

DAY-5 LAB EXPERIMENTS

R PROGRAMMING

EXPERIMENT 1:

Consider the data set and perform the Apriori Algorithm and FP algorithm support:3 and confidence=50%

Customer ID	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

AIM:

To create dataset in the notepad and open the file in weka tool.

MATERIALS REQUIRED:

WEKA TOOL

DATASET:

@relation items

@attribute a{true,false}

@attribute b{true,false}

@attribute c{true,false}

@attribute d{true,false}

@attribute e{true,false}

@data

true false false true true

true true true false true

true true false true true

true false true true true

false true true false true

false true false true true

false false true true false

true true true false false

true false false true true

true true false false true

→ Save this file as **x.arff** in the file explorer path.

OUTPUT:

```
Apriori
=====

Minimum support: 0.35 (3 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 13

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 19
Size of set of large itemsets L(3): 12
Size of set of large itemsets L(4): 2

Best rules found:

1. c=false 5 ==> e=true 5    <conf:(1)> lift:(1.25) lev:(0.1) [0] conv:(1)
2. d=false 4 ==> b=true 4    <conf:(1)> lift:(1.67) lev:(0.16) [1] conv:(1.6)
3. b=false 4 ==> d=true 4    <conf:(1)> lift:(1.67) lev:(0.16) [1] conv:(1.6)
4. a=true c=false 4 ==> e=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
5. a=true d=true 4 ==> e=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
6. c=false d=true 4 ==> e=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
7. a=true d=false 3 ==> b=true 3    <conf:(1)> lift:(1.67) lev:(0.12) [1] conv:(1.2)
8. a=true b=false 3 ==> d=true 3    <conf:(1)> lift:(1.67) lev:(0.12) [1] conv:(1.2)
9. b=false e=true 3 ==> a=true 3    <conf:(1)> lift:(1.43) lev:(0.09) [0] conv:(0.9)
10. a=true b=false 3 ==> e=true 3    <conf:(1)> lift:(1.25) lev:(0.06) [0] conv:(0.6)
```

```

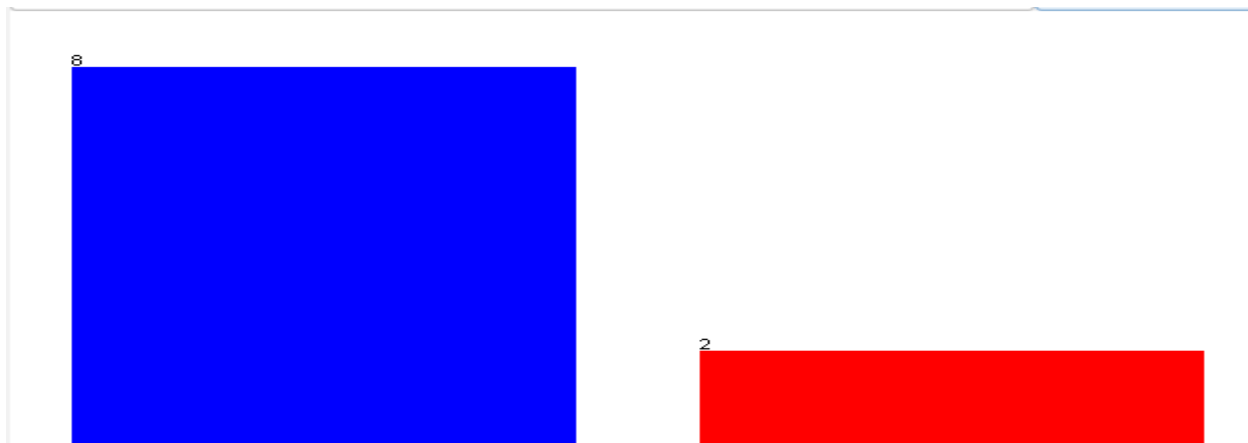
Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:    items
Instances:   10
Attributes:  5
             a
             b
             c
             d
             e

=== Associator model (full training set) ===

FPGrowth found 3 rules (displaying top 3)

1. [b=false, a=false]: 1 ==> [e=false]: 1 <conf:(1)> lift:(5) lev:(0.08) conv:(0.8)
2. [b=false, e=false]: 1 ==> [a=false]: 1 <conf:(1)> lift:(3.33) lev:(0.07) conv:(0.7)
3. [a=false, e=false]: 1 ==> [b=false]: 1 <conf:(1)> lift:(2.5) lev:(0.06) conv:(0.6)

```



EXPERIMENT 2:

Consider the data set and perform the Apriori Algorithm and FP algorithm support:3 and confidence=50%

Consider the market basket transactions shown in the above table.

(a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

(b) What is the maximum size of frequent itemsets that can be extracted
(assuming minsup > 0)?

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

AIM:

To create a dataset in the notepad

MATERIALS REQUIRED:

WEKA TOOL

DATASET:

@relation transaction_data

@attribute Milk {true, false}

@attribute Bread {true, false}

@attribute Butter {true, false}

@attribute Cookies {true, false}

@attribute Beer {true, false}

@attribute Diapers {true, false}

@data

true, true, true, true, true, true

true, true, true, false, true, false

true, false, true, false, true, true
true, true, false, true, false, true
false, false, true, true, true, true
true, false, false, true, true, true
false, true, true, true, false, true
false, true, true, false, true, true
true, true, true, false, true, false
false, true, true, true, true, true

→ Save this file as **x.arff** in the file explorer path.

OUTPUT:

```
Apriori
=====

Minimum support: 0.45 (4 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

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

Best rules found:

1. Cookies=true 6 ==> Diapers=true 6    <conf:(1)> lift:(1.25) lev:(0.12) [1] conv:(1.2)
2. Bread=true Beer=true 5 ==> Butter=true 5    <conf:(1)> lift:(1.25) lev:(0.1) [0] conv:(1)
3. Milk=false 4 ==> Butter=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
4. Milk=false 4 ==> Diapers=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
5. Cookies=false 4 ==> Butter=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
6. Cookies=false 4 ==> Beer=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
7. Milk=true Butter=true 4 ==> Beer=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
8. Milk=false Diapers=true 4 ==> Butter=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
9. Milk=false Butter=true 4 ==> Diapers=true 4    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.8)
10. Milk=false 4 ==> Butter=true Diapers=true 4    <conf:(1)> lift:(1.67) lev:(0.16) [1] conv:(1.6)
```

```

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:    transaction_data
Instances:   10
Attributes:  6
             Milk
             Bread
             Butter
             Cookies
             Beer
             Diapers
=== Associator model (full training set) ===

FPGrowth found 1 rules (displaying top 1)

1. [Diapers=false]: 2 ==> [Cookies=false]: 2   <conf:(1)> lift:(2.5) lev:(0.12) conv:(1.2)

```

EXPERIMENT 3:

Bayes classification and descion tree (using training and test data)

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 ... 40	medium	no	excellent	yes
13	31 ... 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

AIM:

To create a dataset or open the dataset in the inbuilt data

MATERIALS REQUIRED:

WEKA TOOL

- CREATE A DATASET IN THE EXCEL.
- AND OPEN THE DATASET IN THE WEKA

OUTPUT:

```
Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      177           76.9565 %
Incorrectly Classified Instances    53           23.0435 %
Kappa statistic                    0.4581
Mean absolute error                 0.2677
Root mean squared error             0.3863
Relative absolute error             59.5622 %
Root relative squared error         82.7801 %
Total Number of Instances          230

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.842    0.389    0.826     0.842    0.834      0.458    0.845     0.912     tested_n
          0.611    0.158    0.638     0.611    0.624      0.458    0.845     0.752     tested_p
Weighted Avg.   0.770    0.317    0.767     0.770    0.768      0.458    0.845     0.862

=== Confusion Matrix ===

  a    b  <-- classified as
133  25 |  a = tested_negative
 28  44 |  b = tested_positive
```

EXPERIMENT 4:

Implement using WEKA for the given Suppose a database has five_transactions.
Let min sup= 50%(2) and min con f = 80%.

Transactions	Items
T1	(M, O, N, K, E, Y)
T2	(D, O, N, K, E, Y)
T3	(M, A, K, E)
T4	(M, U, C, K, Y)
T5	(C,O, O, K, I ,E)

- Find all frequent item sets using Apriori algorithm
- Also draw FP-Growth Tree

Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic

AIM:

To create a dataset in the notepad

MATERIALS REQUIRED:

WEKA TOOL

DATASET:

@relation transaction_data

@attribute M {true, false}

@attribute O {true, false}

@attribute N {true, false}

@attribute K {true, false}

@attribute E {true, false}

@attribute Y {true, false}

@attribute D {true, false}

@attribute A {true, false}

@attribute U {true, false}

@attribute C {true, false}

@attribute I {true, false}

@data

true, true, true, true, true, true, false, false, false, false, false

false, true, true, true, true, true, true, false, false, false, false

true, false, true, false, true, false, false, true, false, false, false

true, false, false, true, true, false, false, false, true, true, false

false, true, false, true, true, true, false, false, false, true, true

→ Save this file as **x.arff** in the file explorer path.

OUTPUT:

```
Apriori
=====
Minimum support: 0.85 (4 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3

Generated sets of large itemsets:

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

Best rules found:

1. K=true 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. A=false 4 ==> K=true 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
3. K=true 4 ==> A=false 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
4. D=false 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. A=false 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. U=false 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. I=false 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. E=true A=false 4 ==> K=true 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
9. K=true A=false 4 ==> E=true 4    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. K=true E=true 4 ==> A=false 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
```

```

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:    transaction_data
Instances:    5
Attributes:   11
              M
              O
              N
              K
              E
              Y
              D
              A
              U
              C
              I

=== Associator model (full training set) ===

FPGrowth found 61 rules (displaying top 10)

1. [C=false]: 3 ==> [U=false]: 3   <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
2. [M=false]: 2 ==> [U=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
3. [C=false]: 3 ==> [I=false]: 3   <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
4. [Y=false]: 2 ==> [I=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
5. [O=false]: 2 ==> [I=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
6. [Y=false]: 2 ==> [D=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
7. [O=false]: 2 ==> [D=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
8. [N=false]: 2 ==> [D=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
9. [N=false]: 2 ==> [A=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
10. [M=false]: 2 ==> [A=false]: 2   <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)

```

EXPERIMENT 5:

Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

Create the dataset using ARFF file format:

a. Find the **frequent itemsets** and generate **association rules** on this. Assume that minimum support threshold ($s = 33.33\%$) and minimum confidence threshold ($c = 60\%$).

b. List the various rules generated by apriori and FP tree algorithm, mention whether accepted or rejected.

Prediction of Categorical Data using Rule base classification and decision tree classification through WEKA using any datasets. Compare the accuracy using two algorithms and plot the graph

AIM:

To create a data set in the notepad

MATERIALS REQUIRED:

WEKA TOOL

DATASET:

@relation transaction_data

@attribute Hot_Dogs {true, false}

@attribute Buns {true, false}

@attribute Ketchup {true, false}

@attribute Coke {true, false}

@attribute Chips {true, false}

@data

true, true, true, false, false

true, true, false, false, false

true, false, true, true, true

false, false, false, true, true

false, false, true, false, true

true, false, false, true, true

→ Save this file as **x.arff** in the file explorer path.

OUTPUT:

```
Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      2          100   %
Incorrectly Classified Instances    0           0   %
Kappa statistic                     1
Mean absolute error                 0.2381
Root mean squared error             0.2564
Relative absolute error             47.619 %
Root relative squared error         51.2873 %
Total Number of Instances          2

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          1.000    ?        1.000     1.000    1.000     ?        ?         1.000    true
          ?        0.000    ?         ?         ?         ?        ?         ?        false
Weighted Avg.   1.000    ?        1.000     1.000    1.000     ?        ?         1.000

=== Confusion Matrix ===

a b  <-- classified as
2 0 | a = true
0 0 | b = false
```

