

**Name:** Soham Belurgikar

**Roll No.:** 2019130006

**Course:** DA (Data Analytics)

**Experiment No.:** 5

**Name of the Experiment:** Apriori Algorithm and Association rule mining with WEKA

**Objective:** Apply Apriori Algorithm to given dataset

Association Rule Mining with WEKA

**Implementation:**

**Task:** <https://colab.research.google.com/drive/1sANAFvw-k3ia1lqGNKGDzb5E7ZDdy0IF?usp=sharing>

**Exercise 1:**

Transaction	A	B	C	D	E	K
T1	1	1	0	1	0	1
T2	1	1	1	1	1	0
T3	1	1	1	0	1	0
T4	1	1	0	1	0	0

Minimum support of 60% => item has to occur in at least 3 transactions

Item sets containing 1 item:

A 4, 100%

B 4, 100%

C 2, 50%

D 3, 75%

E 2, 50%

K 1, 25%

Item sets containing 2 items:

We only take the item sets from the previous phase whose support is 60% or more.

A B 4, 100%

A D 3, 75%

B D 3, 75%

Item sets containing 3 items:

We only take the item sets from the previous phase whose support is 60% or more.

A B D 3

Form the rules and calculate their confidence (c). We only take the item sets from the previous phases whose support is 60% or more.

Rules:

A  $\rightarrow$  B  $P(B|A) = |B \cap A| / |A| = 4/4$ , c: 100%

B  $\rightarrow$  A c: 100%

A  $\rightarrow$  D c: 75%

D  $\rightarrow$  A c: 100%

B  $\rightarrow$  D c: 75%

D  $\rightarrow$  B c: 100%

AB  $\rightarrow$  D c: 75%

D  $\rightarrow$  AB c: 100%

AD  $\rightarrow$  B c: 100%

B  $\rightarrow$  AD c: 75%

BD  $\rightarrow$  A c: 100%

A  $\rightarrow$  BD c: 75%

The rules with a confidence measure of 75% are pruned, and we are left with the following rule set:

A -> B

B -> A

D -> A

D -> B

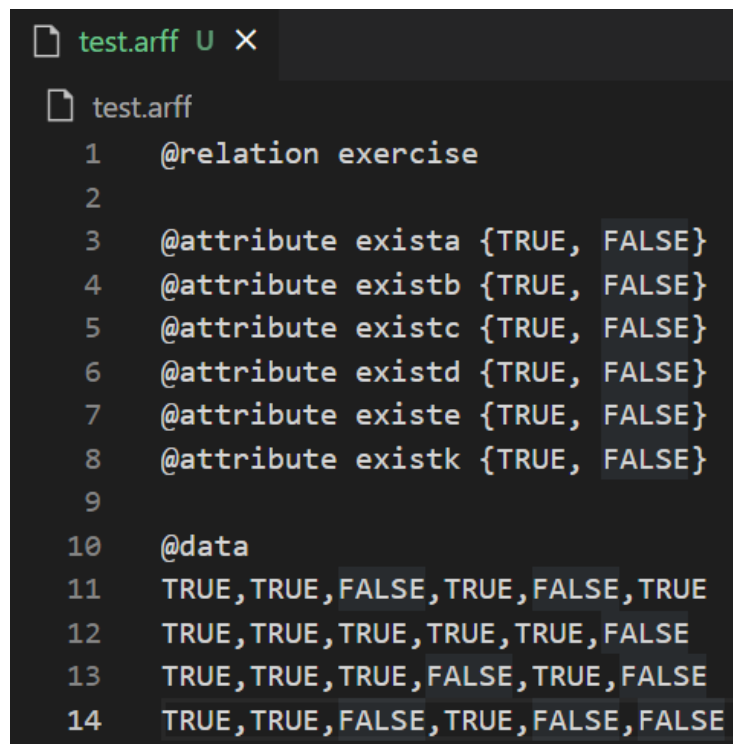
D -> AB

AD-> B

DB-> A


### **Exercise 2:**

Input file: test.arff

A screenshot of a text editor window titled 'test.arff U X'. The editor displays the ARFF file format for a dataset named 'exercise'. It includes attribute declarations for 'exista', 'existb', 'existc', 'existd', 'existe', and 'existk', each with a domain of {TRUE, FALSE}. The data section contains four rows of binary values. The text is as follows:

```
1  @relation exercise
2
3  @attribute exista {TRUE, FALSE}
4  @attribute existb {TRUE, FALSE}
5  @attribute existc {TRUE, FALSE}
6  @attribute existd {TRUE, FALSE}
7  @attribute existe {TRUE, FALSE}
8  @attribute existk {TRUE, FALSE}
9
10 @data
11 TRUE,TRUE,FALSE,TRUE,FALSE,TRUE
12 TRUE,TRUE,TRUE,TRUE,TRUE,FALSE
13 TRUE,TRUE,TRUE,FALSE,TRUE,FALSE
14 TRUE,TRUE,FALSE,TRUE,FALSE,FALSE
```

Weka configuration file:

 weka.gui.GenericObjectEditor ✕

weka.associations.Apriori

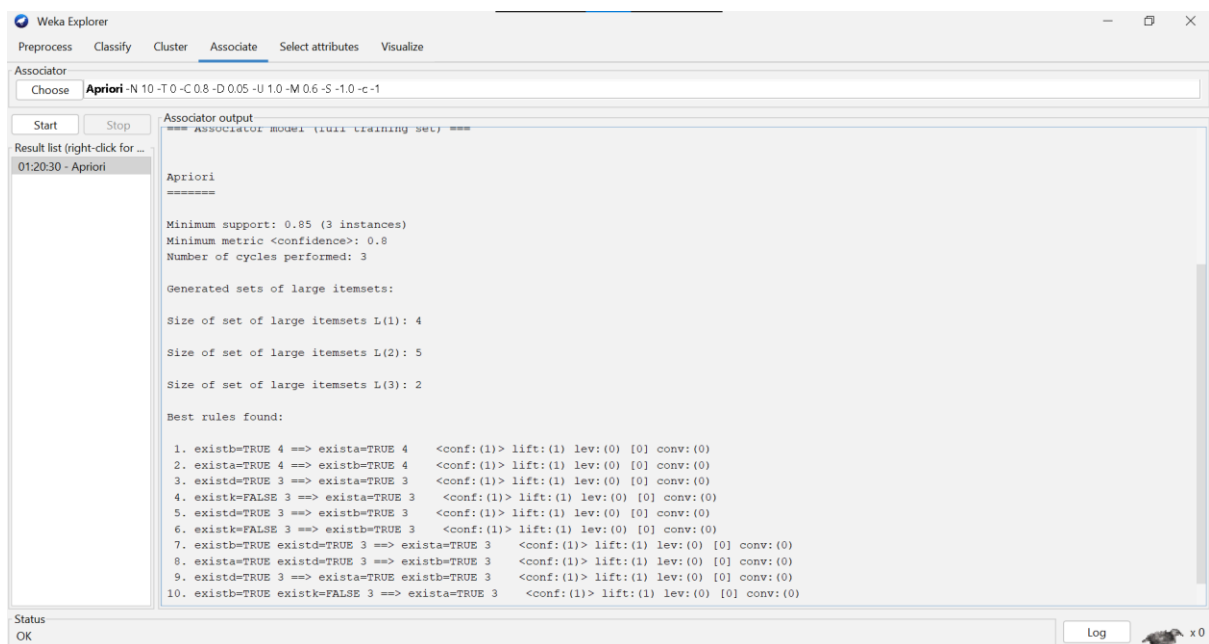
About

Class implementing an Apriori-type algorithm. More  
Capabilities

car	False	▼
classIndex	-1	
delta	0.05	
doNotCheckCapabilities	False	▼
lowerBoundMinSupport	0.6	
metricType	Confidence	▼
minMetric	0.8	
numRules	10	
outputItemSets	False	▼
removeAllMissingCols	False	▼
significanceLevel	-1.0	
treatZeroAsMissing	False	▼
upperBoundMinSupport	1.0	
verbose	False	▼

Open... Save... OK Cancel

## Output:



```
Weka Explorer
Preprocess  Classify  Cluster  Associate  Select attributes  Visualize

Associator
Choose  Apriori -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.6 -S -1.0 -c -1

Start  Stop

Result list (right-click for ...)
01:20:30 - Apriori

Associator output
===== ASSOCIATOR MODEL (full training set) =====

Apriori
=====

Minimum support: 0.85 (3 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 5

Size of set of large itemsets L(3): 2

Best rules found:

1. existb=TRUE 4 ==> exista=TRUE 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. exista=TRUE 4 ==> existb=TRUE 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. existd=TRUE 3 ==> exista=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. existk=FALSE 3 ==> exista=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. existd=TRUE 3 ==> existb=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. existk=FALSE 3 ==> existb=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. existb=TRUE existd=TRUE 3 ==> exista=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. exista=TRUE existd=TRUE 3 ==> existb=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. existd=TRUE 3 ==> exista=TRUE existb=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. existb=TRUE existk=FALSE 3 ==> exista=TRUE 3    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Status
OK
```

This output is consistent with the manual solution in Exercise 1. The file format used was .arff.

## Exercise 3:



```
Weka Explorer
Preprocess  Classify  Cluster  Associate  Select attributes  Visualize

Associator
Choose  Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start  Stop

Result list (right-click for ...)
20:16:52 - Apriori

Associator output
===== ASSOCIATOR MODEL (full training set) =====

Apriori
=====

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3    <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3    <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3    <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
8. temperature=cool play=yes 3 ==> humidity=normal 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2    <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)

Status
OK
```

## Exercise 4:

No.	1: handicapped-infants Nominal	2: water-project-cost-sharing Nominal	3: adoption-of-the-budget-resolution Nominal	4: physician-fee-freeze Nominal	5: el-salvador-aid Nominal	6: religious-groups-in-schools Nominal	7: ant
1	n	y	n	y	y	y	n
2	n	y	n	y	y	y	n
3		y	y		y	y	n
4	n	y	y	n		y	n
5	y	y	y	n	y	y	n
6	n	y	y	n	y	y	n
7	n	y	n	y	y	y	n
8	n	y	n	y	y	y	n
9	n	y	n	y	y	y	n
10	y	y	y	n	n	n	y
11	n	y	n	y	y	n	n
12	n	y	n	y	y	y	n
13	n	y	y	n	n	n	y
14	y	y	y	n	n	y	y
15	n	y	n	y	y	y	n
16	n	y	n	y	y	y	n
17	y	n	y	n	n	y	n
18	y		y	n	n	n	y
19	n	y	n	y	y	y	n
20	y	y	y	n	n	n	y
21	y	y	y	n	n		y
22	y	y	y	n	n	n	y
23	y		y	n	n	n	y

**Weka Explorer**

Preprocess   Classify   Cluster   **Associate**   Select attributes   Visualize

Associator: Choose **Apriori** -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c 1

Start   Stop

Result list (right-click for ...)

- 20:16:52 - Apriori
- 20:20:09 - Apriori

Associator output

Minimum support: 0.45 (196 instances)  
Minimum metric <confidence>: 0.9  
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20  
Size of set of large itemsets L(2): 17  
Size of set of large itemsets L(3): 6  
Size of set of large itemsets L(4): 1

Best rules found:

- adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 <conf: (1)> lift: (1.63) lev: (0.19) [84] conv: (84.58)
- adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198 <conf: (1)> lift: (1.63) lev: (0.19) [80] conv: (40.74)
- physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 <conf: (1)> lift: (1.62) lev: (0.19) [80] conv: (40.74)
- physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 <conf: (1)> lift: (1.62) lev: (0.18) [77] conv: (39.01)
- physician-fee-freeze=n 247 ==> Class=democrat 245 <conf: (0.99)> lift: (1.62) lev: (0.21) [93] conv: (31.8)
- el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 <conf: (0.98)> lift: (1.77) lev: (0.2) [85] conv: (22.18)
- el-salvador-aid=n ==> aid-to-nicaraguan-contras=y 204 <conf: (0.98)> lift: (1.76) lev: (0.2) [88] conv: (18.46)
- adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198 <conf: (0.98)> lift: (1.72) lev: (0.17) [71] conv: (9.85)
- el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 <conf: (0.97)> lift: (1.57) lev: (0.17) [71] conv: (9.85)
- aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210 <conf: (0.96)> lift: (1.7) lev: (0.2) [86] conv: (10.47)

Status: OK

Log

## Exercise 5:

The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Apriori' algorithm is chosen with default parameters: -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1. The 'Start' button has been clicked, and the 'Associator output' pane displays the results.

Associator output

```
number of cycles performed: 17  
  
Generated sets of large itemsets:  
  
Size of set of large itemsets L(1): 44  
  
Size of set of large itemsets L(2): 380  
  
Size of set of large itemsets L(3): 910  
  
Size of set of large itemsets L(4): 633  
  
Size of set of large itemsets L(5): 105  
  
Size of set of large itemsets L(6): 1  
  
Best rules found:  
  
1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf: (0.92)> lift: (1.27) lev: (0.03) [155] conv: (3.35)  
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf: (0.92)> lift: (1.27) lev: (0.03) [149] conv: (3.28)  
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf: (0.92)> lift: (1.27) lev: (0.03) [150] conv: (3.27)  
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf: (0.92)> lift: (1.27) lev: (0.03) [159] conv: (3.26)  
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf: (0.91)> lift: (1.27) lev: (0.04) [164] conv: (3.15)  
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf: (0.91)> lift: (1.26) lev: (0.03) [151] conv: (3.06)  
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf: (0.91)> lift: (1.26) lev: (0.03) [145] conv: (3.01)  
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf: (0.91)> lift: (1.26) lev: (0.04) [179] conv: (3)  
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf: (0.91)> lift: (1.26) lev: (0.03) [156] conv: (3)  
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf: (0.91)> lift: (1.26) lev: (0.04) [179] conv: (2.92)
```

## Conclusion:

- Because it relies on prior knowledge about frequent itemset properties, the algorithm is called Apriori. We employ an iterative strategy, often known as level-wise search, in which we use k-frequent itemsets to find k+1 itemsets.
- Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction.
- In exercise 4, most of the generated rules have similar confidence and lift levels. None of the rules in the default output involve Class = republican. This is because of the lack of significant number of members belonging to republican party in the chosen dataset.