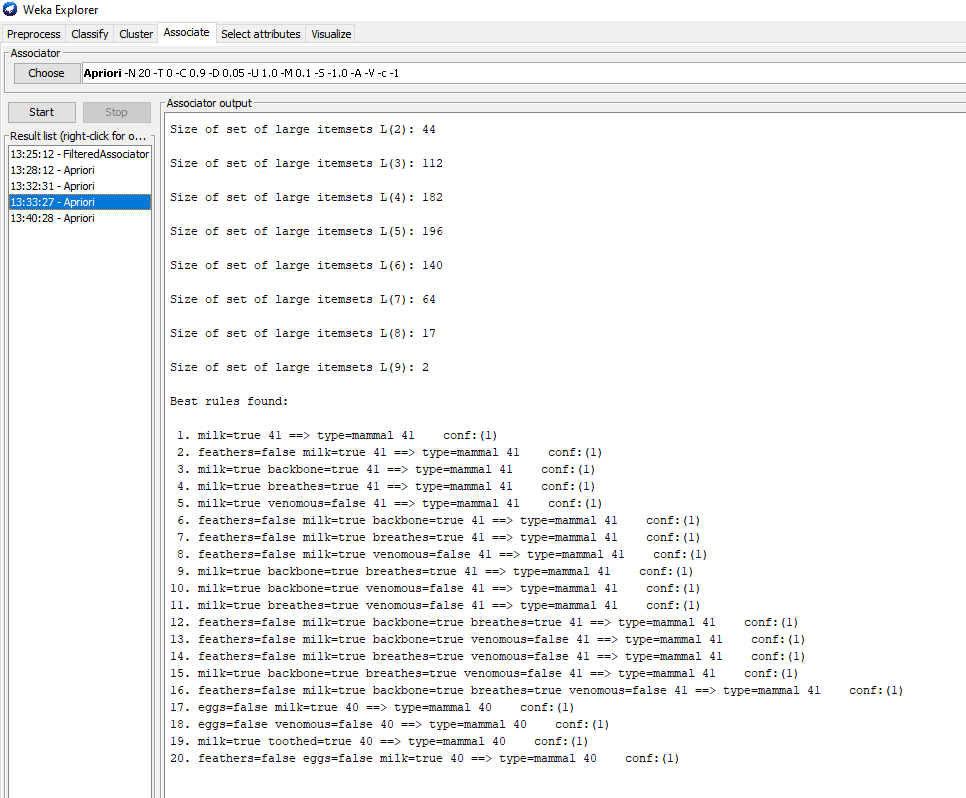
-> All attributes norminal except : legs

- Preprocess data to make sure the dataset can be used directly with Apriori algorithm.

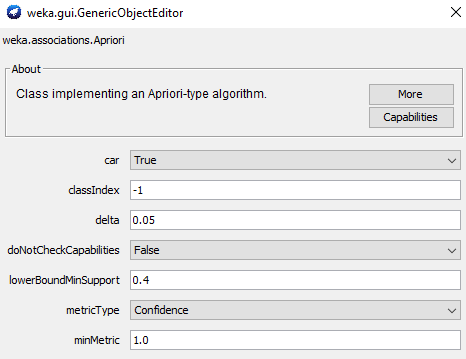
- Use the Apriori algorithm with the default parameters. Record the generated rules.



- Vary the number of rules generated (20, 30, ...). Record how many rules have to be generated before generating a rule containing type=mammal.

******

- Vary the maximum support until a rule containing type=mammal is the top rule generated.Record the maximum support needed.

******

- Select one generated rule that was interesting to you. Why was it interesting? What does it mean?

Check its confidence and support – are they high enough?