



KHAI PHÁ LUẬT APIORI

Giảng viên: Nguyễn Tu Trung
BM HTTT, Khoa CNTT, Trường ĐH Thủy Lợi

Hà Nội, 2021

Nội dung

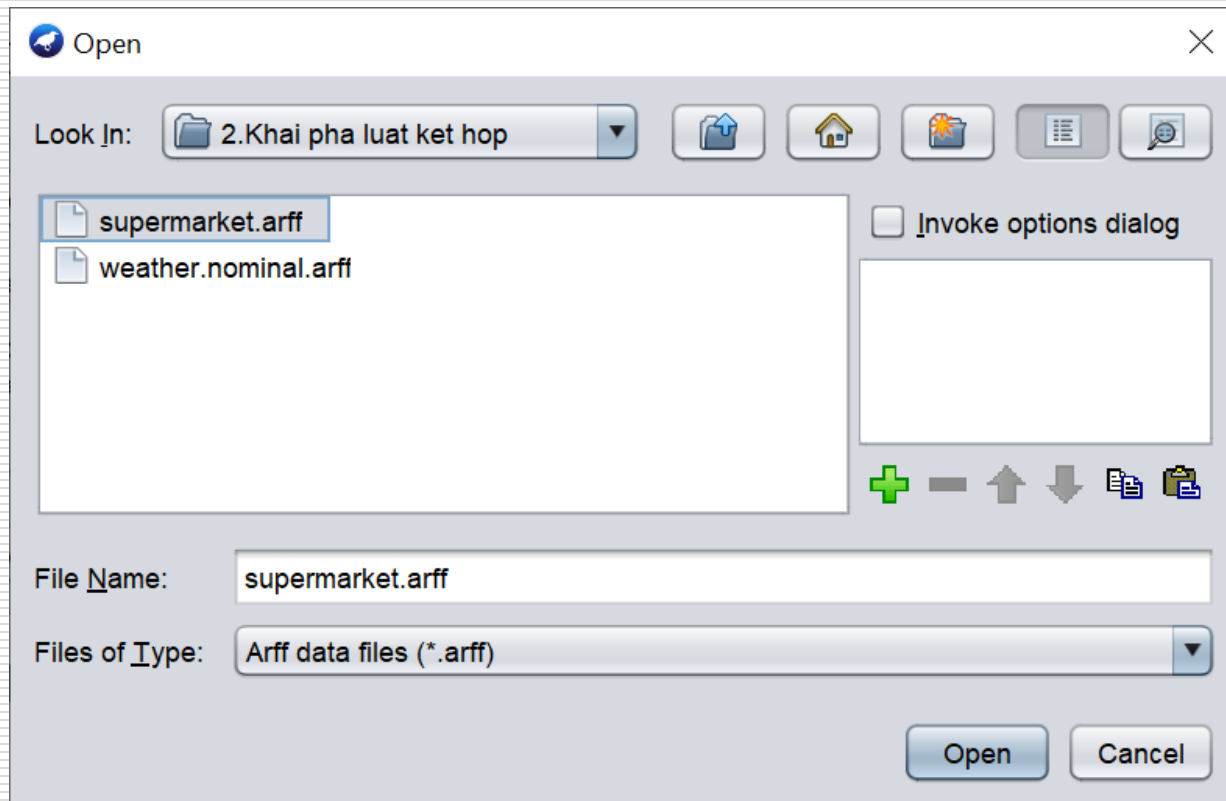
- ❖ Khai phá giỏ hàng
- ❖ Khai phá dữ liệu thời tiết có dữ liệu số

Khai phá giỏ hàng

- ❖ Mở file supermarket.arff
- ❖ Loại bỏ thuộc tính dư thừa
- ❖ Khai phá luật kết hợp

Mở file supermarket.arff

- ❖ Trong Tab Preprocessing, chọn Open File
- ❖ Chọn file supermarket.arff



Loại bỏ thuộc tính dư thừa

- ❖ Tick chọn các trường ko cần thiết và nhấn Remove để loại bỏ:
- ❖ Ví dụ: Các thuộc tính department...

The screenshot shows the Weka Explorer application window. The 'Preprocess' tab is selected. The 'Filter' section shows 'None' is chosen. The 'Current relation' section shows 'Relation: supermarket' and 'Instances: 4627'. The 'Attributes' section shows a list of attributes with checkboxes. The 'Selected attribute' section shows 'Name: department5' and 'Type: Nominal'. The 'Status' section shows 'Remove selected attributes.' and 'OK'.

Current relation

Relation: supermarket
Instances: 4627

Attributes: 217
Sum of weights: 4627

Attributes

No.	Name
1	<input checked="" type="checkbox"/> department1
2	<input checked="" type="checkbox"/> department2
3	<input checked="" type="checkbox"/> department3
4	<input type="checkbox"/> department4
5	<input checked="" type="checkbox"/> department5
6	<input type="checkbox"/> department6
7	<input type="checkbox"/> department7

Selected attribute

Name: department5
Missing: 4452 (96%)
Distinct: 1
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	t	175	175.0

Status

Remove selected attributes.

OK

Loại bỏ thuộc tính dư thừa

❖ Kết quả:

The screenshot shows the Weka Explorer application window. The 'Selected attribute' section displays the following information:

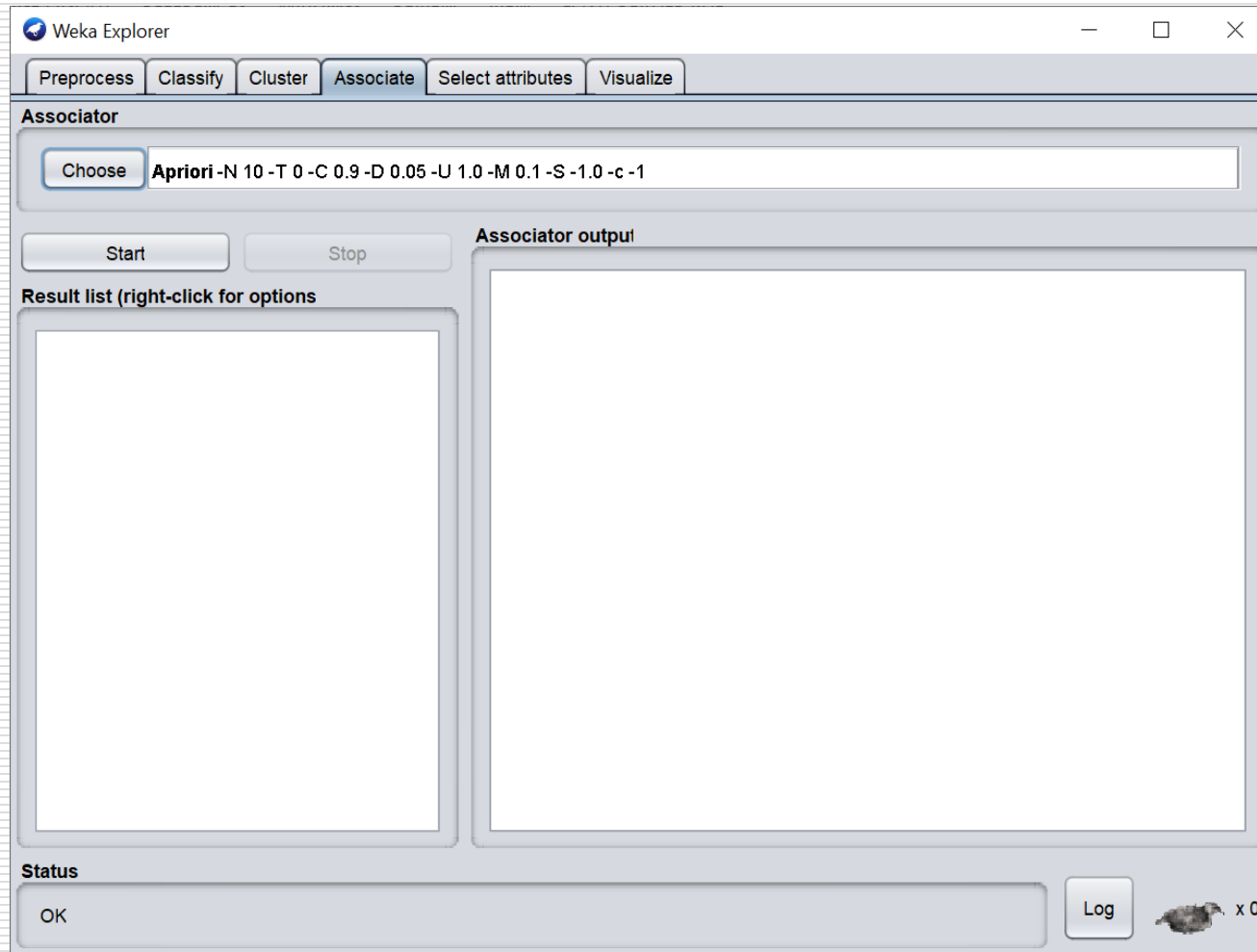
- Name: grocery misc
- Missing: 4449 (96%)
- Distinct: 1
- Type: Nominal
- Unique: 0 (0%)

No.	Label	Count	Weight
1	t	178	178.0

Below the table, the 'Class: total (Nom)' dropdown is set to 'total (Nom)', and the 'Visualize All' button is visible. The visualization area shows a horizontal bar chart with a red bar (top) and a blue bar (bottom), representing the distribution of the selected attribute.

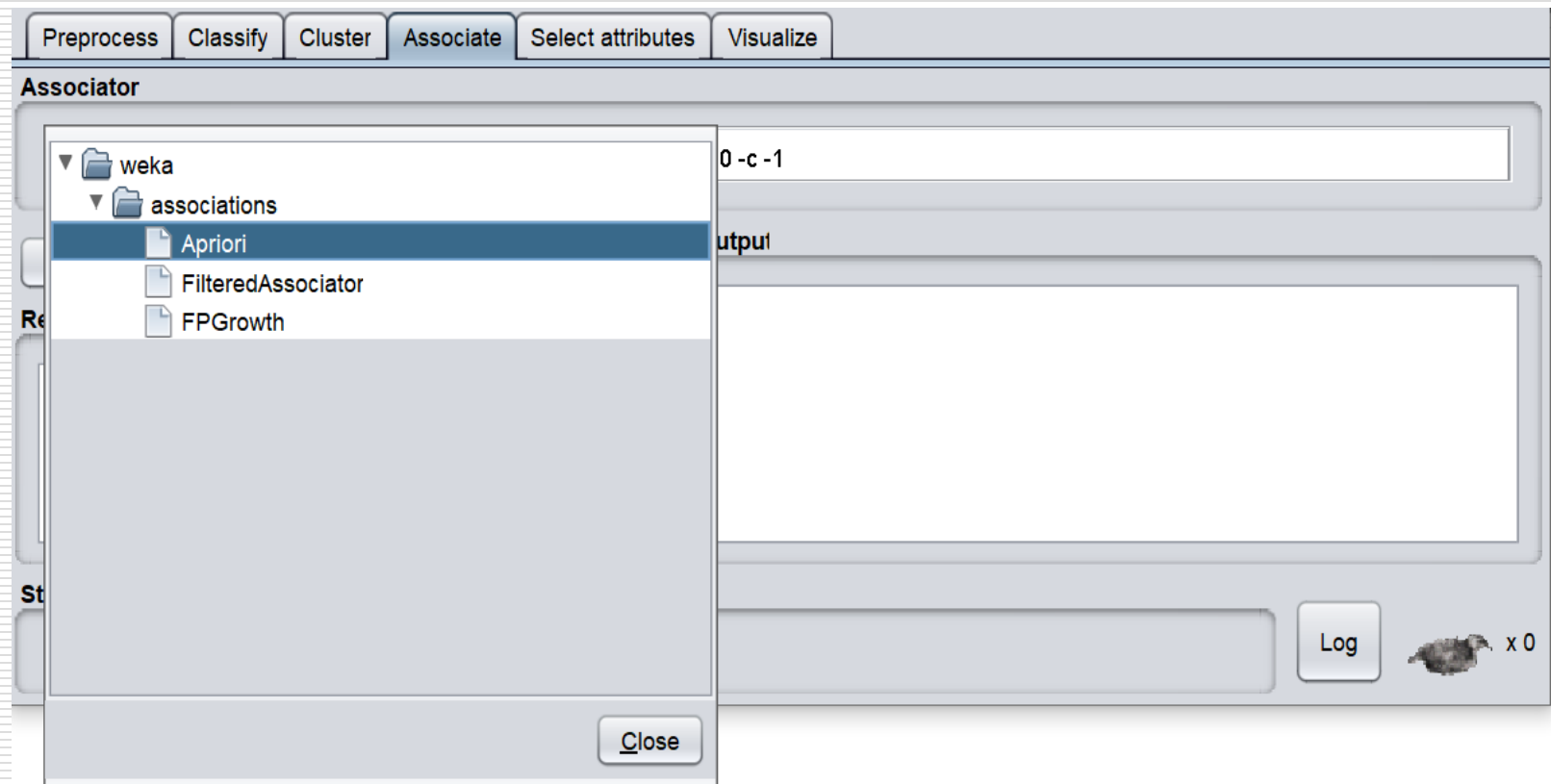
Khai phá luật kết hợp Apriori

❖ Chọn tab Associate



Khai phá luật kết hợp Apriori

- ❖ Nhấn nút Choose để chọn thuật toán
- ❖ Chọn thuật toán Apriori



Khai phá luật kết hợp Apriori

- ❖ Nhấn chuột trái vào Hộp thông tin thuật toán bên cạnh để mở cửa sổ cấu hình thuật toán
- ❖ Thiết lập tham số minSup, minCon, số luật tối đa numRules, nhấn OK để đóng cửa sổ

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm. More Capabilities

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.4

metricType Confidence

minMetric 0.7

numRules 20

outputItemSets False

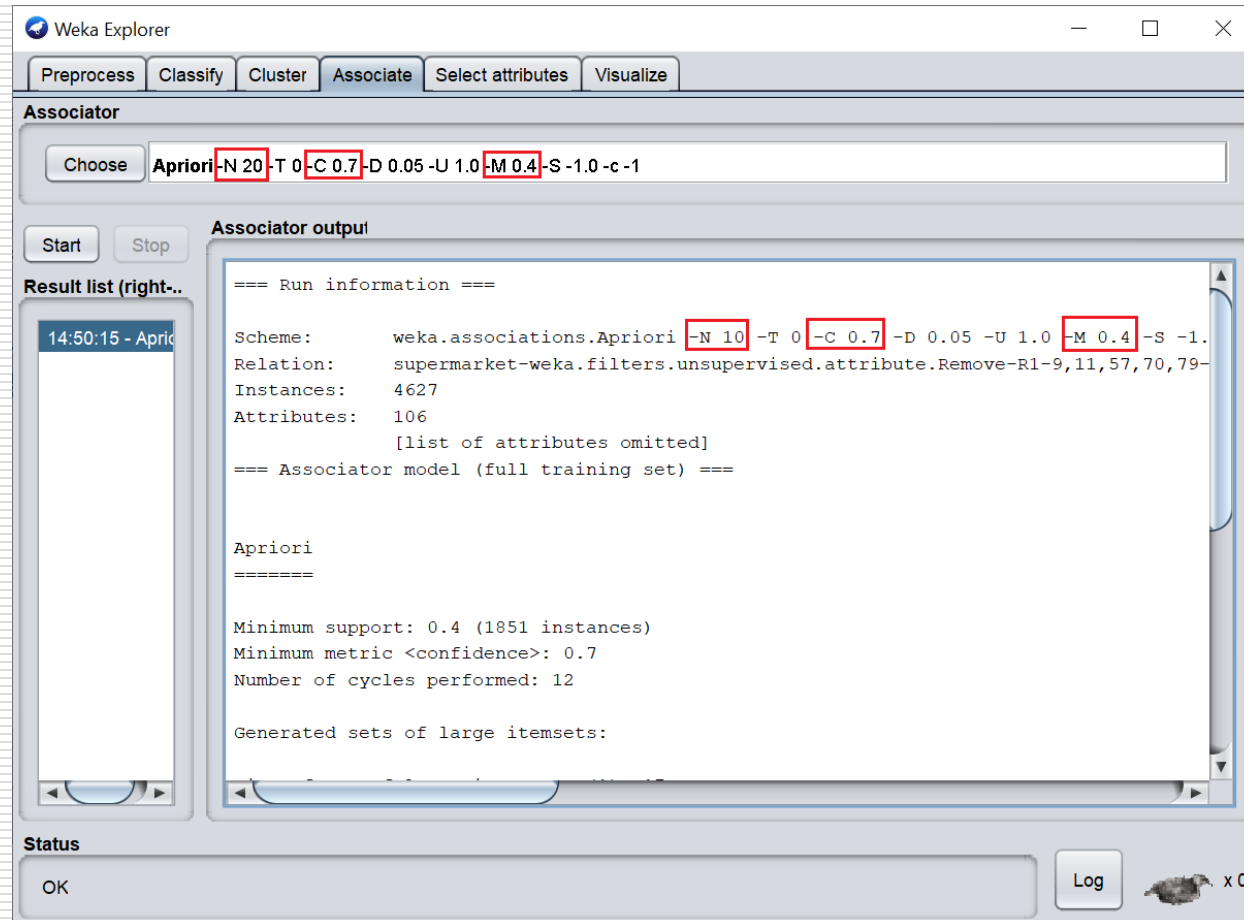
removeAllMissingCols False

significanceLevel -1.0

Open... Save... OK Cancel

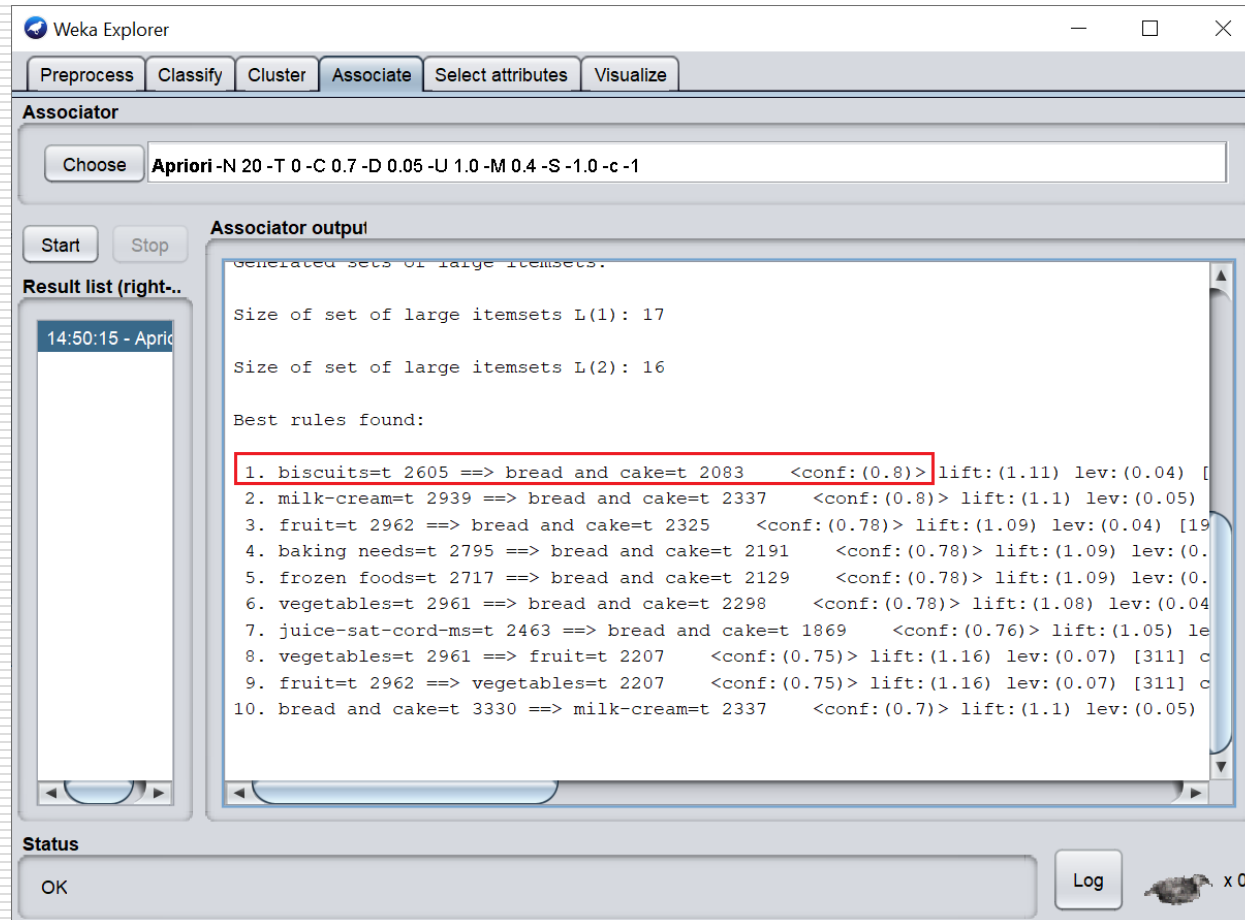
Khai phá luật kết hợp Apriori

❖ Nhấn nút Start và xem kết quả:



Khai phá luật kết hợp Apriori

❖ Kết quả (tiếp):



The screenshot shows the Weka Explorer interface with the 'Associator' tab selected. The 'Associator output' window displays the results of an Apriori association analysis. The output includes the size of the set of large itemsets for L(1) and L(2), and a list of the best rules found. The first rule is highlighted with a red box.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose **Apriori -N 20 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1**

Start Stop

Result list (right-...

14:50:15 - Apriori

Associator output

Generated sets of large itemsets.

Size of set of large itemsets L(1): 17

Size of set of large itemsets L(2): 16

Best rules found:

1. biscuits=t 2605 ==> bread and cake=t 2083 <conf:(0.8)> lift:(1.11) lev:(0.04) [
2. milk-cream=t 2939 ==> bread and cake=t 2337 <conf:(0.8)> lift:(1.1) lev:(0.05)
3. fruit=t 2962 ==> bread and cake=t 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) [19
4. baking needs=t 2795 ==> bread and cake=t 2191 <conf:(0.78)> lift:(1.09) lev:(0.
5. frozen foods=t 2717 ==> bread and cake=t 2129 <conf:(0.78)> lift:(1.09) lev:(0.
6. vegetables=t 2961 ==> bread and cake=t 2298 <conf:(0.78)> lift:(1.08) lev:(0.04
7. juice-sat-cord-ms=t 2463 ==> bread and cake=t 1869 <conf:(0.76)> lift:(1.05) le
8. vegetables=t 2961 ==> fruit=t 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) [311] c
9. fruit=t 2962 ==> vegetables=t 2207 <conf:(0.75)> lift:(1.16) lev:(0.07) [311] c
10. bread and cake=t 3330 ==> milk-cream=t 2337 <conf:(0.7)> lift:(1.1) lev:(0.05)

Status

OK Log x 0

Khai phá luật kết hợp Apriori

❖ Nhận xét:

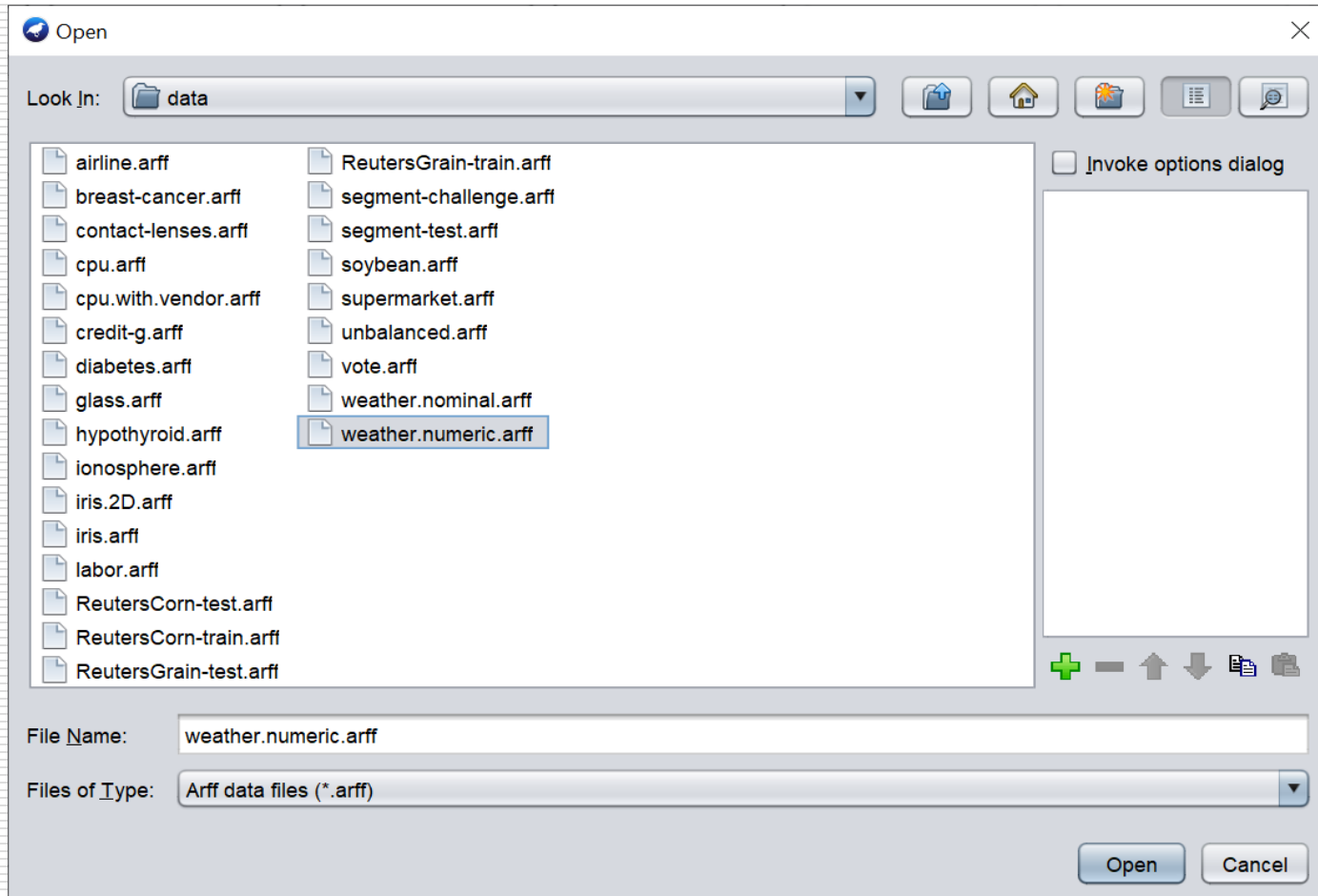
- Các luật sắp xếp theo thứ tự giảm dần của conf
- Trường hợp này chỉ ra được 10 luật thỏa mãn
- Ví dụ luật 1: Nếu mua bánh quy thì sẽ mua thêm bánh mì và bánh ngọt với độ tin cậy là 80%
- Ví dụ luật 2: Nếu mua kem sữa thì sẽ mua thêm bánh mì và bánh ngọt với độ tin cậy là 80%

Khai phá dữ liệu thời tiết có dữ liệu số

- ❖ Mở file weather.numeric.arff
- ❖ Quan sát dữ liệu
- ❖ Chuyển đổi kiểu dữ liệu
- ❖ Khai phá luật kết hợp

Mở file weather.numeric.arff

❖ Mở file weather.numeric.arff



Quan sát dữ liệu

- ❖ 2 thuộc tính kiểu số
- ❖ => Cần chuyển về kiểu Nominal (dạng tên)

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply] [Stop]

Current relation: Relation: weather | Instances: 14 | Attributes: 5 | Sum of weights: 14

Selected attribute: Name: temperature | Missing: 0 (0%) | Distinct: 12 | Type: Numeric | Unique: 10 (71%)

Statistic	Value
Minimum	64
Maximum	85
Mean	73.571
StdDev	6.572

Attributes: [All] [None] [Invert] [Pattern]

No.	Name
1	<input type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

[Remove]

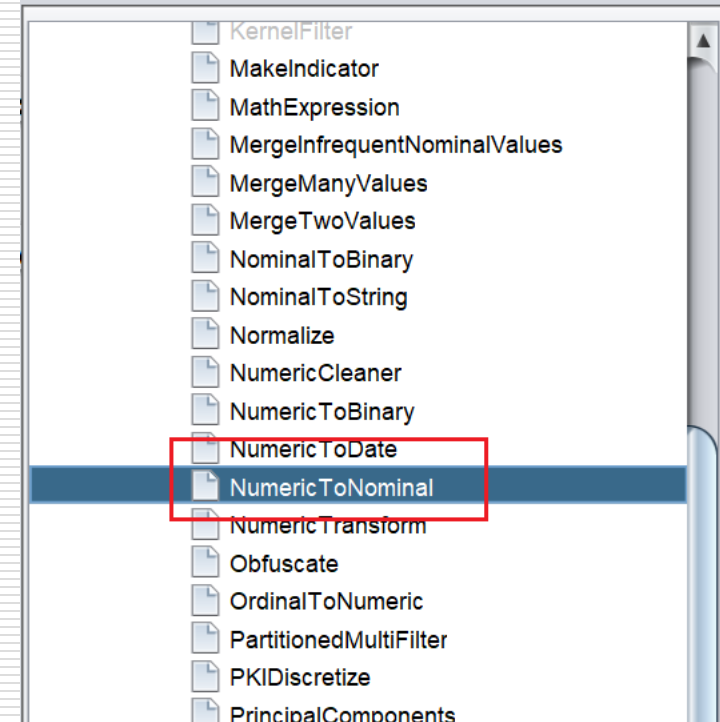
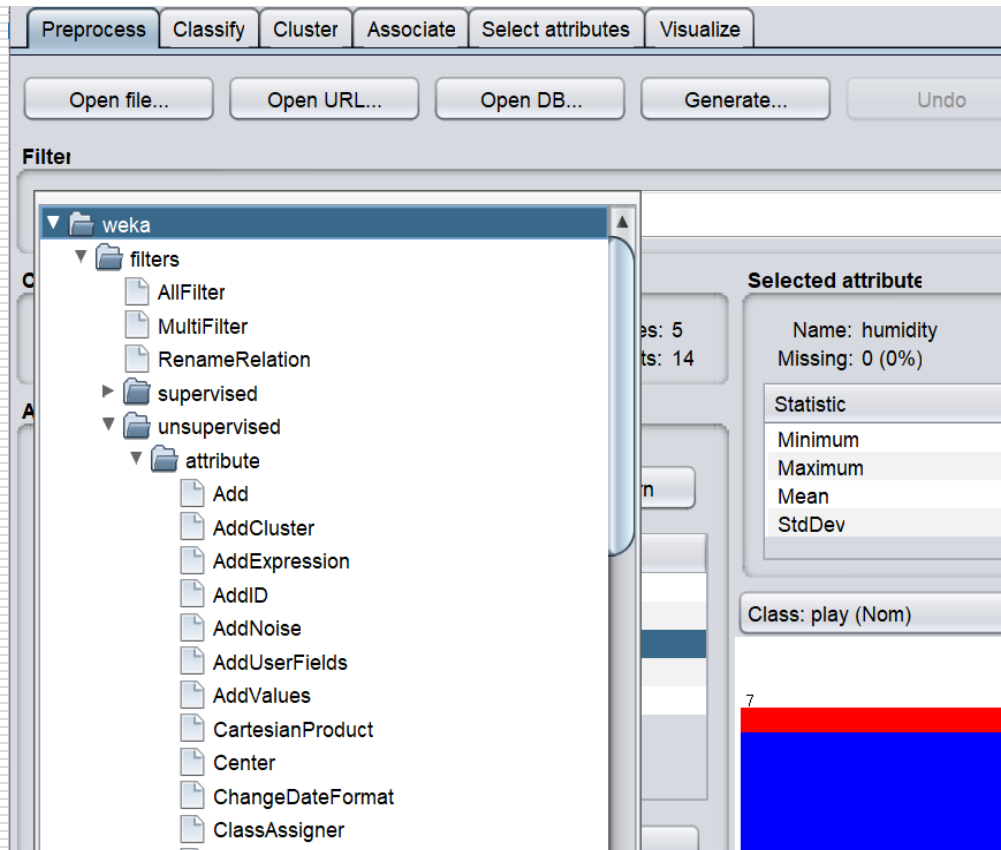
Status: OK

Class: play (Nom) [Visualize All]

Bar chart showing the distribution of 'play' values (Nominal) for different 'temperature' values (Numeric). The x-axis represents temperature (64 to 85), and the y-axis represents the count (0 to 8). The chart shows two bars: a red bar for temperature 64 (count 8) and a blue bar for temperature 85 (count 6).

Chuyển đổi kiểu dữ liệu

- ❖ Nhấn nút Choose -> Weka -> filters -> Unsupervised -> Attribute



Chuyển đổi kiểu dữ liệu

❖ Chọn 2 thuộc tính cần chuyển:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose **NumericToNominal -R first-last** [Apply] [Stop]

Current relation

Relation: weather | Instances: 14 | Attributes: 5 | Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.	Name
1	<input type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input checked="" type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

[Remove]

Selected attribute

Name: humidity | Missing: 0 (0%) | Distinct: 10 | Type: Numeric | Unique: 7 (50%)

Statistic	Value
Minimum	65
Maximum	96
Mean	81.643
StdDev	10.285

Class: play (Nom) [Visualize All]

7 7

65 80.5 96

Status: OK [Log] x 0

Chuyển đổi kiểu dữ liệu

❖ Nhấn nút Apply:

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' section shows 'NumericToNominal -R first-last' applied. The 'Current relation' is 'weather-weka.filters.unsupervise...' with 5 attributes and 14 instances. The 'Selected attribute' is 'temperature', which has been converted to 'Nominal' type. The 'Attributes' list on the left shows 'temperature' selected. The 'Class' is 'play (Nom)'. A bar chart at the bottom visualizes the data distribution for the 'play' class.

Filter: Choose **NumericToNominal -R first-last** **Apply** Stop

Current relation: Relation: weather-weka.filters.unsupervise... Attributes: 5
Instances: 14 Sum of weights: 14

Selected attribute: Name: temperature Type: **Nominal**
Missing: 0 (0%) Distinct: 12 Unique: 10 (71%)

No.	Label	Count	Weight
1	64	1	1.0
2	65	1	1.0
3	68	1	1.0
4	69	1	1.0
5	70	1	1.0

Attributes: All None Invert Pattern

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Class: play (Nom) Visualize All

Status: OK Log x 0

Khai phá luật kết hợp

- ❖ Chuyển qua tab Associate:
minsup=0.3,
minConf=0.6,
numRules=20

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm. [More](#) [Capabilities](#)

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.3

metricType Confidence

minMetric 0.6

numRules 20

outputItemSets False

removeAllMissingCols False

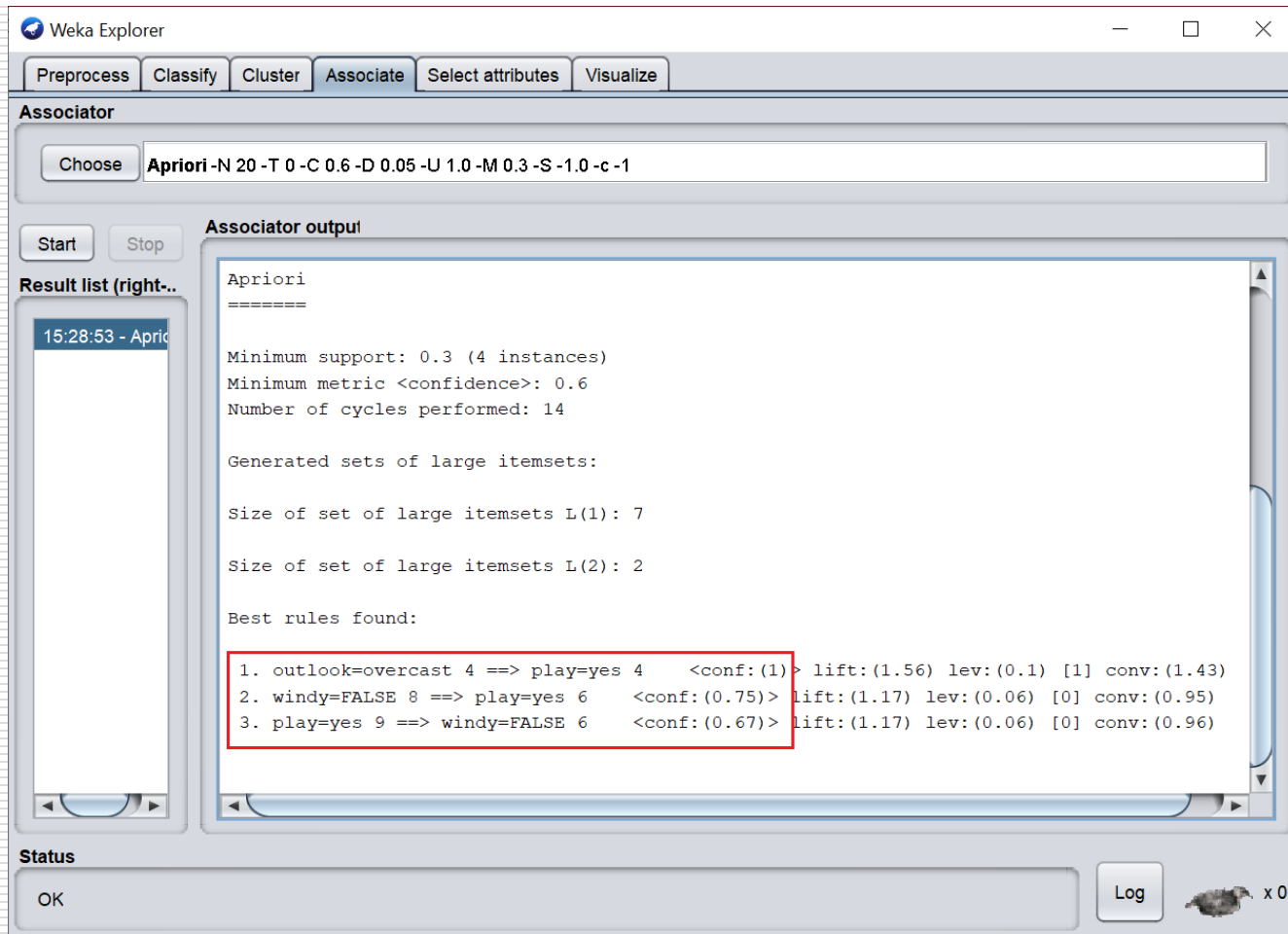
significanceLevel -1.0

treatZeroAsMissing False

Open... Save... OK Cancel

Khai phá luật kết hợp

❖ Kết quả thu được:



The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Associator' dropdown is set to 'Apriori' with parameters '-N 20 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.3 -S -1.0 -c -1'. The 'Associator output' window displays the following text:

```
Apriori
=====
Minimum support: 0.3 (4 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 2

Best rules found:
1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. windy=FALSE 8 ==> play=yes 6    <conf:(0.75)> lift:(1.17) lev:(0.06) [0] conv:(0.95)
3. play=yes 9 ==> windy=FALSE 6    <conf:(0.67)> lift:(1.17) lev:(0.06) [0] conv:(0.96)
```

The third rule is highlighted with a red box. The 'Status' bar at the bottom shows 'OK'.