

DAY-3

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}

INPUT

@relation dataset

@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

OUTPUT

The top screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'FPGrowth' algorithm is chosen, and the 'Run' button has been clicked. The output window displays the following information:

```

=== Run information ===
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:    dataset
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 ==> [a=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)

```

The bottom screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Apriori' algorithm is chosen, and the 'Run' button has been clicked. The output window displays the following information:

```

=== Run information ===
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    dataset
Instances:   10
Attributes:  5
             a
             b
             c
             d
             e

=== Associator model (full training set) ===
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  a=true 3 ==> e=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)

```

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}

INPUT
APRIOR

```

Associator
Choose Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.3 -S -1.0 -c -1

Start Stop
Result list (right-click for ...)
10:05:17 - Apriori

Associator output
=====
Minimum support: 0.55 (5 instances)
Minimum matrix confidence: 0.5
Number of cycles performed: 9

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5
Size of set of large itemsets L(2): 5
Size of set of large itemsets L(3): 1

Best rules found:
1. bread=T 5 ==> beer=F 5 <conf: (1)> lift: (1.67) lev: (0.2) [2] conv: (2)
2. butter=T 5 ==> beer=F 5 <conf: (1)> lift: (1.67) lev: (0.2) [2] conv: (2)
3. butter=T 5 ==> bread=T 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
4. bread=T 5 ==> butter=T 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
5. butter=F 5 ==> bread=F 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
6. bread=F 5 ==> butter=F 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
7. bread=T butter=T 5 ==> beer=F 5 <conf: (1)> lift: (1.67) lev: (0.2) [2] conv: (2)
8. beer=F butter=T 5 ==> bread=T 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
9. beer=F bread=T 5 ==> butter=T 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)
10. butter=T 5 ==> beer=F bread=T 5 <conf: (1)> lift: (2) lev: (0.25) [2] conv: (2.5)

```

FP GROWTH

```

Associator
Choose FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.3

Start Stop
Result list (right-click for ...)
10:05:17 - Apriori
10:06:24 - FPGrowth

Associator output
===== Run information =====

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.3
Relation:    items
Instances:   10
Attributes:  6
             milk
             beer
             diapers
             bread
             butter
             cookies

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

FPGrowth found 4 rules (displaying top 4)

1. [butter=F]: 5 ==> [bread=F]: 5 <conf: (1)> lift: (2) lev: (0.25) conv: (2.5)
2. [bread=F]: 5 ==> [butter=F]: 5 <conf: (1)> lift: (2) lev: (0.25) conv: (2.5)
3. [milk=F, butter=F]: 3 ==> [bread=F]: 3 <conf: (1)> lift: (2) lev: (0.15) conv: (1.5)
4. [milk=F, bread=F]: 3 ==> [butter=F]: 3 <conf: (1)> lift: (2) lev: (0.15) conv: (1.5)

```

3. Bayes classification and decision tree (using training and test data)

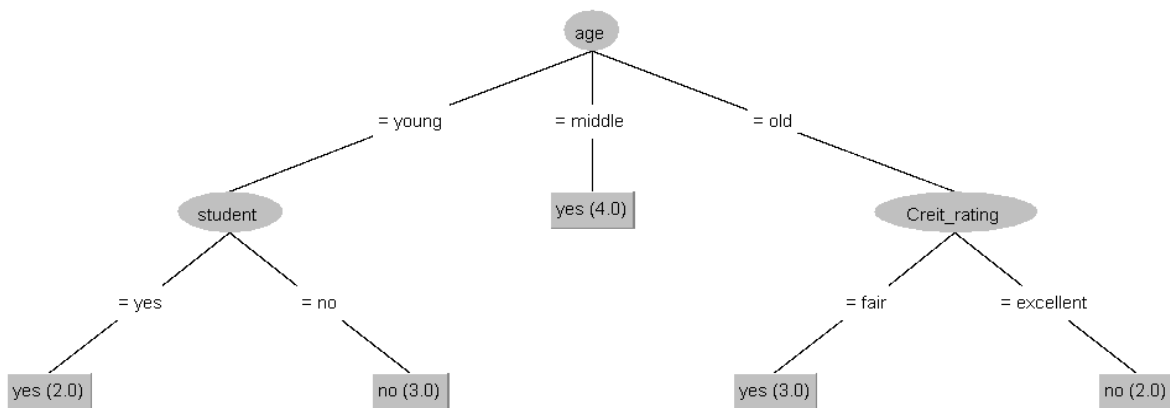
RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40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

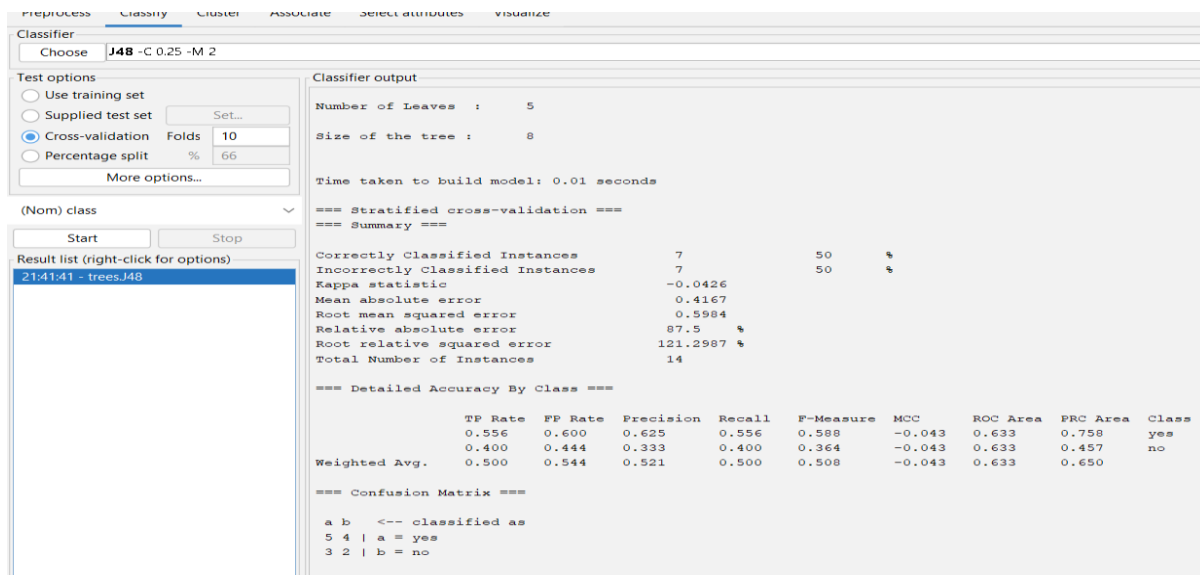
INPUT

```
@relation decision_tree
@attribute age{young,middle,old}
@attribute income{low,medium,high}
@attribute student{yes,no}
@attribute Creit_rating{fair,excellent}
@attribute class{yes,no}
@data
```

```
young high no fair no
young high no excellent no
middle high no fair yes
old medium no fair yes
old low yes fair yes
old low yes excellent no
middle low yes excellent yes
young medium no fair no
young low yes fair yes
old medium yes fair yes
young medium yes excellent yes
middle medium no excellent yes
middle high yes fair yes
old medium no excellent no
```

OUTPUT





4. Analysis the dataset “diabetes. csv” how the diabetes trend is for different age people, using linear regression and multiple regression.

INPUT

```
data<-read.csv("C:/Users/Hari Naidu/Desktop/POM/download papers/diabetes.csv")
```

```
data
```

```
relation<-lm(data$Age~data$Outcome)
```

```
relation
```

```
relation<-lm(data$Age~data$Outcome+data$BMI)
```

OUTPUT

```

[ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> relation<-lm(data$Age~data$Outcome+data$BMI)
> relation

```

```
Call:
```

```
lm(formula = data$Age ~ data$Outcome + data$BMI)
```

```
Coefficients:
```

```

(Intercept)  data$Outcome  data$BMI
  32.84734      6.14177    -0.05469

```

```
> relation<-lm(data$Age~data$Outcome)
```

```
> relation
```

```
Call:
```

```
lm(formula = data$Age ~ data$Outcome)
```

```
Coefficients:
```

```

(Intercept)  data$Outcome
  31.190      5.877

```

5. 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

OUTPUT

APRIOR

Associator

Choose **Apriori** -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1

Start Stop

Result list (right-click for ...)

17:11:38 - Apriori

Associator output

```

y
d
a
u
c
i

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

Apriori
=====
Minimum support: 0.85 (4 instances)
Minimum metric <confidence>: 0.8
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. e=T 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. d=F 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. a=F 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. u=F 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. i=F 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. u=F 4 ==> e=T 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
7. e=T 4 ==> u=F 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
8. e=T u=F 4 ==> k=T 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. k=T u=F 4 ==> e=T 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
10. k=T e=T 4 ==> u=F 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)

```

FPGROWTH

Associator

Choose **FPGrowth** -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.5

Start Stop

Result list (right-click for ...)

17:11:38 - Apriori

17:13:15 - FPGrowth

Associator output

```

=== Run information ===

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.5
Relation:     database
Instances:    5
Attributes:   11
              m
              o
              n
              k
              e
              y
              d
              a
              u
              c
              i

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

FPGrowth found 7 rules (displaying top 7)

1. [c=F]: 3 ==> [u=F]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
2. [c=F]: 3 ==> [i=F]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
3. [n=F]: 3 ==> [d=F]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
4. [u=F, i=F]: 3 ==> [c=F]: 3 <conf:(1)> lift:(1.67) lev:(0.24) conv:(1.2)
5. [c=F]: 3 ==> [u=F, i=F]: 3 <conf:(1)> lift:(1.67) lev:(0.24) conv:(1.2)
6. [u=F, c=F]: 3 ==> [i=F]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
7. [i=F, c=F]: 3 ==> [u=F]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)

```

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

OUTPUT TREE

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier
Choose **J48 -C 0.25 -M 2**

Test options
☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % **66**
 More options...

(Nom) class
 Start Stop

Result list (right-click for options)
 12:51:54 - bayes.NaiveBayes
 12:58:01 - bayes.NaiveBayes
 13:06:38 - trees.J48
 13:09:14 - functions.Logistic
 13:10:16 - functions.Logistic
13:11:08 - trees.J48

Classifier output
 Number of Leaves : 20
 Size of the tree : 39
 Time taken to build model: 0.01 seconds
 === Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	567	73.8281 %
Incorrectly Classified Instances	201	26.1719 %
Kappa statistic	0.4164	
Mean absolute error	0.3158	
Root mean squared error	0.4463	
Relative absolute error	69.4841 %	
Root relative squared error	93.6293 %	
Total Number of Instances	768	

 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.814	0.403	0.790	0.814	0.802	0.417	0.751	0.811	tested_neg
	0.597	0.186	0.632	0.597	0.614	0.417	0.751	0.572	tested_pos
Weighted Avg.	0.738	0.327	0.735	0.738	0.736	0.417	0.751	0.727	

 === Confusion Matrix ===

```

    a  b  <-- classified as
407  93 | a = tested_negative
108 160 | b = tested_positive
  
```

PREPROCESSOR

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: pima_diabetes
Instances: 768
Attributes: 9 Sum of weights: 768

Attributes
 All None Invert Pattern

No.	Name
1	preg
2	plas
3	pres
4	skin
5	insu
6	mass
7	pedi
8	age
9	class

Remove

Selected attribute
Name: preg
Missing: 0 (0%)
Distinct: 17
Type: Numeric
Unique: 2 (0%)

Statistic	Value
Minimum	0
Maximum	17
Mean	3.845
StdDev	3.37

Class: class (Nom) Visualize All

LOGISTIC

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic - R 1.0E-8 - M -1 -num-decimal-places 4**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds

☐ Percentage split %

More options...

Classifier output

```

mass          0.9142
pedi          0.3886
age           0.9852

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      593           77.2135 %
Incorrectly Classified Instances    175           22.7865 %
Kappa statistic                    0.4734
Mean absolute error                 0.3094
Root mean squared error             0.3954
Relative absolute error              68.0818 %
Root relative squared error         82.9651 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.880    0.429    0.793     0.880    0.834      0.480    0.832     0.892    tested_neg.
                0.571    0.120    0.718     0.571    0.636      0.480    0.832     0.715    tested_pos.
Weighted Avg.   0.772    0.321    0.767     0.772    0.765      0.480    0.832     0.831

=== Confusion Matrix ===
  a  b  <-- classified as
440  60 |  a = tested_negative
115 153 |  b = tested_positive

```

(Nom) class

Start Stop

Result list (right-click for actions)

12:51:54 - bayes.NaiveBayes

12:58:01 - bayes.NaiveBayes

13:06:38 - trees.J48

13:09:14 - functions.Logistic

13:10:16 - functions.Logistic

7. Create the dataset using ARFF file format:

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

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

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

INPUT

@relation hotdogs

@attribute hotdogs {t,f}

@attribute buns {t,f}

@attribute ketchup {t,f}

@attribute coke {t,f}

@attribute chips {t,f}

@data

t t t f f

t t f f f

t f f t t

f f f t t

f f t f t

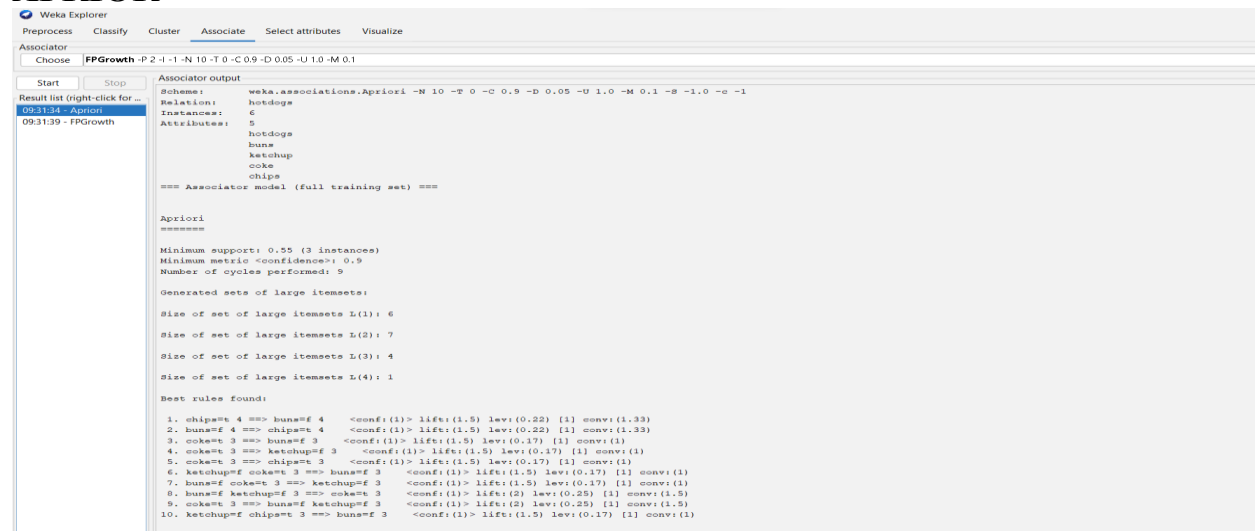
t f f t t

OUTPUT

FP GROWTH



APRIOR



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

INPUT

@relation decision_tree

@attribute age{young,middle,old}

@attribute income{low,medium,high}

@attribute student{yes,no}

@attribute Creit_rating{fair,excellent}

@attribute class{yes,no}

@data

young high no fair no

young high no excellent no

middle high no fair yes

old medium no fair yes

old low yes fair yes

old low yes excellent no

middle low yes excellent yes

young medium no fair no

young low yes fair yes

old medium yes fair yes

young medium yes excellent yes

middle medium no excellent yes

middle high yes fair yes

old medium no excellent no

OUTPUT

RULE BASED CLASSIFICATION

The screenshot displays the Weka Explorer interface with the 'Classify' tab selected. The classifier used is 'J48 -C 0.25 -M 2'. The test options are set to 'Cross-validation' with 10 folds and 66% split. The classifier output shows the following details:

Classifier output

```
=====
age = young
| student = yes: yes (2.0)
| student = no: no (3.0)
age = middle: yes (4.0)
age = old
| Creit_rating = fair: yes (3.0)
| Creit_rating = excellent: no (2.0)

Number of Leaves :    5
Size of the tree :    8

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      7           50 %
Incorrectly Classified Instances    7           50 %
Kappa statistic                    -0.0426
Mean absolute error                 0.4167
Root mean squared error             0.5954
Relative absolute error             87.5 %
Root relative squared error        121.2967 %
Total Number of Instances         14

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
0.400    0.444    0.333    0.400    0.364    -0.043    0.633    0.457    yes
Weighted Avg.    0.500    0.544    0.521    0.500    0.500    -0.043    0.633    0.650

=== Confusion Matrix ===
  a b   <-- classified as
 5 4 | a = yes
```

The 'Result list' on the left shows the file '09.17.07 - trees.j48'.

DECESION TREE

