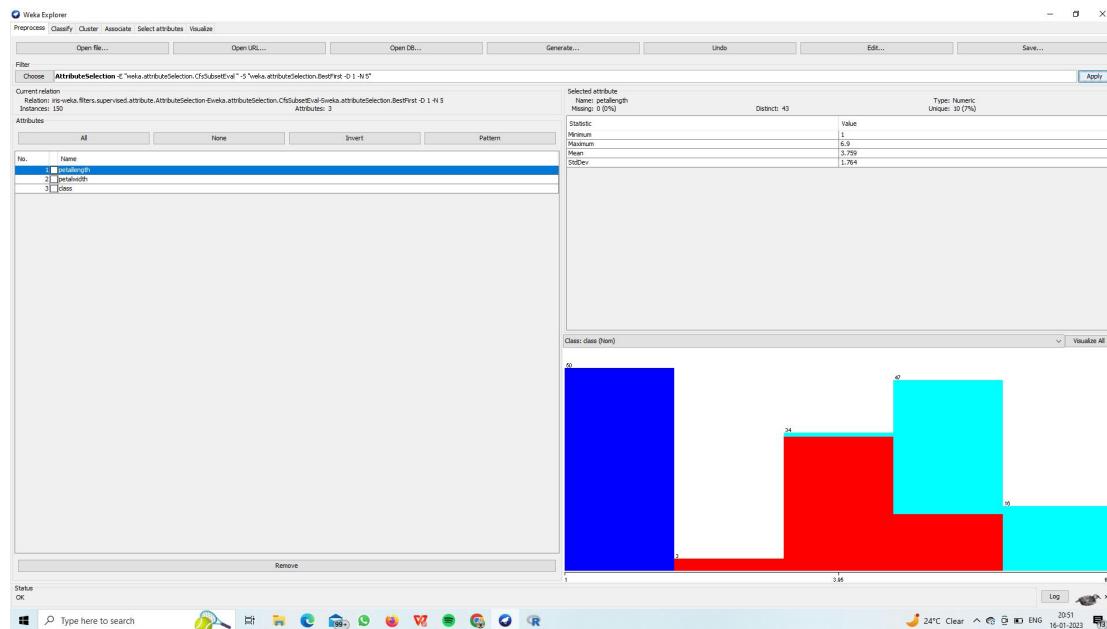


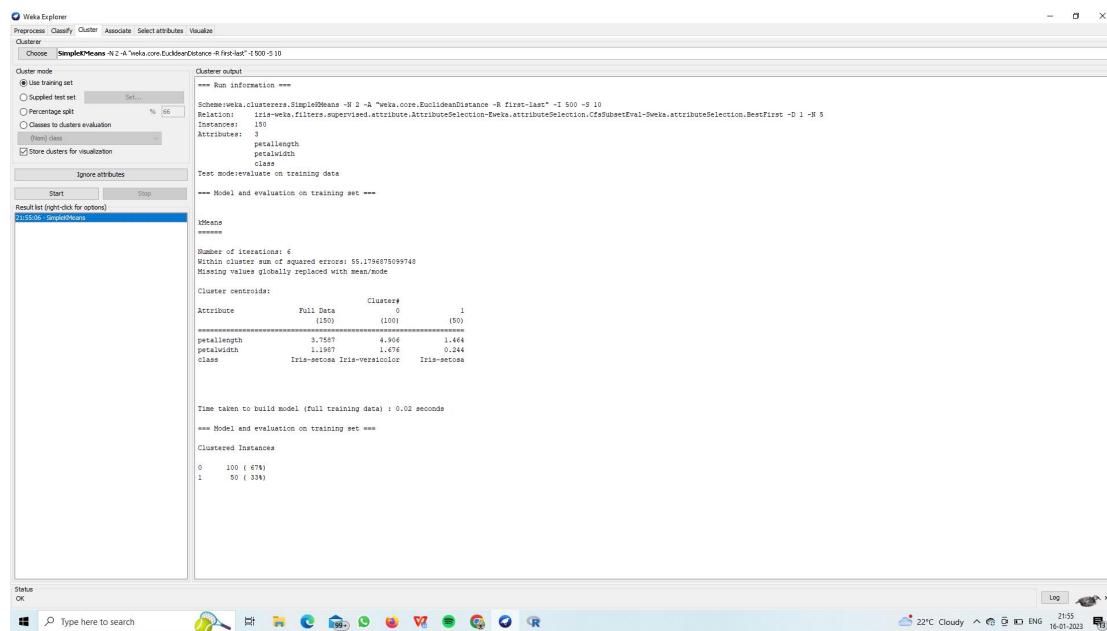
# CSA1668-DATAWAREHOUSING AND DATA MINING

E.SRIHARI  
192124168

## 1. DATA PREPROCESSING AND PREPARATION FOR KNOWLEDGE ANALYSIS USING WEKA.

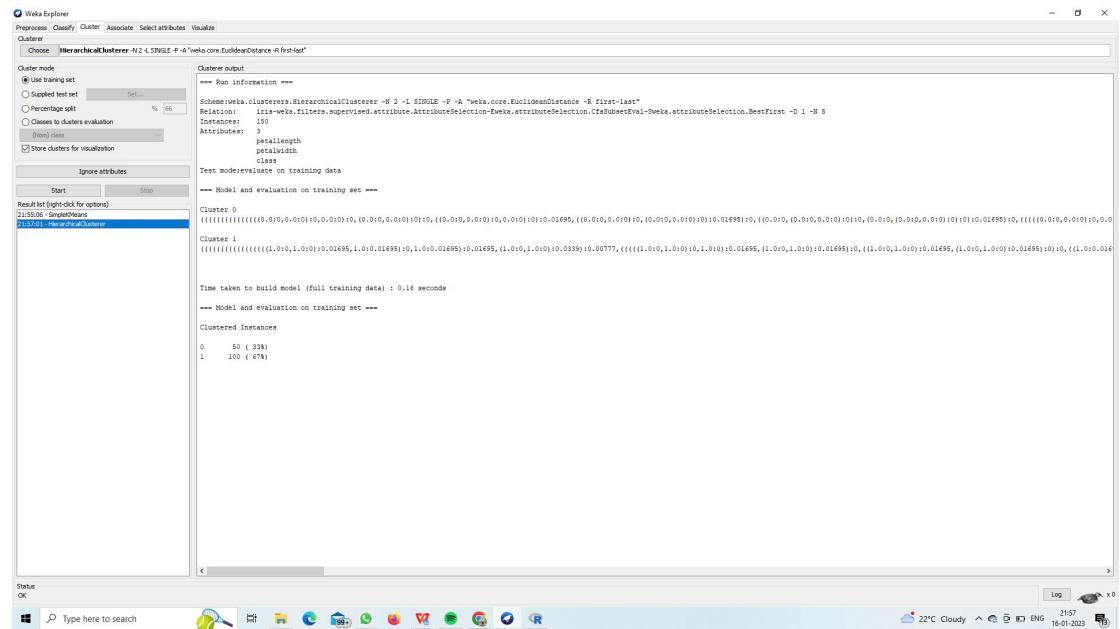


## 2.K-MEANS CLUSTER ANALYSIS USING WEKA.



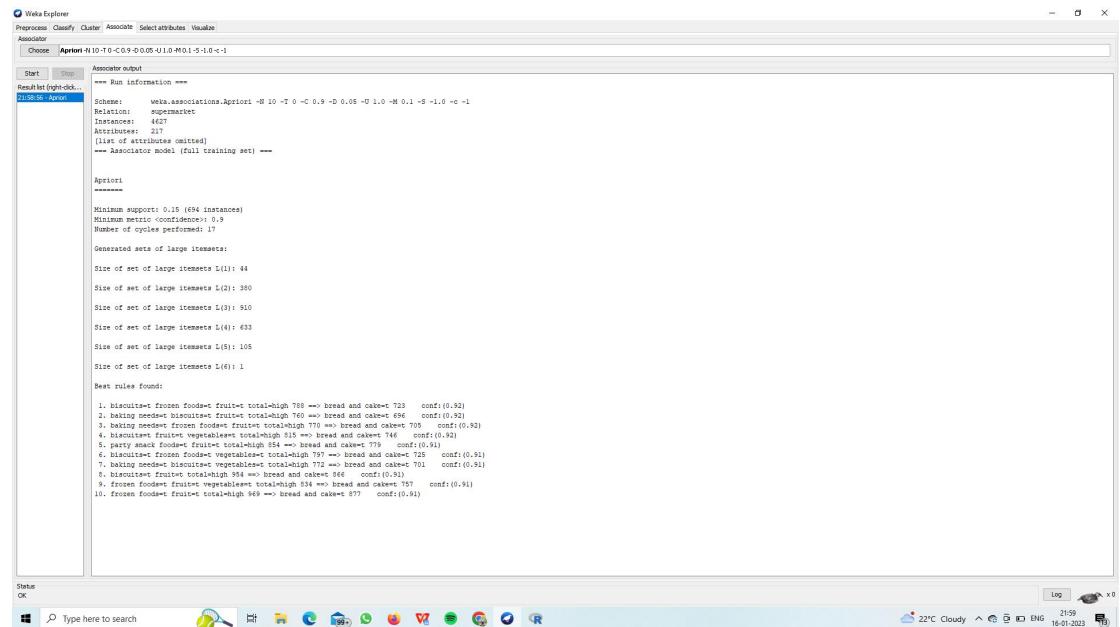
CSA1668-DATAWAREHOUSING AND DATA MINING

### **3. DATA ANALYSIS BY COBWEB-HIERARCHAL CLUSTERING ALGORITHM USING WEKA.**



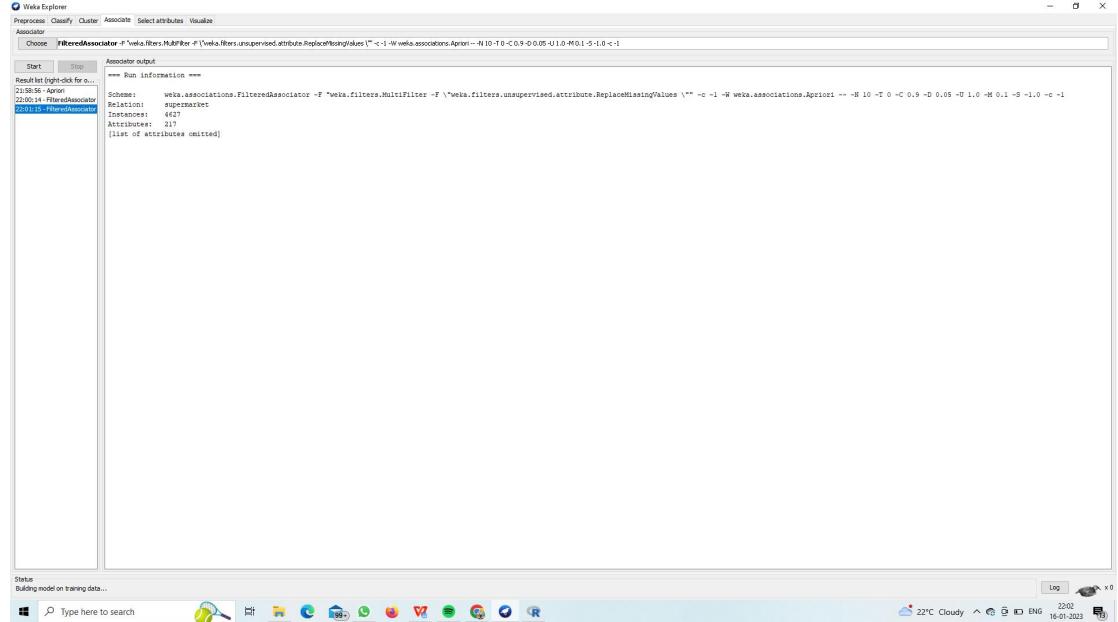
#### **4. KNOWLEDGE MINING USING ASSOCIATION RULE USING WEKA.**

**APRIORI:-**



# CSA1668-DATAWAREHOUSING AND DATA MINING

## FILTERED ASSOSIATOR:-

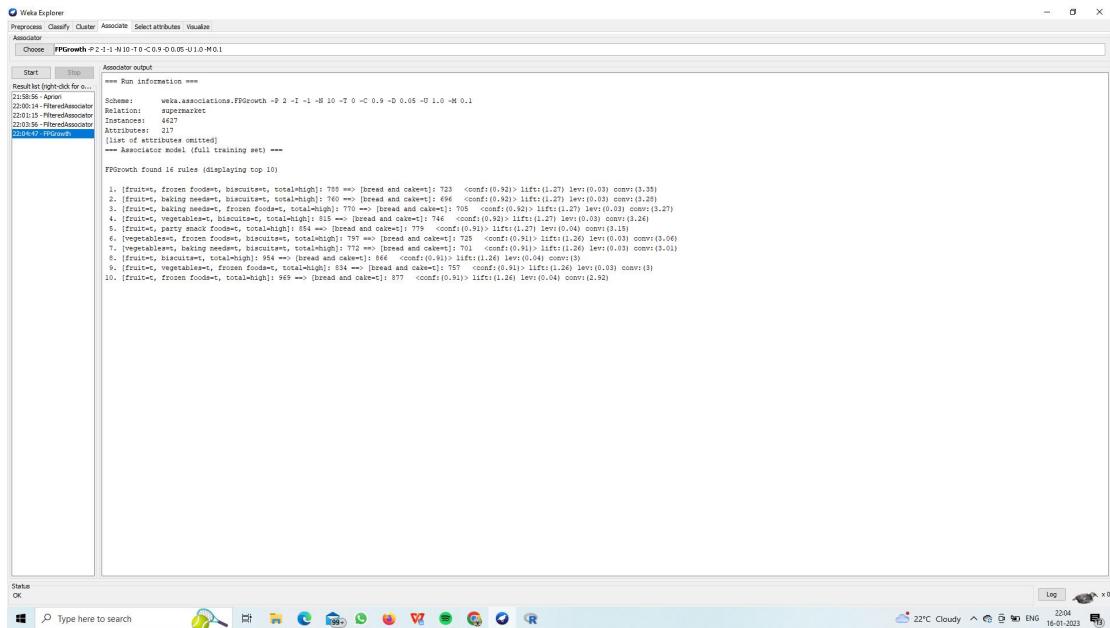


## FP GROWTH:-



# CSA1668-DATAWAREHOUSING AND DATA MINING

## 5. FP GROWTH ALGORITHM USING WEKA.

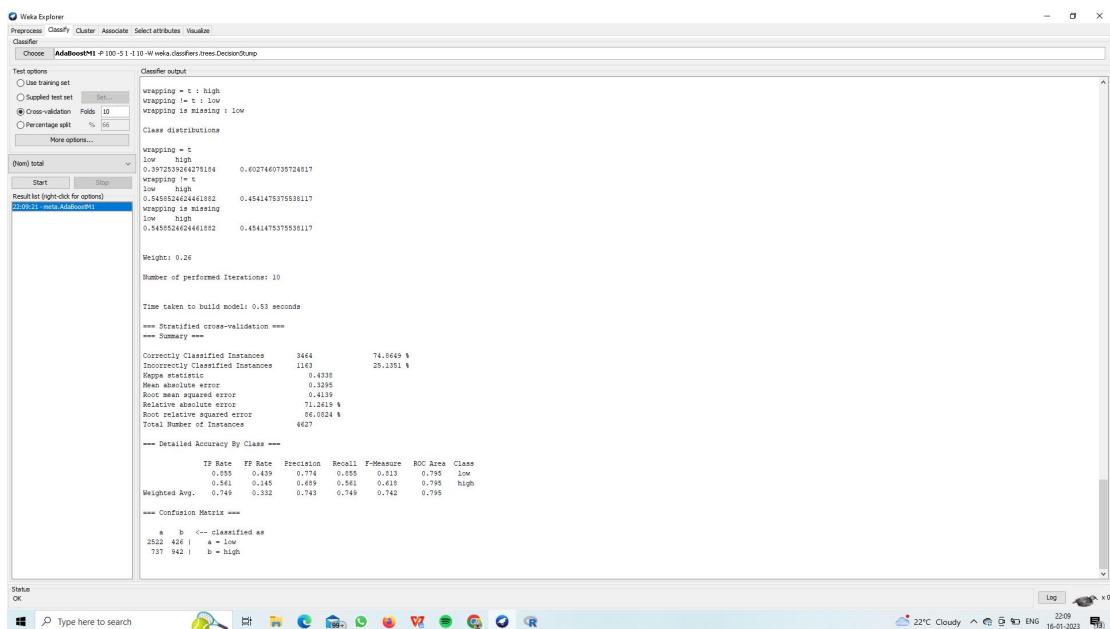


The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The classifier chosen is 'FPGrowth'. The log output displays the following information:

```
21:58:56 - Akron
Schemes: weka.associations.FPGrowth -P 2 -I -l -N 10 -T 0 -C 0.9 -D 0.05 -G 1.0 -M 0.1
Relation: supermarket
Instances: 4427
Attributes: 7
[list of attributes omitted]
*** Associate model (full training set) ***
FPGrowth found 14 rules (displaying top 10)

1. [fruit=+, frozen food=+, biscuits=+, total=high]: 785 => [bread and cake]=1; 723 <conf:(0.9)> lift:(1.27) lev:(0.03) conv:(3.38)
2. [fruit=+, baking need=+, biscuits=+, total=high]: 760 => [bread and cake]=1; 694 <conf:(0.9)> lift:(1.27) lev:(0.03) conv:(3.28)
3. [fruit=+, baking need=+, frozen food=+, total=high]: 770 => [bread and cake]=1; 705 <conf:(0.9)> lift:(1.27) lev:(0.03) conv:(3.27)
4. [fruit=+, vegetables=+, biscuits=+, total=high]: 801 => [bread and cake]=1; 796 <conf:(0.9)> lift:(1.27) lev:(0.03) conv:(3.24)
5. [vegetables=+, frozen food=+, biscuits=+, total=high]: 784 => [bread and cake]=1; 778 <conf:(0.9)> lift:(1.27) lev:(0.03) conv:(3.15)
6. [vegetables=+, frozen food=+, biscuits=+, total=high]: 735 => [bread and cake]=1; 725 <conf:(0.9)> lift:(1.26) lev:(0.03) conv:(3.06)
7. [vegetables=+, baking need=+, biscuits=+, total=high]: 772 => [bread and cake]=1; 701 <conf:(0.9)> lift:(1.26) lev:(0.03) conv:(3.01)
8. [fruit=+, vegetables=+, biscuits=+, total=high]: 757 => [bread and cake]=1; 748 <conf:(0.9)> lift:(1.26) lev:(0.03) conv:(2.92)
9. [fruit=+, vegetables=+, frozen food=+, total=high]: 834 => [bread and cake]=1; 757 <conf:(0.9)> lift:(1.26) lev:(0.03) conv:(3)
10. [fruit=+, frozen food=+, total=high]: 969 => [bread and cake]=1; 877 <conf:(0.9)> lift:(1.26) lev:(0.04) conv:(2.92)
```

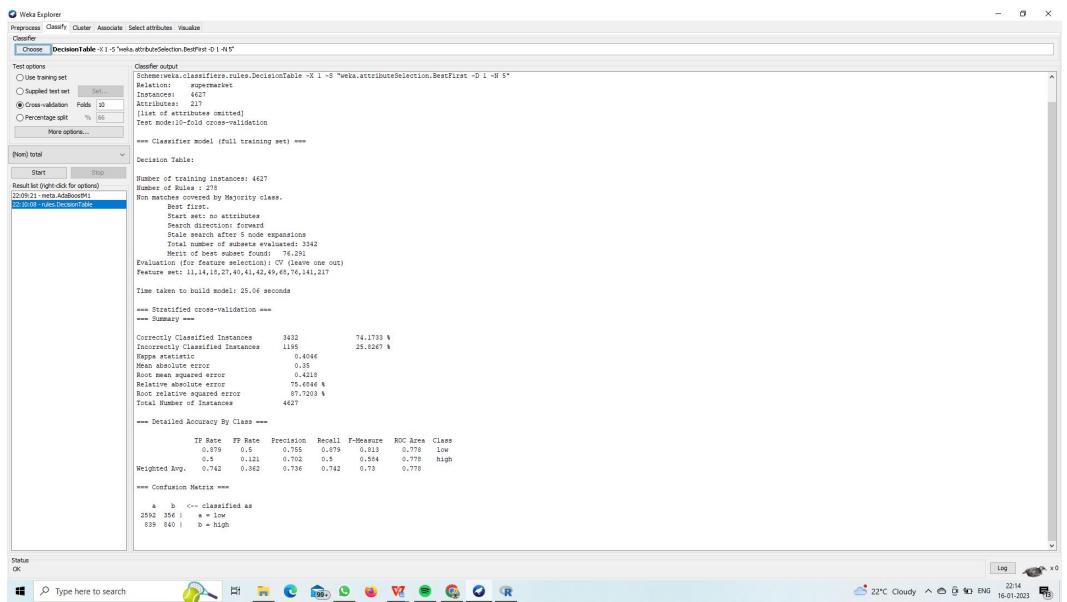
## 6. PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORITHM USING WEKA.



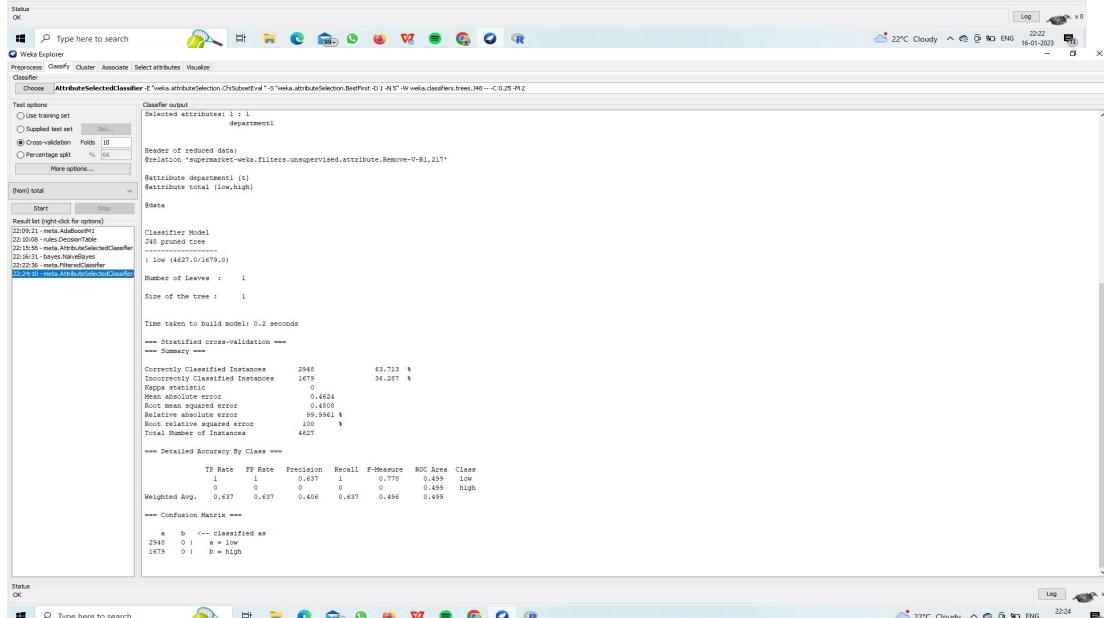
The screenshot shows the Weka Explorer interface with the 'Classifier' tab selected. The classifier chosen is 'AdaBoostM1'. The log output displays the following information:

```
23:09:21 - weka.classifiers.trees.Decision stump
Test options
 Use training set
 Supplied test set
 Cross-validation Folder: 10
 Percentage split %: 66
More options...
Classifier output
wrapping = t : high
wrapping != t : low
wrapping is missing : low
Class distributions
wrapping = t
low : 0.3972539242791814
high : 0.4027460735724817
wrapping != t
low : 0.5488524624441882
high : 0.4541475375538117
wrapping is missing
low : 0.4548524624441882
high : 0.4541475375538117
Weight: 0.26
Number of performed iterations: 10
Time taken to build model: 0.53 seconds
*** Stratified cross-validation ***
*** Summary ***
Correctly Classified Instances 3444 74.0649 %
Incorrectly Classified Instances 1163 25.1351 %
Kappa statistic 0.4338
Mean absolute error 0.3455
Root mean squared error 0.4139
Relative absolute error 71.2619 %
Root relative squared error 84.0824 %
Total Number of Instances 4627
*** Detailed Accuracy By Class ***
           TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
           0.755    0.439    0.774    0.675    0.617    0.795    low
           0.755    0.439    0.774    0.675    0.617    0.795    high
Weighted Avg.  0.749    0.532    0.749    0.749    0.742    0.795
*** Confusion Matrix ***
   a   b   <- classified as
2522 426 |   a = low
737 942 |   b = high
```

CSA1668-DATAWAREHOUSING AND DATA MINING

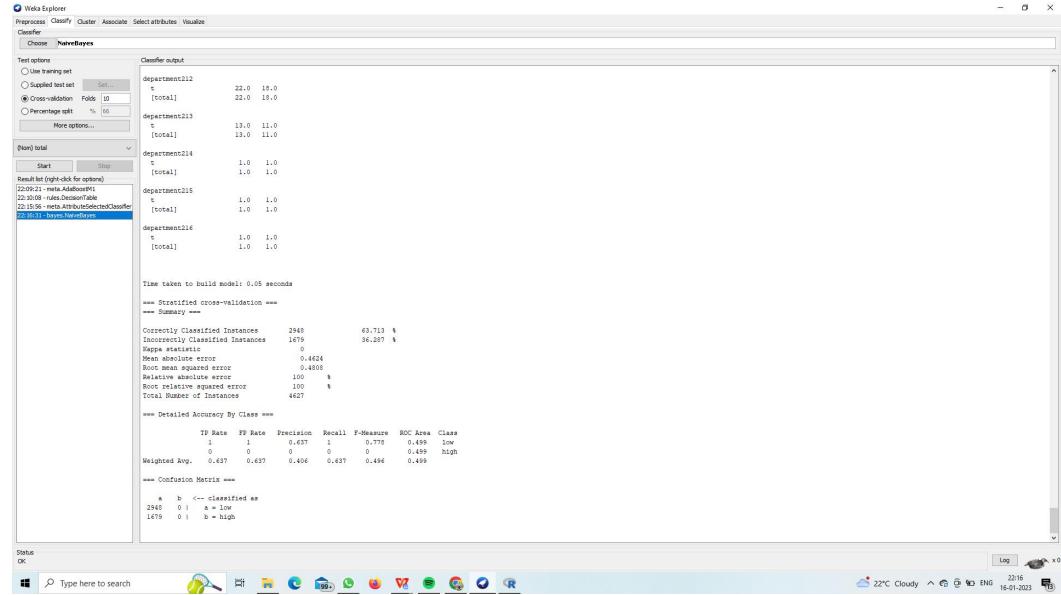


## **7. EVALUATING THE ACCURACY OF THE CLASSIFIERS USING WEKA.**

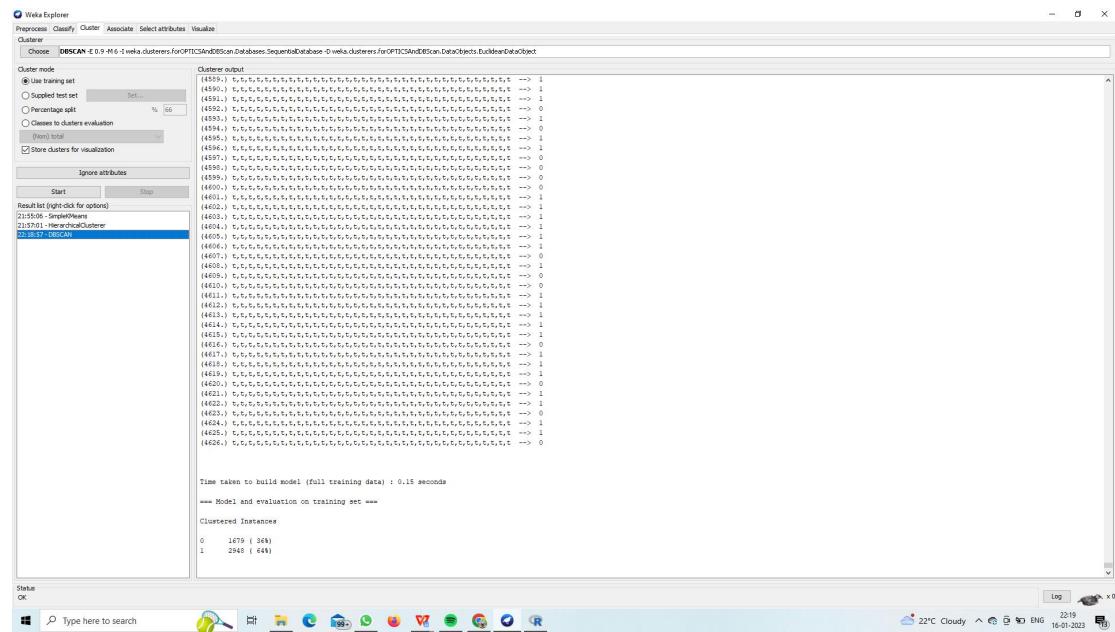


# CSA1668-DATAWAREHOUSING AND DATA MINING

## 8. PREDICTION OF CATEGORICAL DATA USING BAYESIAN ALGORITHM USING WEKA.



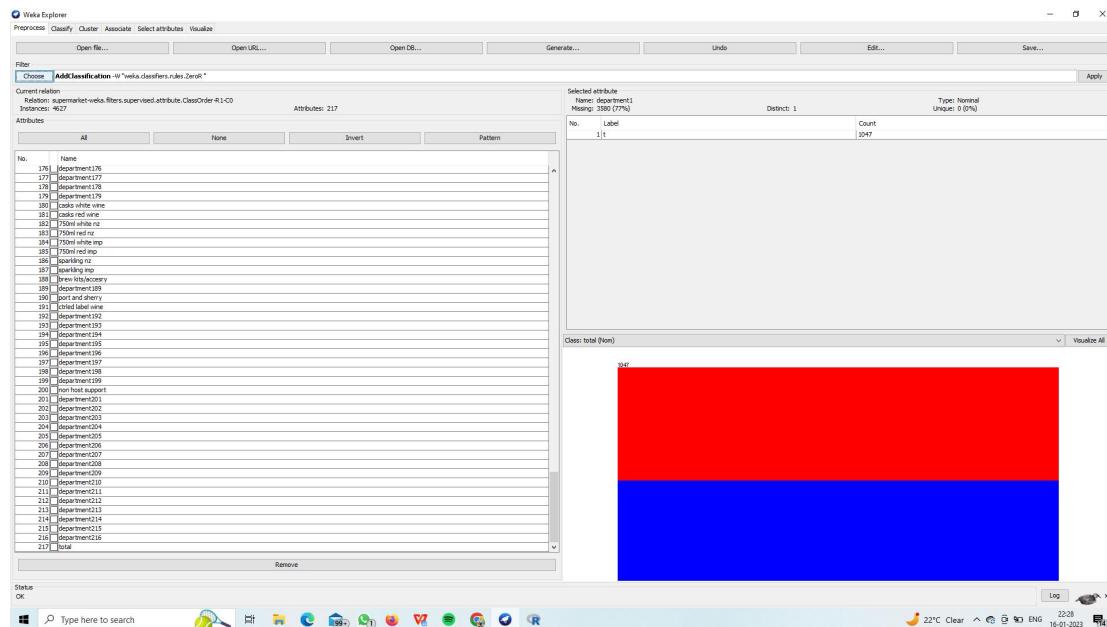
## 9. DATA ANALYSIS BY DENSITY BASED CLUSTERING ALGORITHM USING WEKA.



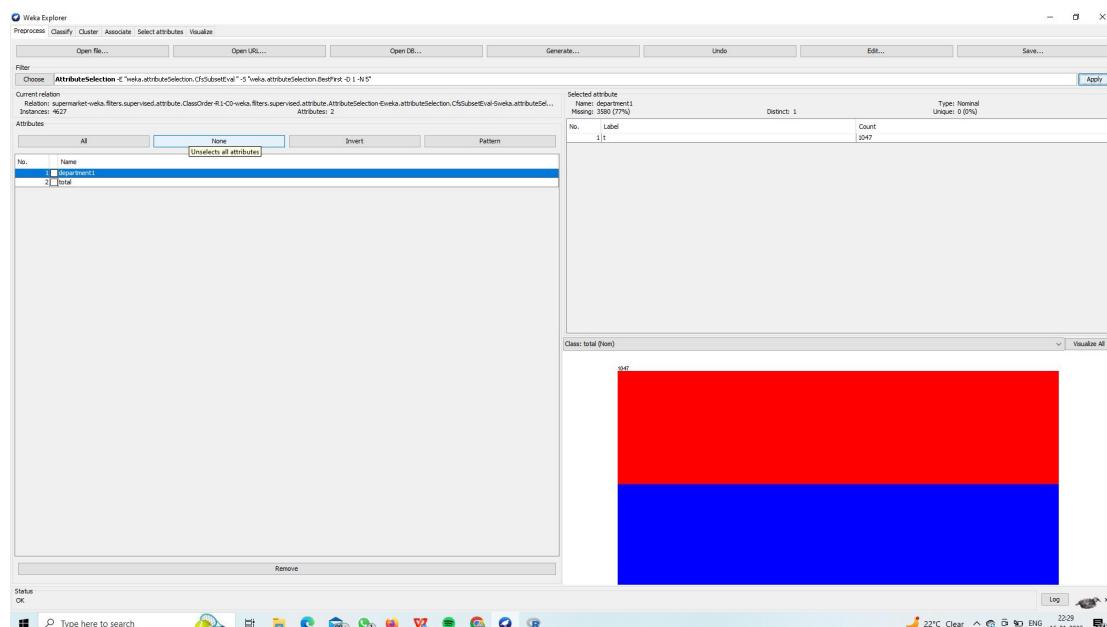
# CSA1668-DATAWAREHOUSING AND DATA MINING

## DATA PREPROCESSING (SUPERVISED)

### ADD CLASSIFICATION:-

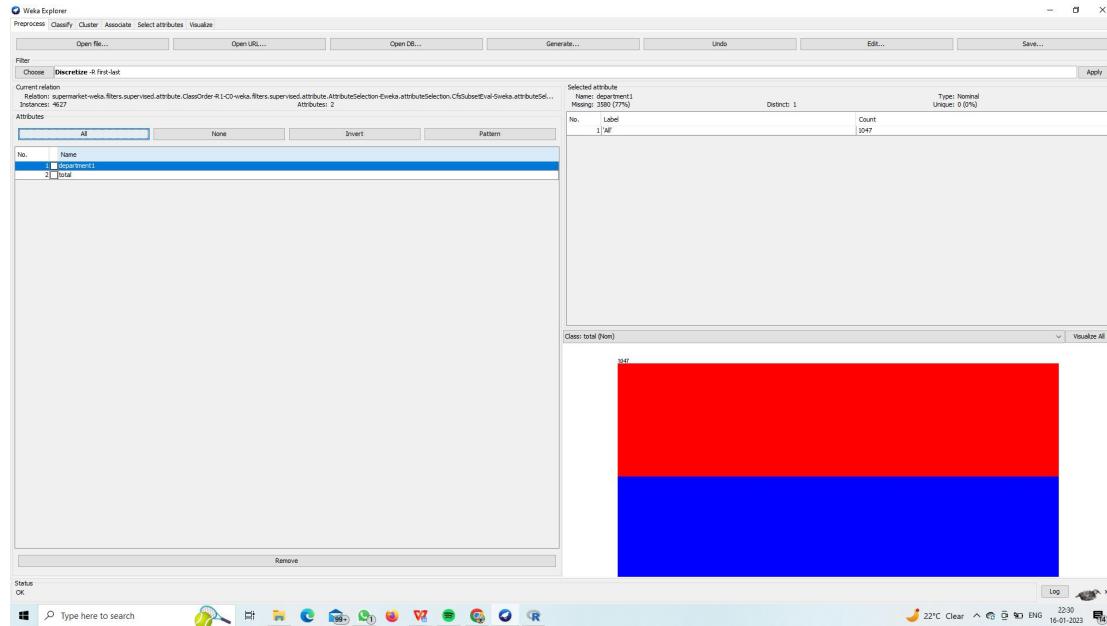


### ATTRIBUTE SELECTION:-

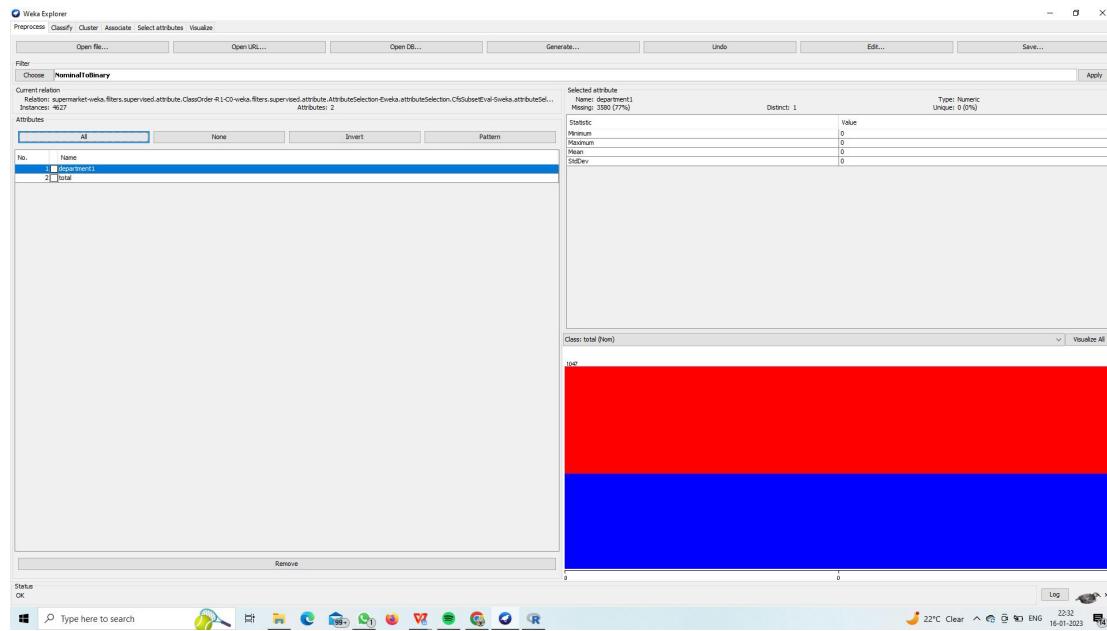


# CSA1668-DATAWAREHOUSING AND DATA MINING

## DISCRETIZE:-

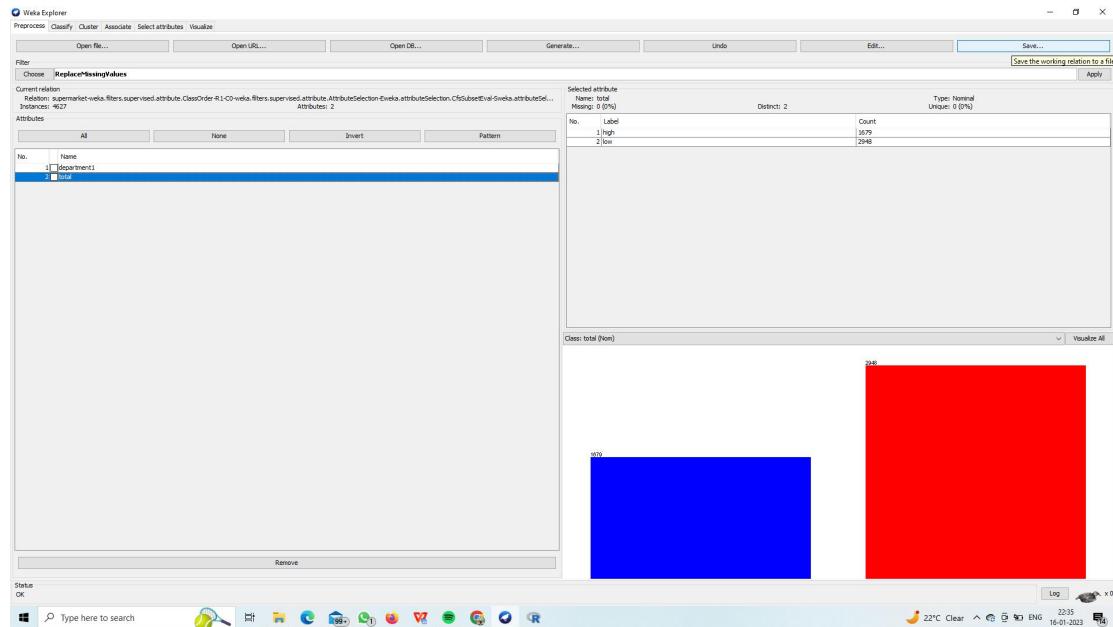


## NOMINAL TO BINARY:-

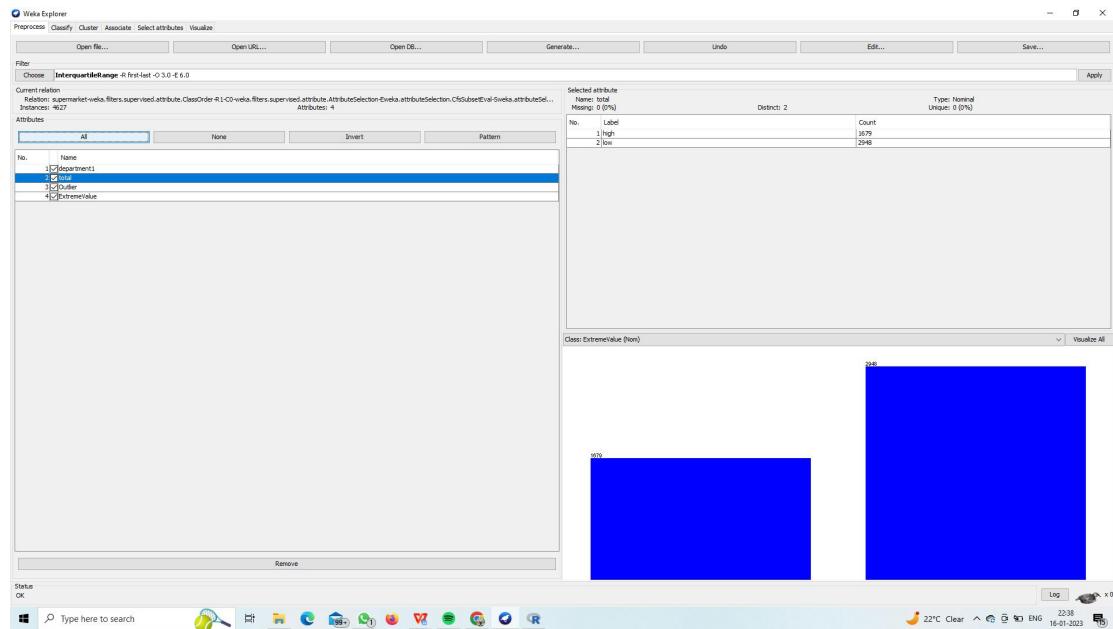


# CSA1668-DATAWAREHOUSING AND DATA MINING

(UNSUPERVISED)  
REPLACINGMISSING VALUES:-

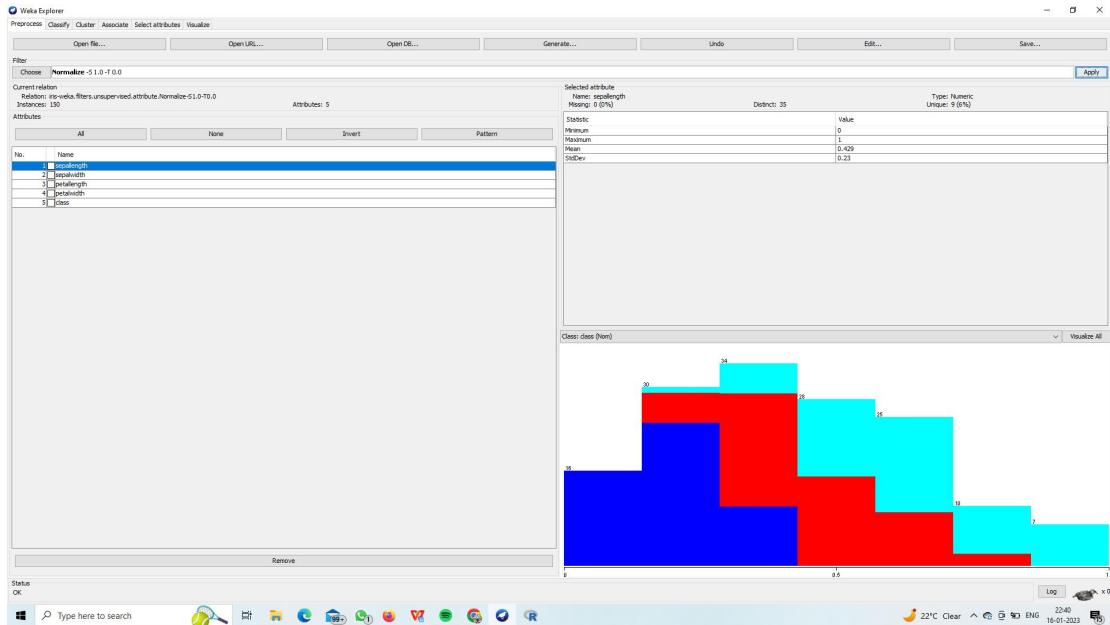


IQR:-



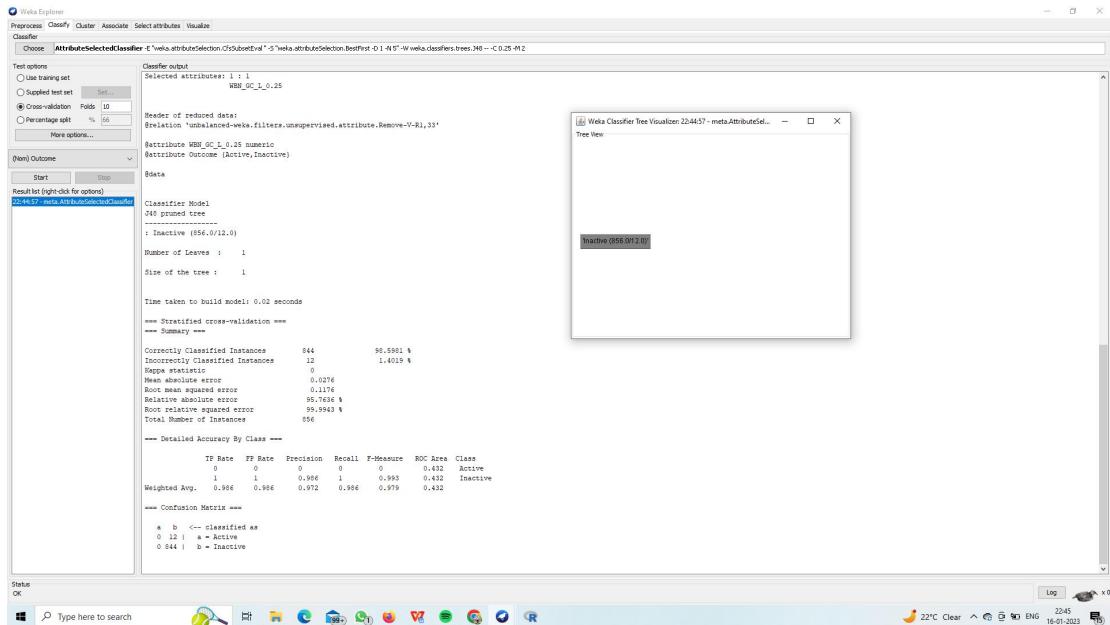
# CSA1668-DATAWAREHOUSING AND DATA MINING

## NORMALIZE:-

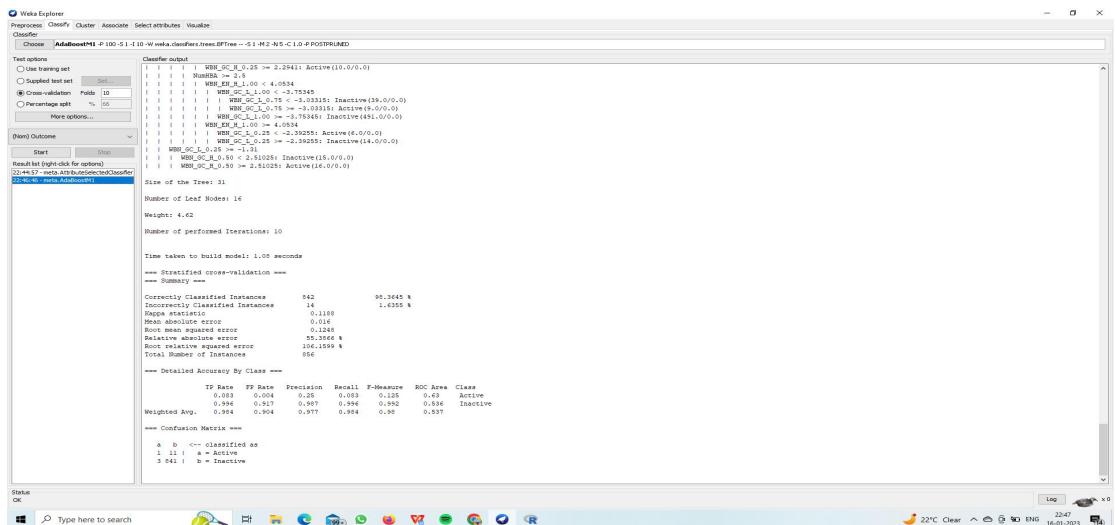


## CLASSIFICATIONS:-

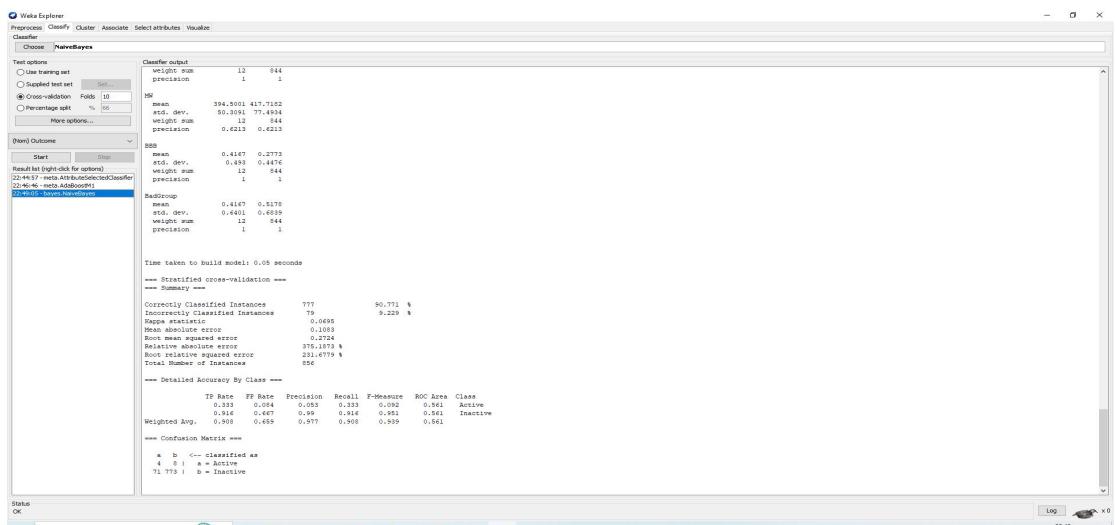
### ATTRIBUTE SELECTION CLASSIFIER:-\



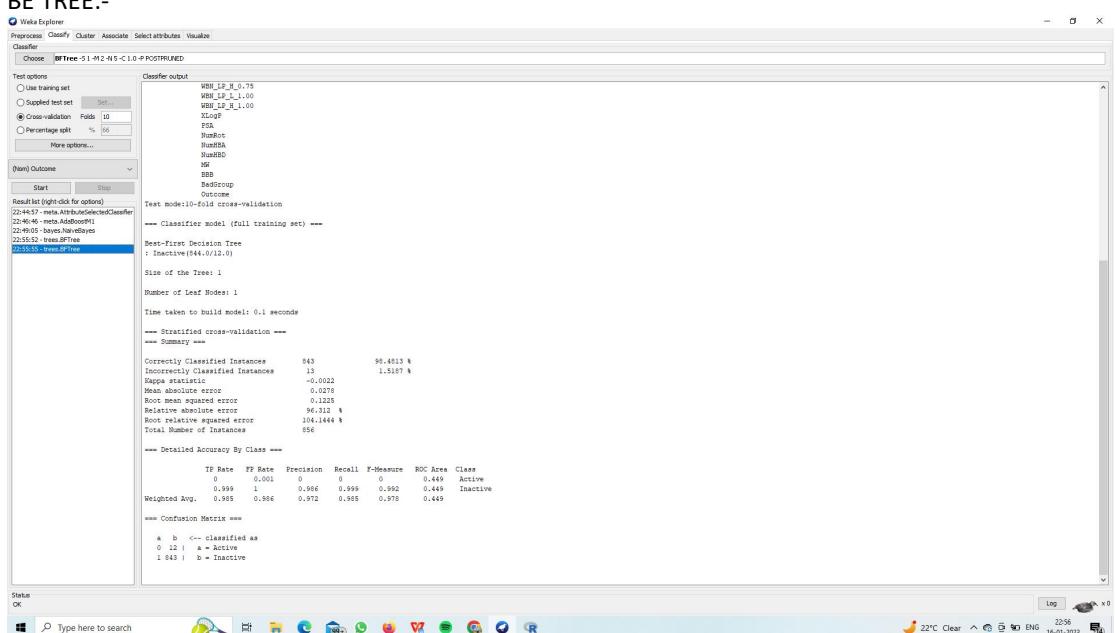
CSA1668-DATAWAREHOUSING AND DATA MINING



## NAIVEBAYES:



BE TRUE



CSA1668-DATAWAREHOUSING AND DATA MINING

## DECISION TREE:-

The screenshot shows the Weka Explorer interface with the following details:

- Test options:** Use training set, Supplied test set (selected), Cross-validation (set to 10), Percentage split (set to 60), More options...
- Classifier output:** WINE\_ID\_1\_1.00, WINE\_ID\_1\_1.00, XLogit, PRA, Buseck, NaiveBk, Random, NB, BBR, Radecup, Outcome.
- Test mode:** 10-fold cross-validation.
- Result list (right-click for options):** J48 (3-class classification tree) (selected), DecisionTable, PostPruning, ROC, ConfusionMatrix, Evaluation, FeatureImportance, FeatureWeights, ModelSummary, Log.
- Decision Table:** (displayed)
- Number of training instances:** 856
- Number of Rules:** 3
- Rules recorded by Majority Class:**
- Evaluation (for feature selection):** CV (leave one out)
- Feature set:** 33
- Time taken to build model:** 9.4 seconds
- Stratified cross-validation:** (displayed)
- Classification results:**

	Correctly Classified Instances	Incorrectly Classified Instances	%
Overall	544	312	60.5881 %
By Class			
0	544	0	100.0000 %
1	0	312	100.0000 %
- Kappa statistic:** 0
- Mean absolute error:** 0.0288
- Root mean squared error:** 0.1176
- Relative absolute error:** 66.4317 %
- Root relative squared error:** 79.9979 %
- Total Number of Instances:** 856
- Detailed Accuracy By Class:**

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.445	0.445	Active
1	1	0.3956	0.3956	1	0.993	0.449	Inactive
- Weighted Avg.:** 0.592, 0.862, 0.972, 0.596, 0.579, 0.449
- Confusion Matrix:**

	0	1
0	544	0
1	0	312
- Legend:** 0 = B (Inactive), 1 = A (Active).

### FILTERED CLSIFIER:-

The screenshot shows the Weka Explorer interface with the following details:

- Progress:** Classify Cluster Associate Select attributes Visualize
- Classifier:** **FilteredClassifier** (selected) > weka.libsvm.LibSVM, specified attribute: Distance - R first last > weka.classifiers.trees.J48 -- C=0.25 M=2
- Test options:**
  - Use training set
  - Supplied test set
  - Cross-validation** Folds 10
  - Percentage % 66
  - More options...
- (Name) Outcome:**
- Start**
- Reset**  **Cancel**  **Run**
- Result list (click for options):**
  - 22:45:57: recta.AdditiveSmootherClassifier
  - 22:46:46: recta.AdditiveModel1
  - 22:47:33: tree.J48Tree
  - 22:47:52: tree.JPTree
  - 22:50:01: rules.RTRuleSet
  - 22:50:30: rules.RTRuleSet
  - 22:50:30: weka.FilteredClassifier
- Classifier output:**

```
Attributes XLogit ("\\All\\")  
Attributes YLogit ("\\All\\")  
Attributes NumLogit ("\\All\\")  
Attributes NumHab ("\\All\\")  
Attributes SepalLength ("\\All\\")  
Attributes SepalWidth ("\\All\\")  
Attributes PetalLength ("\\All\\")  
Attributes PetalWidth ("\\All\\")  
Attributes SepalGroup ("\\All\\")  
Attributes Outcome (Active,Inactive)
```

**Results:**

```
22:45:57 recta.AdditiveSmootherClassifier  
22:46:46 recta.AdditiveModel1  
22:47:33 tree.J48Tree  
22:47:52 tree.JPTree  
22:50:01 rules.RTRuleSet  
22:50:30 rules.RTRuleSet  
22:50:30 weka.FilteredClassifier  
  
Classifier Model:  
J48 pruned tree  
=====  
Invasive (0.50/0.12,0)  
Number of Leaves : 1  
Size of the tree : 1  
Size of the root : 1  
  
Time taken to build model: 0.01 seconds  
  
== Stratified cross-validation ==  
== Summary ==  
  
Correctly Classified Instances 544 50.000 %  
Incorrectly Classified Instances 12 1.000 %  
Kappa statistic 0  
Mean absolute error 0.0374  
Root mean squared error 0.1176  
Relative standard deviation 94.740 %  
Root relative squared error 99.040 %  
Total Number of Instances 556  
  
== Detailed Accuracy By Class ==  


|   | TP Rate | FP Rate | Precision | Recall | F-Measure | ECC Rate | Class    |
|---|---------|---------|-----------|--------|-----------|----------|----------|
| 0 | 1       | 0       | 0.992     | 1      | 0.993     | 0.432    | Active   |
| 1 | 0       | 1       | 0.996     | 0      | 0.993     | 0.432    | Inactive |



Weighted Avg. 0.995 0.996 0.992 0.996 0.979 0.432

  
Confusion Matrix ==

|   | a   b | <- Classified as |
|---|-------|------------------|
| a | 544   | a = Active       |
| b | 12    | b = Inactive     |

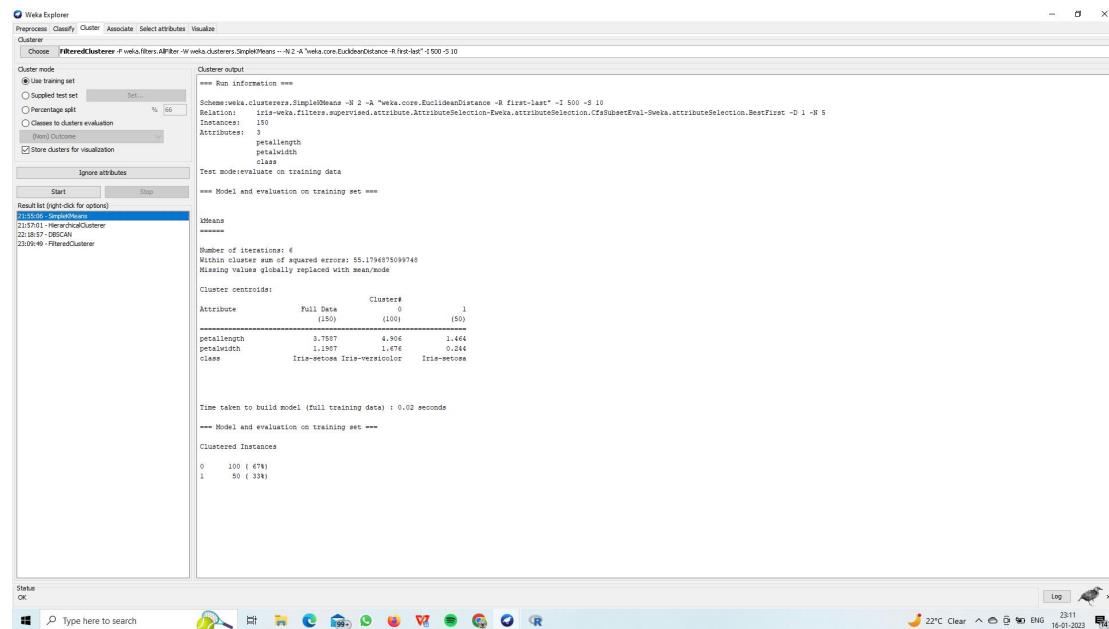

```
- Status:** OK
- Log:**
- Type here to search:**
- Icons:**
- System Tray:** 22°C Clear ENG 16:40 2015

## **CLUSTERING:-**

### **FILTERED CLUSTRING:-**

CSA1668-DATAWAREHOUSING AND DATA MINING

## **SIMPLEKMEANS:-**



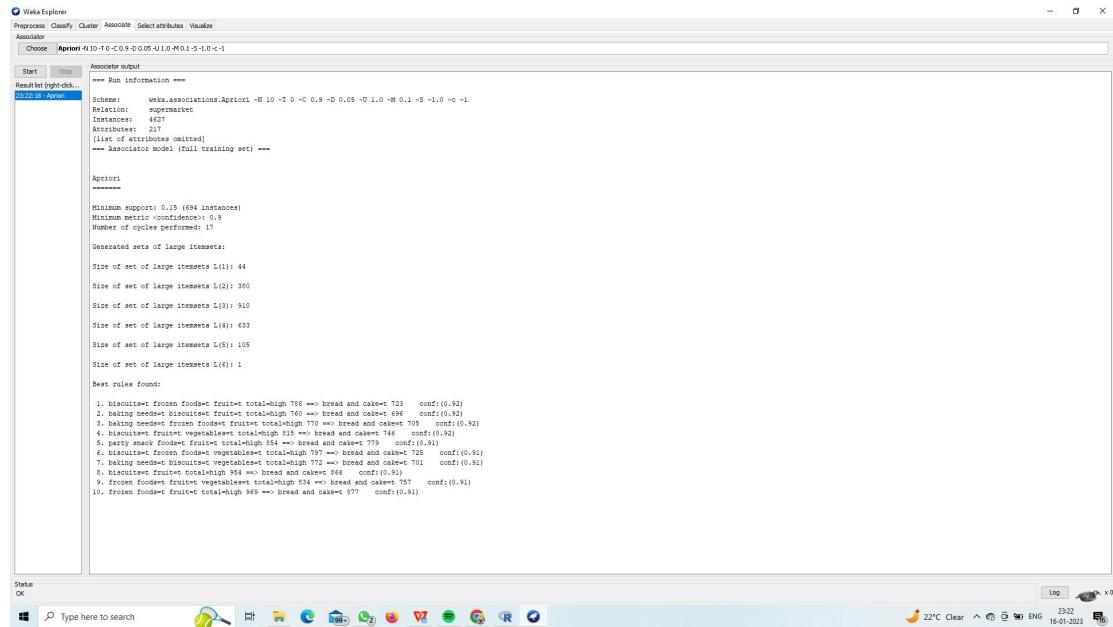
### **HIERARCHICAL CLUSTERING:-**



# CSA1668-DATAWAREHOUSING AND DATA MINING

## ASSOCIATIVE:-

### APRIORI ALGORITHM:-



The screenshot shows the Weka Explorer interface with the "Associate" tab selected. The "Choose" dropdown is set to "Apriori -N 10 -T 0.9 -D 0.05 -U 1.0 -M 0.1 -S 1.0 -c 1". The "Start" button is highlighted. The output window displays the following information:

```
Weka Explorer
Process Classify Cluster Associate Select attributes Visualize
Choose Apriori -N 10 -T 0.9 -D 0.05 -U 1.0 -M 0.1 -S 1.0 -c 1
Result list (right-click for details)
23:23:14 - April
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: supermarket
Instances: 4627
Attributes: 217
[111 of 217 attributes omitted]
*** Associate model (full training set) ***

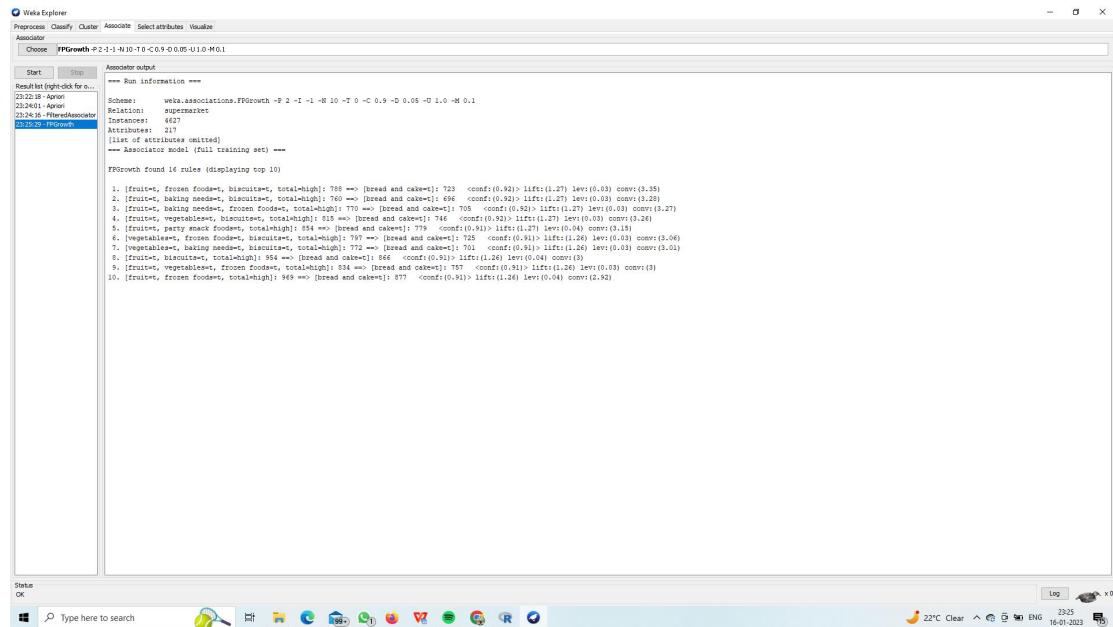
Apriori
-----
Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

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=frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 conf:(0.92)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 conf:(0.92)
3. baking needs=frozen foods=t fruit=t total=high 771 ==> bread and cake=t 759 conf:(0.92)
4. biscuits=frozen foods=t vegetables=t fruit=t total=high 744 ==> bread and cake=t 701 conf:(0.92)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 conf:(0.91)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 conf:(0.91)
7. biscuits=t frozen foods=t vegetables=t total=high 854 ==> bread and cake=t 779 conf:(0.91)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 conf:(0.91)
9. frozen foods=fruit vegetables=t total=high 834 ==> bread and cake=t 757 conf:(0.91)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 conf:(0.91)

Status OK
Log 22°C Clear 23:22 ENG 16-01-2023
```

### FP GROWTH ALGORITHM:-



The screenshot shows the Weka Explorer interface with the "Associate" tab selected. The "Choose" dropdown is set to "FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1". The "Start" button is highlighted. The output window displays the following information:

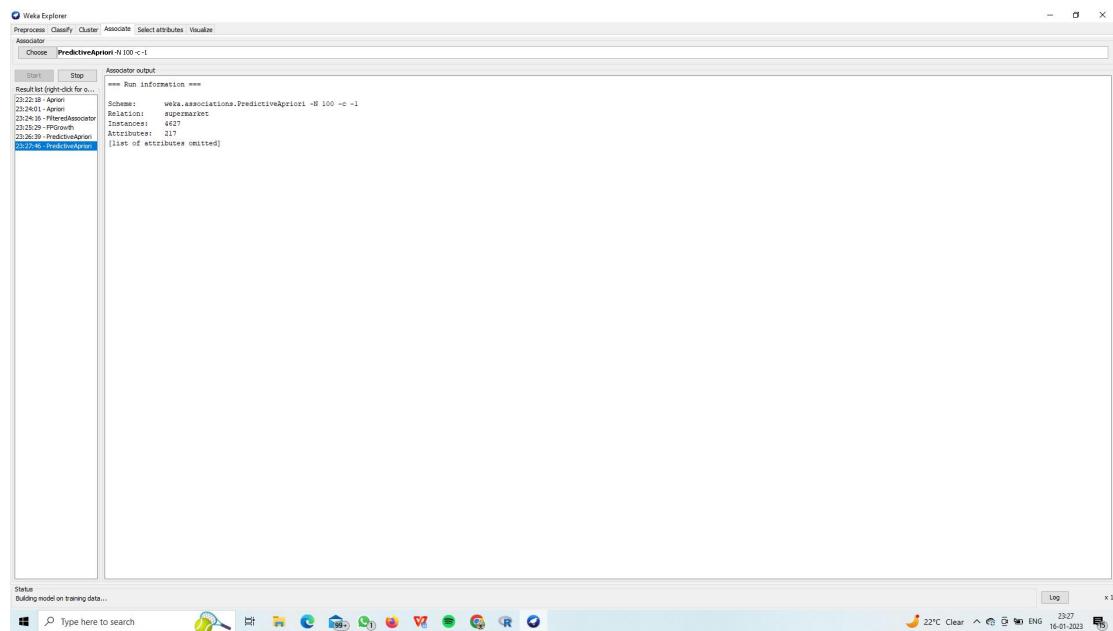
```
Weka Explorer
Process Classify Cluster Associate Select attributes Visualize
Choose FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Result list (right-click for details)
23:22:18 - April
23:22:19 - April
23:24:16 - April
23:24:16 - FPGrowth
23:25:29 - FPGrowth
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: supermarket
Instances: 4627
Attributes: 217
[111 of attributes omitted]
*** Associate model (full training set) ***

FPGrowth found 14 rules (displaying top 10)
1. [fruit=t, frozen foods=t, biscuits=t, total=high]: 788 ==> [bread and cake=t]: 723 conf:(0.92)> lift:(1.27) lev:(0.03) conv:(1.35)
2. [fruit=t, baking needs=t, biscuits=t, total=high]: 760 ==> [bread and cake=t]: 696 conf:(0.92)> lift:(1.27) lev:(0.03) conv:(1.20)
3. [fruit=t, baking needs=t, frozen foods=t, total=high]: 770 ==> [bread and cake=t]: 759 conf:(0.92)> lift:(1.27) lev:(0.03) conv:(1.27)
4. [fruit=t, biscuits=t, vegetables=t, total=high]: 744 ==> [bread and cake=t]: 701 conf:(0.92)> lift:(1.27) lev:(0.03) conv:(1.24)
5. [fruit=t, party snack foods=t, total=high]: 854 ==> [bread and cake=t]: 779 conf:(0.91)> lift:(1.27) lev:(0.04) conv:(1.15)
6. [vegetables=t, frozen foods=t, biscuits=t, total=high]: 797 ==> [bread and cake=t]: 725 conf:(0.91)> lift:(1.26) lev:(0.03) conv:(1.06)
7. [fruit=t, biscuits=t, total=high]: 854 ==> [bread and cake=t]: 779 conf:(0.91)> lift:(1.26) lev:(0.03) conv:(1.01)
8. [fruit=t, biscuits=t, total=high]: 954 ==> [bread and cake=t]: 866 conf:(0.91)> lift:(1.24) lev:(0.04) conv:(0.9)
9. [fruit=t, vegetables=t, frozen foods=t, total=high]: 834 ==> [bread and cake=t]: 757 conf:(0.91)> lift:(1.24) lev:(0.03) conv:(0.9)
10. [fruit=t, frozen foods=t, total=high]: 969 ==> [bread and cake=t]: 877 conf:(0.91)> lift:(1.24) lev:(0.04) conv:(0.9)

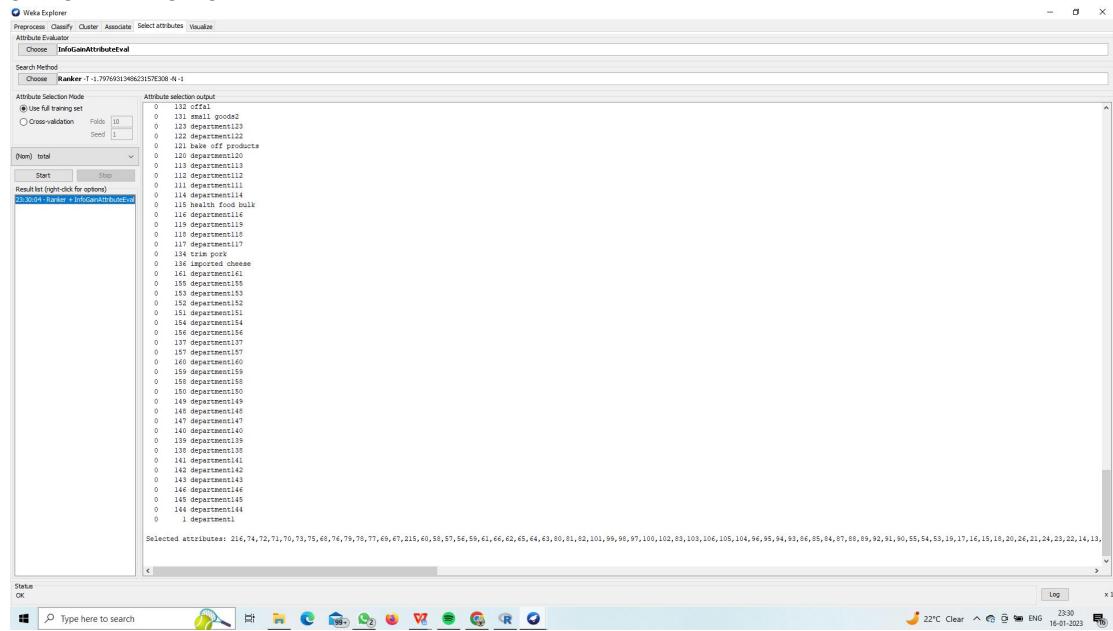
Status OK
Log 22°C Clear 23:25 23:25 ENG 16-01-2023
```

# CSA1668-DATAWAREHOUSING AND DATA MINING

## PREDICTIVE APRIORI:-

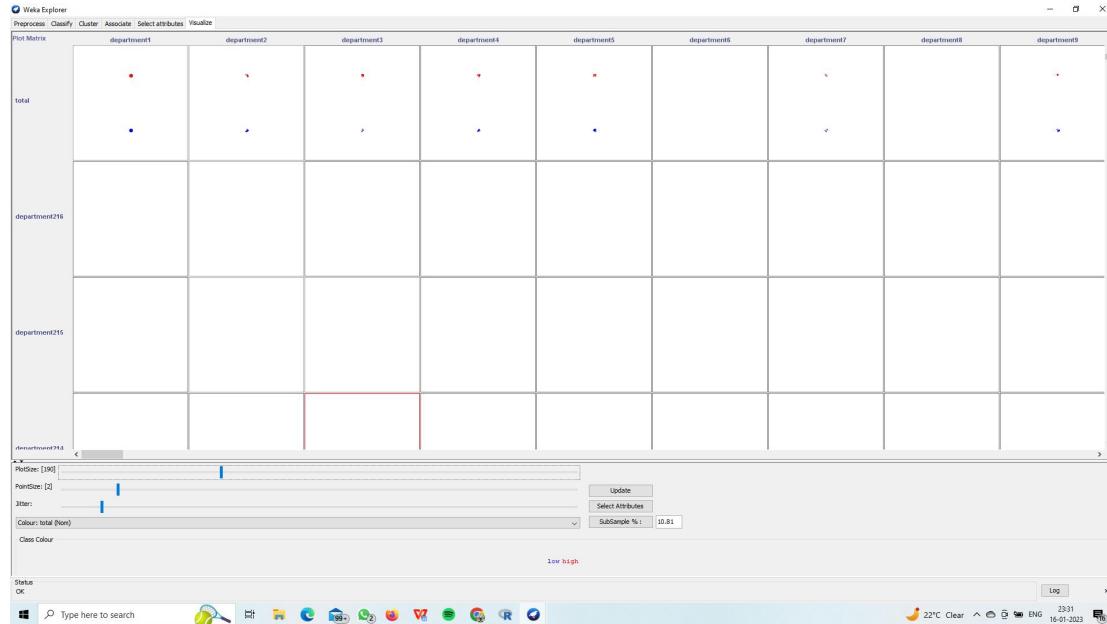


## SELECTATTRIBUTES:



# CSA1668-DATAWAREHOUSING AND DATA MINING

## VISUALIZATION:-



## 11. IN CLSSIFICATION,RANDOM TREE VISUALIZATION.

