Experiment 4: Classification of data using Bayesian approach

AIM: To apply naïve bayes classifier on a given data set.

Description:

In machine learning, Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes‟ Theorem with strong (naïve) independence assumptions between the features

Example:

.

AGE	INCOME	STUDENT	CREDIT_RATING	BUYS_COMPUTER
<30	High	No	Fair	No
<30	High	No	Excellent	No
31-40	High	No	Fair	Yes
>40	Mediu m	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31-40	Mediu m	Yes	Excellent	Yes
<=30	Low	No	Fair	No
<=30	Mediu m	Yes	Fair	Yes
>40	Mediu m	Yes	Fair	Yes
<30	Mediu m	Yes	Excellent	Yes
31-40	Mediu m	No	Excellent	Yes
31-40	High	Yes	Fair	Yes
>40	Mediu m	No	Excellent	No

CLASS:

C1:buys_com puter = ‘yes’

C2:buys_com puter=’no’

DATA TO BECLASSIFIED :

X= (age<=30, income=Medium, Student=Yes, credit_rating=Fair)

- P(C1): P(buys_computer=”yes”)= 9/14 =0.643
  P (buys_computer=”no”) =5/14=0.357

- Compute P(X/C1) and p(x/c2) weget:

1. P(age=”<=30” |buys_computer=”yes”)=2/9
2. P(age=”<=30”|buys_computer=”no”)=3/5
3. P(income=”medium”|buys_computer=”yes”)=4/9
4. P(income=”medium”|buys_computer=”no”)=2/5
5. P(student=”yes”|buys_computer=”yes”)=6/9
6. P(student=”yes” |buys_computer=”no”)=1/5=0.2
7. P(credit_rating=”fair ”|buys_computer=”yes”)=6/9
8. P(credit_rating=”fair” |buys_computer=”no”)=2/5

- X=(age<=30, income=medium, student=yes, credit_rating=fair) P(X/C1): P (X/buys_computer=”yes”)=2/9*4/9*6/9*6/9= 32/1134
  P(X/C2):P(X/buys_computer=”no”)=3/5*2/5*1 /5*2/5=12/125

  P(C1/X)=P(X/C1)*P(C1)

  P(X/buys_computer=”yes”)*P(buys_computer=”yes”)=(32/1134)*(9/14)=0.019

  P(C2/X)=p(x/c2)*p(c2)

  P (X/buys_computer=”no”)*P(buys_computer=”no”)=(12/125)*(5/14)=0.007

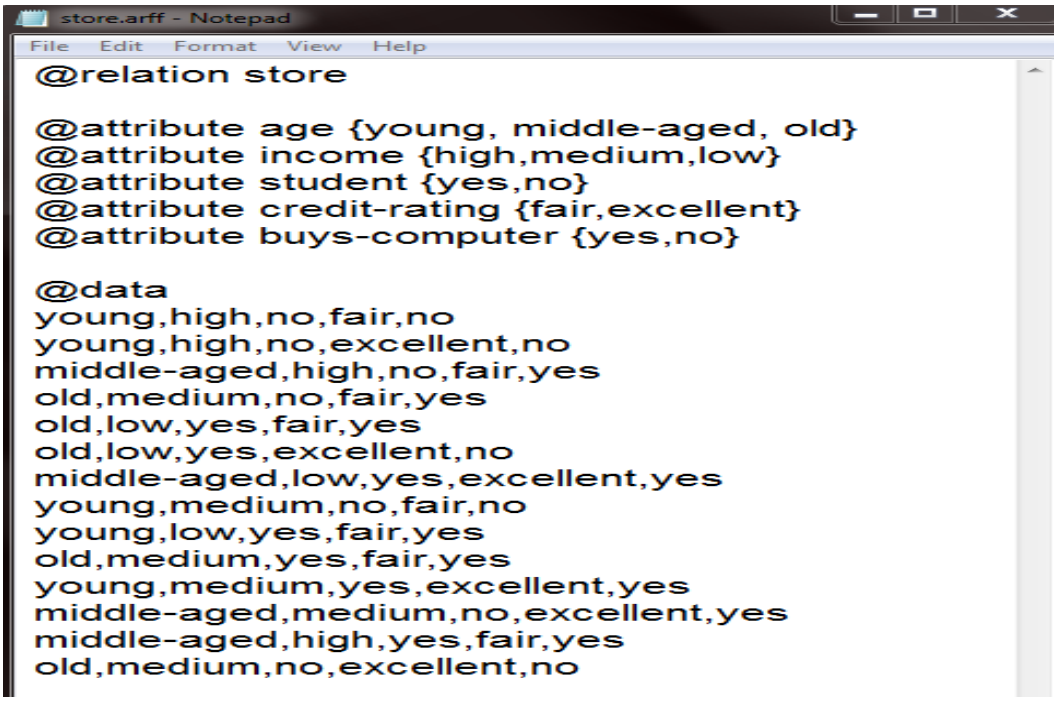
  Therefore, conclusion is that the given data belongs to C1 since P(C1/X)>P(C2/X)

Checking the result in the WEKA tool:

In order to check the result in the tool we need to follow a procedure.

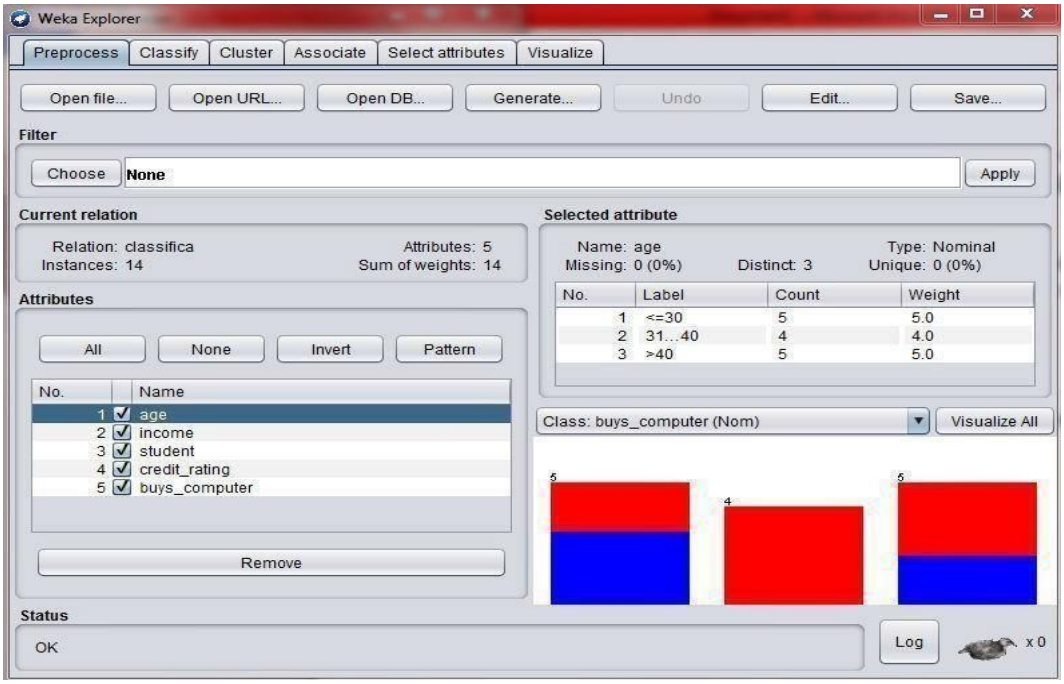
Step 1:

Create a csv file with the above table considered in the example. the arfffile will look as shown below:



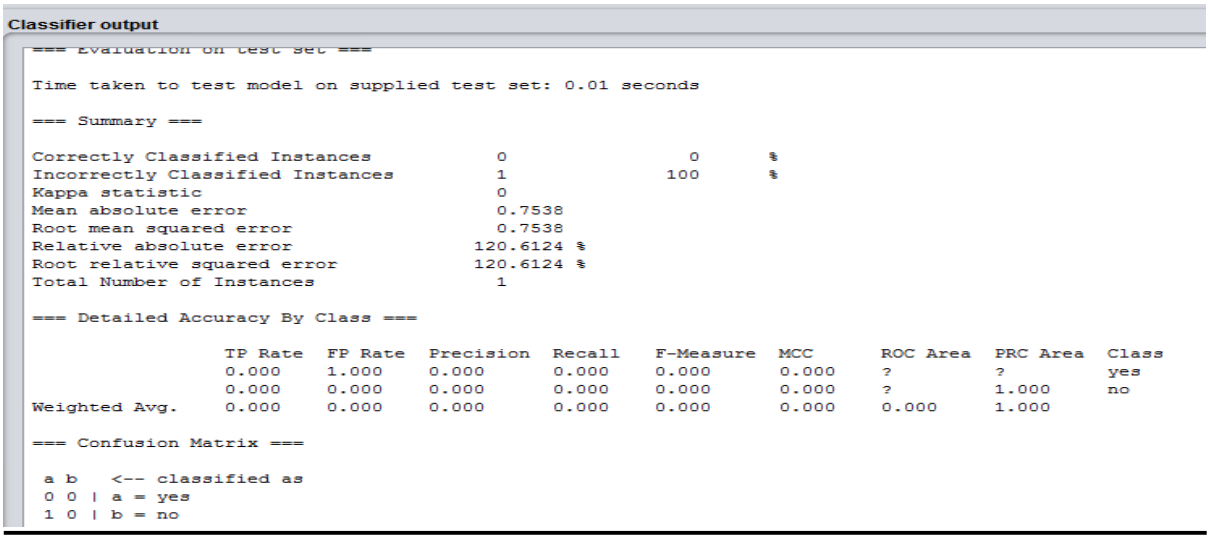
Step 2:

Now open weka explorer and then select all the attributes in the table.



Step 3:

Select the classifier tab in the tool and choose baye‟s folder and then naïve baye‟s classifier to see the result as shown below.



Classifier output

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances           0               0      %
Incorrectly Classified Instances         1             100      %
Kappa statistic                          0
Mean absolute error                      0.7538
Root mean squared error                  0.7538
Relative absolute error                120.6124 %
Root relative squared error            120.6124 %
Total Number of Instances                1

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    1.000    0.000      0.000    0.000      0.000    ?         ?         yes
                 0.000    0.000    0.000      0.000    0.000      0.000    ?         1.000     no
Weighted Avg.    0.000    0.000    0.000      0.000    0.000      0.000    0.000     1.000

=== Confusion Matrix ===

 a b   <-- classified as
 0 0 | a = yes
 1 0 | b = no
```

Exercise

1. Classify data (lung cancer/ diabetes /liver disorder) using Bayesian approach .