

Experiment 2.2

Student Name: Yash Gupta
Branch: BE-CSE
Semester: 6
Subject Name: DM LAB

UID: 20BCS5009
Section/Group: 20BCS_DM-716 B
Date of Performance: 28/03/23
Subject Code: 20CSP_376

AIM :-

Performing classification using Bayesian classification algorithm.

Theory :-

Naive Bayes is a Supervised Non-linear classification algorithm in R Programming. Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Baye's theorem with strong(Naive) independence assumptions between the features or variables. The Naive Bayes algorithm is called "Naive" because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

where,

$P(A|B)$ = Conditional probability of A given B.

$P(B|A)$ = Conditional probability of B given A.

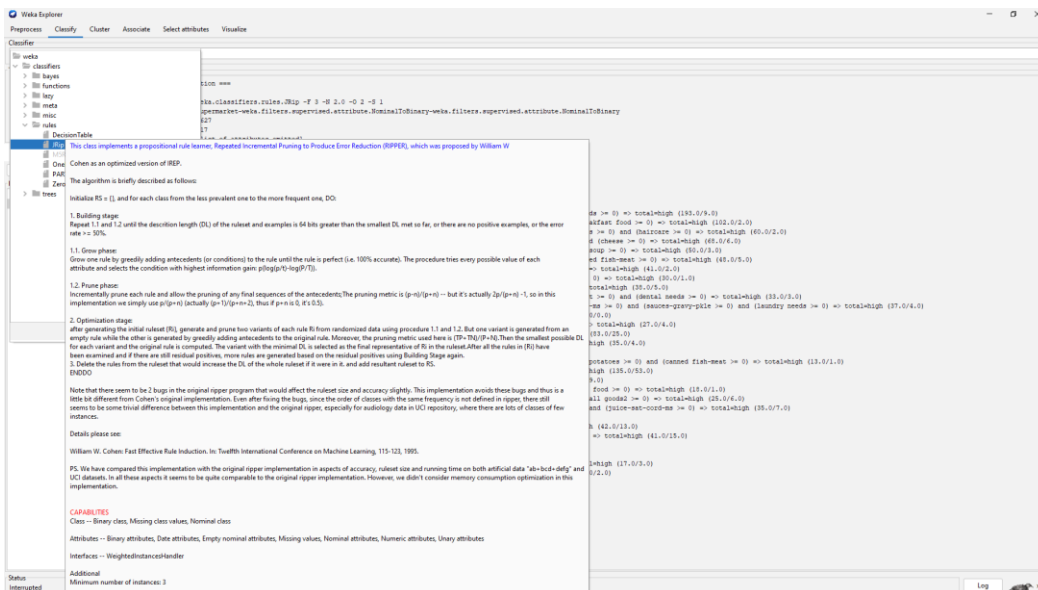
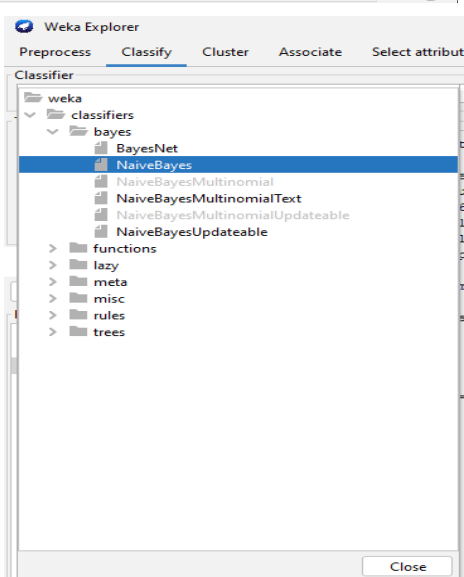
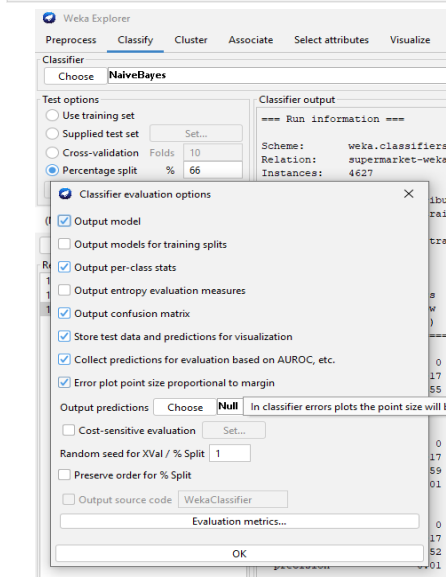
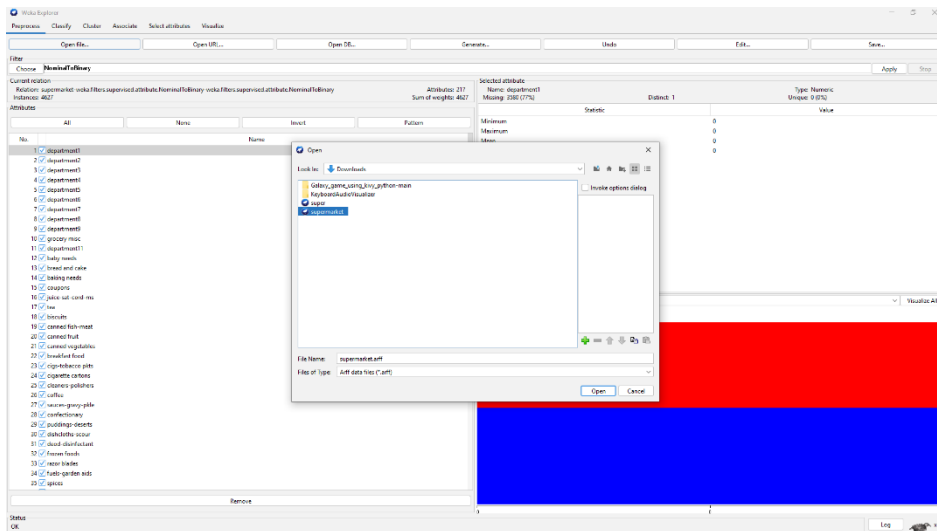
$P(A)$ = Probability of event A.

$P(B)$ = Probability of event B.

For many predictors, we can formulate the posterior probability as follows:

$$P(A|B) = P(B1|A) * P(B2|A) * P(B3|A) * P(B4|A)$$

Output :-



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classify

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds 10

☒ Percentage split % 66

More options...

(None) total

Start Stop

Result list (right click for options)

151212 - rules.Rip

151218 - rules.Rip

151242 - bayes.NaiveBayes

Classifier output

=== Run information ===

Subset: weka.classifiers.bayes.NaiveBayes

Relation: supermarket-weka.filters.supervised.attribute.NominalToBinary

Instances: 4427

Attributes: 317

(List of attributes omitted)

Test mode: split 66.66 train, remainder test

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

Naive Bayes Classifier

Attribute	Class	low	high
		(0.44)	(0.56)
department1	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	550	492
	precision	0.92	0.92
department2	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	88	72
	precision	0.93	0.93
department3	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	32	20
	precision	0.93	0.93
department4	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	47	37
	precision	0.92	0.92
department5	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	87	88
	precision	0.93	0.93
department6	mean	0	0
	std. dev.	0.0017	0.0017
	weight sum	0	2
	precision	0.93	0.93

Status OK Log

Classifier output

```

precision                0.01    0.01

department214
mean                    0        0
std. dev.              0.0017  0.0017
weight sum             0        0
precision              0.01    0.01

department215
mean                    0        0
std. dev.              0.0017  0.0017
weight sum             0        0
precision              0.01    0.01

department216
mean                    0        0
std. dev.              0.0017  0.0017
weight sum             0        0
precision              0.01    0.01

Time taken to build model: 0.06 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.04 seconds

=== Summary ===

Correctly Classified Instances      986          62.6828 %
Incorrectly Classified Instances    587          37.3172 %
Kappa statistic                    0
Mean absolute error                0.4639
Root mean squared error            0.4839
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          1573

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              1.000    1.000    0.627     1.000    0.771      ?       0.503    0.629    low
              0.000    0.000    ?         0.000    ?         ?       0.529    0.388    high
Weighted Avg.   0.627    0.627    ?         0.627    ?         ?       0.513    0.539

=== Confusion Matrix ===

  a  b  <-- classified as
986  0 |  a = low
587  0 |  b = high

```