# SET-01

## Create a Banking Table with the help of Data Mining Tool WEKA

1) Open Start  Programs  Accessories  Notepad

2) Type the following training data set with the help of Notepad for Banking Table. @relation bank

@attribute cust {male,female}

@attribute accno {0101,0102,0103,0104,0105,0106,0107,0108,0109,0110,0111,0112,0113,0114,0115}

@attribute bankname {sbi,hdfc,sbh,ab,rbi}

@attribute location {hyd,jmd,antp,pdtr,kdp}

@attribute deposit {yes,no}

@data

male,0101,sbi,hyd,yes

female,0102,hdfc,jmd,no

male,0103,sbh,antp,yes

male,0104,ab,pdtr,yes

female,0105,sbi,jmd,no

male,0106,ab,hyd,yes

female,0107,rbi,jmd,yes

female,0108,hdfc,kdp,no

male,0109,sbh,kdp,yes

male,0110,ab,jmd,no

female,0111,rbi,kdp,yes

male,0112,sbi,jmd,yes

female,0113,rbi,antp,no

male,0114,hdfc,pdtr,yes

female,0115,sbh,pdtr,no

3) After that the file is saved with .arff file format.

4) Minimize the arff file and then open Start  Programs  weka-3-4.

5) Click on weka-3-4, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on explorer.

7) Explorer shows many options. In that click on ‘open file’ and select the arff file

8) Click on edit button which shows banking table on weka.

## Perform the basic pre-processing operations on data relation such as removing an attribute and filter attribute bank data

**Aim** Perform the basic pre-processing operations on data relation such as removing an attribute and filter attribute bank data.

**Dataset:**

@relation bank

@attribute cust {male,female}

@attribute accno {0101,0102,0103,0104,0105,0106,0107,0108,0109,0110,0111,0112,0113,0114,0115}

@attribute bankname {sbi,hdfc,sbh,ab,rbi}

@attribute location {hyd,jmd,antp,pdtr,kdp}

@attribute deposit {yes,no}

@data

male,0101,sbi,hyd,yes

female,0102,hdfc,jmd,no

male,0103,sbh,antp,yes

male,0104,ab,pdtr,yes

female,0105,sbi,jmd,no

male,0106,ab,hyd,yes

female,0107,rbi,jmd,yes

female,0108,hdfc,kdp,no

male,0109,sbh,kdp,yes

male,0110,ab,jmd,no

female,0111,rbi,kdp,yes

male,0112,sbi,jmd,yes

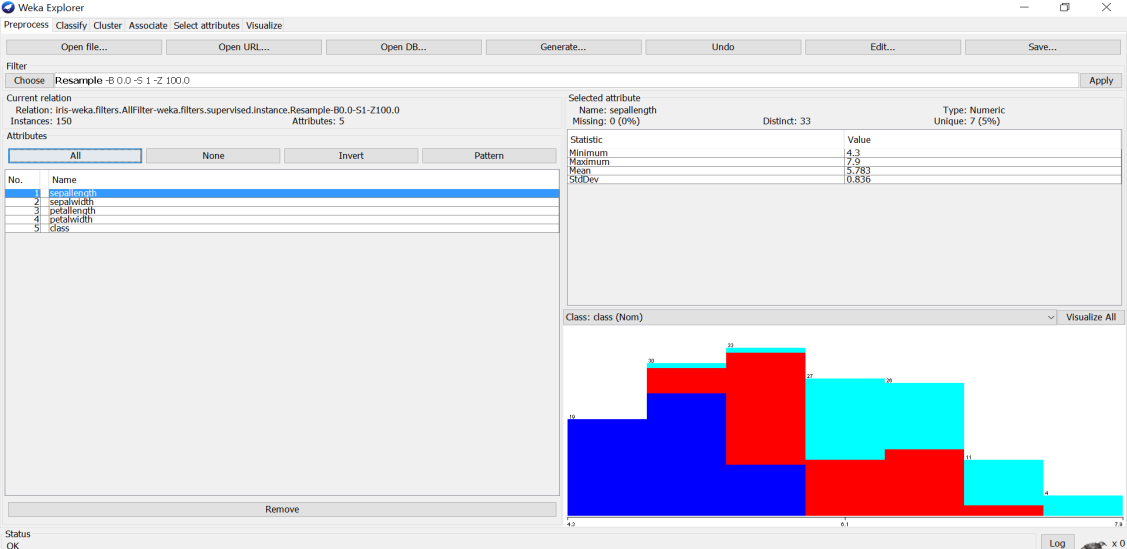
female,0113,rbi,antp,no

male,0114,hdfc,pdtr,yes

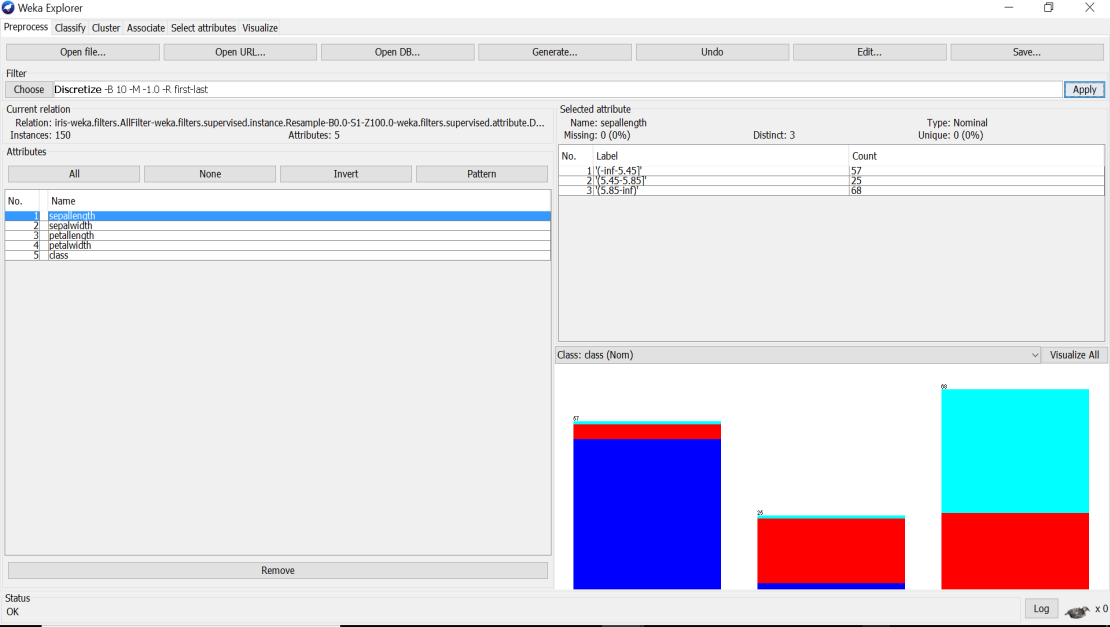
female,0115,sbh,pdtr,no

**Procedure:**

* + 1. For preprocessing the data after selecting the dataset (IRis.arff).
    2. Select Filter option & apply the resample filter & see the below results.



1. Select another filter option & apply the discretization filter, see the below results



Likewise, we can apply different filters for preprocessing the data & see the results in different dimensions.

# SET-02

## **1***.*Create a Customer Table with the help of Data Mining Tool WEKA.

Creation of Customer Table:

Procedure:

1) Open Start  Programs  Accessories  Notepad

2) Type the following training data set with the help of Notepad for Customer Table. @relation customer

@attribute name {x,y,z,u,v,l,w,q,r,n}

@attribute age {youth,middle,senior}

@attribute income {high,medium,low}

@attribute class {A,B}

@data

x,youth,high,A

y,youth,low,B

z,middle,high,A

u,middle,low,B

v,senior,high,A

l,senior,low,B

w,youth,high,A

q,youth,low,B

r,middle,high,A

n,senior,high,A

3) After that the file is saved with .arff file format.

4) Minimize the arff file and then open Start  Programs  weka-3-4.

5) Click on weka-3-4, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on explorer.

7) Explorer shows many options. In that click on ‘open file’ and select the arff file

8) Click on edit button which shows customer table on weka.

## 2. Load each dataset into Weka and perform Naive-bayes classification (contact-lenses). Interpret the results obtained

**Dataset:**

@relation contact-lenses

@attribute age {young, pre-presbyopic, presbyopic}

@attribute spectacle-prescrip {myope, hypermetrope}

@attribute astigmatism {no, yes}

@attribute tear-prod-rate {reduced, normal}

@attribute contact-lenses {soft, hard, none}

@data

young,myope,no,reduced,none

young,myope,no,normal,soft

young,myope,yes,reduced,none

young,myope,yes,normal,hard

young,hypermetrope,no,reduced,none

young,hypermetrope,no,normal,soft

young,hypermetrope,yes,reduced,none

young,hypermetrope,yes,normal,hard

pre-presbyopic,myope,no,reduced,none

pre-presbyopic,myope,no,normal,soft

pre-presbyopic,myope,yes,reduced,none

pre-presbyopic,myope,yes,normal,hard

pre-presbyopic,hypermetrope,no,reduced,none

pre-presbyopic,hypermetrope,no,normal,soft

pre-presbyopic,hypermetrope,yes,reduced,none

pre-presbyopic,hypermetrope,yes,normal,none

presbyopic,myope,no,reduced,none

presbyopic,myope,no,normal,none

presbyopic,myope,yes,reduced,none

presbyopic,myope,yes,normal,hard

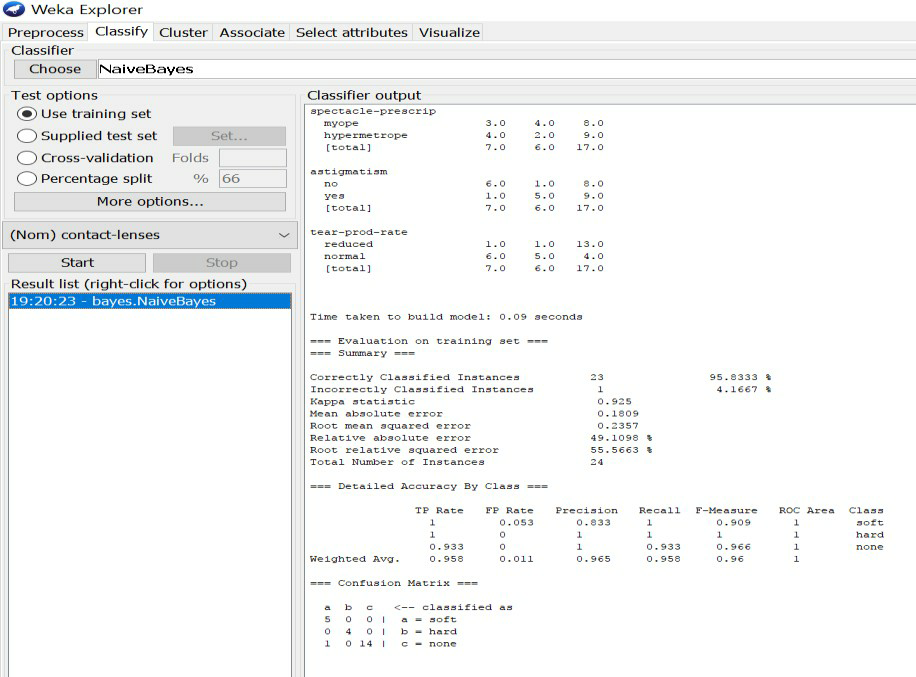
presbyopic,hypermetrope,no,reduced,none

presbyopic,hypermetrope,no,normal,soft

presbyopic,hypermetrope,yes,reduced,none

presbyopic,hypermetrope,yes,normal,none

**Procedure for Naïve-Bayes:**

1. Load the dataset (Contact-Lenses. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under bayes section.
3. In which we selected Naïve-Bayes algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists of F-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below

# SET-03

## 1. Create a Weather Table with the help of Data Mining Tool WEKA.

Aim:

Create a Weather Table with the help of Data Mining Tool WEKA.

Description:

We need to create a Weather table with training data set which includes attributes like outlook, temperature, humidity, windy, play.

Procedure:

Steps:

1) Open Start  Programs  Accessories  Notepad

2) Type the following training data set with the help of Notepad for Weather Table.

@relation weather

@attribute outlook {sunny,rainy,overcast}

@attribute temparature numeric

@attribute humidity numeric

@attribute windy {true,false}

@attribute play {yes,no}

@data

sunny,85.0,85.0,false,no

overcast,80.0,90.0,true,no

sunny,83.0,86.0,false,yes

rainy,70.0,86.0,false,yes

rainy,68.0,80.0,false,yes

rainy,65.0,70.0,true,no

overcast,64.0,65.0,false,yes

sunny,72.0,95.0,true,no

sunny,69.0,70.0,false,yes

rainy,75.0,80.0,false,yes

3) After that the file is saved with .arff file format.

4) Minimize the arff file and then open Start  Programs  weka-3-4.

5) Click on weka-3-4, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on explorer.

7) Explorer shows many options. In that click on ‘open file’ and select the arff file

8) Click on edit button which shows weather table on weka.

## 2. To list all the categorical (or nominal) attributes and the real valued attributes using Weka mining (German credit data)

AIM :To list all the categorical (or nominal) attributes and the real valued attributes using WEKA mining tool.

Tools/ Apparatus: Weka Mining tool.

Procedure:

 Steps for identifying categorical attributes

1.Double click on credit-g.arff file.

2.Select all categorical attributes.

3.Click on invert.

4.Then we get all real valued attributes selected

5.Click on remove

6.Click on visualize all.

Steps for identifying real valued attributes

1.Double click on credit-g.arff file.Select all real valued attributes.

2.Click on invert.

3.Then we get all categorial attributes selected

4.Click on remove

5.Click on visualize all

**Output:**

**Categorical/ Nominal Attributes:**

1. Checking\_status

2.Credit\_history

3.Purpose

4.Savings\_status

5. Employment

6. Personal\_status

7.Other\_parties

8.Property\_magnitude

9. Other\_payment\_plans

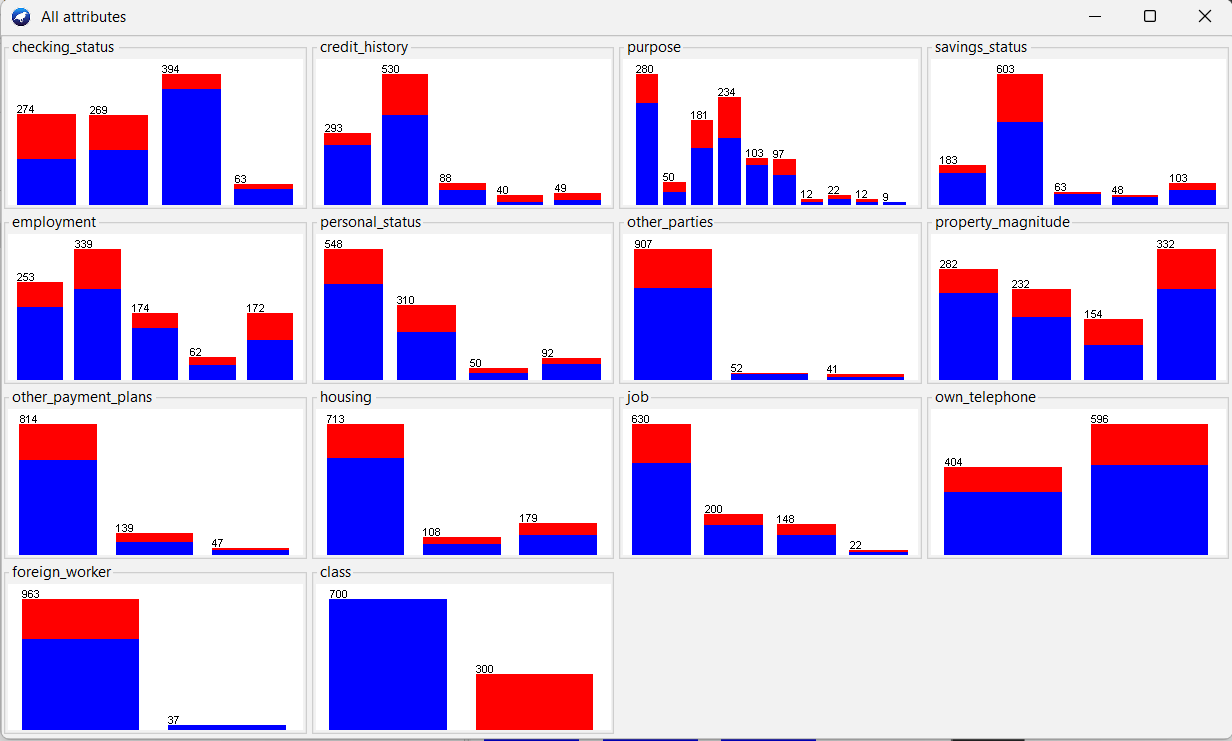
10. Housing

11.Job

12.own\_telephone

13. foreign\_worker

14. Class



**Numeric** **Attributes:**

1. Duration

2. Credit\_amount

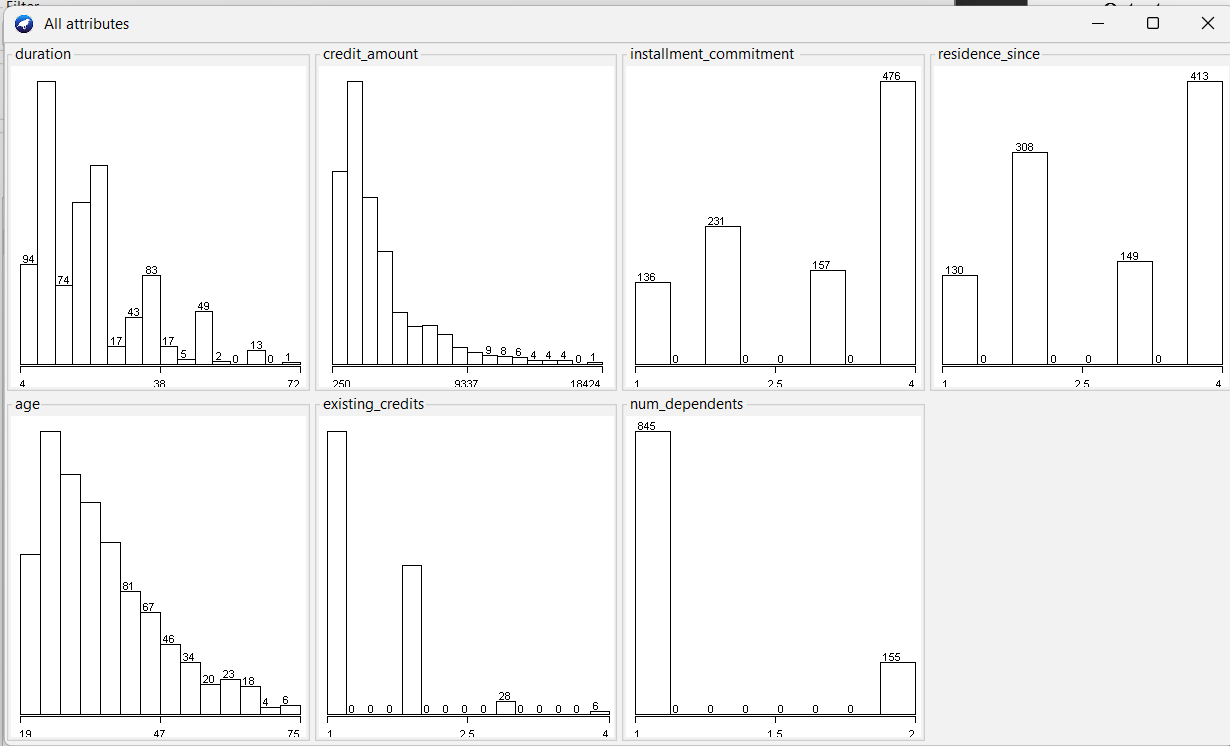
3. Installment\_commitment

4. Residece\_since

5. Age

6. Existing\_credits

7. Num\_dependents.



Result:Hence all categorical and Numerical attributes are displayed

# SET-04

## 1. Create an Iris Table with the help of Data Mining Tool WEKA.

@RELATION iris

@ATTRIBUTE sepallength NUMERIC

@ATTRIBUTE sepalwidth NUMERIC

@ATTRIBUTE petallength NUMERIC

@ATTRIBUTE petalwidth NUMERIC

@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA

5.1,3.5,1.4,0.2,Iris-setosa

4.9,3.0,1.4,0.2,Iris-setosa

4.7,3.2,1.3,0.2,Iris-setosa

4.6,3.1,1.5,0.2,Iris-setosa

5.0,3.6,1.4,0.2,Iris-setosa

5.4,3.9,1.7,0.4,Iris-setosa

4.6,3.4,1.4,0.3,Iris-setosa

5.0,3.4,1.5,0.2,Iris-setosa

4.4,2.9,1.4,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

7.0,3.2,4.7,1.4,Iris-versicolor

6.4,3.2,4.5,1.5,Iris-versicolor

6.9,3.1,4.9,1.5,Iris-versicolor

5.5,2.3,4.0,1.3,Iris-versicolor

6.5,2.8,4.6,1.5,Iris-versicolor

5.7,2.8,4.5,1.3,Iris-versicolor

6.3,3.3,4.7,1.6,Iris-versicolor

4.9,2.4,3.3,1.0,Iris-versicolor

6.6,2.9,4.6,1.3,Iris-versicolor

5.2,2.7,3.9,1.4,Iris-versicolor

6.3,3.3,6.0,2.5,Iris-virginica

5.8,2.7,5.1,1.9,Iris-virginica

7.1,3.0,5.9,2.1,Iris-virginica

6.3,2.9,5.6,1.8,Iris-virginica

6.5,3.0,5.8,2.2,Iris-virginica

7.6,3.0,6.6,2.1,Iris-virginica

4.9,2.5,4.5,1.7,Iris-virginica

7.3,2.9,6.3,1.8,Iris-virginica

6.7,2.5,5.8,1.8,Iris-virginica

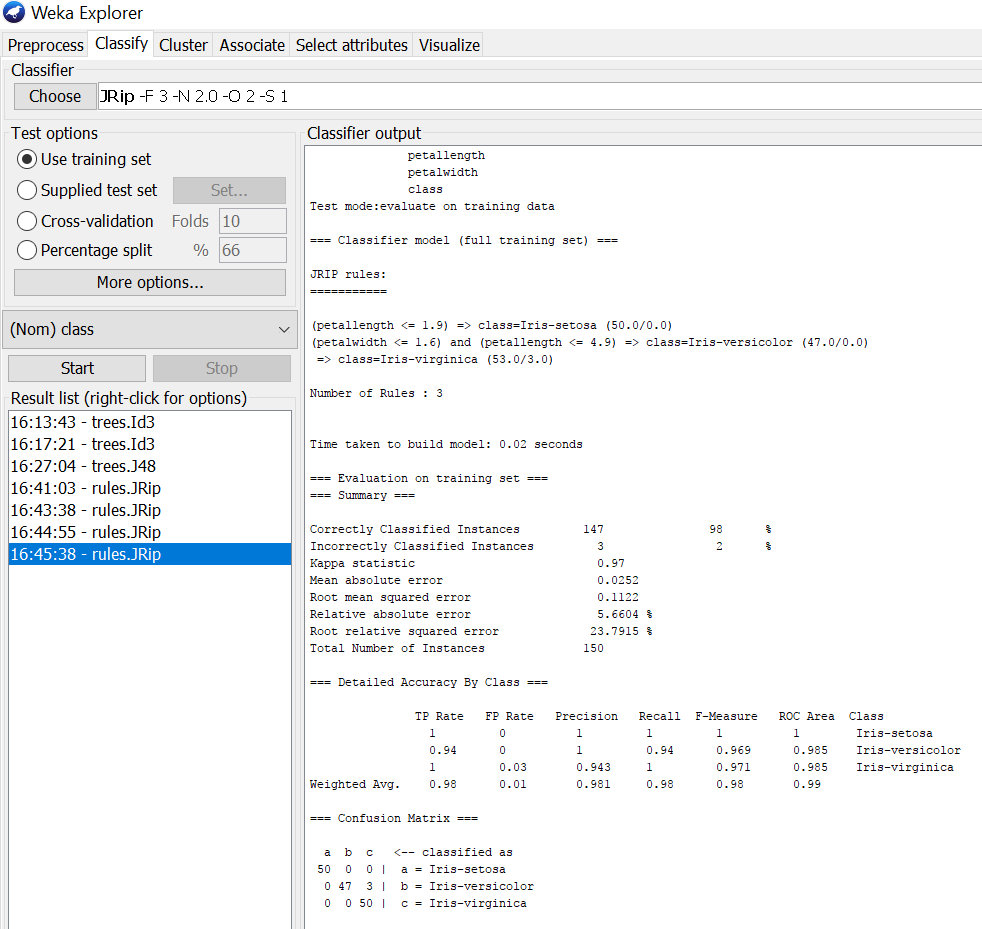
7.2,3.6,6.1,2.5,Iris-virginica

6.5,3.2,5.1,2.0,Iris-virginica

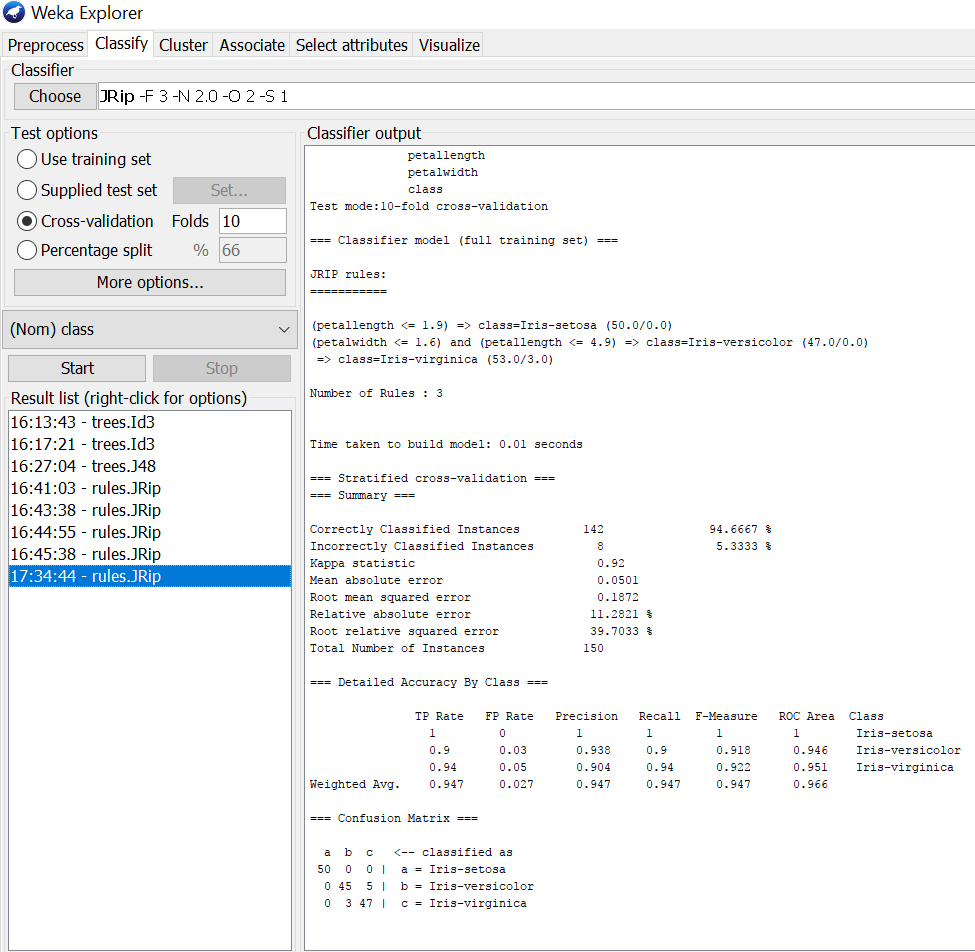
## 2. Demonstrate performing classification on data set Extract if-then rules from the decision tree generated by the classifier, Observe the confusion matrix and derive Accuracy, F-measure, TPrate, FPrate, Precision and Recall values. Apply cross-validation strategy with various fold levels and compare the accuracy results.(iris data)

Procedure**:**

1. Load the dataset (Iris-2D. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under rules section.
3. In which we selected JRip (If-then) algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists ofF-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below.

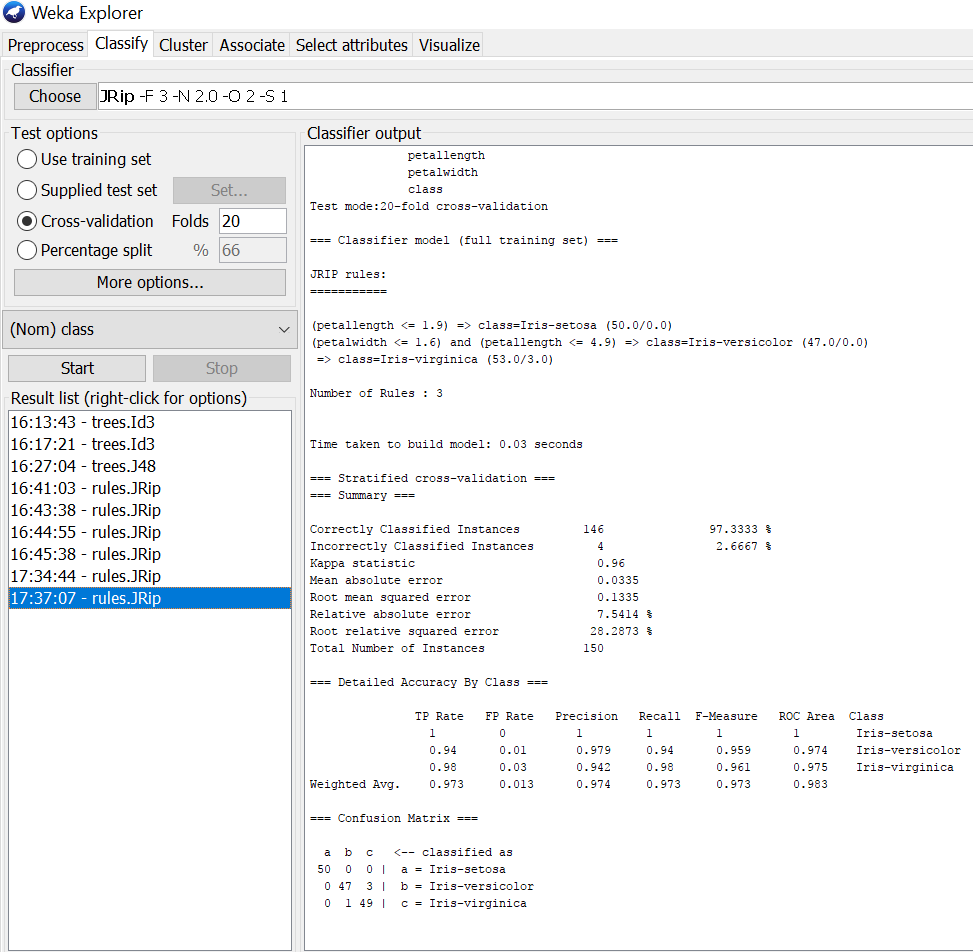


# Using Cross-Validation Strategy with 10 folds:

Here, we enabled cross-validation test option with 10 folds & clicked start button as represented below.

## Using Cross-Validation Strategy with 20 folds:

Here, we enabled cross-validation test option with 20 folds & clicked start button as represented below.



If we see the above results of cross validation with 10 folds & 20 folds. As per our observation the error rate is lesser with 20 folds got 97.3% correctness when compared to 10 folds got 94.6% correctness.

# SET-05

## 1. Create a Buying Table with the help of Data Mining Tool WEKA

1) Open Start  Programs  Accessories  Notepad

2) Type the following training data set with the help of Notepad for Buying Table. @relation buying

@attribute age {L20,20-40,G40}

@attribute income {high,medium,low}

@attribute stud {yes,no}

@attribute creditrate {fair,excellent}

@attribute buyscomp {yes,no}

@data

L20,high,no,fair,yes

20=40,low,yes,fair,yes

G40,medium,yes,fair,yes

L20,low,no,fair,no

G40,high,no,excellent,yes

L20,low,yes,fair,yes

20-40,high,yes,excellent,no

G40,low,no,fair,yes

L20,high,yes,excellent,yes

G40,high,no,fair,yes

L20,low,yes,excellent,no

G40,high,yes,excellent,o

20-40,medium,yes,excellent,yes

L20,medium,yes,fair,yes

G40,high,yes,excellent,yes

3) After that the file is saved with .arff file format.

4) Minimize the arff file and then open Start  Programs  weka-3-4.

5) Click on weka-3-4, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on explorer.

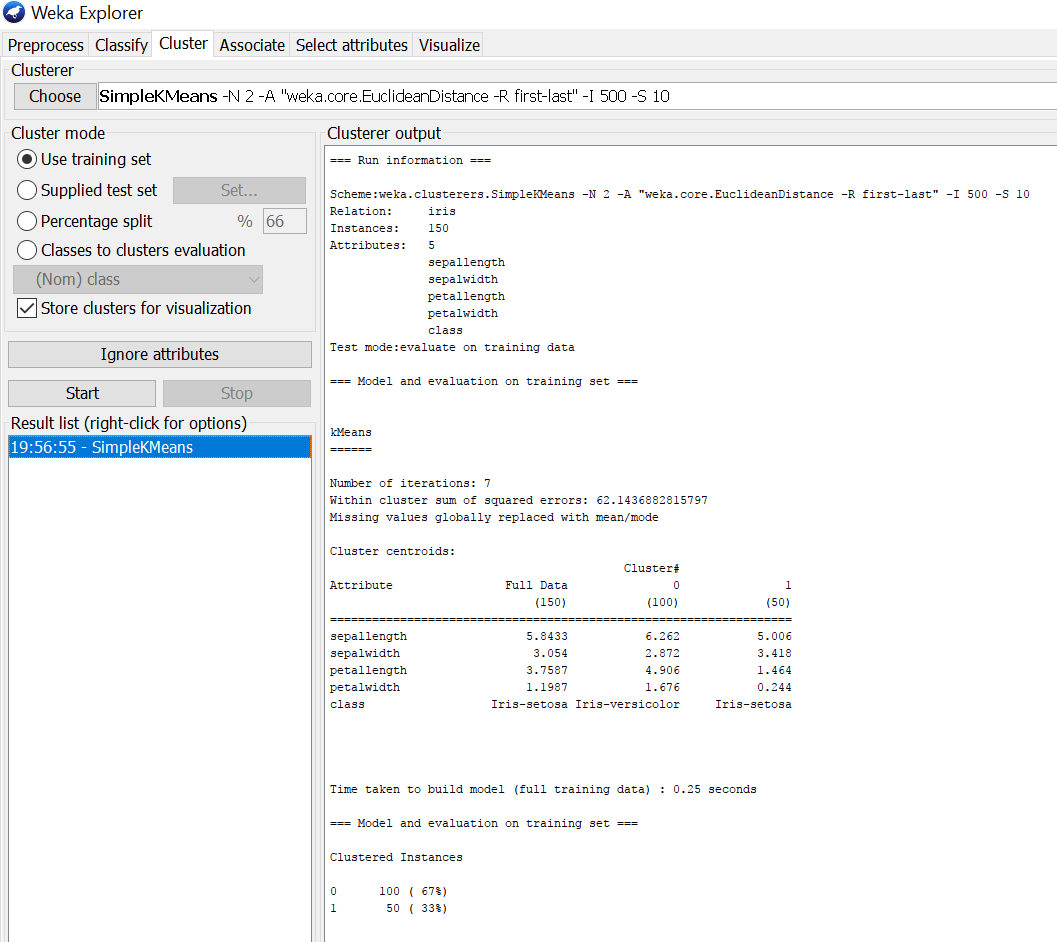
7) Explorer shows many options. In that click on ‘open file’ and select the arff file

8) Click on edit button which shows buying table on weka.

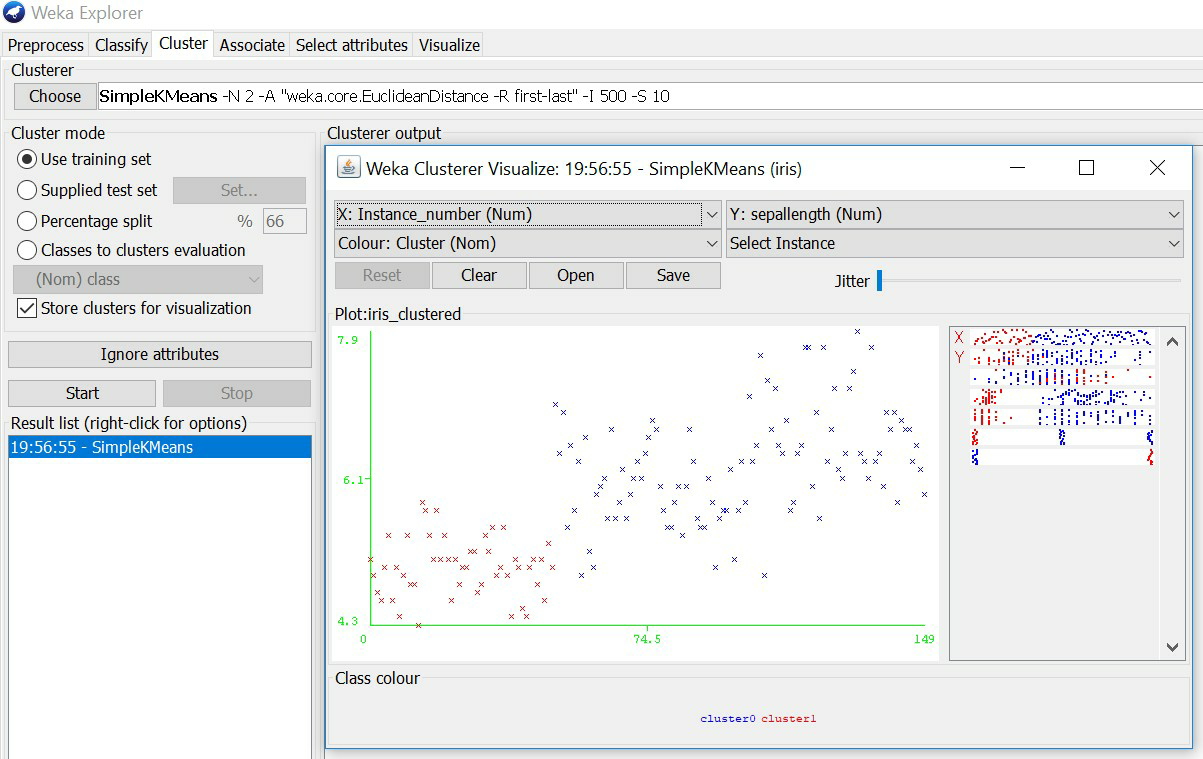
## 2. Demonstrate performing clustering on data sets Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights(Buying).

**Procedure:**

1. Load the dataset (Iris.arff) into weka tool
2. Go to cluster option & in left-hand navigation bar we can see different clustering algorithms under lazy section.
3. In which we selected Simple K-Means algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get the sum of squared errors, centroids, No. of iterations & clustered instances as represented below.



1. If we right click on simple k means, we will get more options in which ―Visualize cluster assignments‖ should be selected for getting cluster visualization as shown below.



# SET-06

## 1. Create an Employee Table with the help of Data Mining Tool WEKA.

Aim:

Create an Employee Table with the help of Data Mining Tool WEKA.

Description:

We need to create an Employee Table with training data set which includes attributes like name, id, salary, experience, gender, phone number.

Procedure:

**Steps:**

1. Open Start Programs Accessories Notepad
2. Type the following training data set with the help of Notepad for Employee Table.

@relation employee

@attributename{x,y,z,a,}

@attribute id numeric

@attribute salary {low,medium,high}

@attribute exp numeric

@attribute gender {male,female}

@attribute phone numeric

@data

x,101,low,2,male,250311

y,102,high,3,female,251665

z,103,medium,1,male,240238

a,104,low,5,female,200200

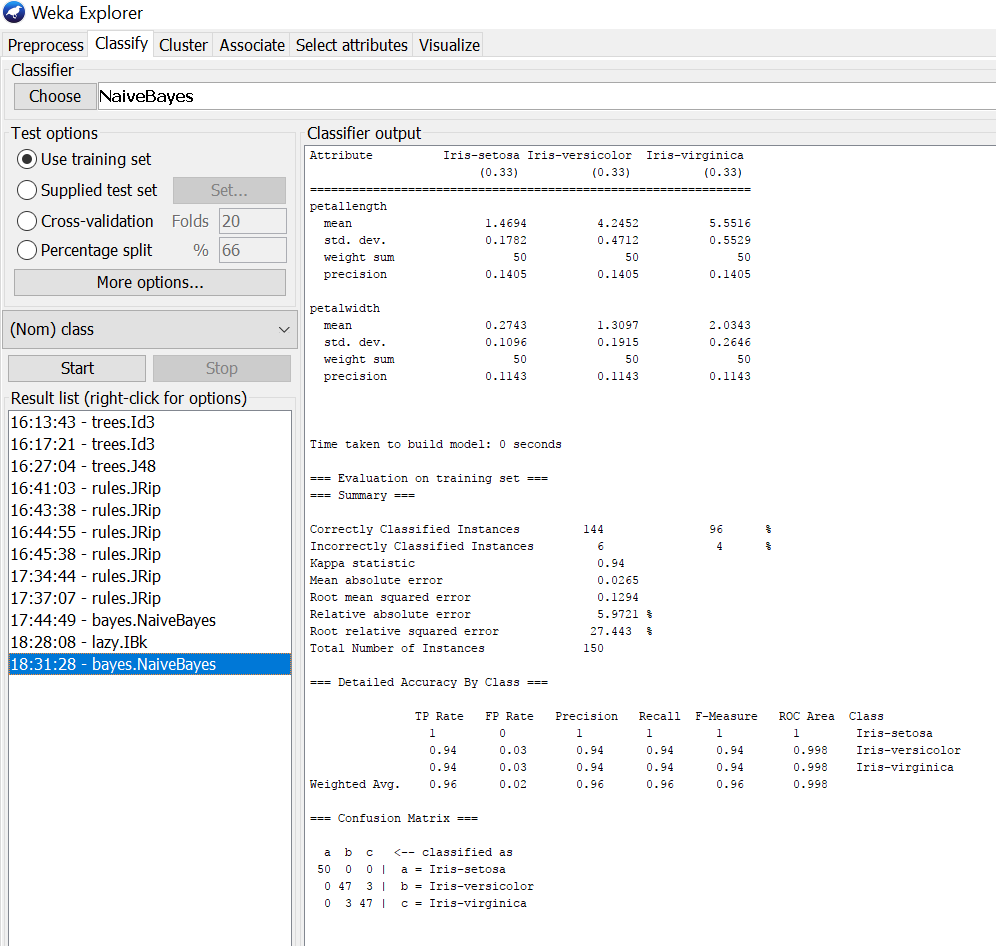
b,105,high,2,male,240240

1. After that the file is saved with **.arff** file format.
2. Minimize the arff file and then open Start Programs weka-3-4.
3. Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
4. In that dialog box there are four modes, click on **explorer**.
5. Explorer shows many options. In that click on **‘open file’** and select the arff file
6. Click on **edit button** which shows employee table on weka.

## 2. Write a Program to implement Bayes classification technique(contact lenses)

**Procedure for Naïve-Bayes:**

1. Load the dataset (Iris-2D. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under bayes section.
3. In which we selected Naïve-Bayes algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists of F-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below.



# SET-07

## 1. Create an Student Table with the help of Data Mining Tool WEKA.

@relation student

@attribute rno numeric

@attribute Firstname string

@attribute Lastname string

@attribute dept {cse,ece,eee,it,civil}

@attribute class {I,II,III,IV}

@attribute CGPA numeric

@attribute status {PASS,FAIL}

@attribute age numeric

@attribute mobilenumber numeric

@data

501, Ramya,chowdary,cse,IV,9.0,PASS,23,8074260766

502, Rahul,Singh,eee,II,7.0,PASS,20,8074260767

503, Yamini,Rajvas,cse,IV,9.0,PASS,23,8074260768

504, Karthik,Medarmetla,ece,III,6.0,PASS,22,8074260769

505, Harsha,Reddy,ece,IV,4.0,FAIL,23,8074260710

507, Joshna,Yadav,civil,III,3.0,FAIL,26,8074260711

508, Pavan,chowdary,it,I,8.5,PASS,27,8074260712

## 2. Use options cross-validation and percentage split and repeat running the Linear Regression Model. Observe the results and meaningful results.(CPU)

Aim: Use options cross-validation and percentage split and repeat running the Linear Regression Model. Observe the results and derive meaningful results.

@relation 'cpu'

@attribute MYCT real

@attribute MMIN real

@attribute MMAX real

@attribute CACH real

@attribute CHMIN real

@attribute CHMAX real

@attribute class real

@data

125,256,6000,256,16,128,198

29,8000,32000,32,8,32,269

29,8000,32000,32,8,32,220

29,8000,32000,32,8,32,172

29,8000,16000,32,8,16,132

26,8000,32000,64,8,32,318

23,16000,32000,64,16,32,367

23,16000,32000,64,16,32,489

23,16000,64000,64,16,32,636

23,32000,64000,128,32,64,1144

400,1000,3000,0,1,2,38

400,512,3500,4,1,6,40

60,2000,8000,65,1,8,92

50,4000,16000,65,1,8,138

350,64,64,0,1,4,10

200,512,16000,0,4,32,35

167,524,2000,8,4,15,19

143,512,5000,0,7,32,28

143,1000,2000,0,5,16,31

110,5000,5000,142,8,64,120

143,1500,6300,0,5,32,30

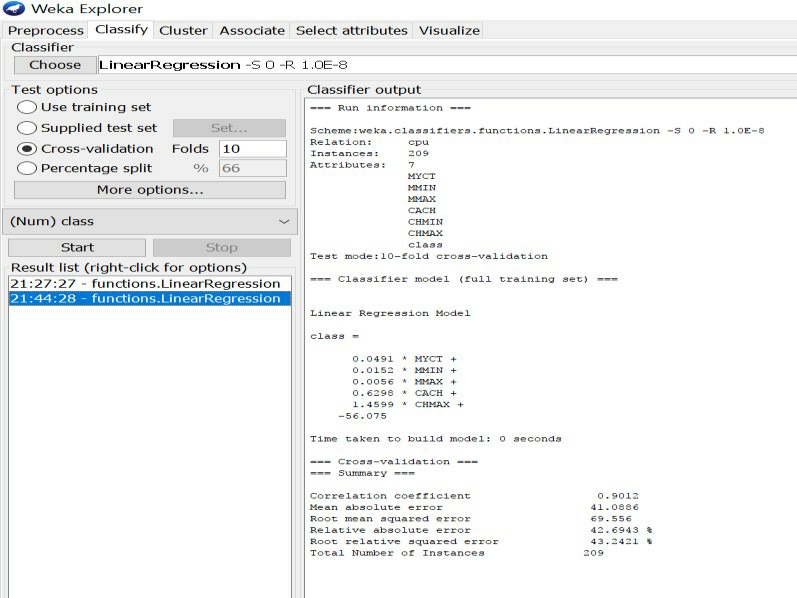
143,3100,6200,0,5,20,33

143,2300,6200,0,6,64,61

110,3100,6200,0,6,64,76

320,128,6000,0,1,12,23

**Procedure for cross-validation:**

1. Load the dataset (Cpu.arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see different classification algorithms under functions section.
3. In which we selected Linear Regression algorithm & click on start option with cross validation option with 10 folds.
4. Then we will get regression model & its result as shown below.

# Procedure for percentage split:

1. Load the dataset (Cpu.arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see different classification algorithms under functions section.
3. In which we selected Linear Regression algorithm & click on start option with percentage split option with 66% split.
4. Then we will get regression model & its result as shown below.

