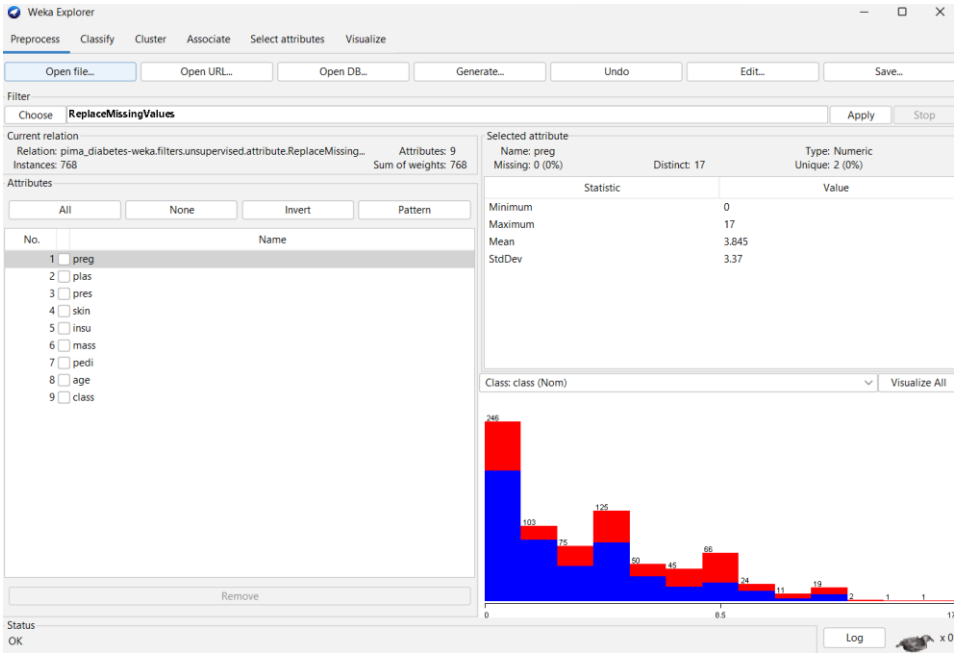
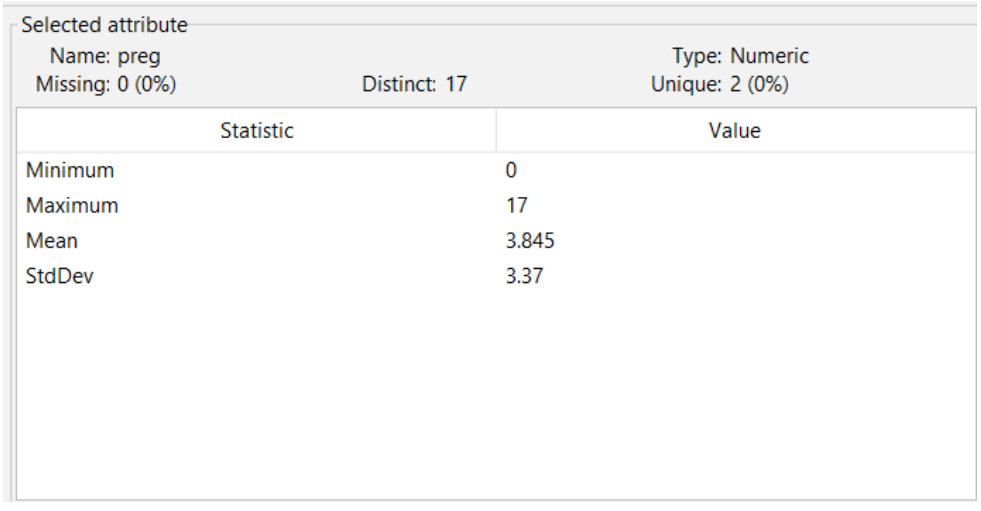
Course Code and Name: CSI0803 Data Mining

PRACTICAL 10

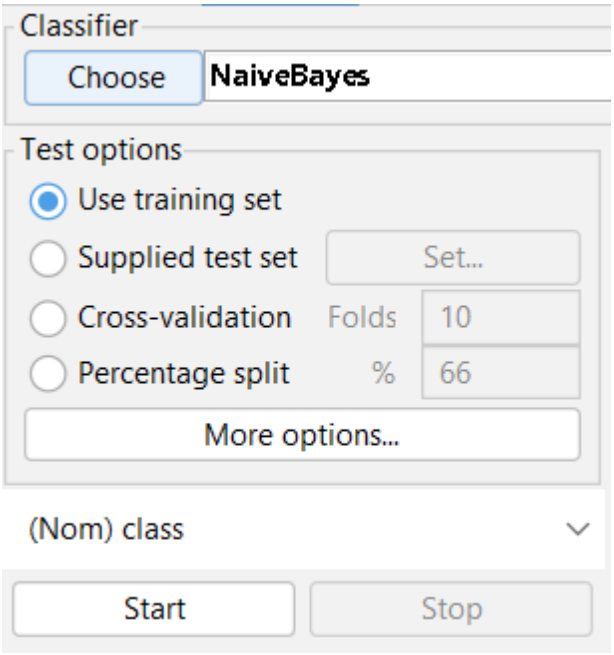
Opening dataset for classification:



Filling missing values:



Using Naïve Bayes Algorithm:



OUTPUT:

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     pima_diabetes-weka.filters.unsupervised.attribute.ReplaceMissingValues
Instances:    768
Attributes:   9
              preg
              plas
              pres
              skin
              insu
              mass
              pedi
              age
              class
Test mode:    evaluate on training data

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

Naive Bayes Classifier
```

Classifier output

Attribute	Class tested_negative (0.65)	tested_positive (0.35)
preg		
mean	3.4234	4.9795
std. dev.	3.0166	3.6827
weight sum	500	268
precision	1.0625	1.0625
plas		
mean	109.9541	141.2581
std. dev.	26.1114	31.8728
weight sum	500	268
precision	1.4741	1.4741
pres		
mean	68.1397	70.718
std. dev.	17.9834	21.4094
weight sum	500	268
precision	2.6522	2.6522
skin		
mean	19.8356	22.2824
std. dev.	14.8974	17.6992
weight sum	500	268
precision	1.98	1.98
insu		
mean	68.8507	100.2812
std. dev.	98.828	138.4883
weight sum	500	268
precision	4.573	4.573
mass		
mean	30.3009	35.1475
std. dev.	7.6833	7.2537
weight sum	500	268
precision	0.2717	0.2717
pedi		
mean	0.4297	0.5504
std. dev.	0.2986	0.3715
weight sum	500	268
precision	0.0045	0.0045
age		
mean	31.2494	37.0808
std. dev.	11.6059	10.9146
weight sum	500	268
precision	1.1765	1.1765

```
Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances         586               76.3021 %
Incorrectly Classified Instances       182               23.6979 %
Kappa statistic                          0.4674
Mean absolute error                      0.2811
Root mean squared error                  0.4133
Relative absolute error                 61.8486 %
Root relative squared error             86.7082 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.842    0.384    0.803      0.842    0.822      0.469    0.825     0.902     tested_negative
                 0.616    0.158    0.676      0.616    0.645      0.469    0.825     0.684     tested_positive
Weighted Avg.    0.763    0.305    0.759      0.763    0.760      0.469    0.825     0.826

=== Confusion Matrix ===

   a   b   <-- classified as
 421  79 |   a = tested_negative
 103 165 |   b = tested_positive
```

Using Decision Tree:

```
Decision Stump

Classifications

plas <= 127.5 : tested_negative
plas > 127.5 : tested_positive
plas is missing : tested_negative

Class distributions

plas <= 127.5
tested_negative tested_positive
0.8061855670103093      0.19381443298969073
plas > 127.5
tested_negative tested_positive
0.38515901060070673     0.6148409893992933
plas is missing
tested_negative tested_positive
0.6510416666666666      0.3489583333333333
```

```
Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         565               73.5677 %
Incorrectly Classified Instances       203               26.4323 %
Kappa statistic                          0.4257
Mean absolute error                      0.3719
Root mean squared error                  0.4312
Relative absolute error                 81.8217 %
Root relative squared error             90.4671 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.782    0.351    0.806      0.782    0.794      0.426    0.716     0.772     tested_negative
                 0.649    0.218    0.615      0.649    0.632      0.426    0.716     0.522     tested_positive
Weighted Avg.    0.736    0.304    0.739      0.736    0.737      0.426    0.716     0.685

=== Confusion Matrix ===

   a   b   <-- classified as
 391 109 |   a = tested_negative
  94 174 |   b = tested_positive
```