DAY-3

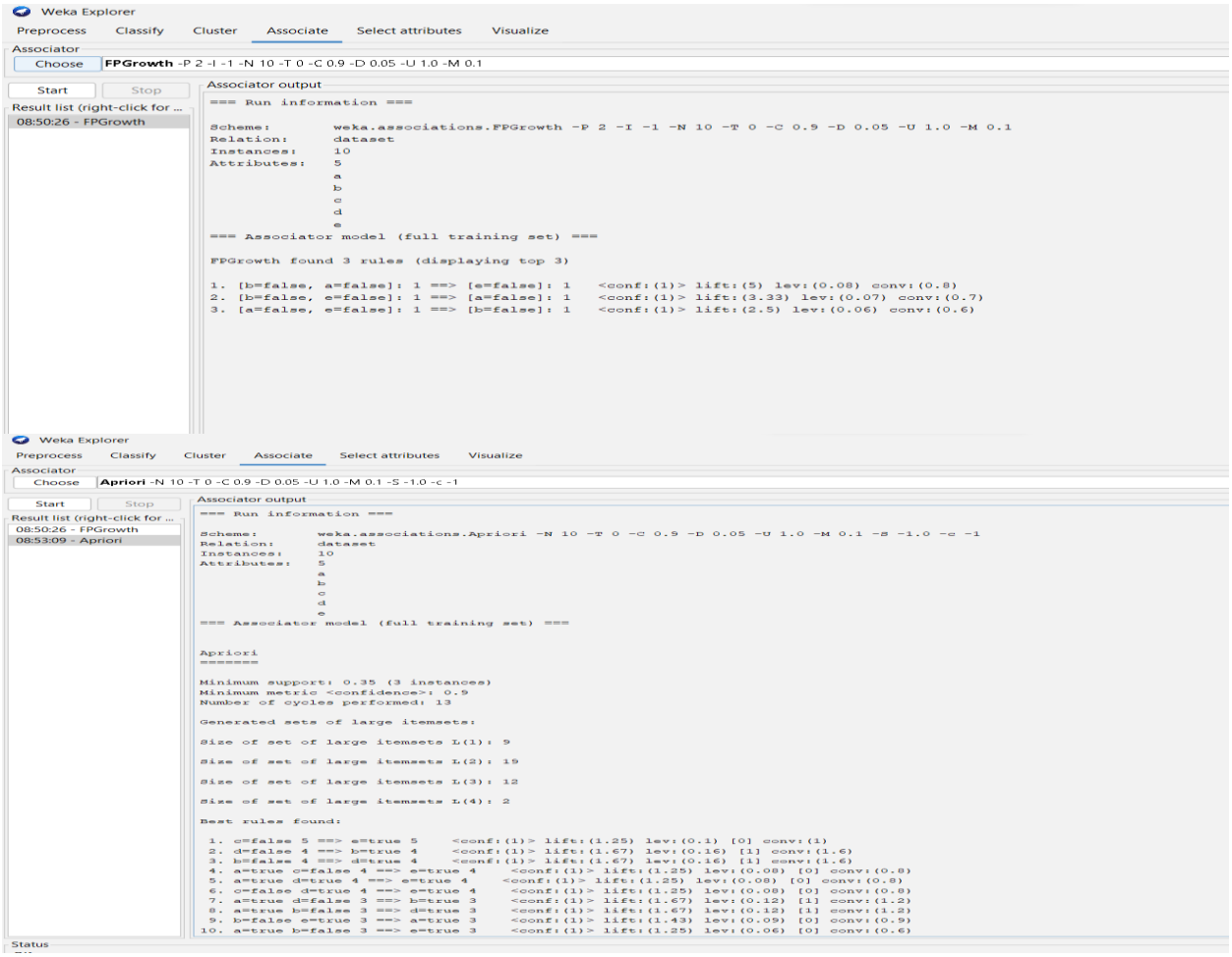
1.Consider the data set and perform the Apriori Algorithm and FP algorithm support:3 and confidence=50%

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



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

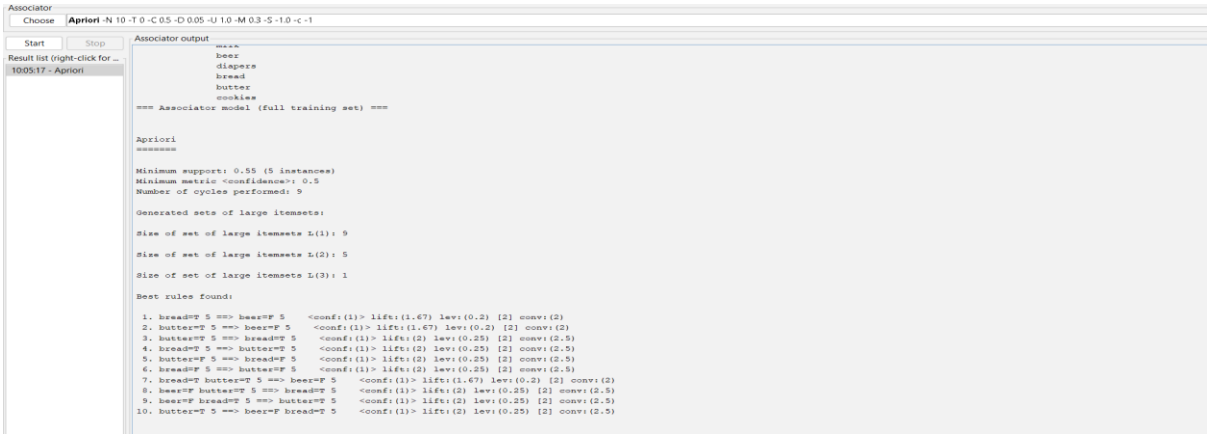
Consider the market basket transactions shown in the above table.

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

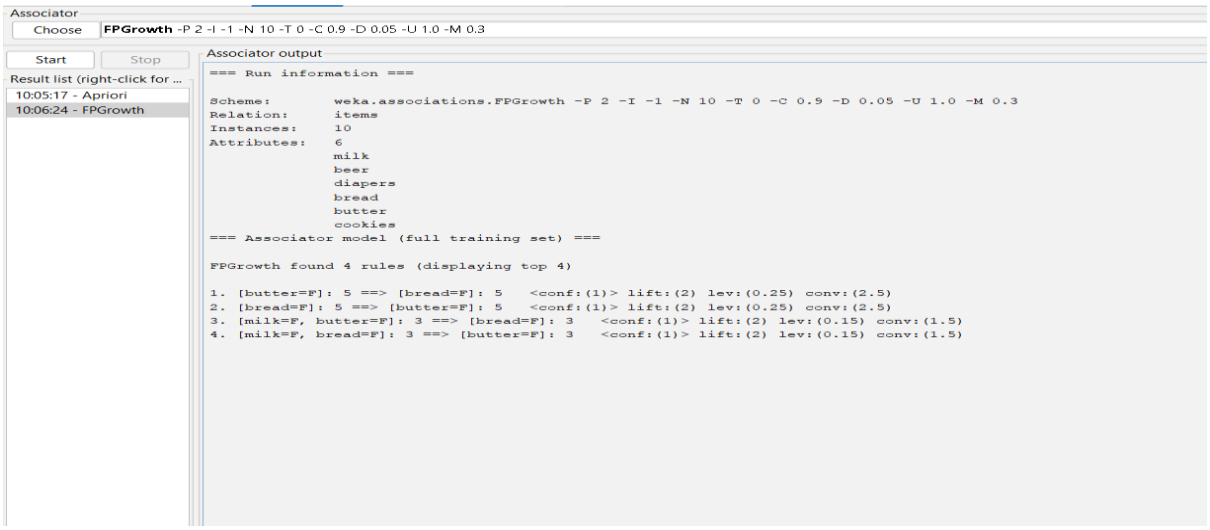
(b) What is the maximum size of frequent itemsets that can be extracted (assuming minsup > 0)?

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



FP GROWTH



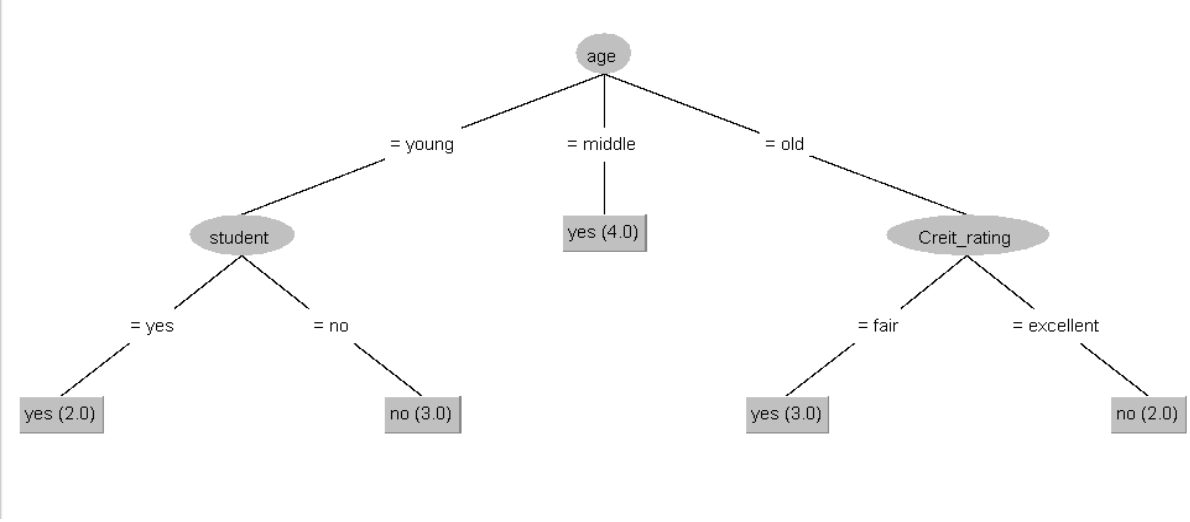
3.Bayes classification and descion tree (using training and test data)

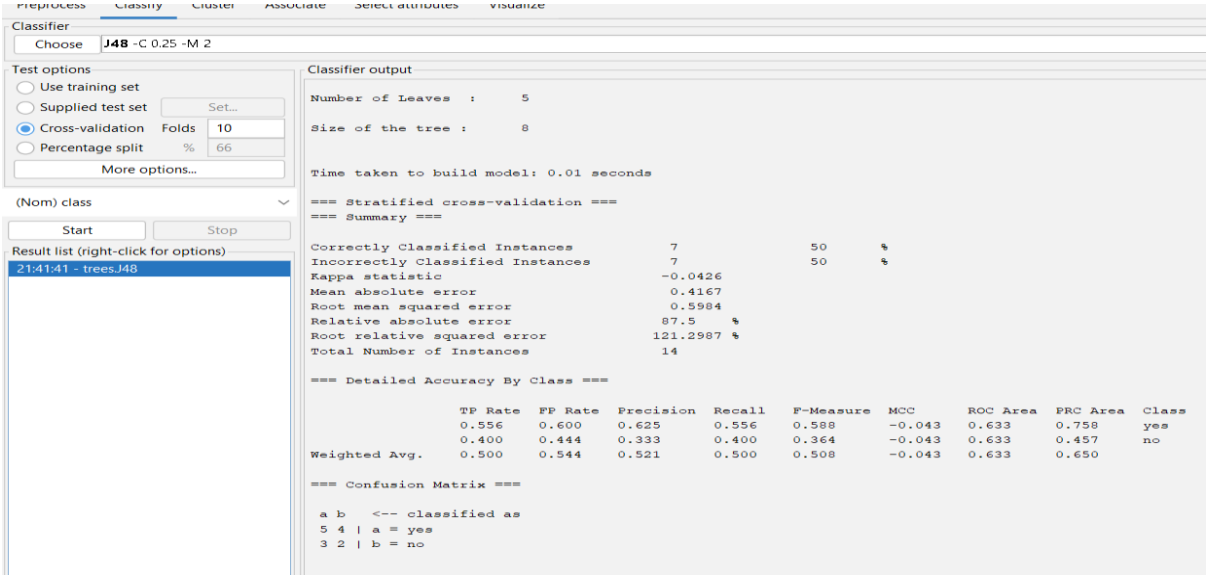
RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10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

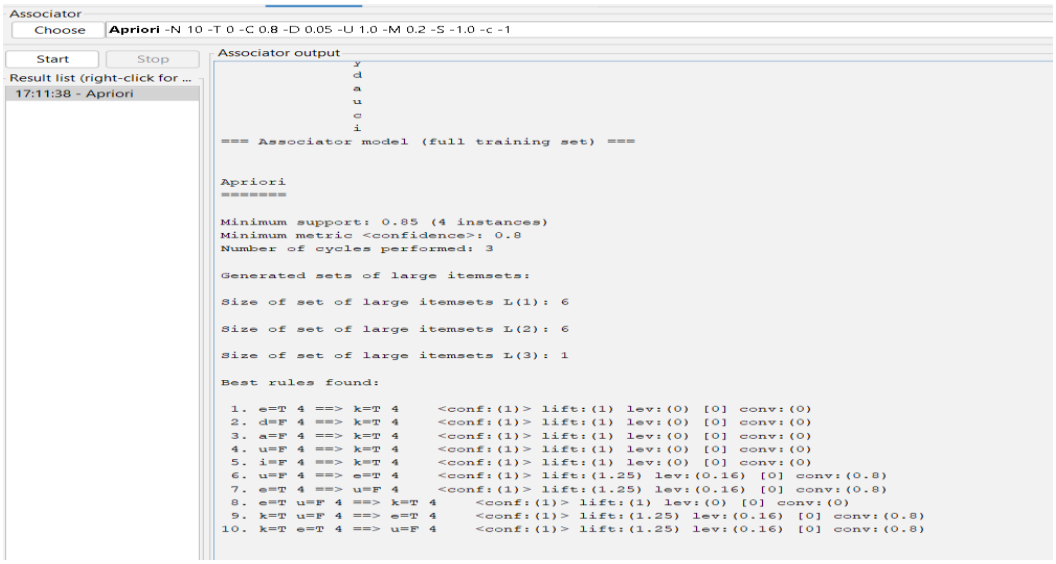
5.Implement using WEKA for the given Suppose a database has five transactions. Let min sup= 50%(2) and min con f = 80%.

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)

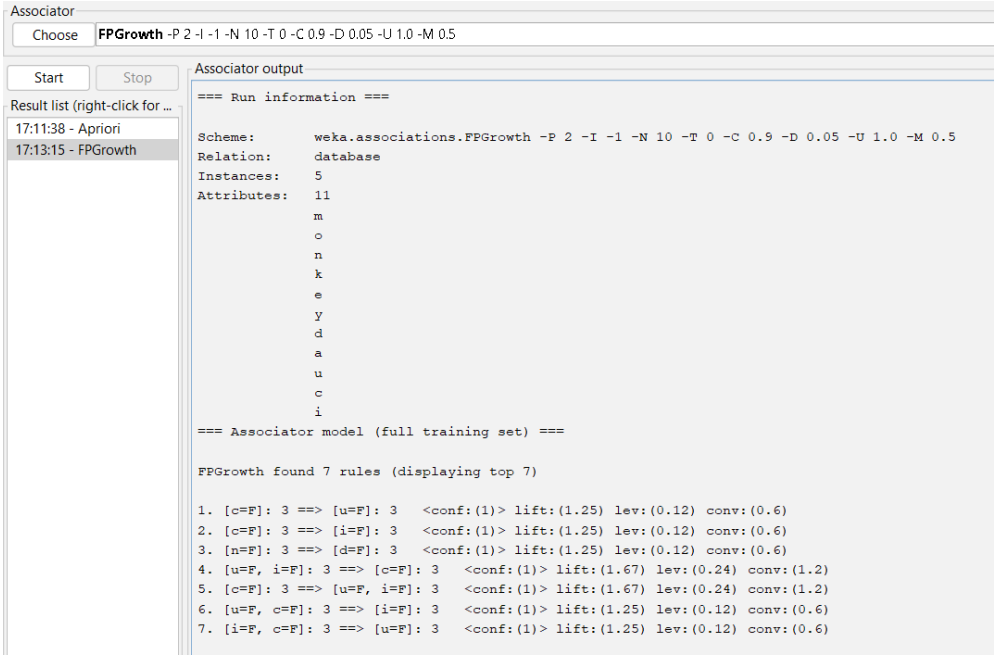
- Find all frequent item sets using Apriori algorithm
- Also draw FP-Growth Tree

OUTPUT

APRIOR



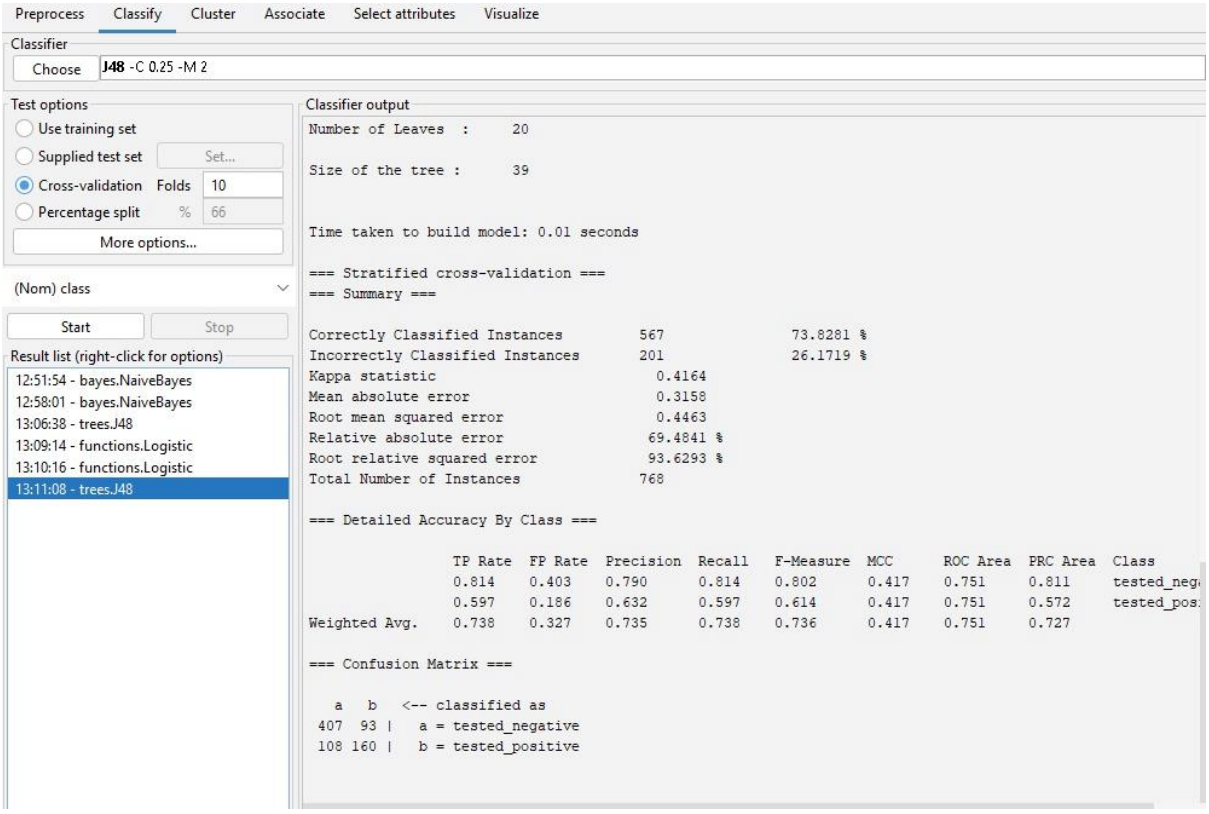
FPGROWTH



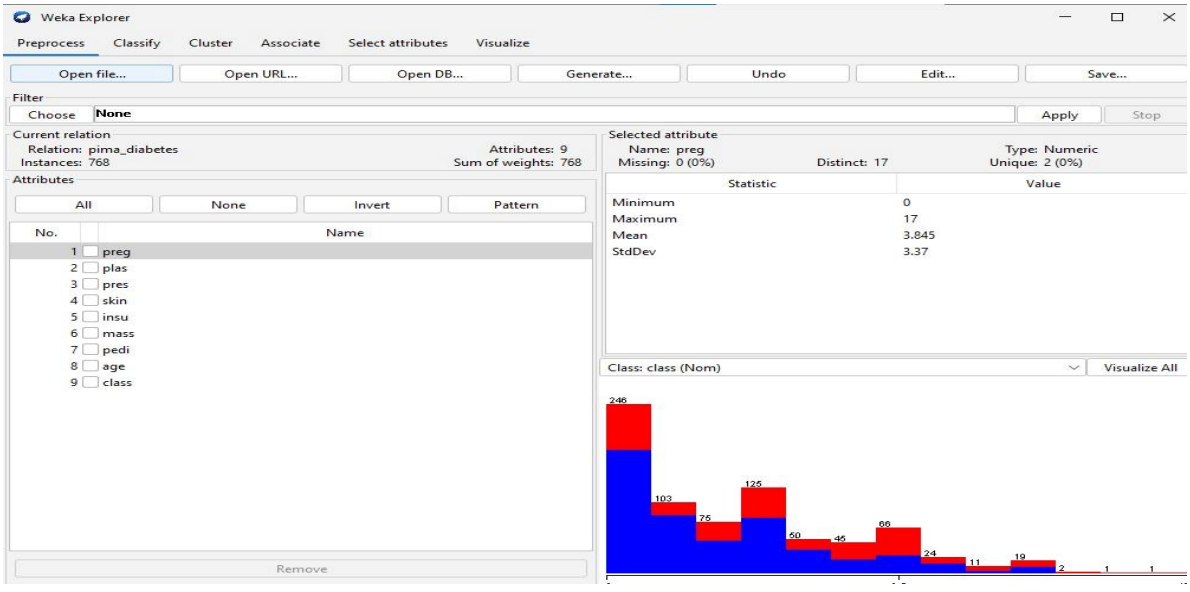
6. Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic

OUTPUT

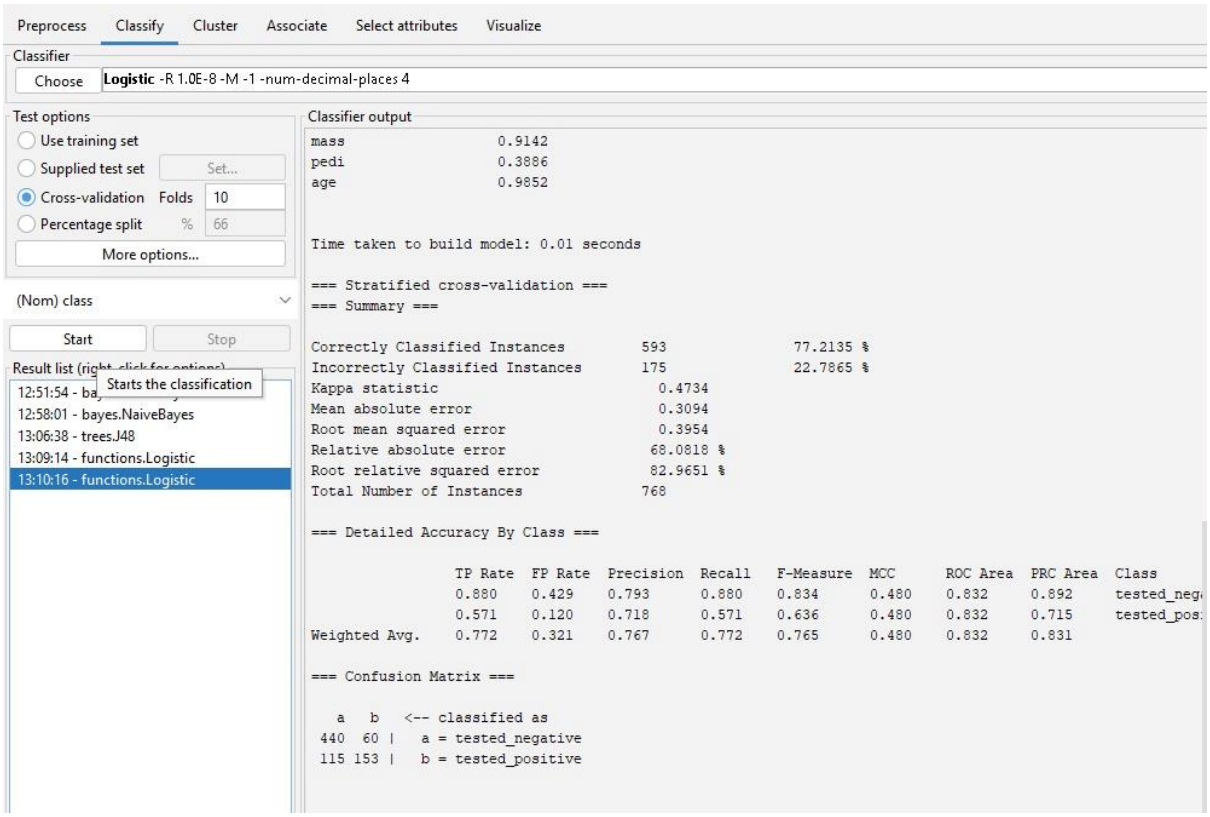
TREE



PREPROCESSOR



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 algorthim ,mention wheather accepted or rejcted.

INPUT

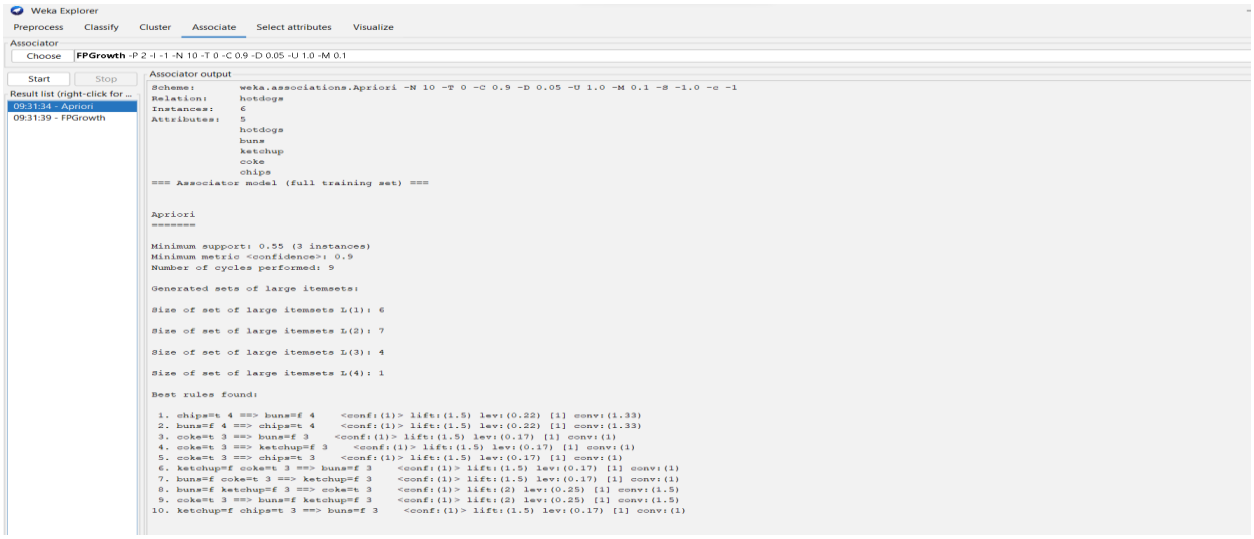
@relation hotdogs
@attribute hotdogs{t,f}
@attribute buns{t,f}
@attribute ketchup{t,f}
@attribute coke{t,f}
@attribute chips{t,f}

@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

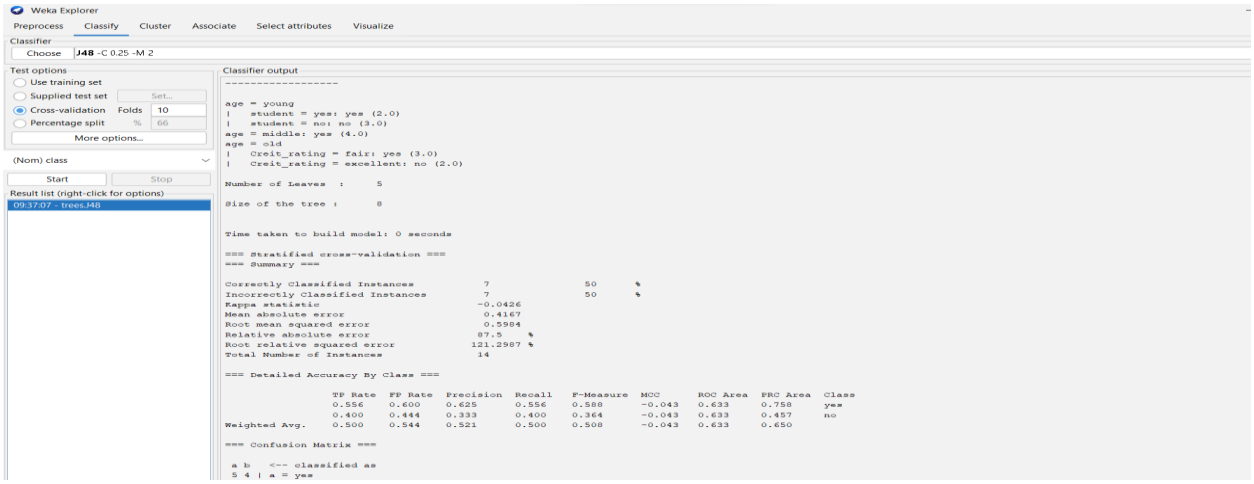
INPUT

@relation decision_tree
@attribute age{young,middle,old}
@attribute income{low,medium,high}
@attribute student{yes,no}

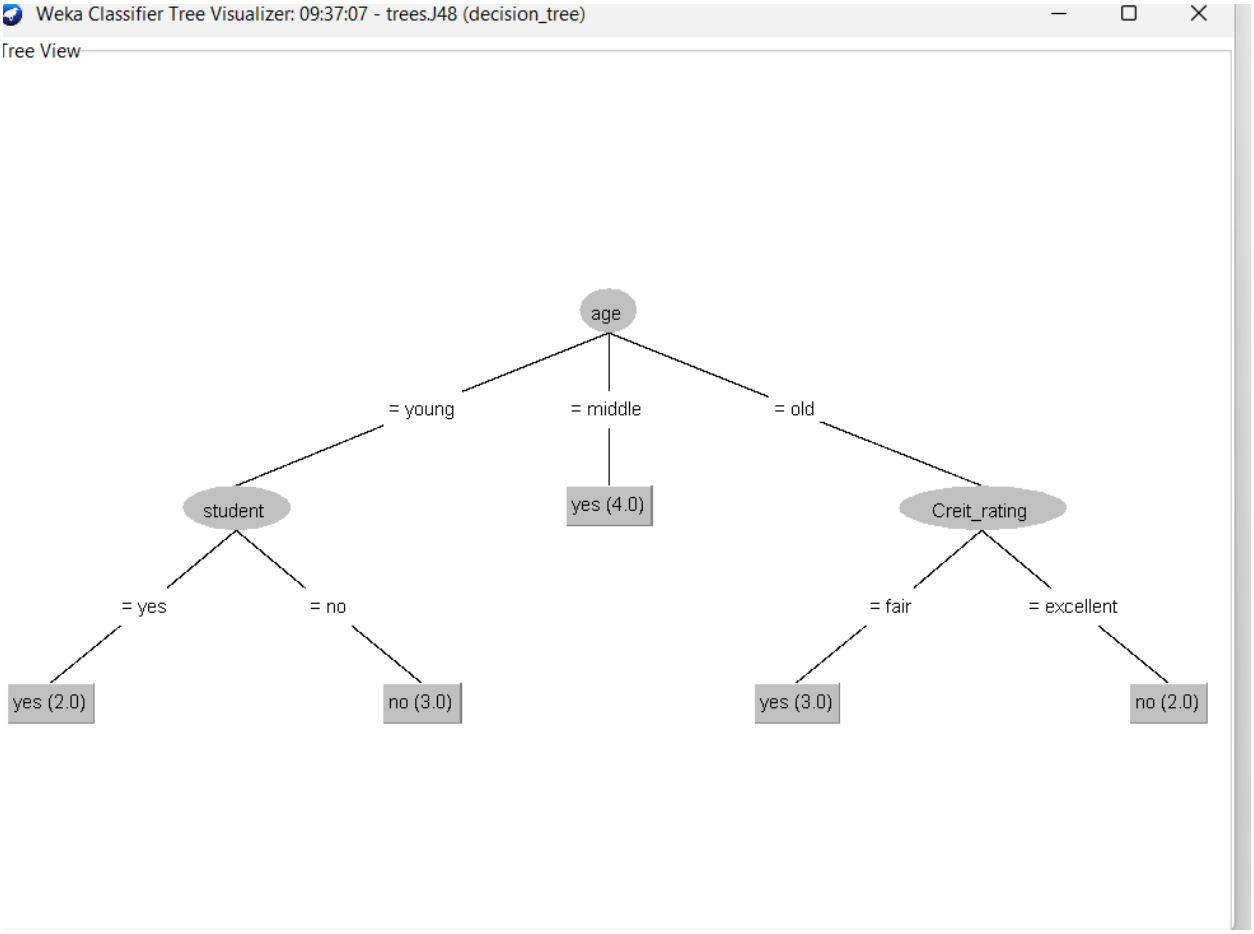
@attribute Creit_rating{fair,excellent}
@attribute class{yes,no}
@data
young high no fair no
young high no excellent no
middle high no fair yes
old medium no fair yes
old low yes fair yes
old low yes excellent no
middle low yes excellent yes
young medium no fair no
young low yes fair yes
old medium yes 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



DECESION TREE


Weka Classifier Tree Visualizer: 09:37:07 - trees.J48 (decision_tree)
Tree View
age
= young
= middle
= old
student
yes (4.0)
Creit_rating
= yes
= no
= fair
= excellent
yes (2.0)
no (3.0)
yes (3.0)
no (2.0)