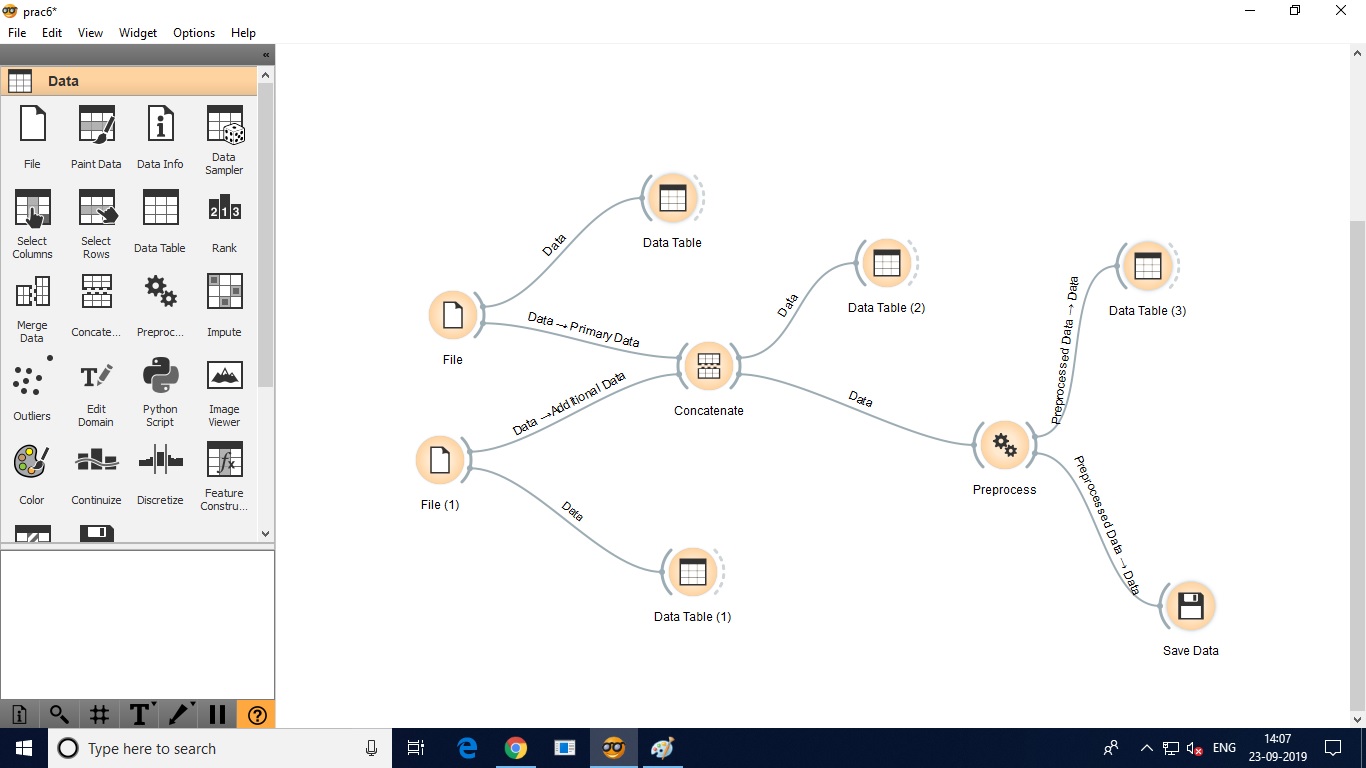
PRACTICAL 6(A)

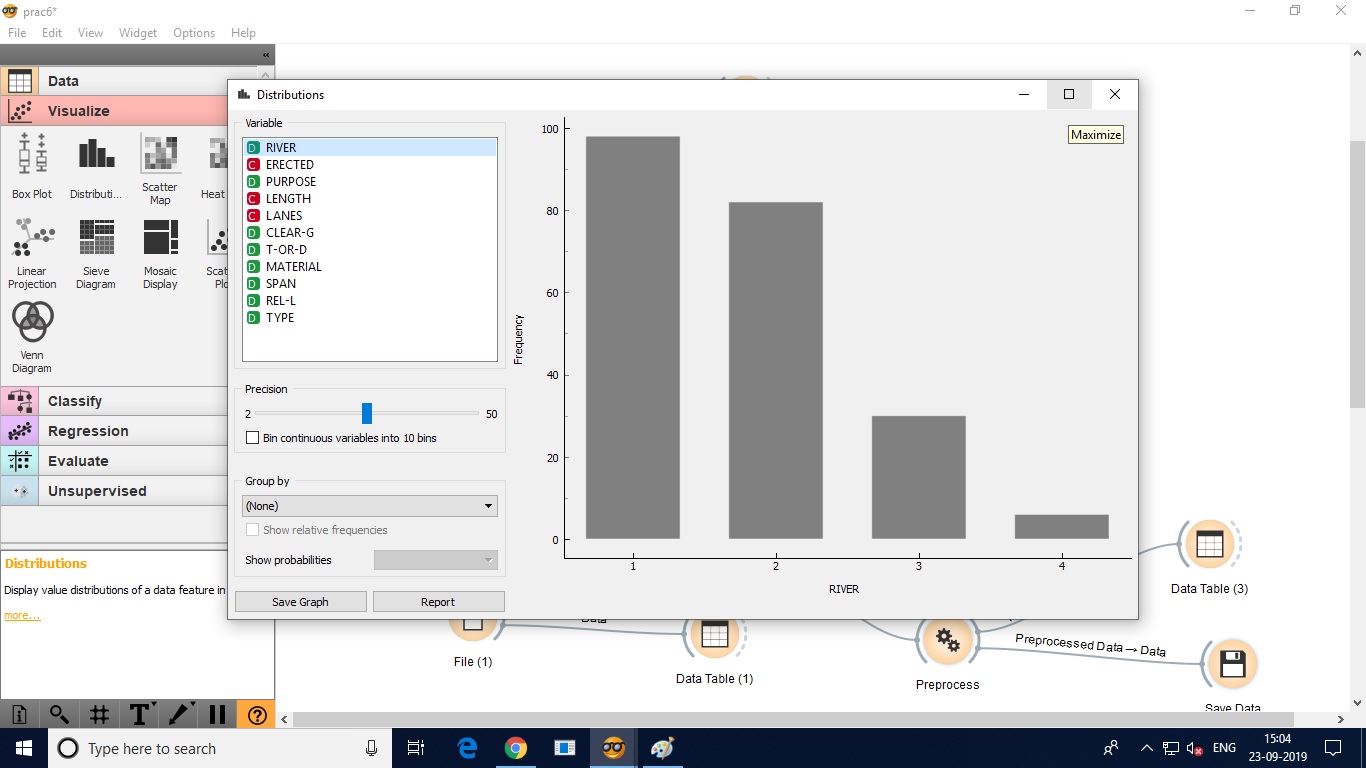
Q.1

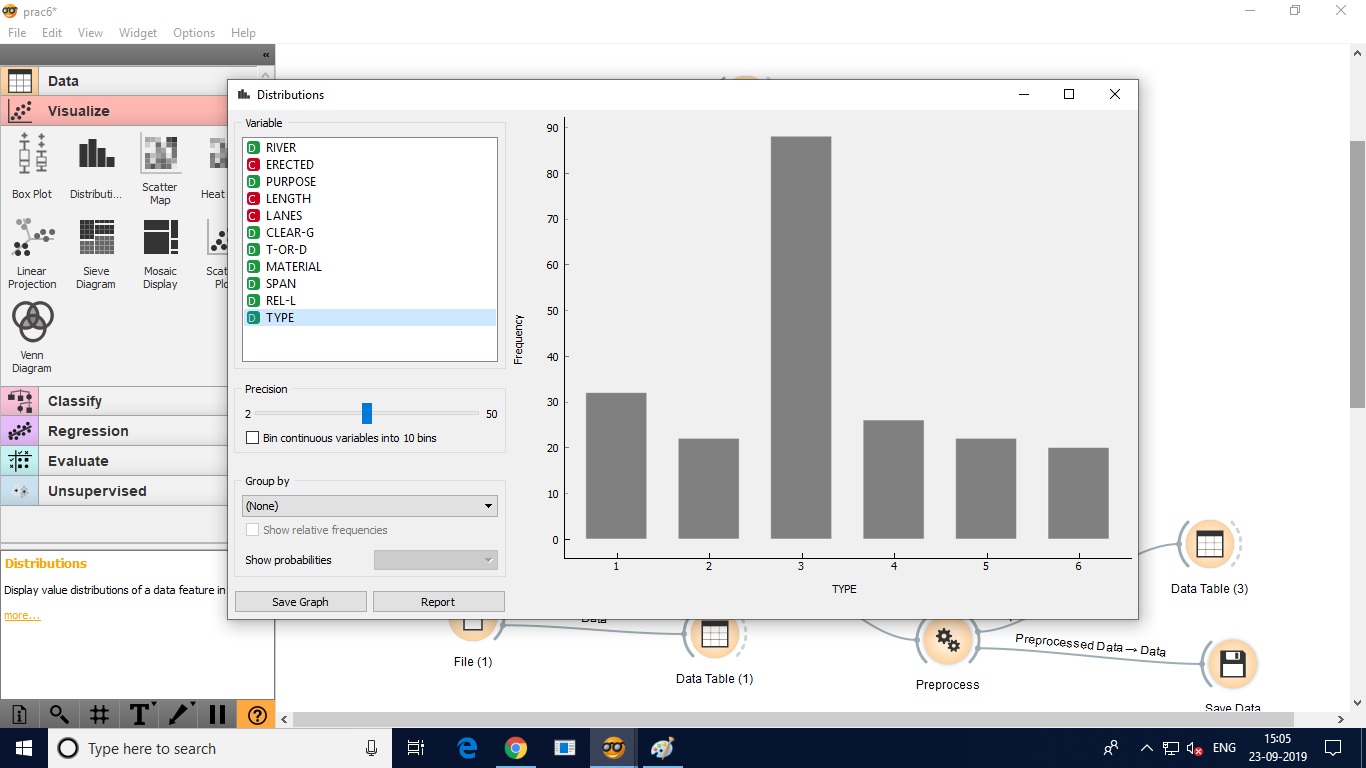
Ans.

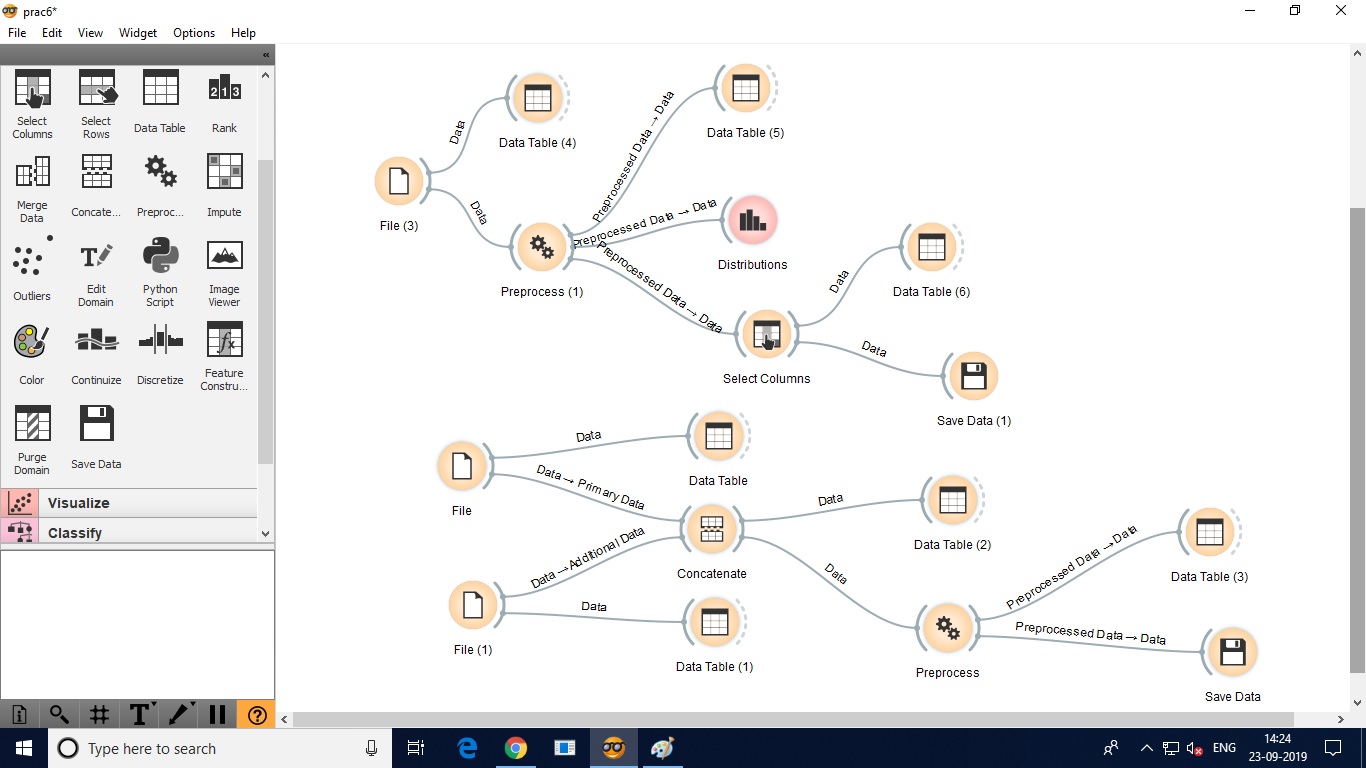


Q.2

Ans.

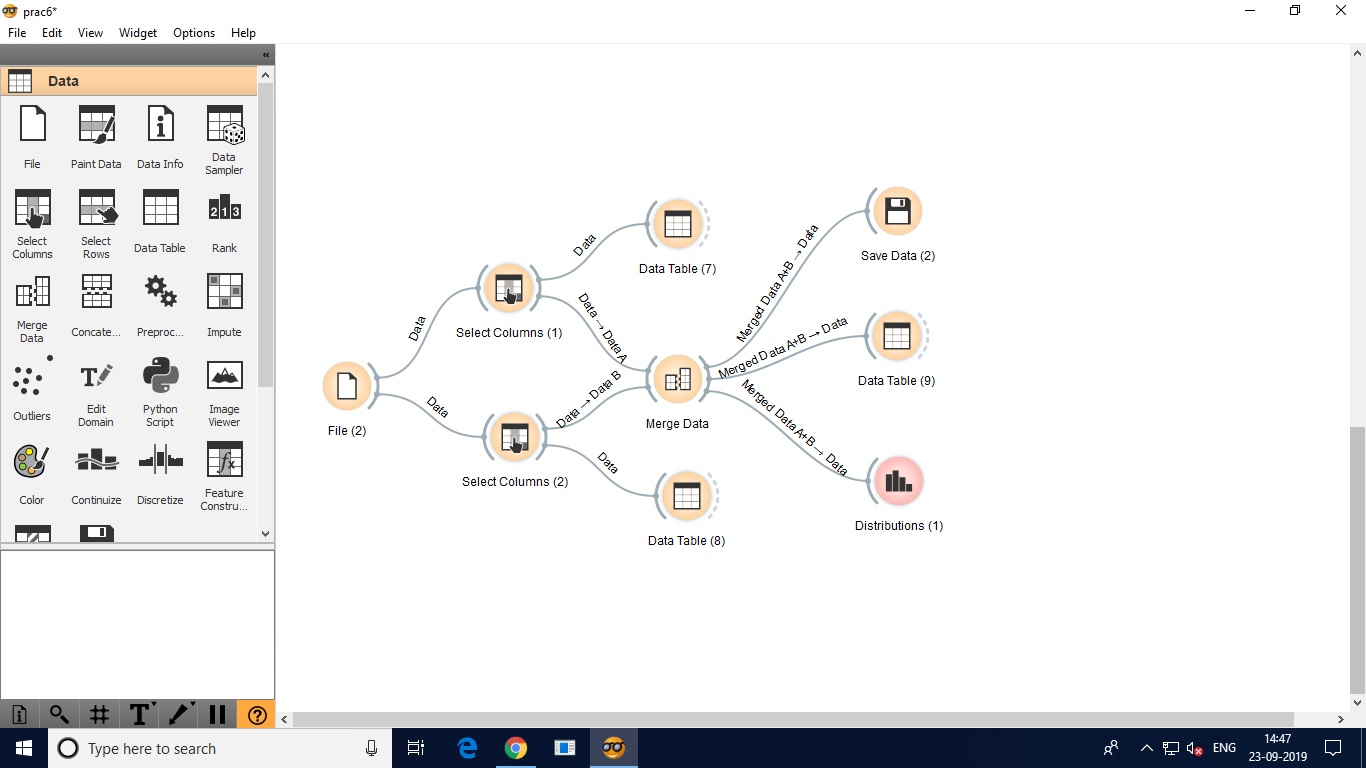


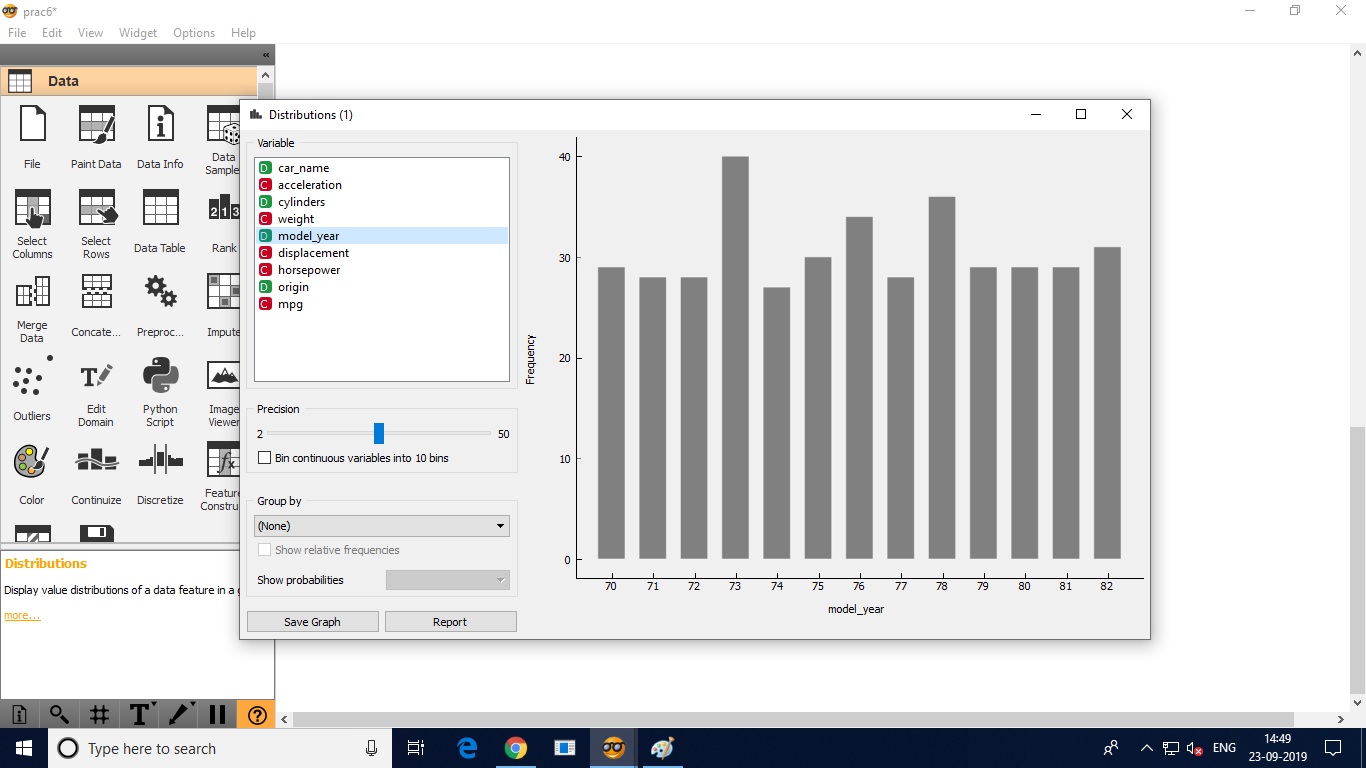




Q.3

Ans.





1.1972

2.max value is:

455.000

index of max value is:

9

[cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, car\_name | mpg]

4.0

455.0

225.0

4425.0

10.0

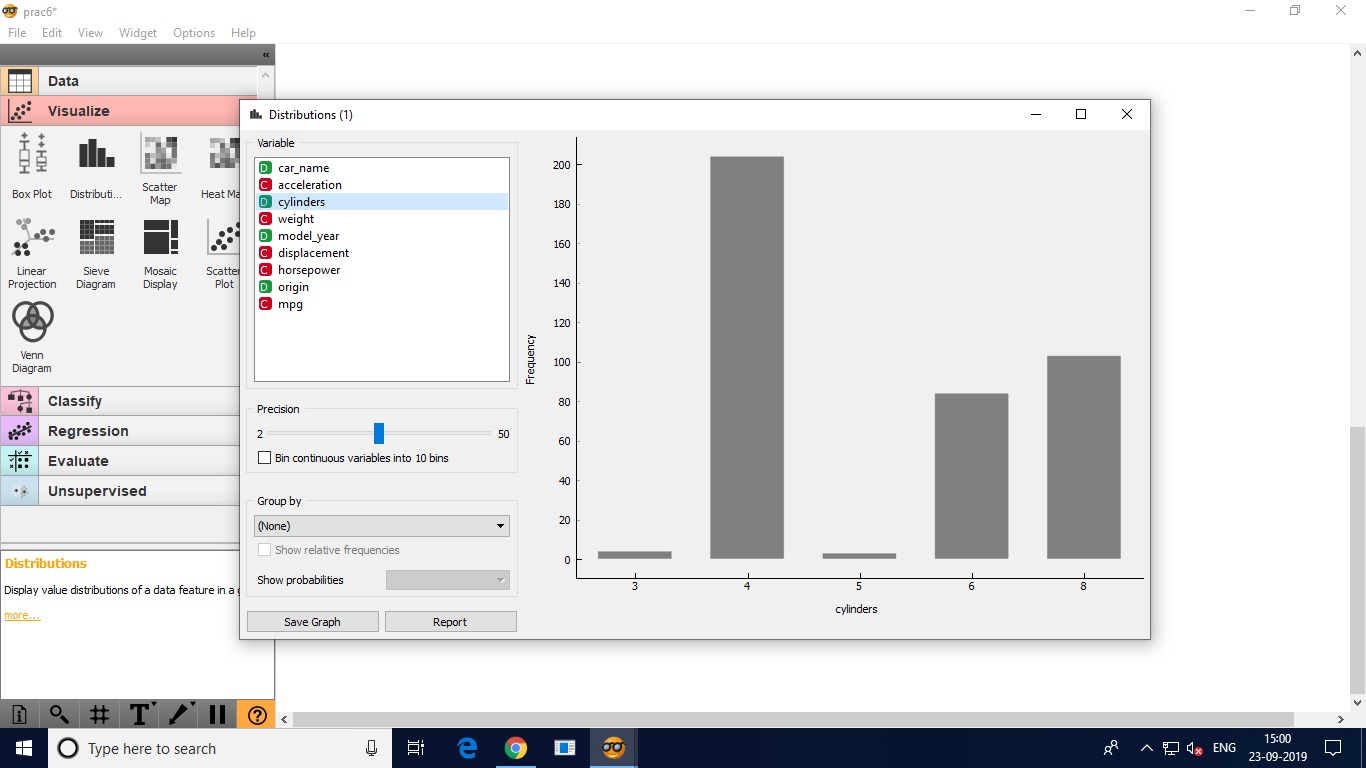
0.0

0.0

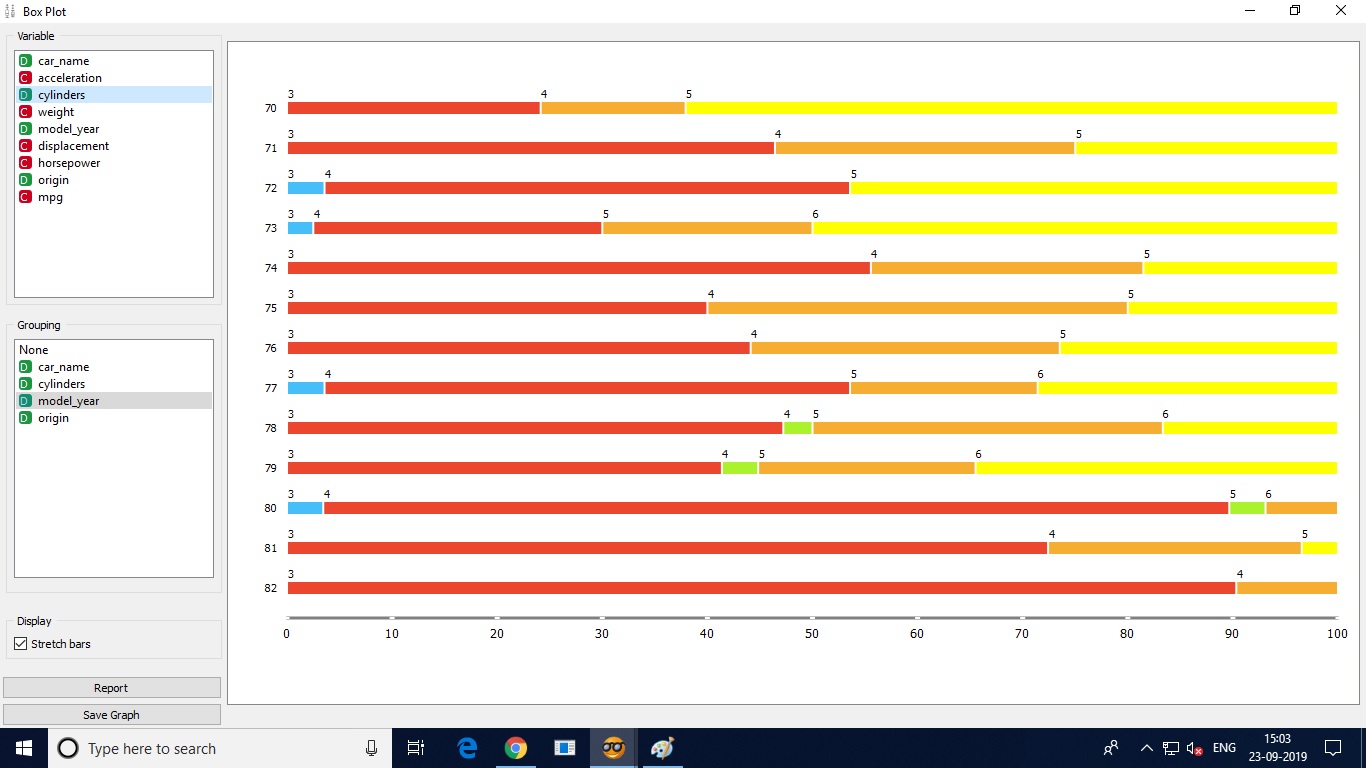
241.0

14.0

4.



5.



PRACTICAL 6(B)

Q.1.1

Ans. 14

Q.1.2

Ans. Outlook: sunny,overcast,rainy

Temperature: hot,mild,cool

Humidity:high,normal

Windy: true,false

Play: yes, no

Q.1.3

Ans. Play

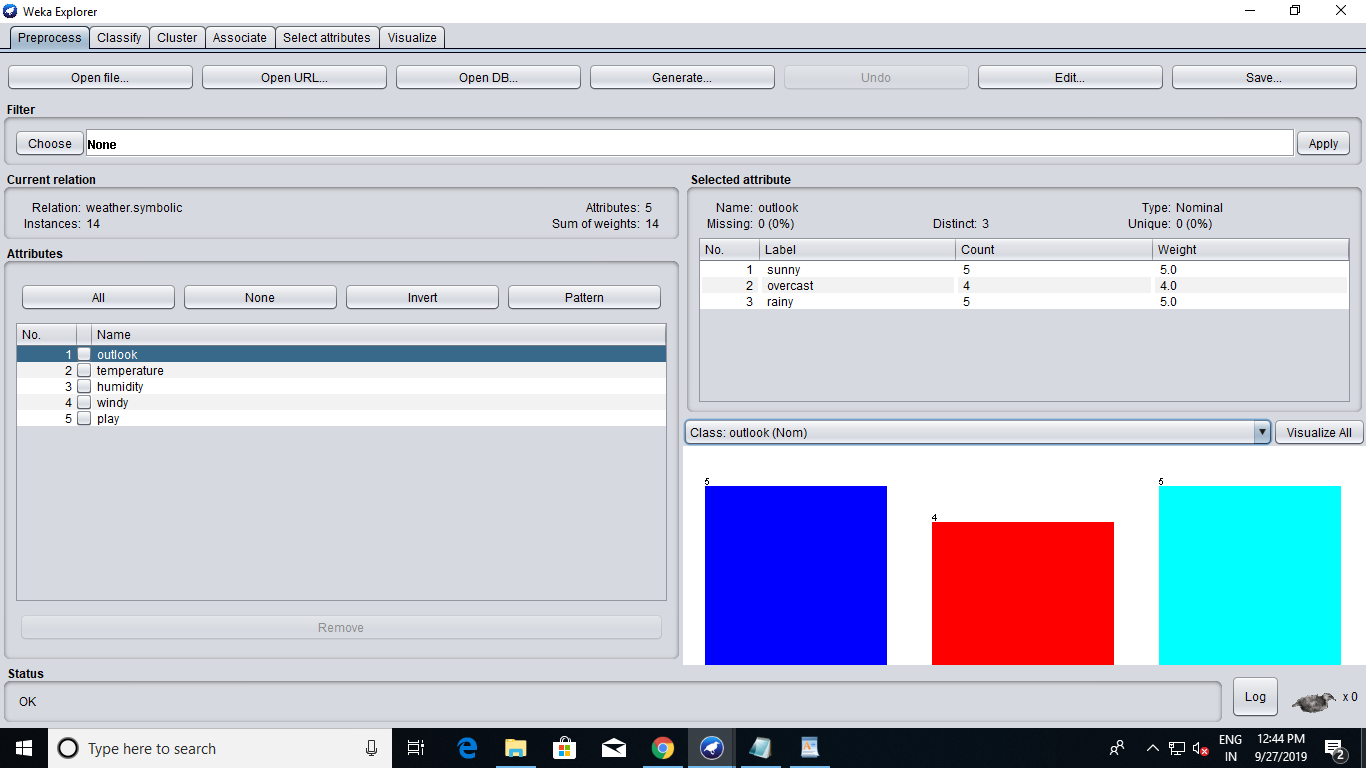
Q.1.4

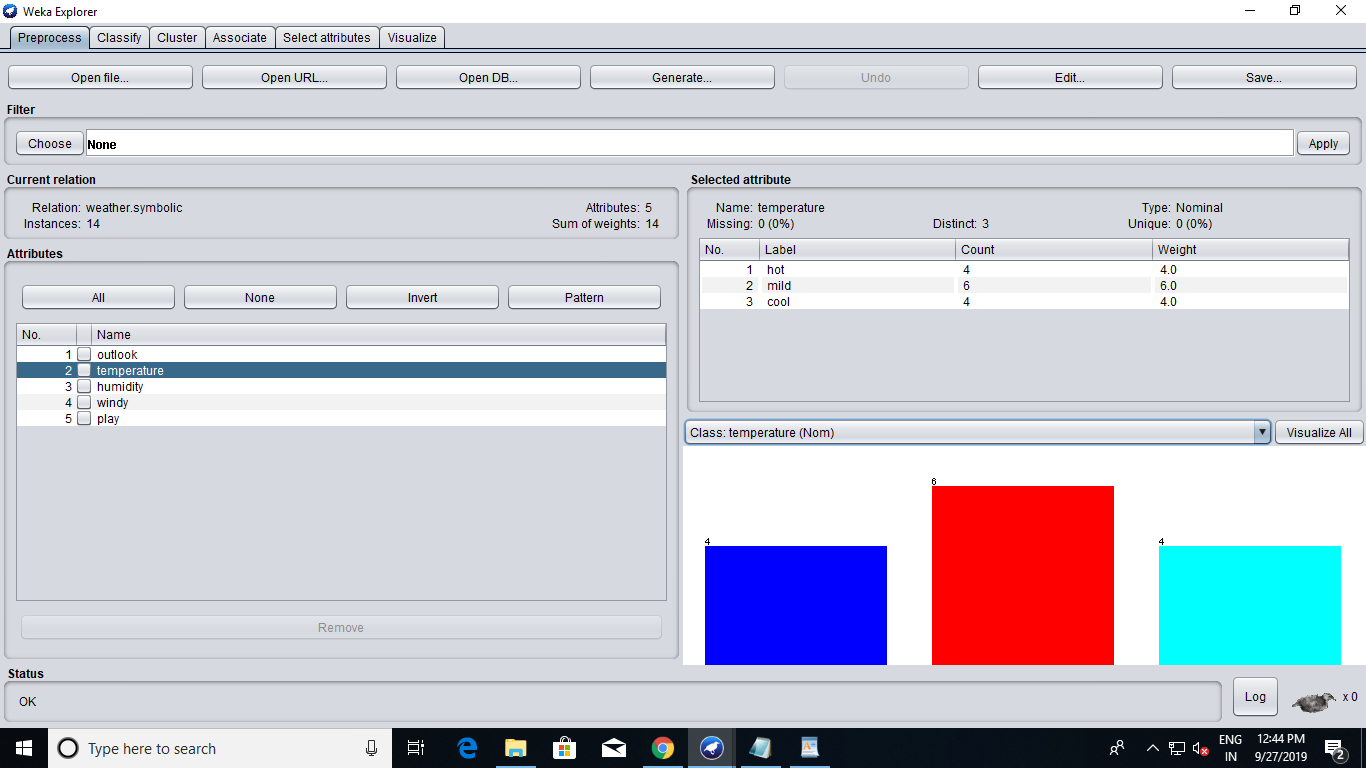
Ans. Y-axis: frequncy count

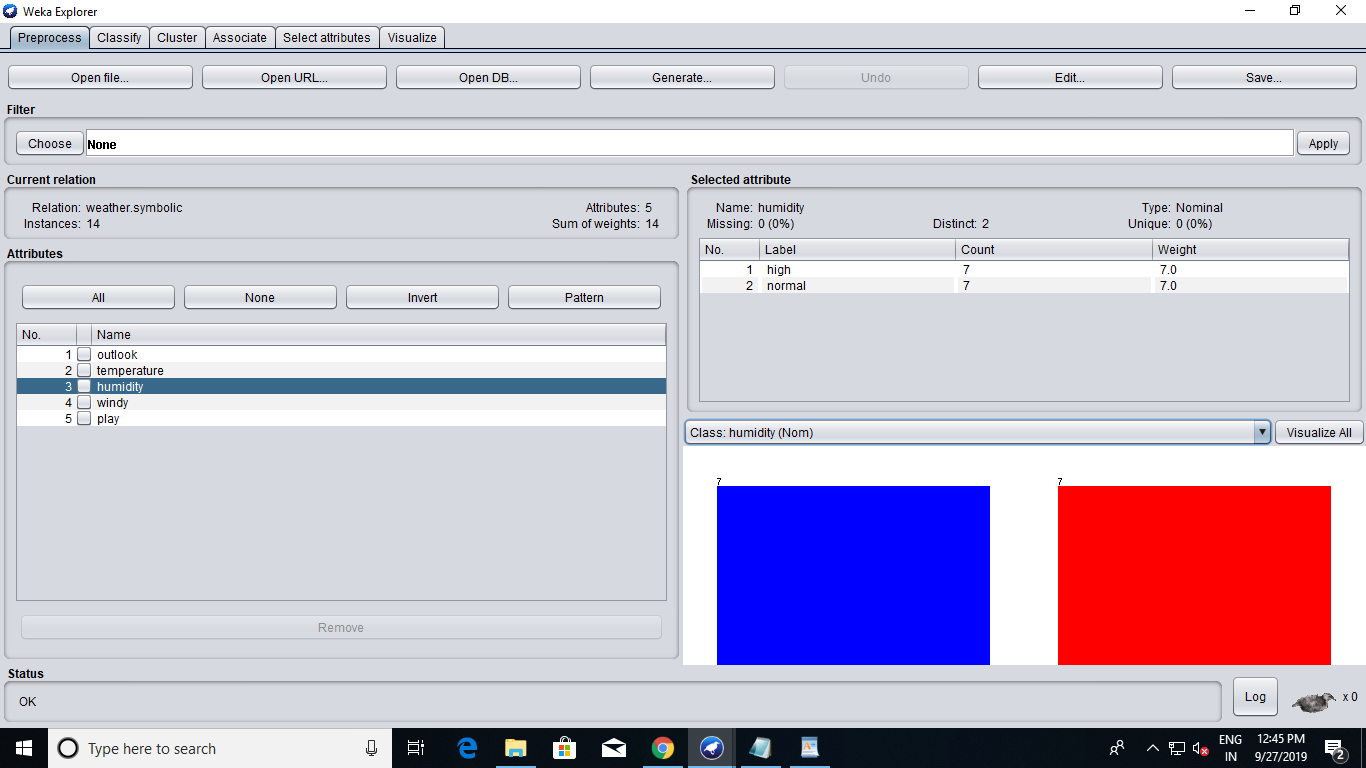
X-axis: variable

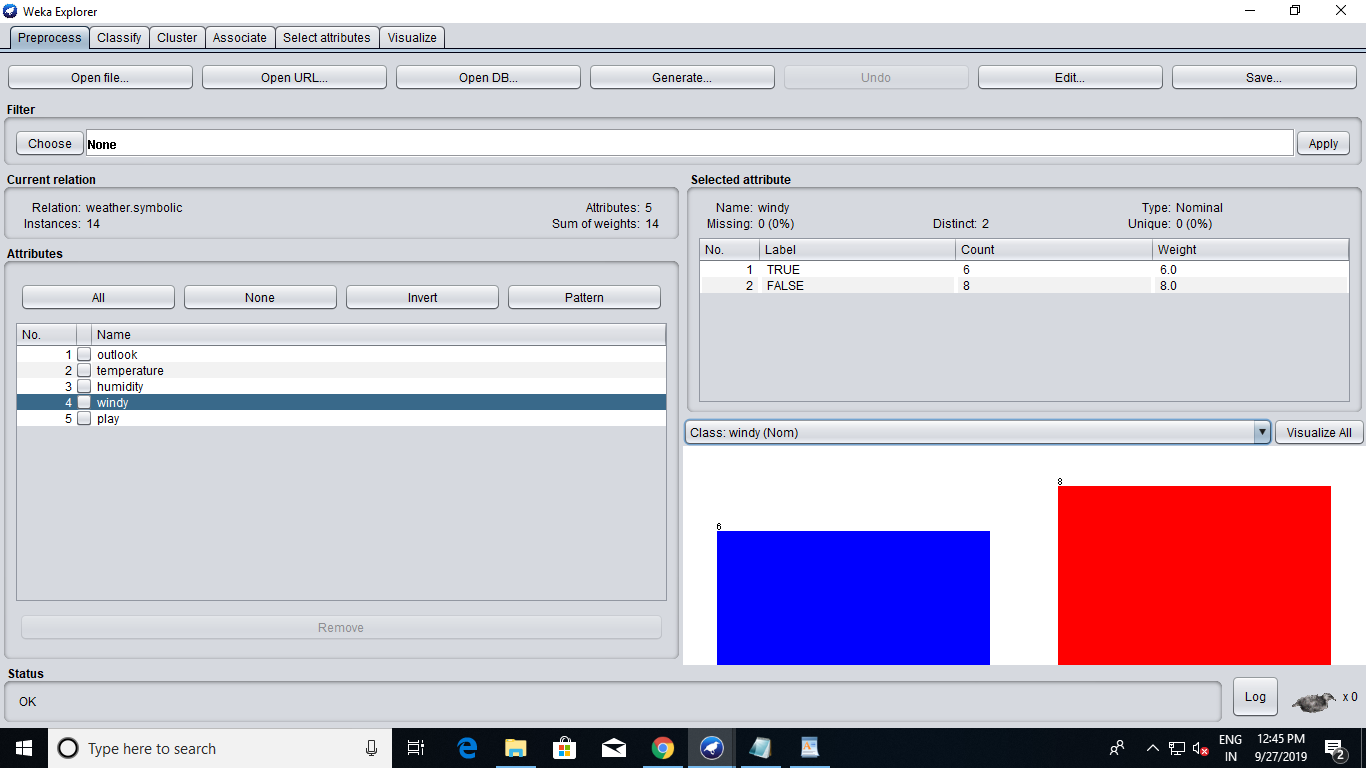
Q.1.5

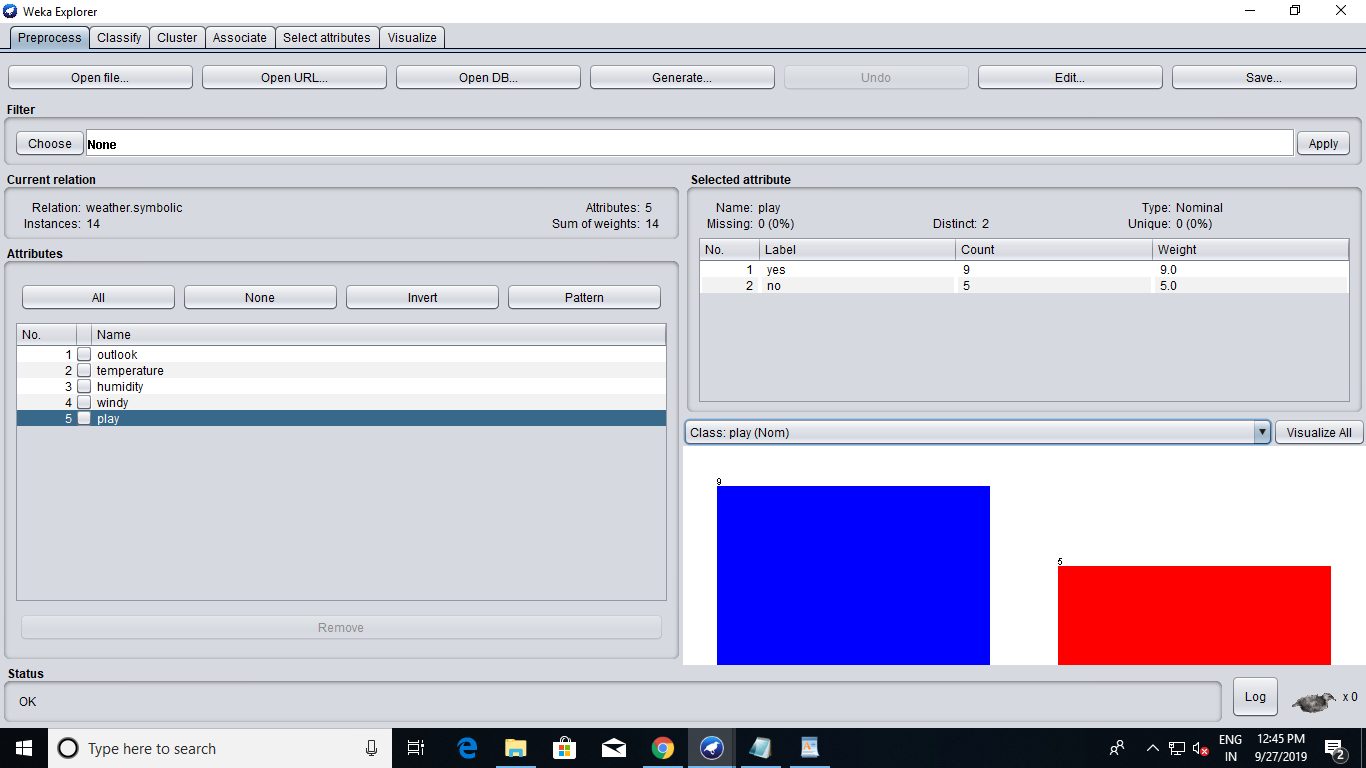
Ans.





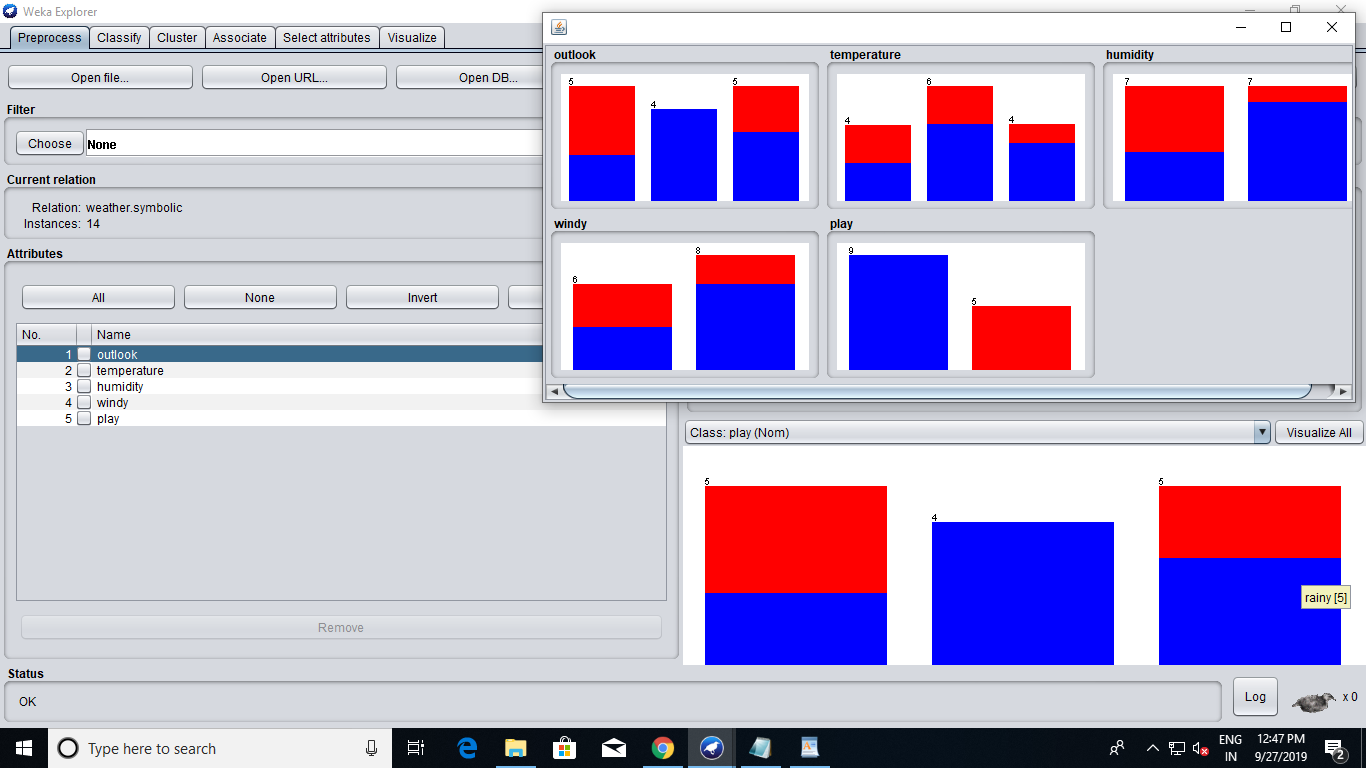






Q.1.6

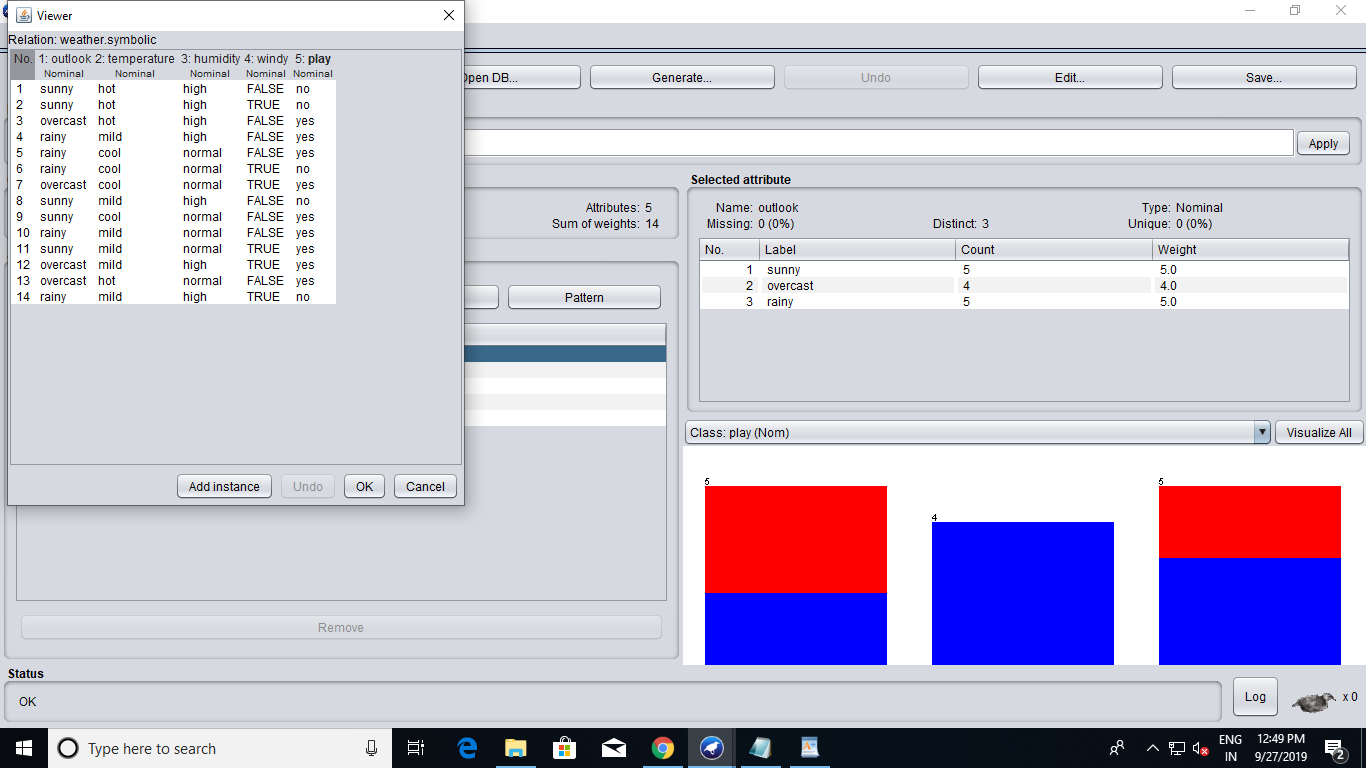
Ans.



Q.1.7

Ans.

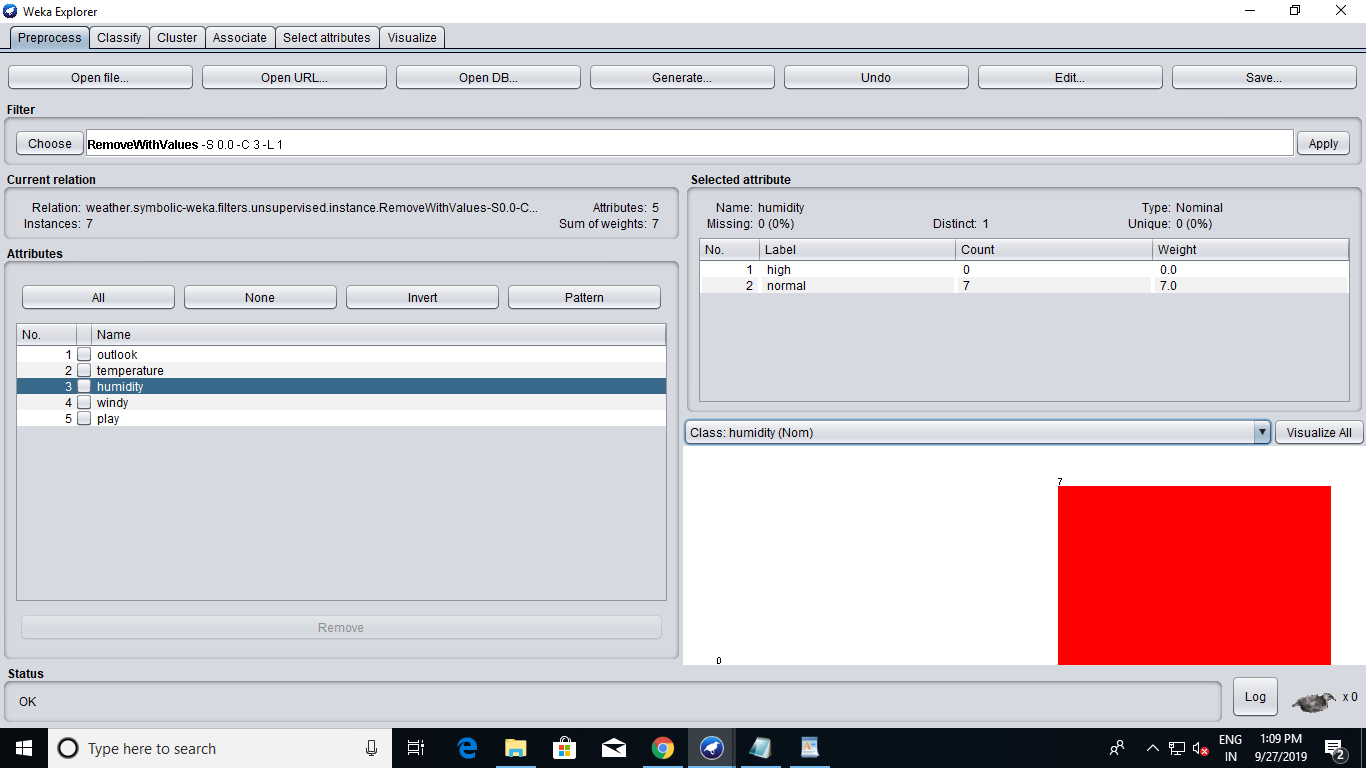
By clicking Edit



Changes can be made by clicking add instance, ok, and save.

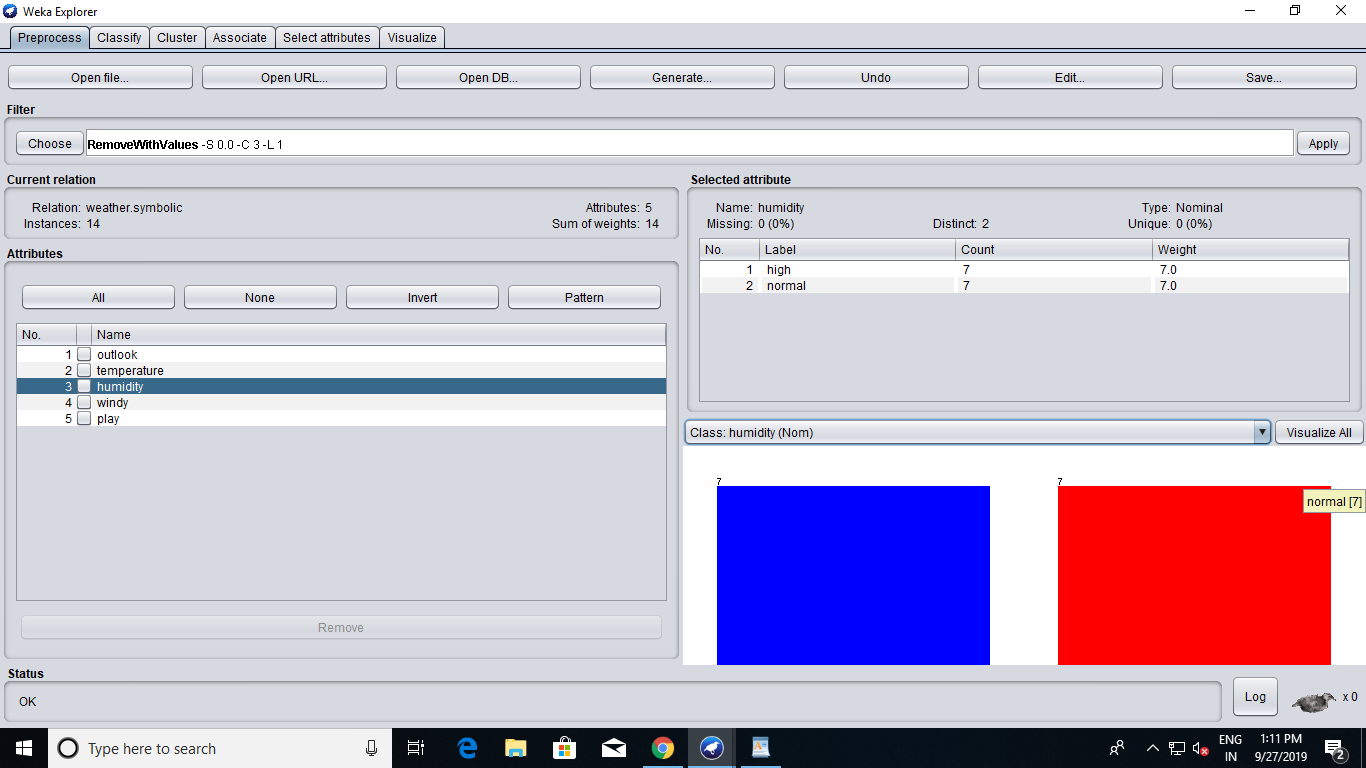
Q.2

Ans 2.1



After applying filter, left click change index to 3 and first-last to 1.

Q.2.2



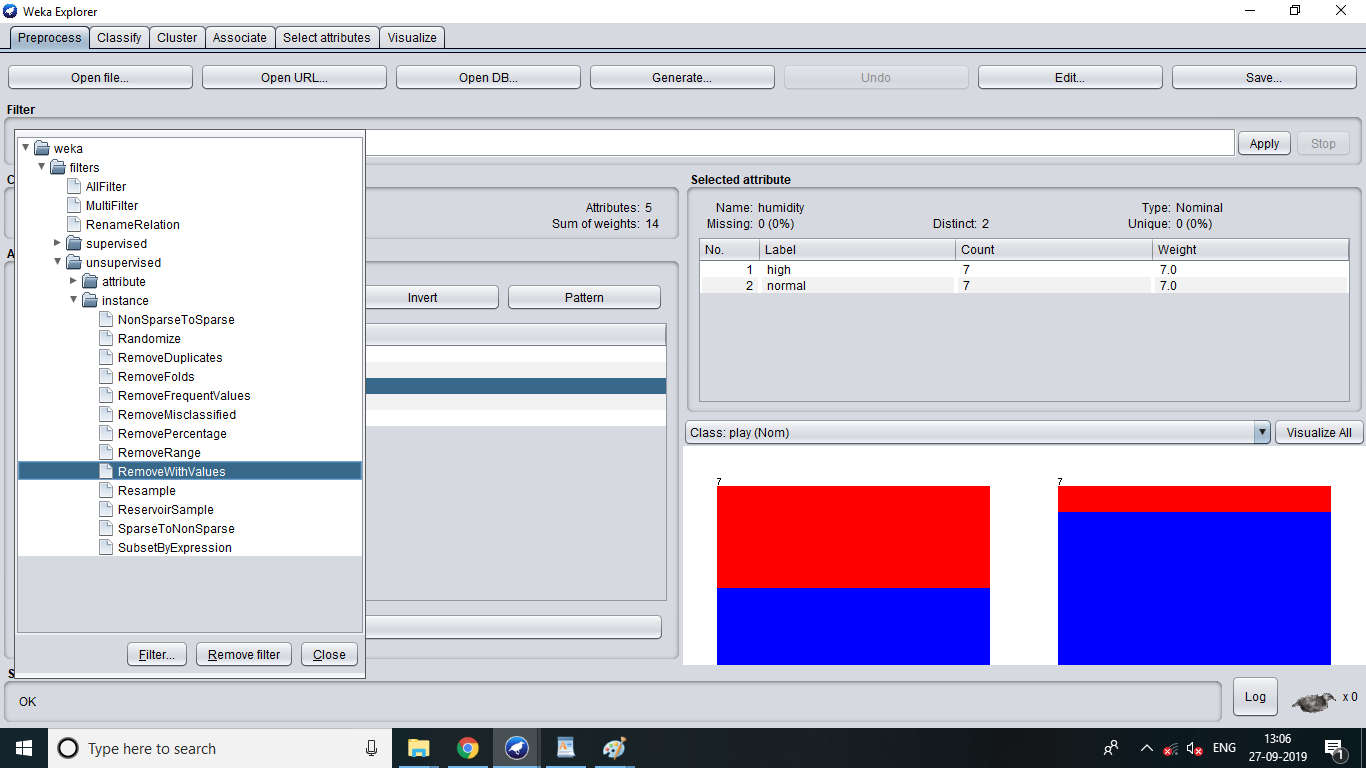
click undo(top right).

Q.2.3.1

Ans.

Weka include many filters that can be used before invoking a classifier to clean up the dataset, or alter it in some way. Filters help with data preparation. For example, you can easily remove an attribute. Or you can remove all instances that have a certain value for an attribute (e.g. instances for which humidity has the value high).

Q.2.3.2



Q.2.3.3

Ans: The two main types of filters in weka are supervised and unsupervised.

Q.2.3.4 What is the difference between the two types of filters? What is the difference betweenand attribute filter and an instance filter?

Supervised filters - in generally - takes in consideration the class value, while the unsupervised filters don't. i.e. supervised 'discretize' filter uses the number of classes as the discretization parameter, while for th

A filter that uses a density-based clusterer to generate cluster membership values; filtered instances are composed of these values plus the class attribute (if set in the input data). An instance filter that copies a range of attributes in the dataset. A filter for turning date attributes into numeric ones.e unsupervised 'discretize' filter you will provide the number of bins ('classes')

Q.3.1

Ans. Loaded dataset sick.

Q3.2

Ans. 3772

Q.3.3

Ans. 30

Q.3.4

Ans . the class attribute is Class and it classifies the person as sick and negative.

Q.3.5

Ans. Age(1),TSH(18),T3(20),TT4(22),T4U(24),FTI(26),TBG(28).

Q.3.6

Ans.

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: sick

Instances: 3772

Attributes: 30

age

sex

on thyroxine

query on thyroxine

on antithyroid medication

sick

pregnant

thyroid surgery

I131 treatment

query hypothyroid

query hyperthyroid

lithium

goitre

tumor

hypopituitary

psych

TSH measured

TSH

T3 measured

T3

TT4 measured

TT4

T4U measured

T4U

FTI measured

FTI

TBG measured

TBG

referral source

Class

Test mode: user supplied test set: size unknown (reading incrementally)

=== Classifier model (full training set) ===

Naive Bayes Classifier

Class

Attribute negative sick

(0.94) (0.06)

==============================================

age

mean 50.99 64.0881

std. dev. 20.2327 16.3603

weight sum 3540 231

precision 4.9348 4.9348

sex

F 2345.0 137.0

M 1054.0 90.0

[total] 3399.0 227.0

on thyroxine

f 3090.0 220.0

t 453.0 13.0

[total] 3543.0 233.0

query on thyroxine

f 3496.0 228.0

t 47.0 5.0

[total] 3543.0 233.0

on antithyroid medication

f 3499.0 232.0

t 44.0 1.0

[total] 3543.0 233.0

sick

f 3420.0 207.0

t 123.0 26.0

[total] 3543.0 233.0

pregnant

f 3489.0 232.0

t 54.0 1.0

[total] 3543.0 233.0

thyroid surgery

f 3489.0 232.0

t 54.0 1.0

[total] 3543.0 233.0

I131 treatment

f 3484.0 231.0

t 59.0 2.0

[total] 3543.0 233.0

query hypothyroid

f 3336.0 204.0

t 207.0 29.0

[total] 3543.0 233.0

query hyperthyroid

f 3313.0 224.0

t 230.0 9.0

[total] 3543.0 233.0

lithium

f 3525.0 231.0

t 18.0 2.0

[total] 3543.0 233.0

goitre

f 3510.0 230.0

t 33.0 3.0

[total] 3543.0 233.0

tumor

f 3448.0 230.0

t 95.0 3.0

[total] 3543.0 233.0

hypopituitary

f 3542.0 231.0

t 1.0 2.0

[total] 3543.0 233.0

psych

f 3365.0 225.0

t 178.0 8.0

[total] 3543.0 233.0

TSH measured

t 3175.0 230.0

f 368.0 3.0

[total] 3543.0 233.0

TSH

mean 5.0964 4.1999

std. dev. 25.142 13.2259

weight sum 3174 229

precision 1.8531 1.8531

T3 measured

t 2776.0 229.0

f 767.0 4.0

[total] 3543.0 233.0

T3

mean 2.108 0.8921

std. dev. 0.7813 0.4392

weight sum 2775 228

precision 0.1551 0.1551

TT4 measured

t 3311.0 232.0

f 232.0 1.0

[total] 3543.0 233.0

TT4

mean 109.4196 92.6175

std. dev. 35.6301 31.3848

weight sum 3310 231

precision 1.7833 1.7833

T4U measured

t 3163.0 224.0

f 380.0 9.0

[total] 3543.0 233.0

T4U

mean 1.0063 0.8352

std. dev. 0.194 0.1351

weight sum 3162 223

precision 0.0143 0.0143

FTI measured

t 3165.0 224.0

f 378.0 9.0

[total] 3543.0 233.0

FTI

mean 110.446 110.8756

std. dev. 33.0156 34.1984

weight sum 3164 223

precision 1.6867 1.6867

TBG measured

f 3542.0 232.0

[total] 3542.0 232.0

TBG

mean 0 0

std. dev. 0.0017 0.0017

weight sum 0 0

precision 0.01 0.01

referral source

SVHC 378.0 10.0

other 2169.0 34.0

SVI 849.0 187.0

STMW 113.0 1.0

SVHD 37.0 4.0

[total] 3546.0 236.0

Time taken to build model: 0.02 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.17 seconds

=== Summary ===

Correctly Classified Instances 3507 92.9745 %

Incorrectly Classified Instances 265 7.0255 %

Kappa statistic 0.5433

Mean absolute error 0.0869

Root mean squared error 0.227

Relative absolute error 75.4633 %

Root relative squared error 94.6774 %

Total Number of Instances 3772

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.939 0.212 0.985 0.939 0.962 0.567 0.930 0.992 negative

0.788 0.061 0.457 0.788 0.579 0.567 0.930 0.672 sick

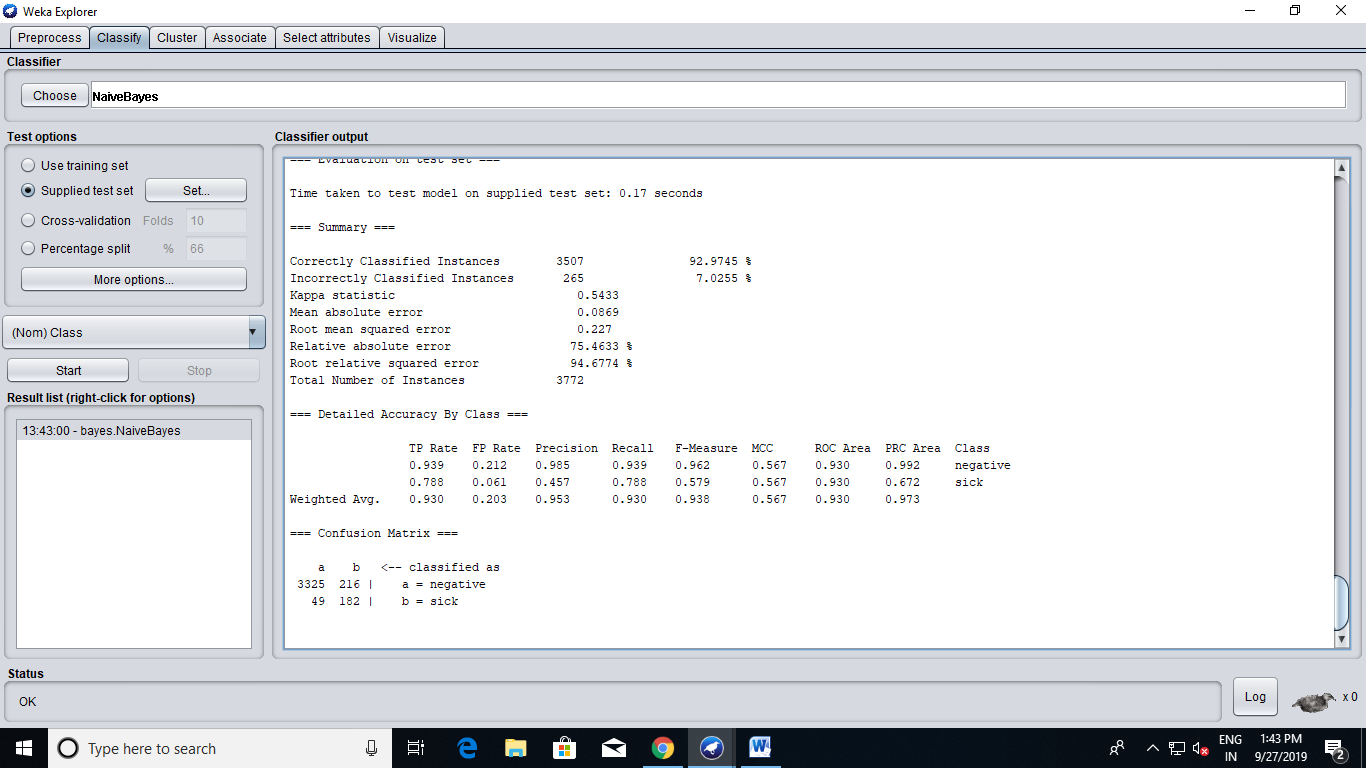
Weighted Avg. 0.930 0.203 0.953 0.930 0.938 0.567 0.930 0.973

=== Confusion Matrix ===

a b <-- classified as

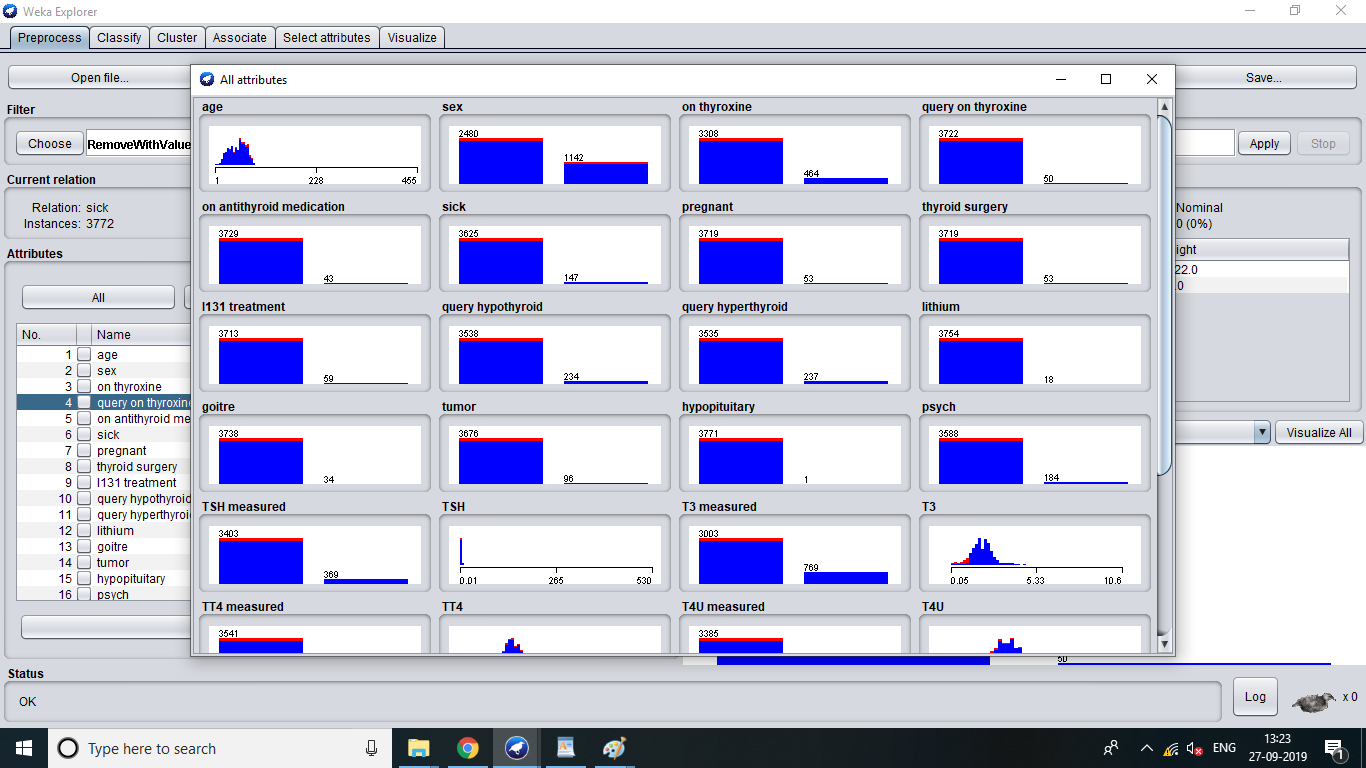
3325 216 | a = negative

49 182 | b = sick



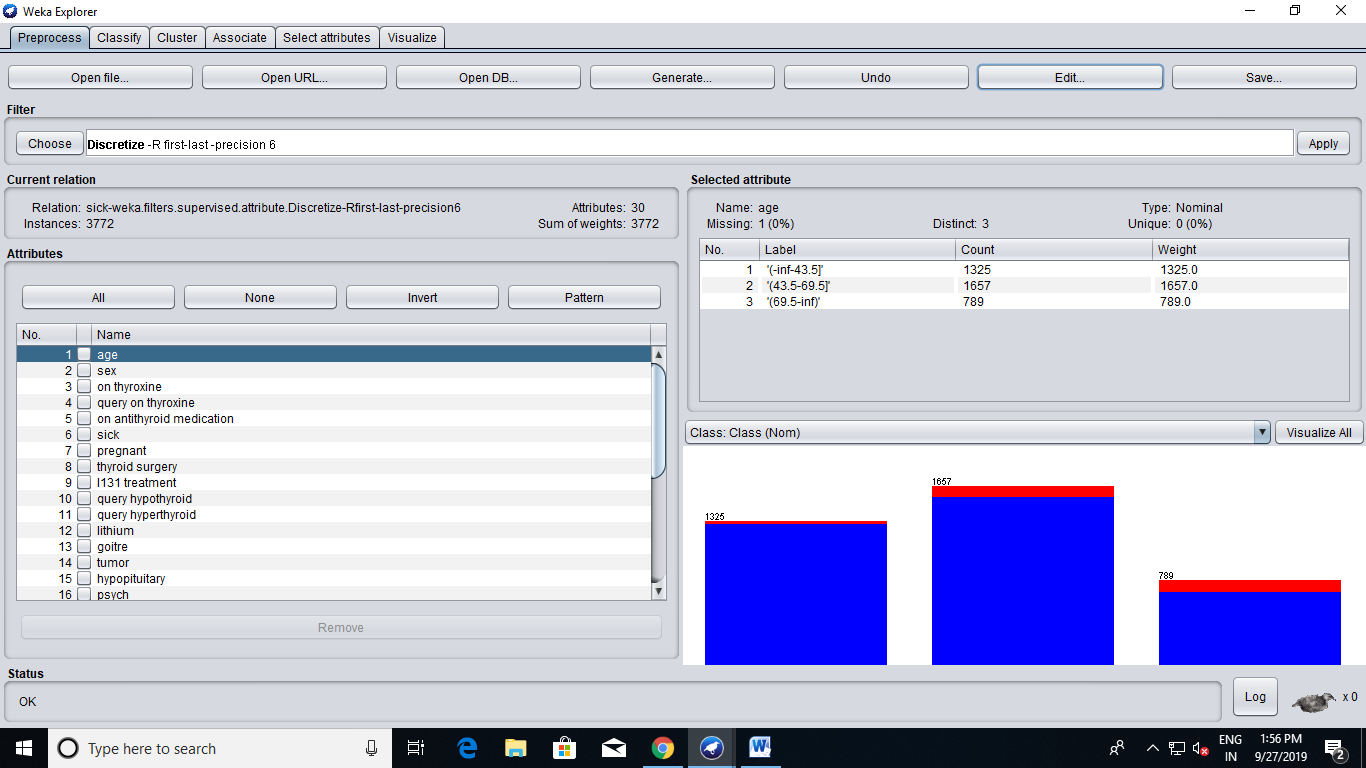
Q.2.1

Ans.Loaded

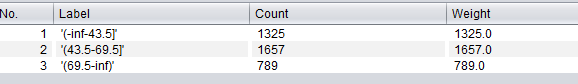


Q.2.2.

Ans: Applied



Q.2.3.



Q.2.4.

Ans. 3.

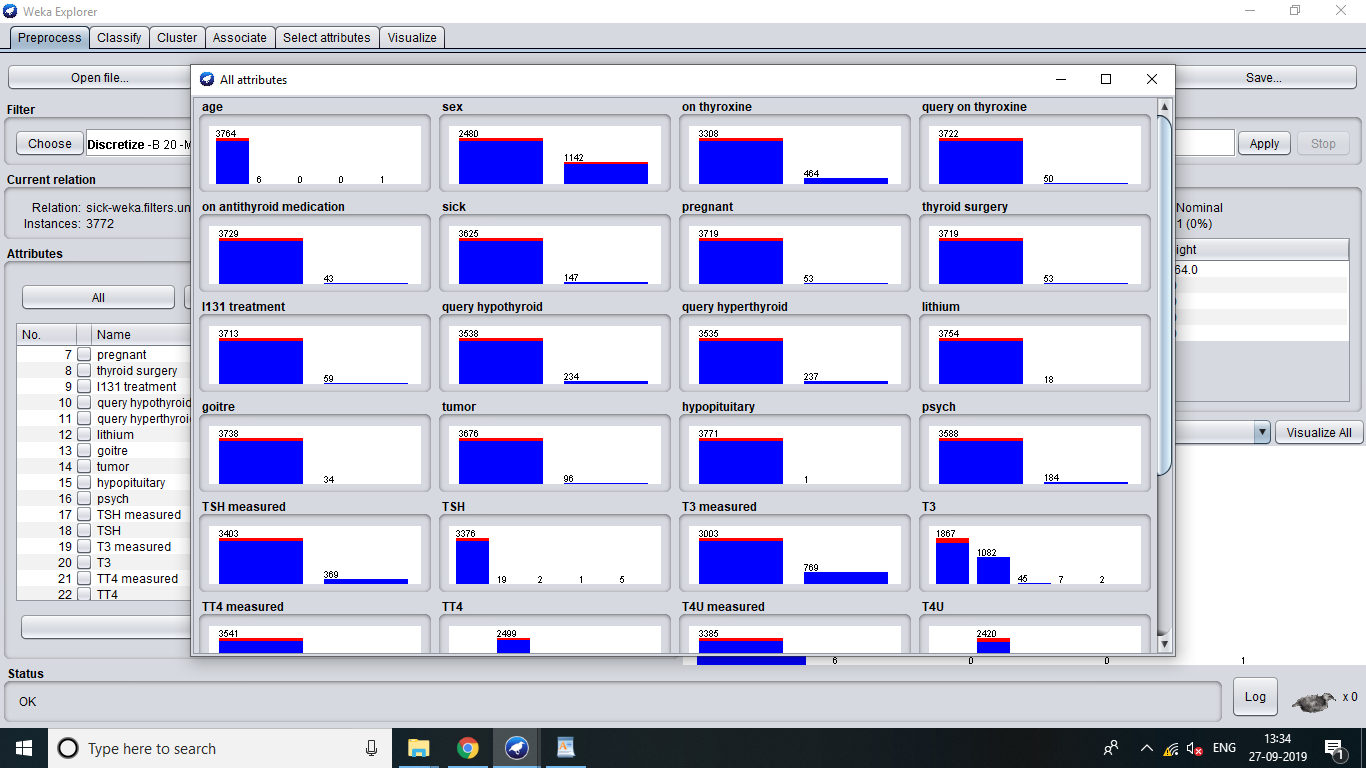


5. Undo the filter applied in the previous step.

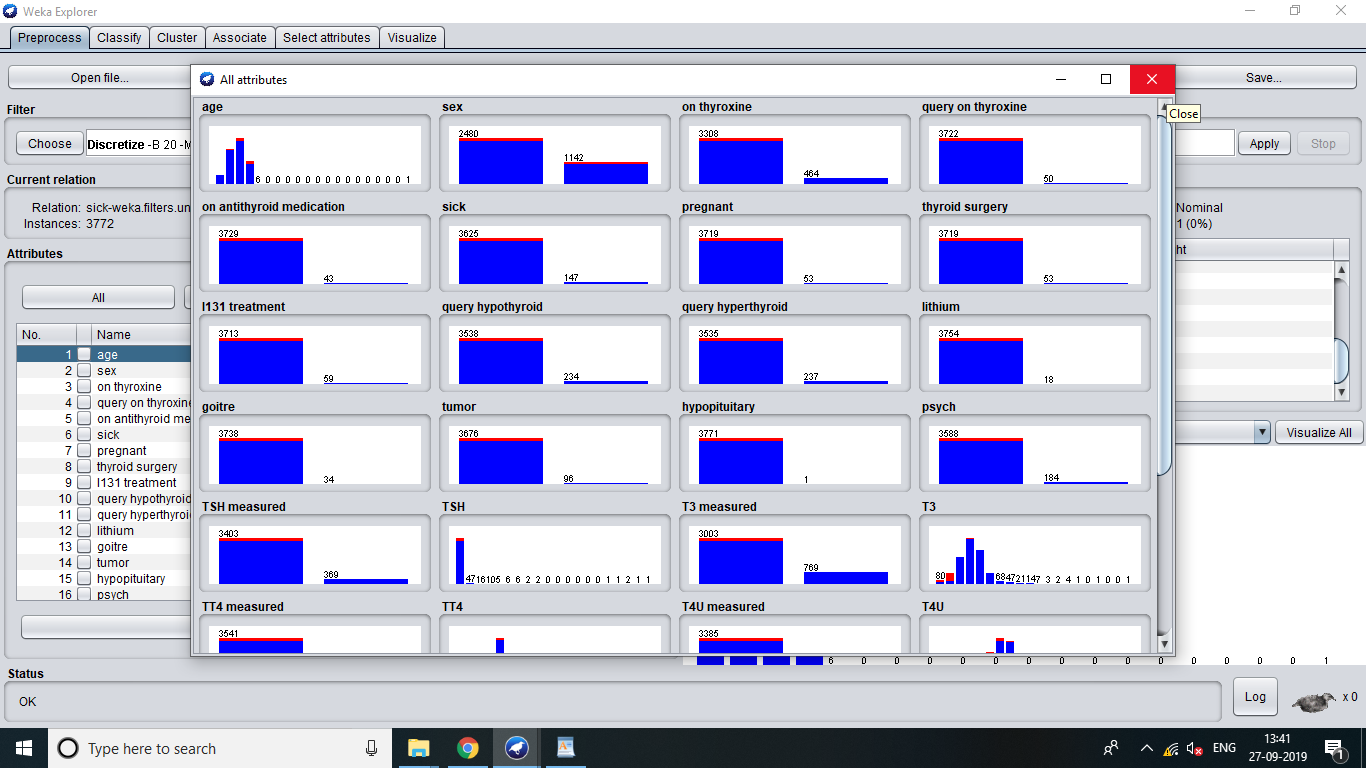
done

6. Apply the unsupervised discretization filter. Do this twice:

1. In this step, set 'bins'=5



2. In this step, set 'bins'=10

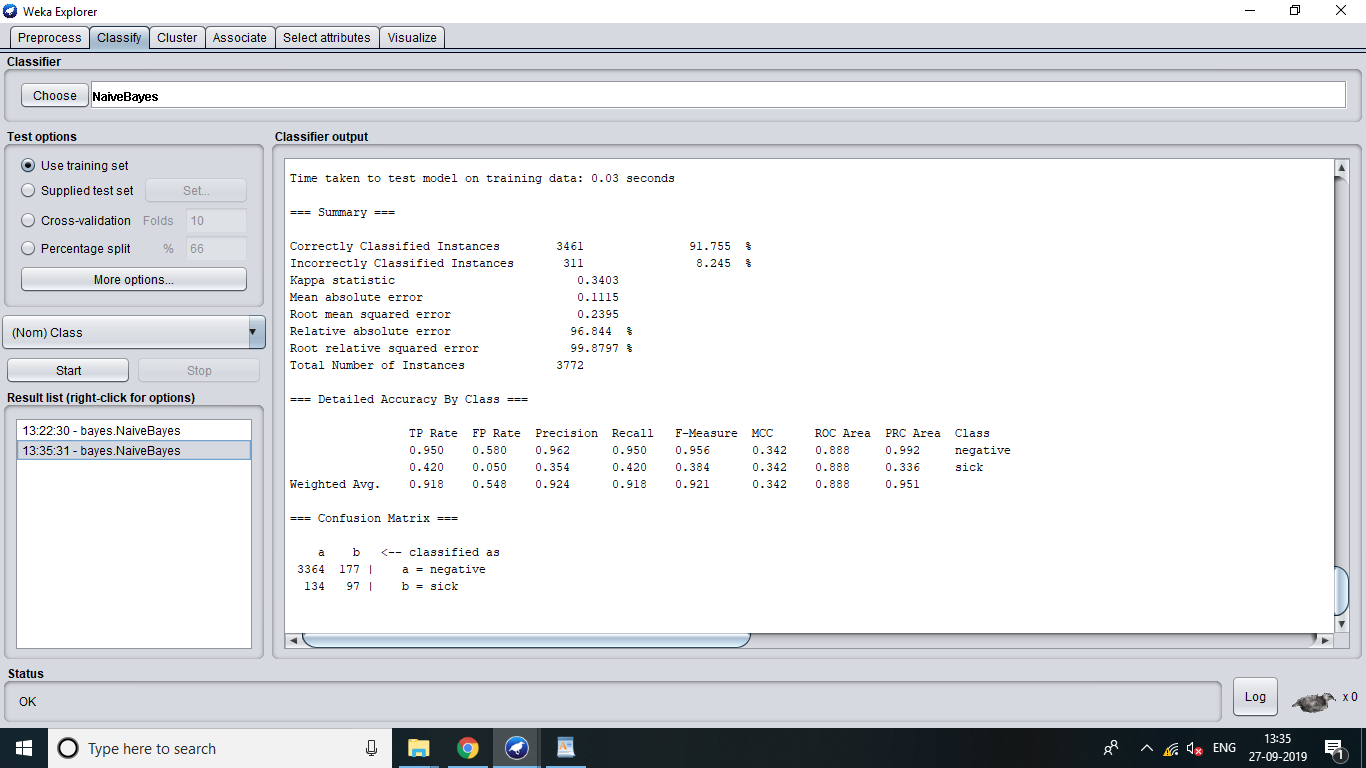


3. What is the effect of the unsupervised filter filter on the datset?

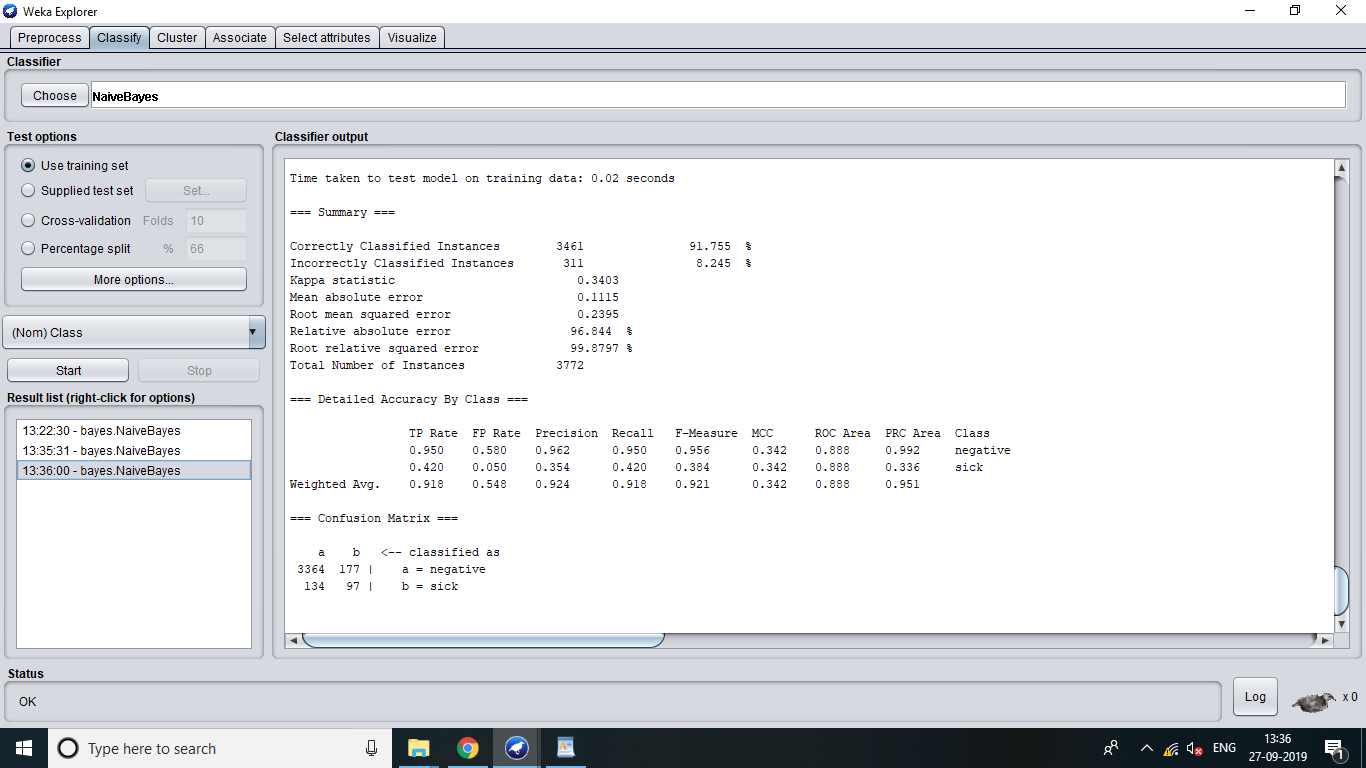
it just divides the table in 5 bins

7. Run the the Naive Bayes classifier after apply the following filters

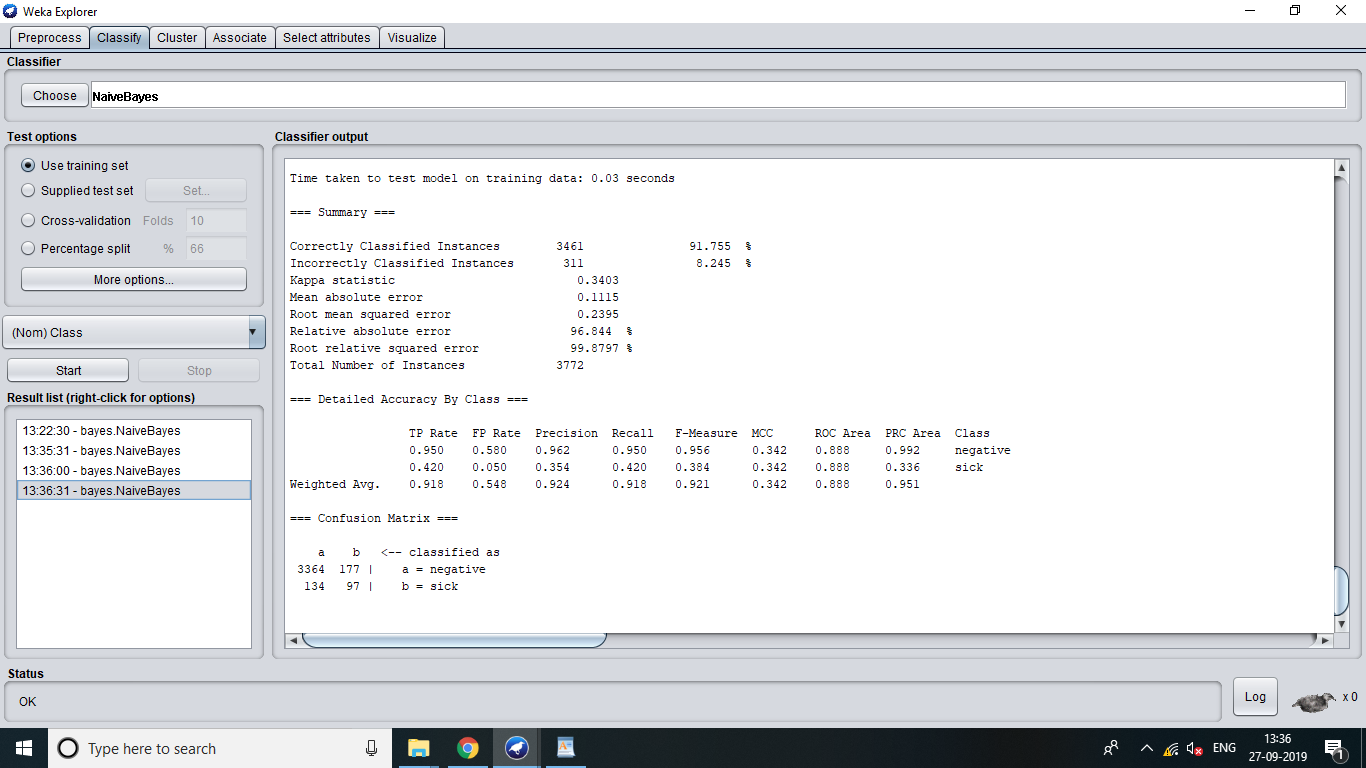
1. Unsupervised discretized with 'bins'=5



2. Unsupervised discretized with 'bins'=10



3. Unsupervised discretized with 'bins''=20.



8. Compare the accuracy of the following cases

1. Naive Bayes without discretization filters

92.9745 without the discretization filters

2. Naive Bayes with a supervised discretization filter

91.155

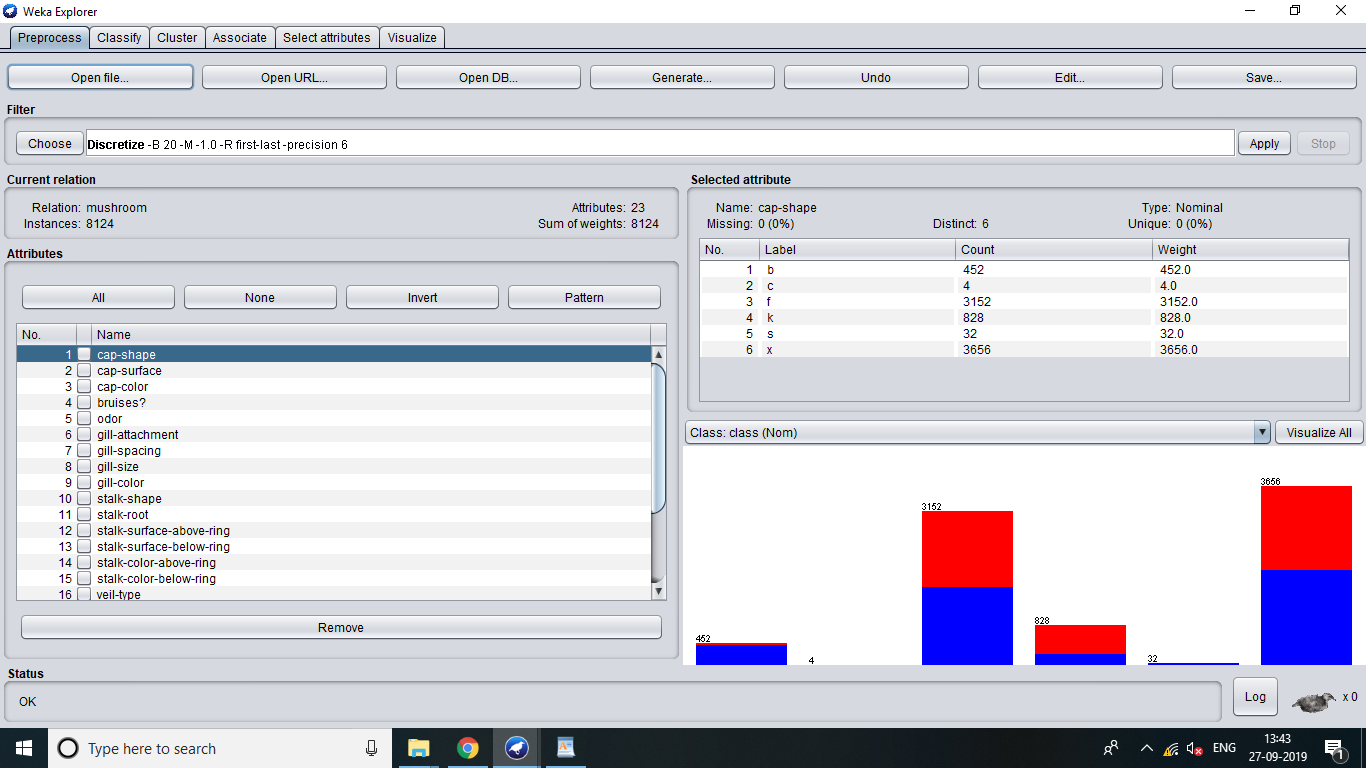
3. Naive Bayes with an unsupervised discretization filter with different values for the 'bins attributes.

Part II: Attribute Selection

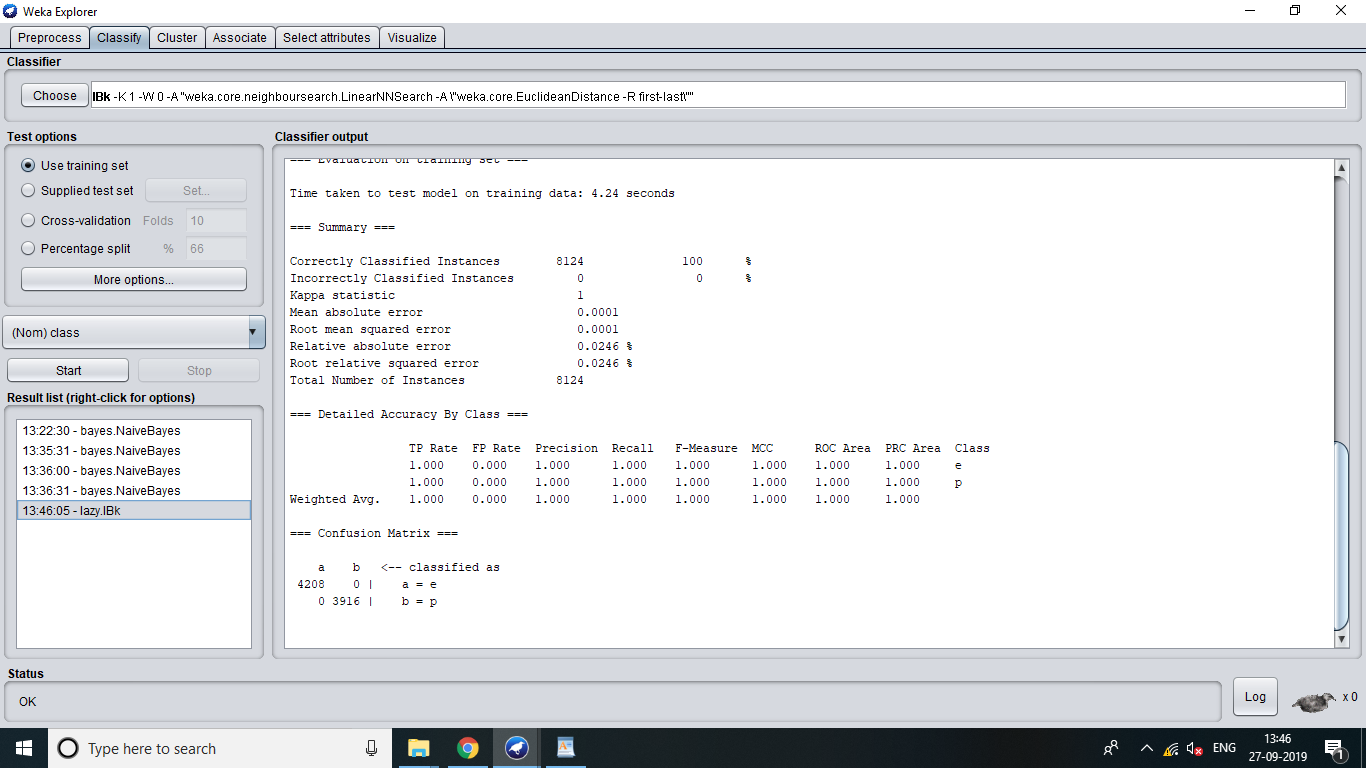
1. Perform the following tasks:

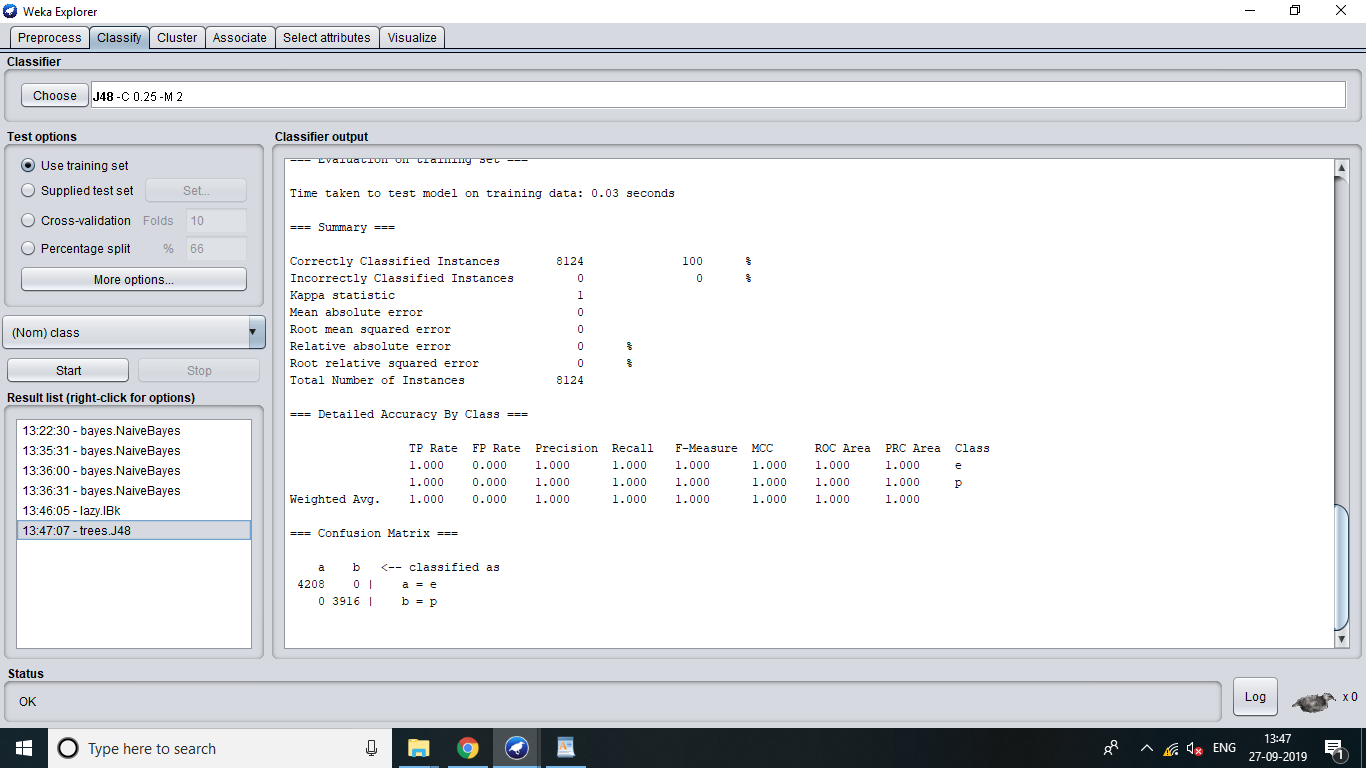
1. Load the 'mushroom.arff' dataset

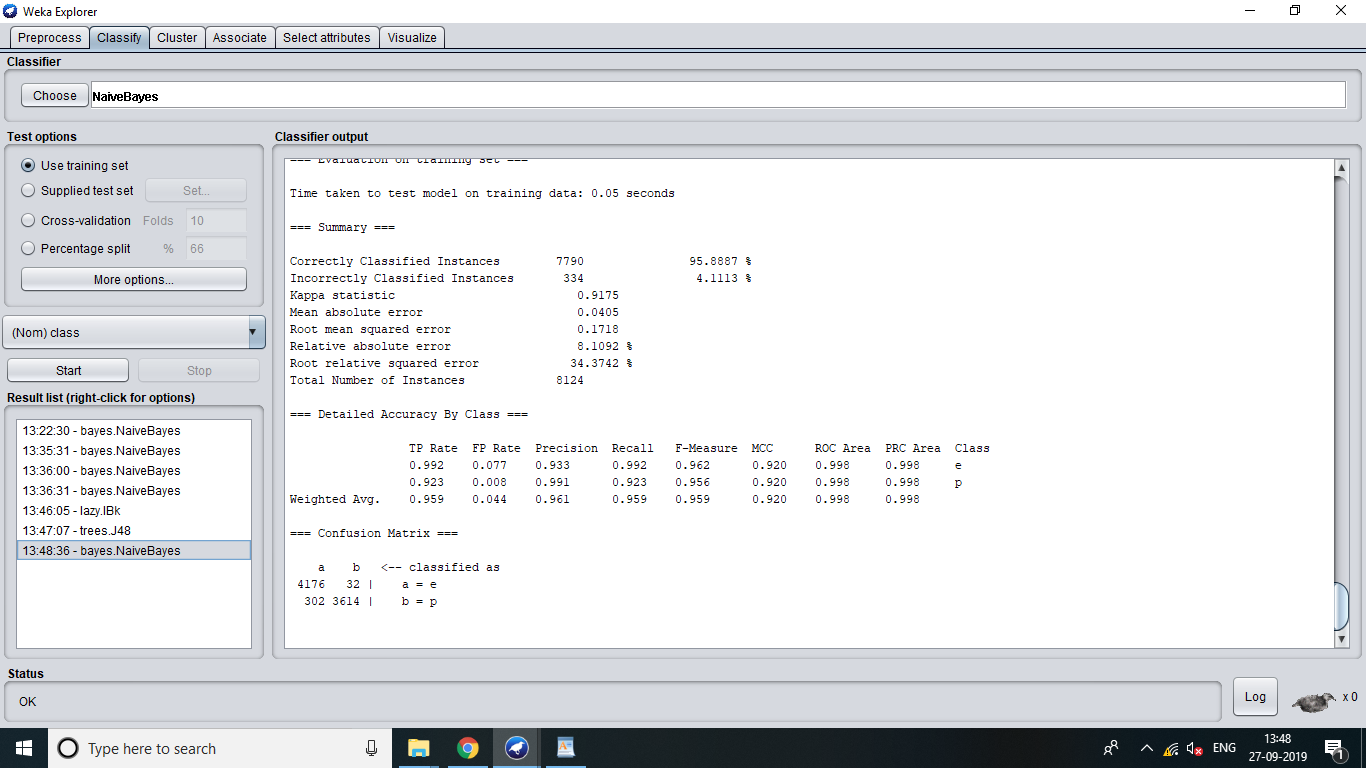
loaded



2. Run the J48, 1Bk, and the Naive Bayes classifiers.







3. What is the accuracy of each of these classifiers?

100% for j48

100% for Ibk

95.888% for NAIVE BAYES

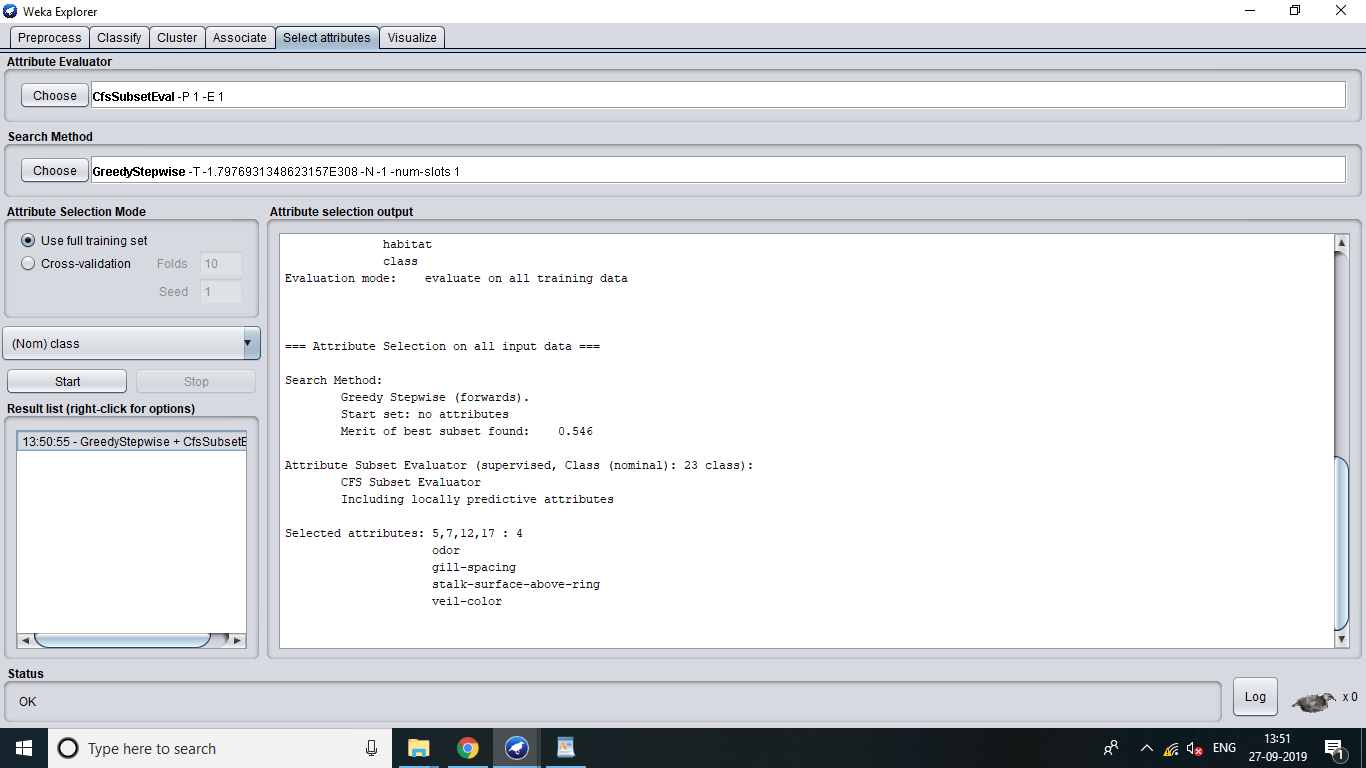
2. Perform the following tasks:

1. Go to the 'Select Attributes' panel

2. Set attribute evaluator to CFSSubsetEval

3. Set the search method to 'Greedy Stepwise'

4. Analyze the results window



5. Record the attribute numbers of the most important attributes

5,7,12,17 : 4

odor

gill-spacing

stalk-surface-above-ring

veil-color

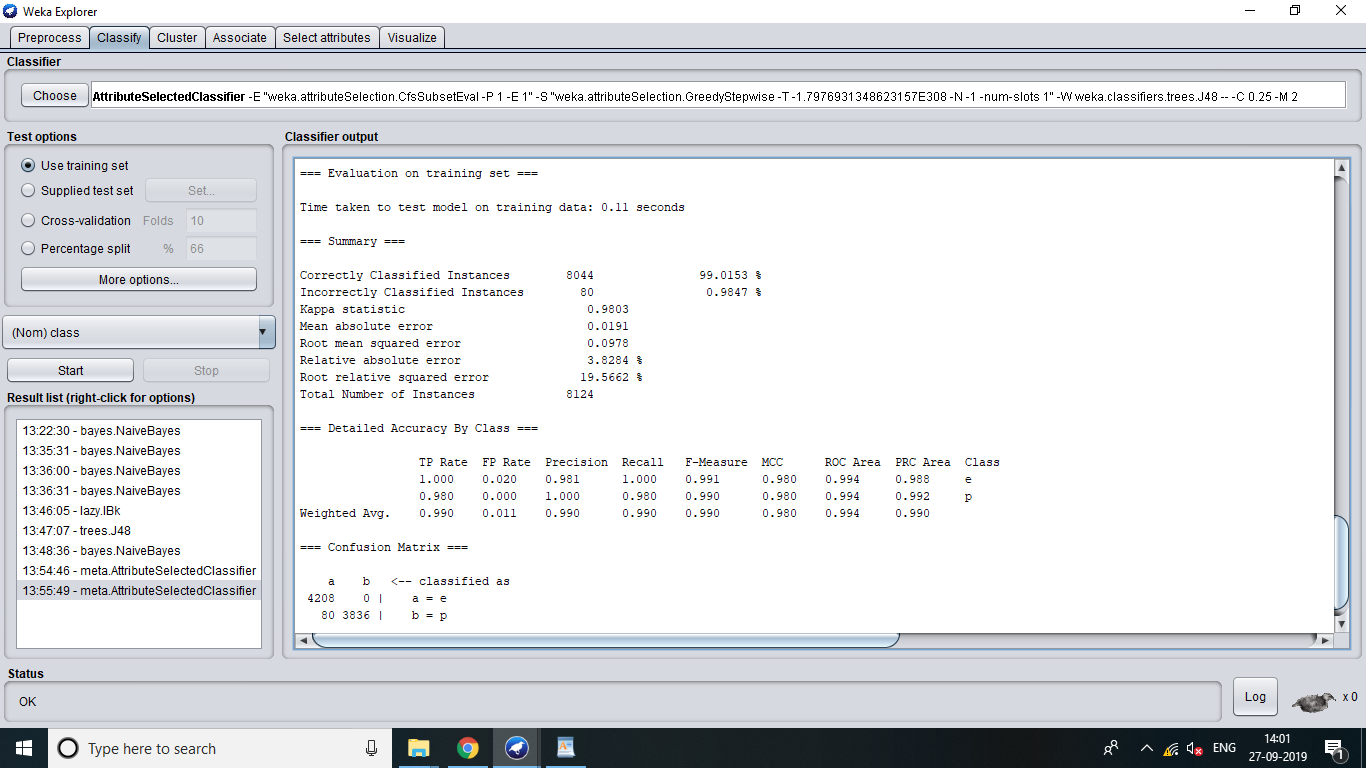
6. Run the meta classifier AttributeSelectedClassifier using the following:

1. CFSSubsetEval

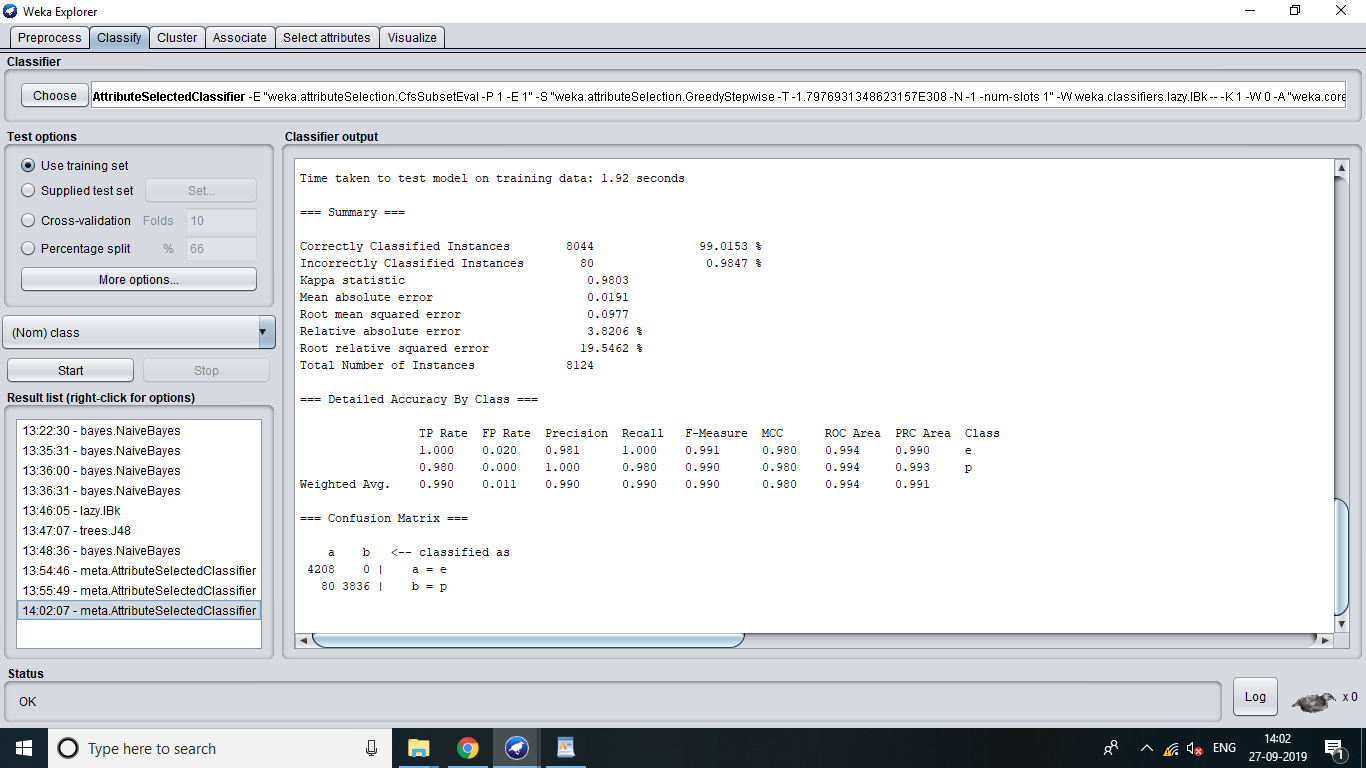
2. GreedStepwise

3. J48, 1Bk, and NaiveBayes

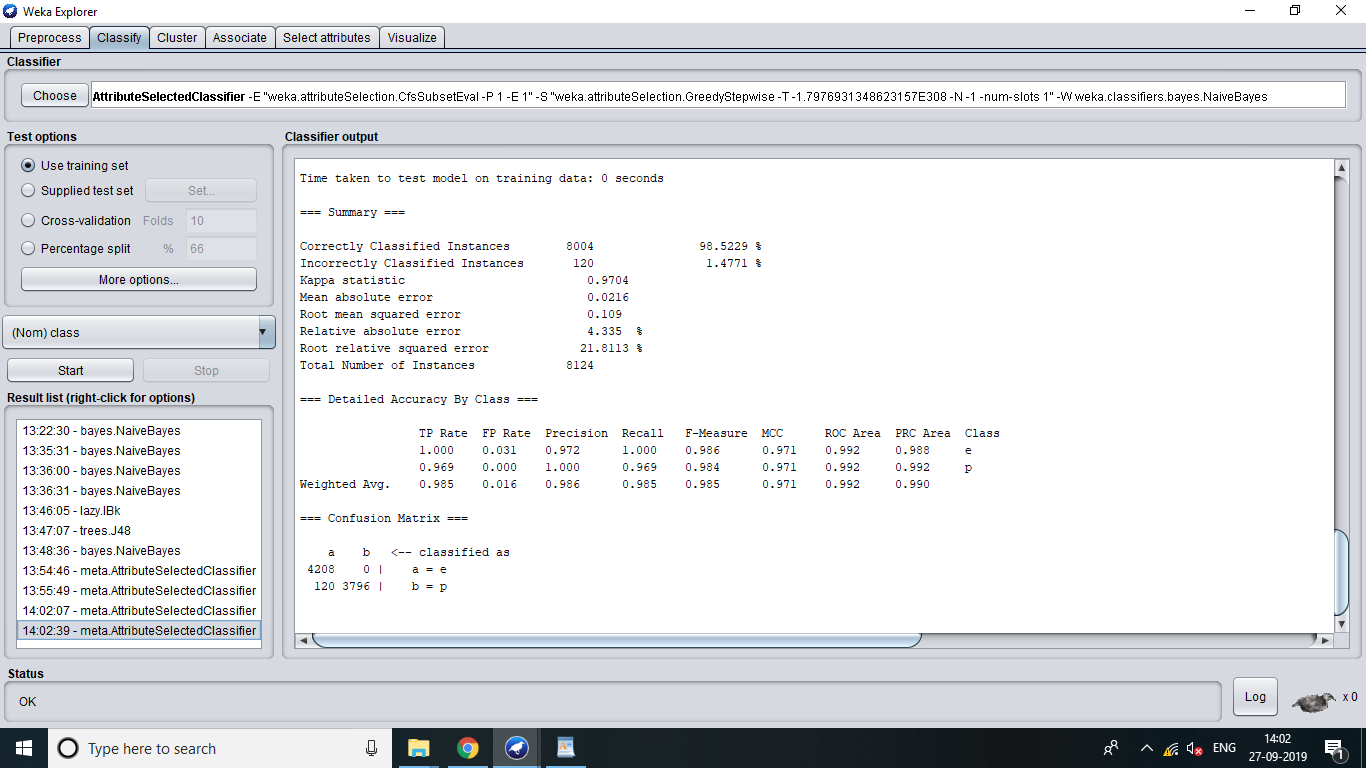
J48



1bk



NAIVE BAYES



8. What are the benefits of attribute selection?

Ans. The benefits of attribute section is that it helps to give classification.