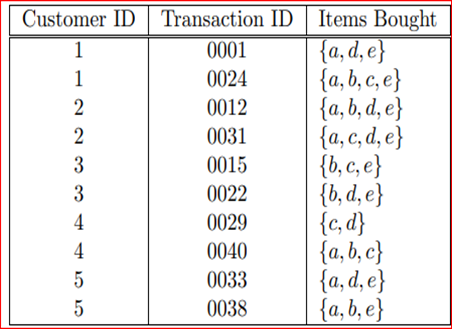
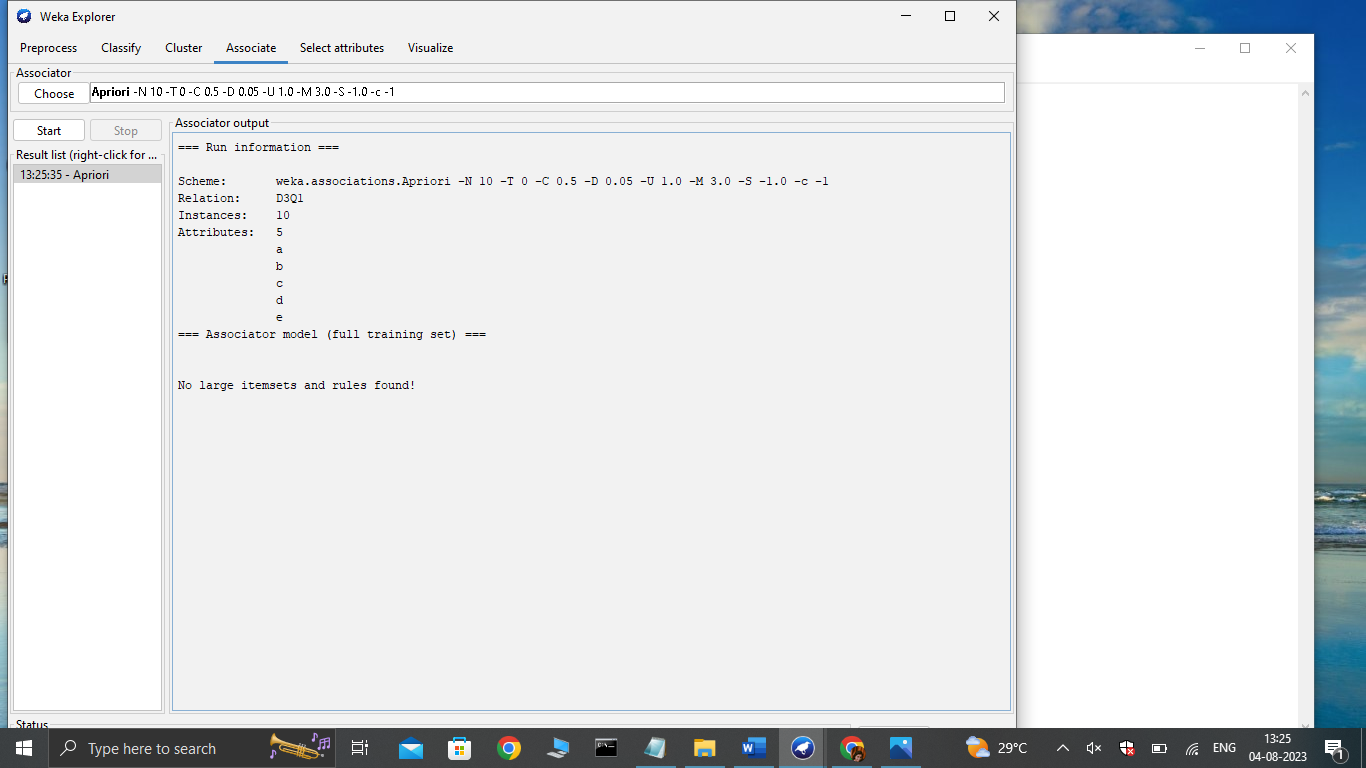
DATA WAREHOUSING AND DATA MINING

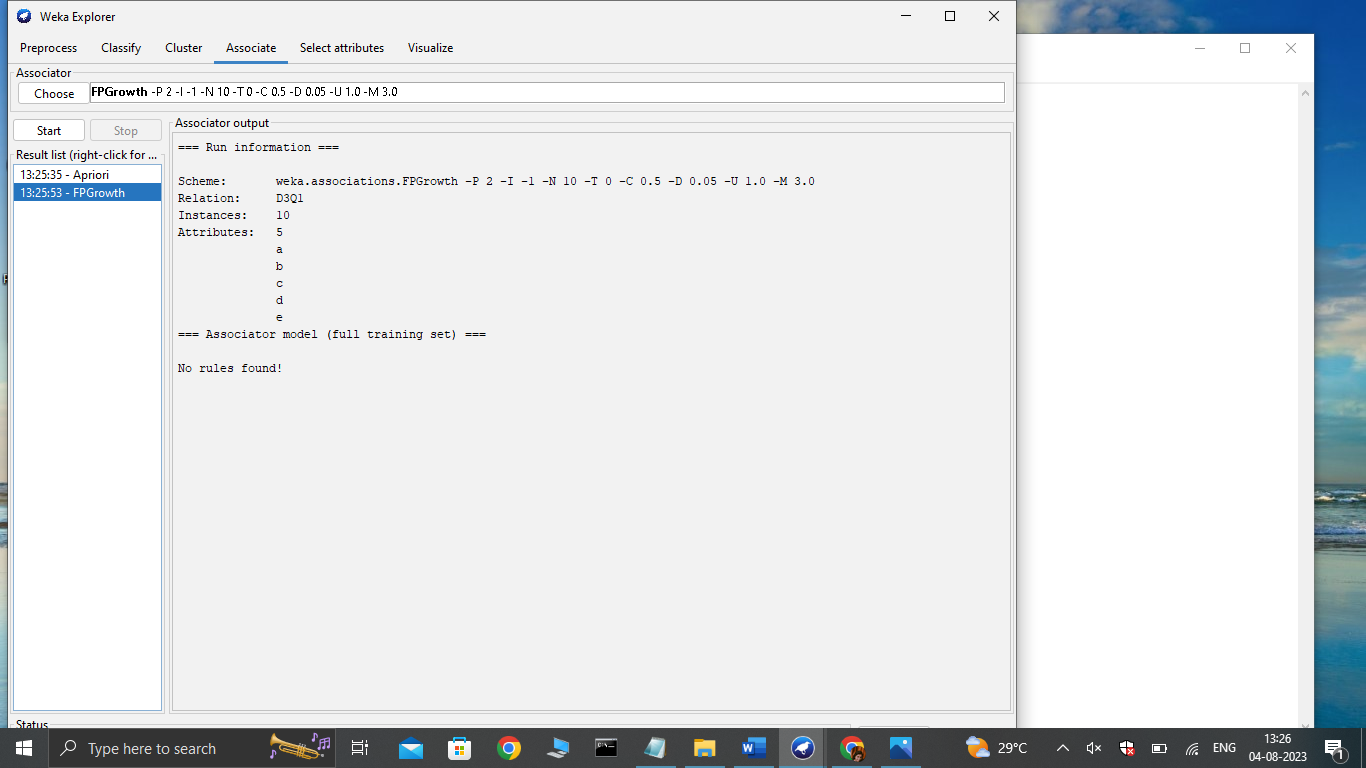
DAY – 3

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



**ANSWERS:**





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

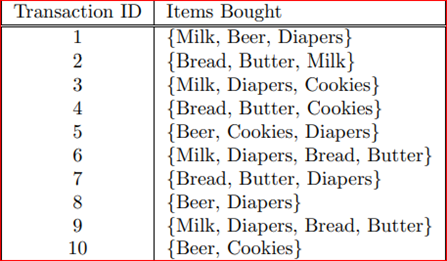
Consider the market basket transactions shown in the above table.

(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)?



**ANSWER:**

@relation D3Q2

@attribute milk{true, false}

@attribute beer{true, false}

@attribute diapers{true, false}

@attribute bread{true, false}

@attribute butter{true, false}

@attribute cookies{true, false}

@data

true true true false false false

true false false true true false

true false true false false true

false false false true true true

false true true false false true

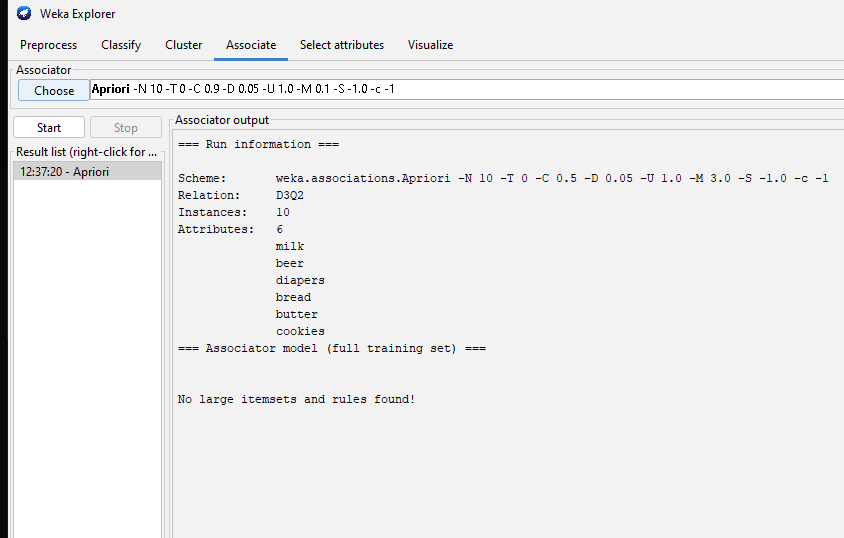
true false true true true false

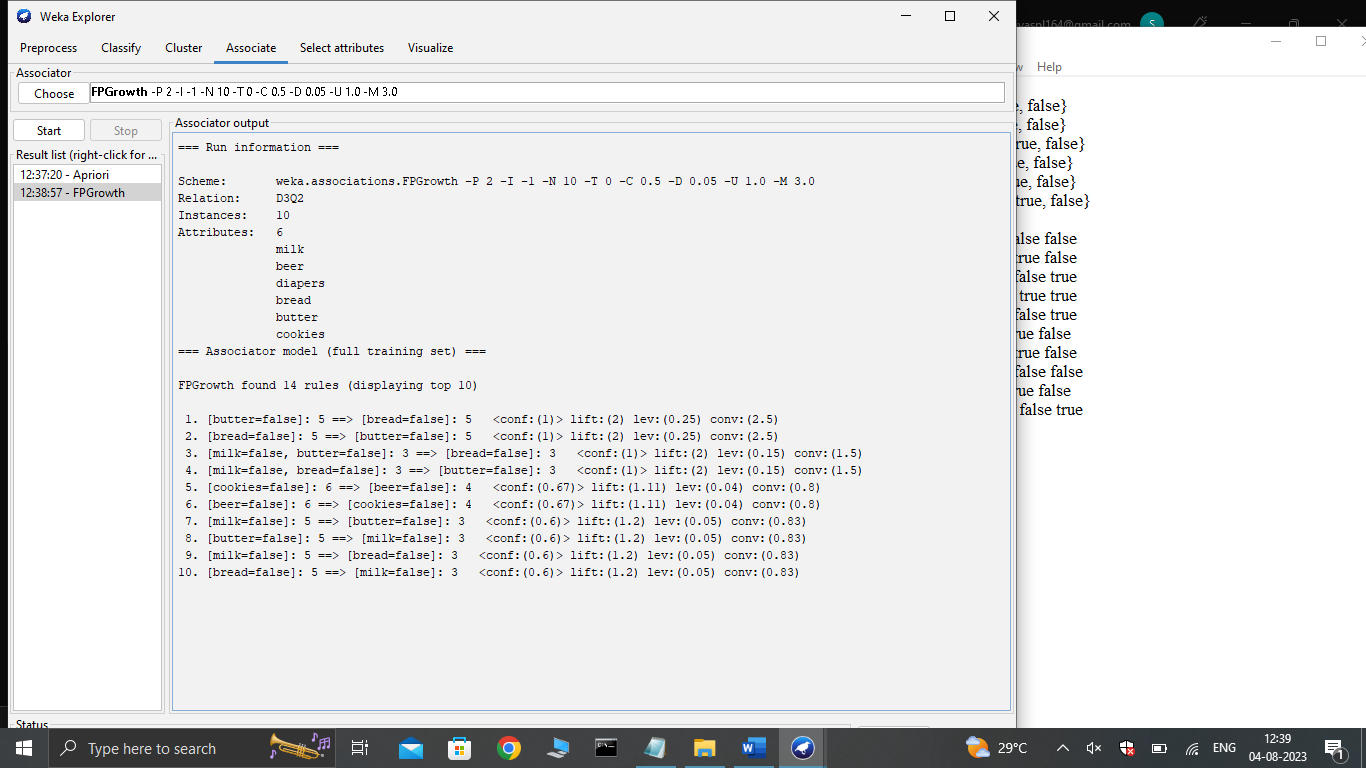
false false true true true false

false true true false false false

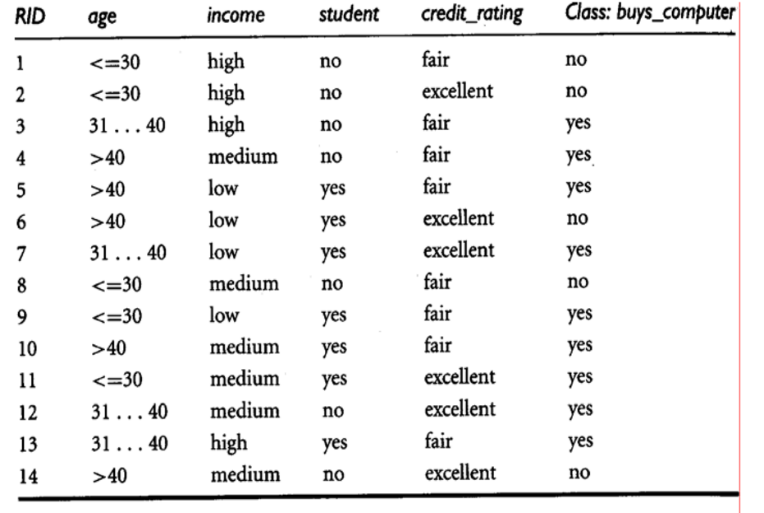
true false true true true false

false true false false false true





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



**ANSWERS:**

@relation D3Q3

@attribute age{young,middle,old}

@attribute income{high,medium,low}

@attribute student{no,yes}

@attribute credit{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

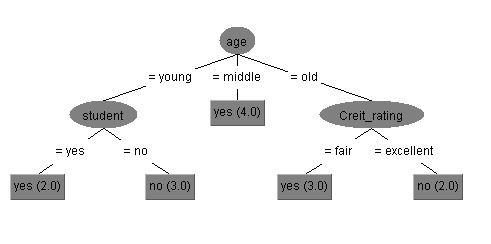
old medium yes fair yes

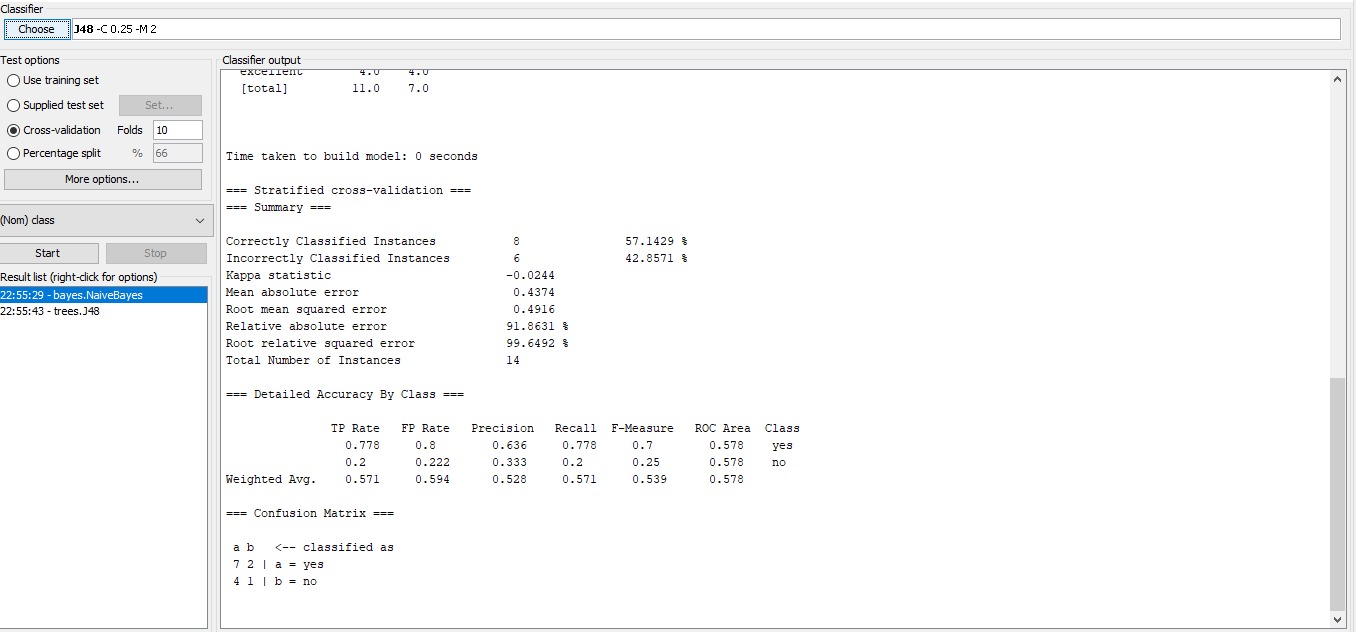
young medium yes excellent yes

middle medium no excellent yes

middle high yes fair yes

old medium no excellent no





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

**ANSWERS:**

diabetes <- read.csv("diabetes.csv")

View(diabetes)

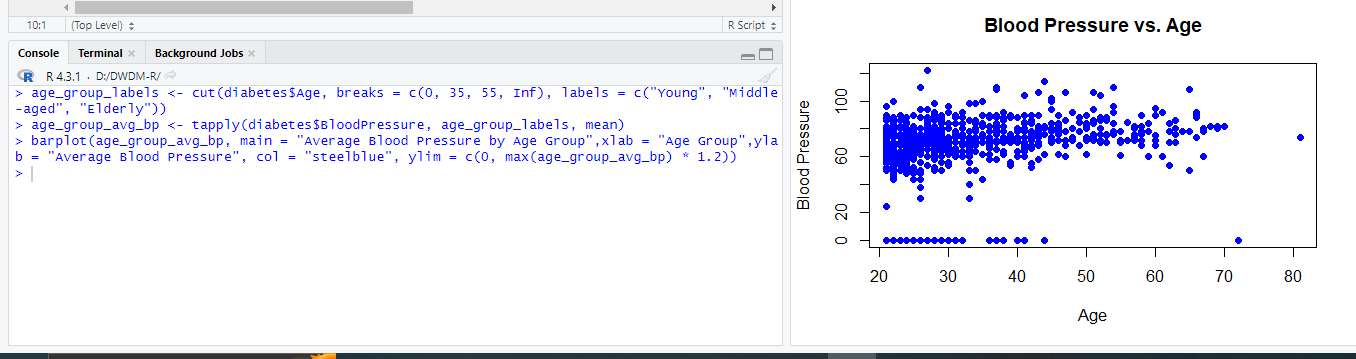
getwd()

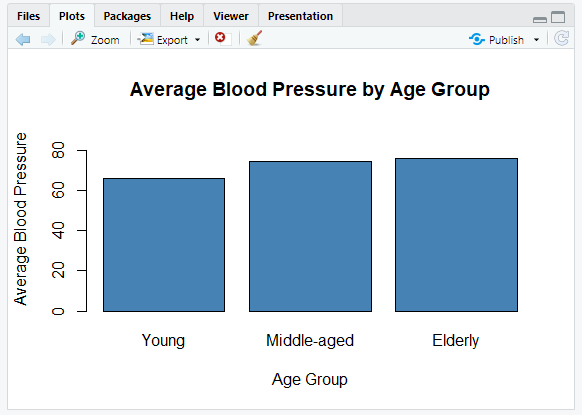
plot(diabetes$Age, diabetes$BloodPressure, xlab = "Age", ylab = "Blood Pressure", main = "Blood Pressure vs. Age", col = "blue",pch = 16)

age\_group\_labels <- cut(diabetes$Age, breaks = c(0, 35, 55, Inf), labels = c("Young", "Middle-aged", "Elderly"))

age\_group\_avg\_bp <- tapply(diabetes$BloodPressure, age\_group\_labels, mean)

barplot(age\_group\_avg\_bp, main = "Average Blood Pressure by Age Group",xlab = "Age Group",ylab = "Average Blood Pressure", col = "steelblue", ylim = c(0, max(age\_group\_avg\_bp) \* 1.2))





1. Implement using WEKA for the given Suppose a database has five transactions. Let min sup= 50% 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

**ANSWER:**

@relation D3Q5

@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

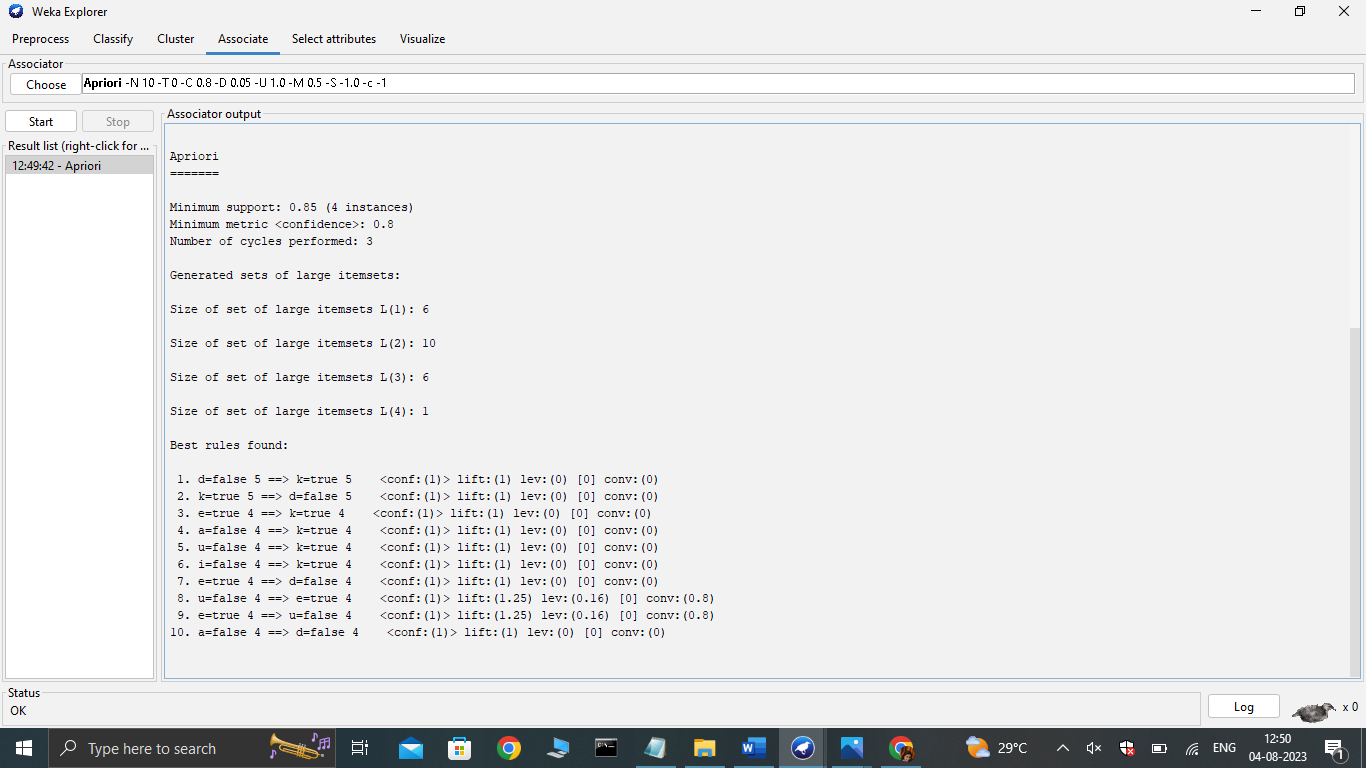
true true true true true true false false false false false

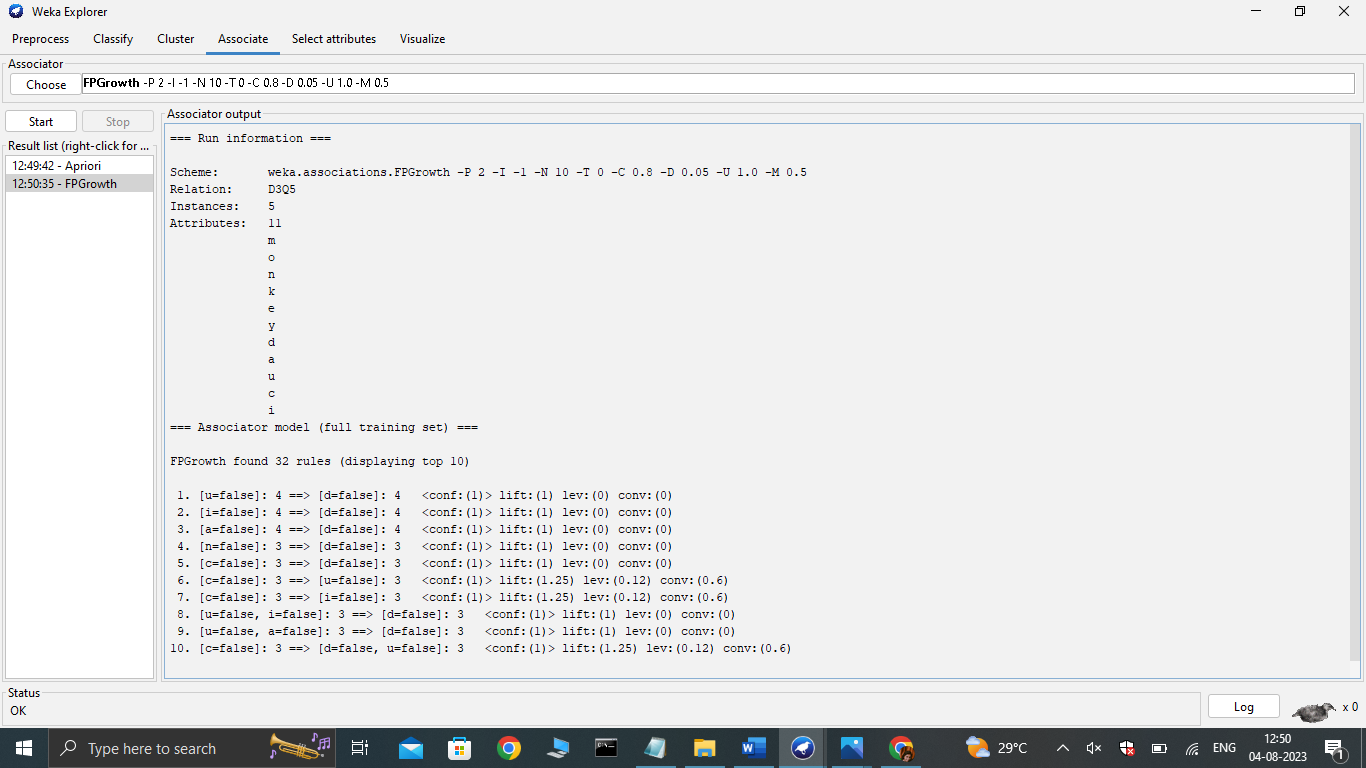
false true true true true true false false false false false

true false false true true false false true false false false

true false false true false true false false true true false

false true false true true false false false false true true

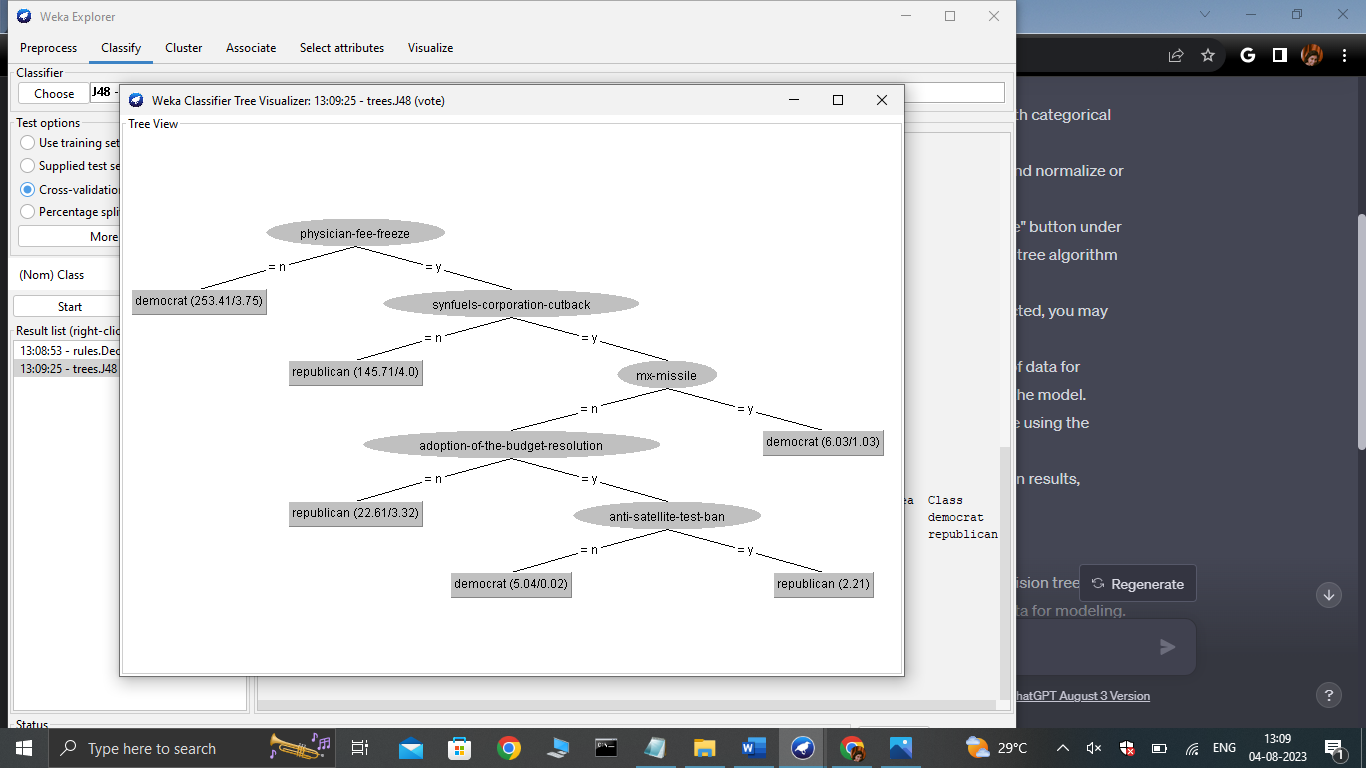




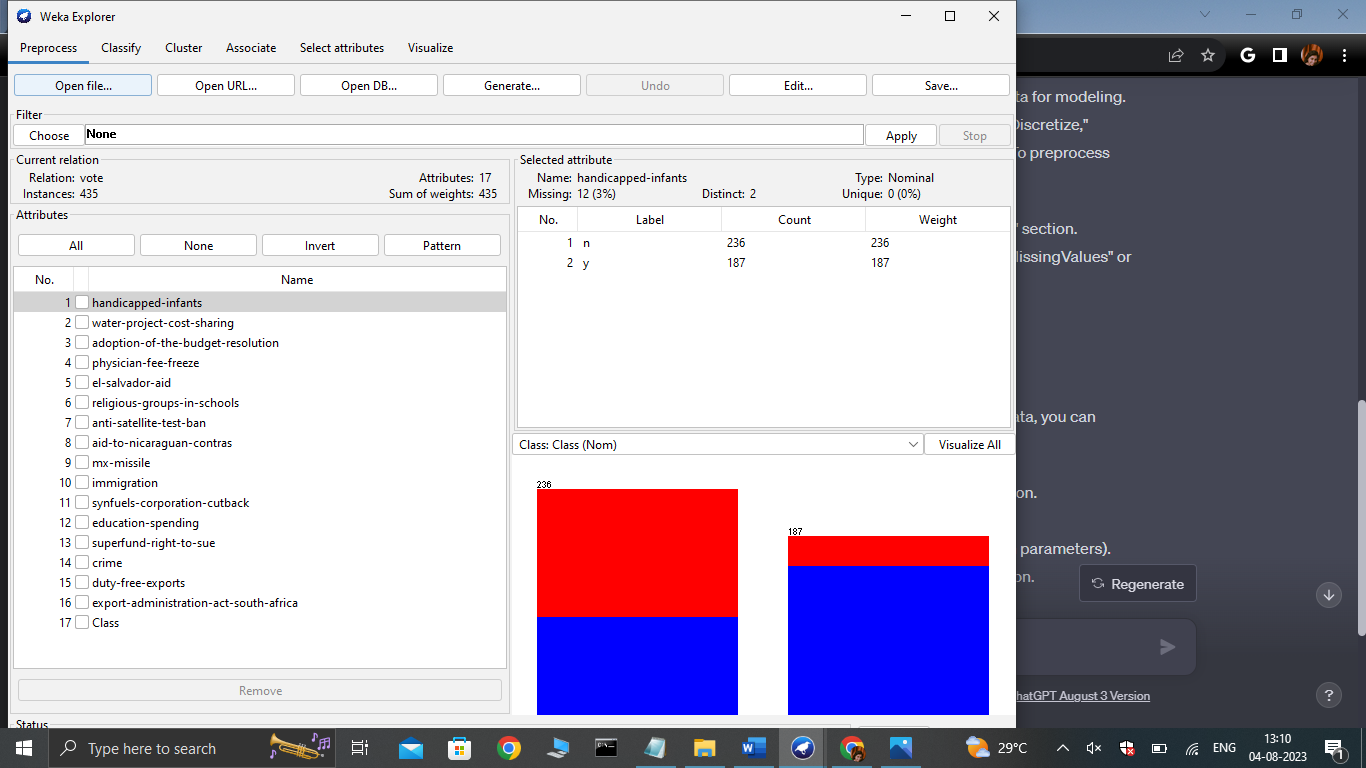
1. Prediction of Categorical Data using Decision Tree Algorithm through WEKA using any datasets. a) Tree b) Preprocess c) Logistic

**ANSWER:**

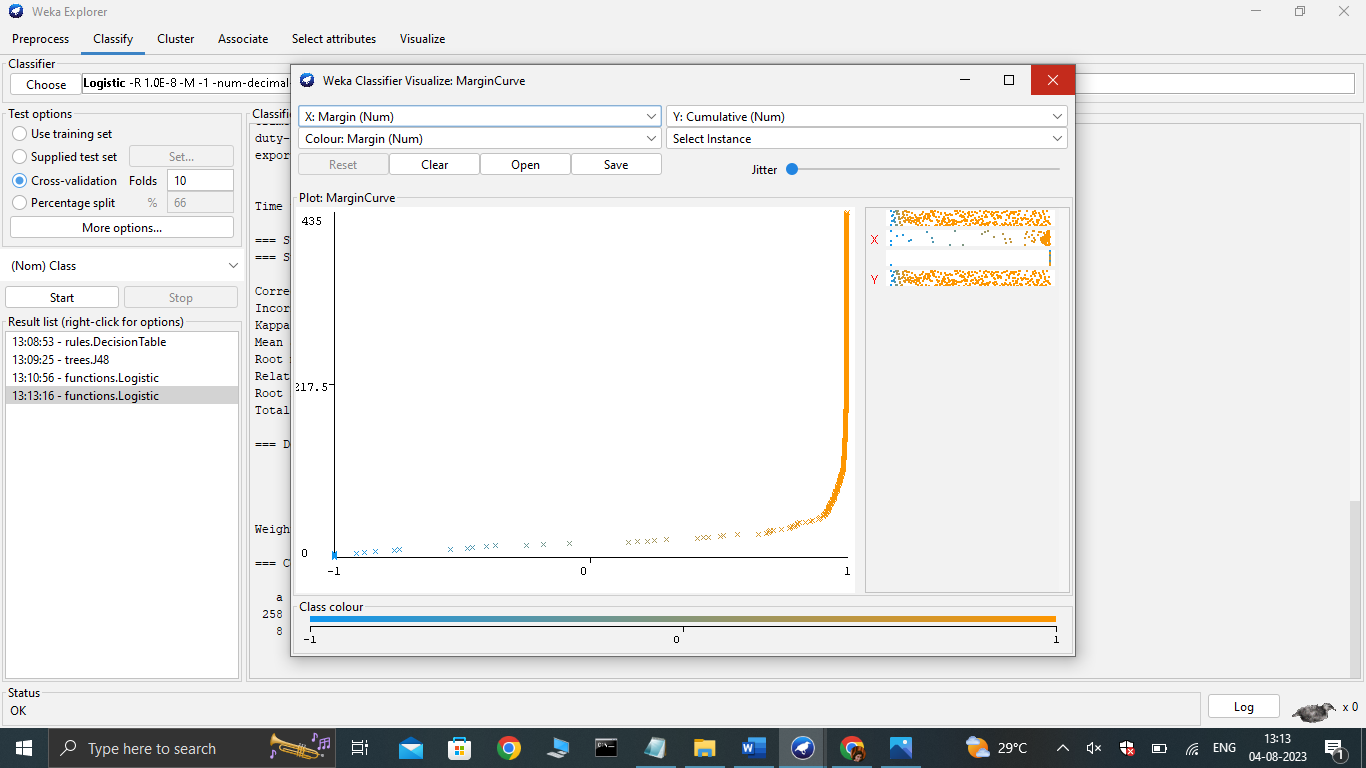
(i)

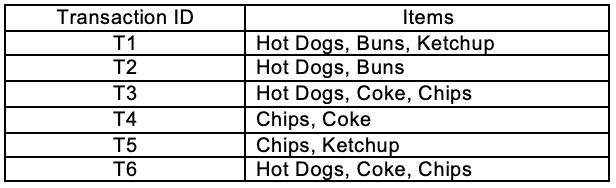


(ii)



(iii)



7. 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 confident threshold (c = 60%).

b.List the various rule generated by apriori and FP tree algorthim ,mention wheather accepted or rejcted.

**ANSWER:**

@relation D3Q7

@attribute hotdog{true,false}

@attribute bun{true,false}

@attribute ketchup{true,false}

@attribute chips{true,false}

@attribute coke{true,false}

@data

true true true false false

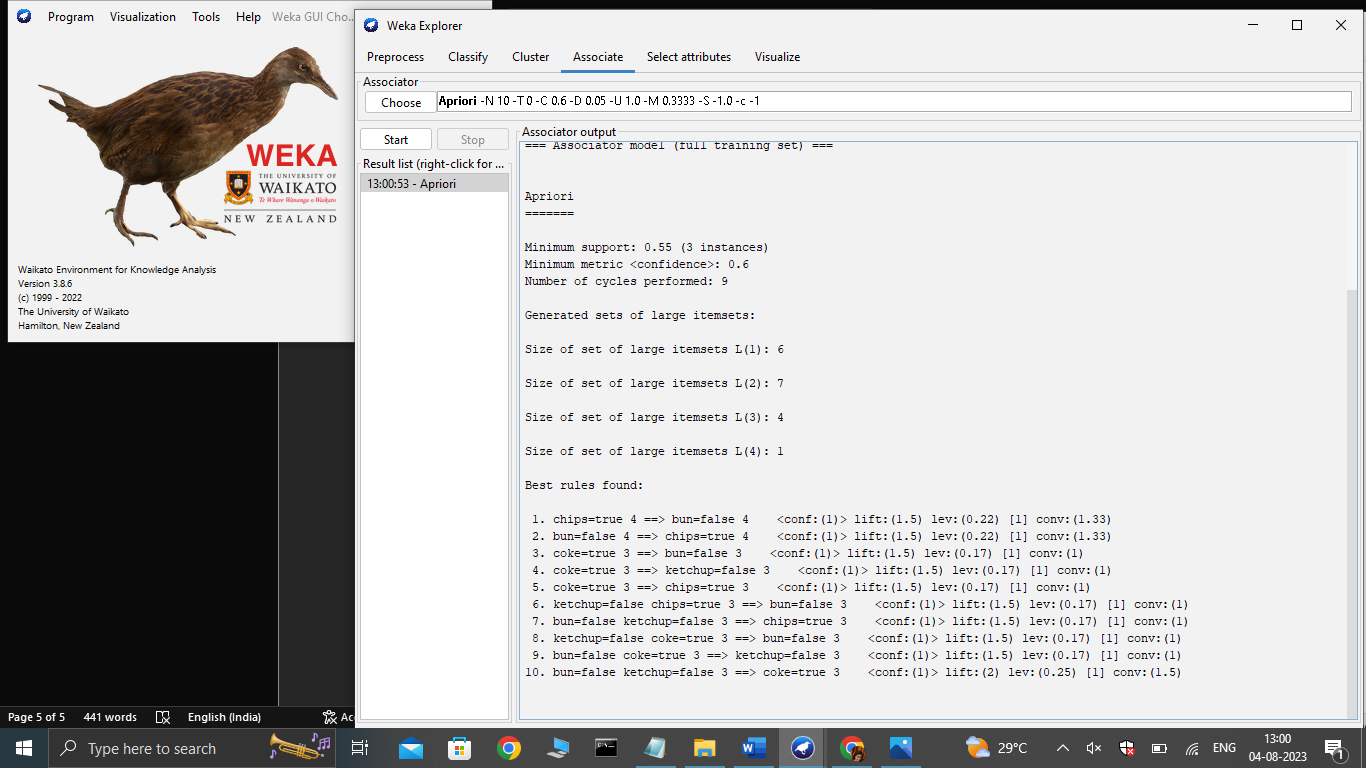
true true false false false

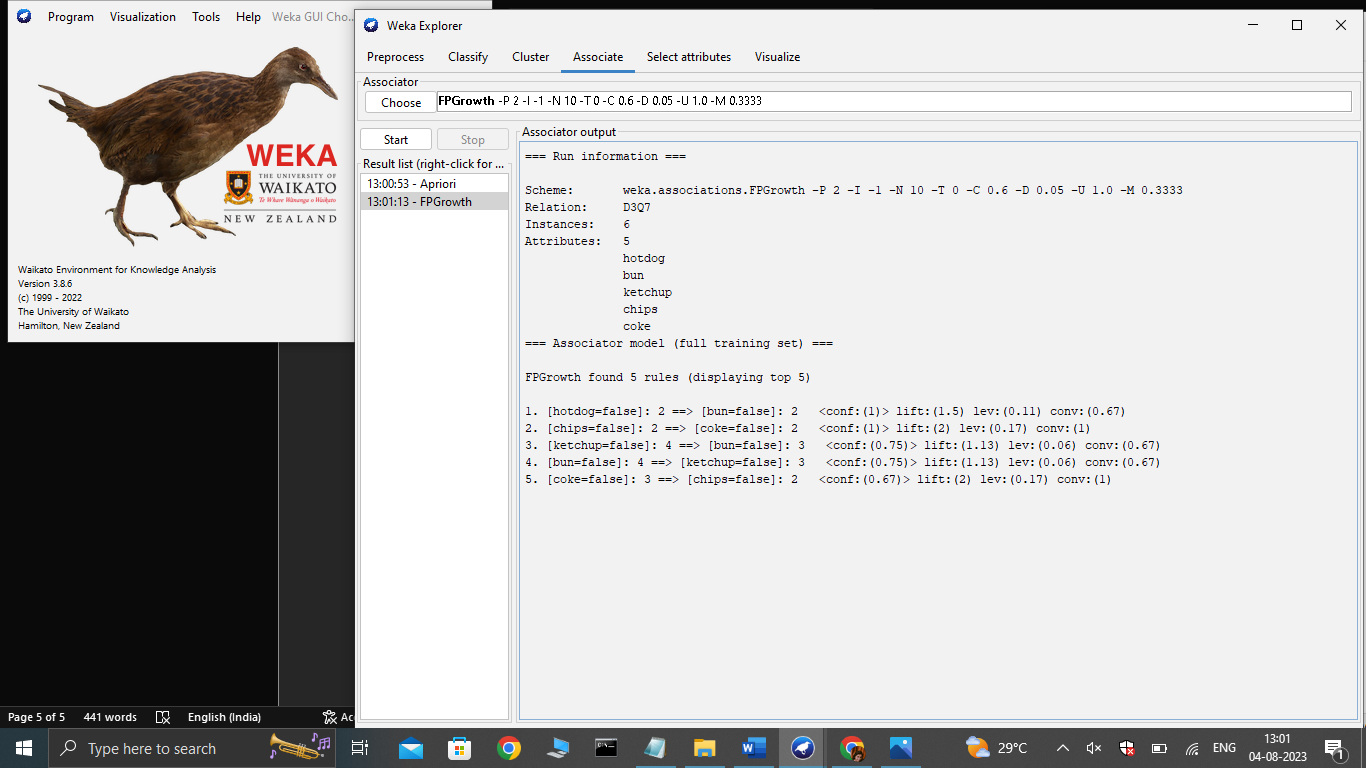
true false false true true

false false false true true

false false true true false

true false false true true





1. 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

**ANSWERS:**

