Abdul Saboor (20L-1113 | BDS-6A1)

Created	@March 2, 2023
Course	Data Mining

```
−0.367fractal_dimension_mean−0.28fractal_dimension_se−0.275fractal_dimension_worst+0.234radius_mean−0.233compa
  mean
  std. dev.
  weight sum
  precision


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         519               91.2127 %
Incorrectly Classified Instances        50                8.7873 %
Kappa statistic                          0.8099
Mean absolute error                      0.1204
Root mean squared error                  0.2536
Relative absolute error                 25.7542 %
Root relative squared error             52.4474 %
Total Number of Instances              569

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.854    0.053    0.905      0.854   0.879      0.811  0.972     0.943     M
                 0.947    0.146    0.916      0.947   0.931      0.811  0.972     0.984     B
Weighted Avg.    0.912    0.112    0.912      0.912   0.912      0.811  0.972     0.969

=== Confusion Matrix ===

   a   b   <-- classified as
 181  31 |   a = M
  19 338 |   b = B
```
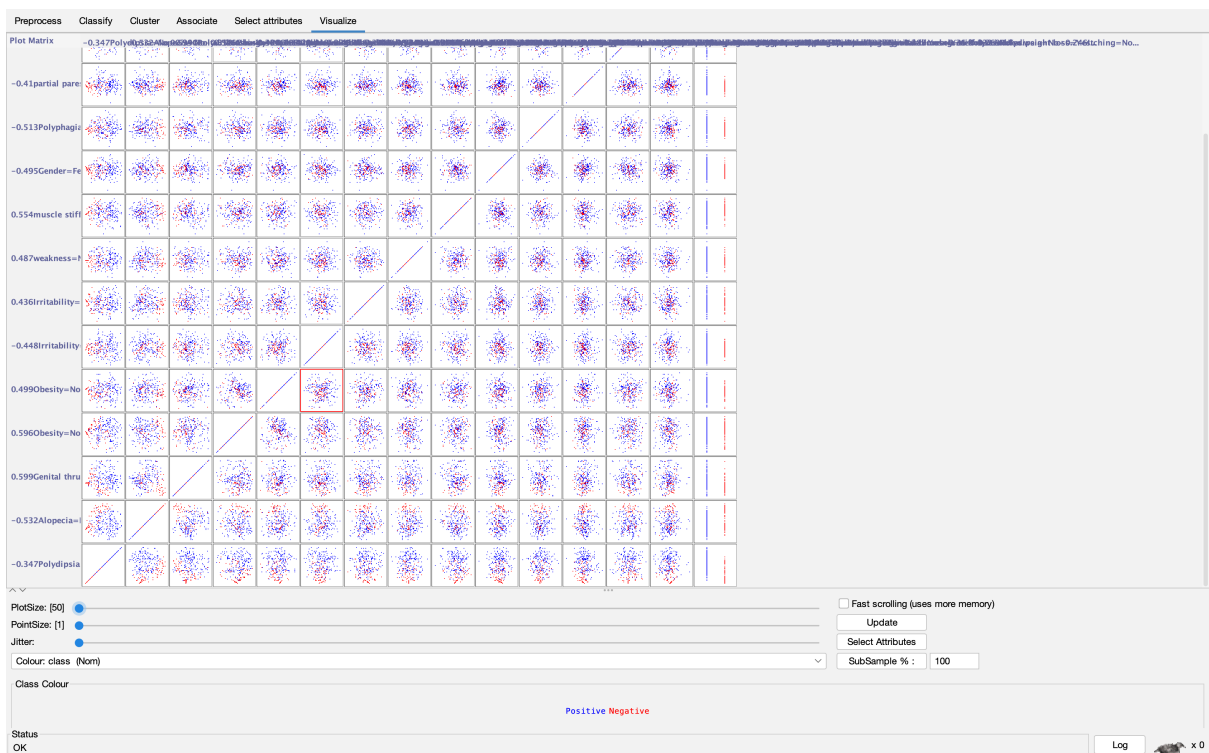
Diabetes Dataset

PCA with correlation matrix

- variance is 0.95

- variance is 0.5

Classification Algorithms

- we remove the **last 3 attributes** as it had no effect on overall data

Decision Trees

```
|  |  |  |     -0.495Polyphagia=Yes+0.494partial paresis=Yes+0.42Itching=No-0.309delayed healing=No+0.287Ger
|   -0.515Polyuria=Yes-0.39weakness=No+0.377Genital thrush=Yes+0.361Alopecia=No+0.315Polyphagia=Yes... > 0.943
0.349Polydipsia=No-0.342Polyuria=Yes-0.338partial paresis=Yes-0.317Polyphagia=Yes-0.307visual blurring=Yes...

Number of Leaves  :     26

Size of the tree :     51


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         474               91.1538 %
Incorrectly Classified Instances        46                8.8462 %
Kappa statistic                          0.6854
Mean absolute error                      0.1011
Root mean squared error                  0.2893
Relative absolute error                 35.8202 %
Root relative squared error             77.1604 %
Total Number of Instances              520

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.739    0.053    0.739      0.739   0.739      0.685    0.863     0.643     Yes
                 0.947    0.261    0.947      0.947   0.947      0.685    0.863     0.950     No
Weighted Avg.    0.912    0.226    0.912      0.912   0.912      0.685    0.863     0.898

=== Confusion Matrix ===

   a   b   <-- classified as
  65  23 |   a = Yes
  23 409 |   b = No
```

Random Forest

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.17 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         493               94.8077 %
Incorrectly Classified Instances        27                5.1923 %
Kappa statistic                          0.796
Mean absolute error                      0.1067
Root mean squared error                  0.2111
Relative absolute error                 37.8372 %
Root relative squared error             56.2951 %
Total Number of Instances              520

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.727    0.007    0.955      0.727   0.826      0.806    0.952     0.894     Yes
                 0.993    0.273    0.947      0.993   0.969      0.806    0.952     0.986     No
Weighted Avg.    0.948    0.228    0.948      0.948   0.945      0.806    0.952     0.971

=== Confusion Matrix ===

   a   b   <-- classified as
  64  24 |   a = Yes
   3 429 |   b = No
```

Naive Bayes

```
-0.551delayed healing=No-0.296partial paresis=Yes+0.284visual blurring=Yes-0.268Polydipsia=No+0.246Itching=No.
  mean
  std. dev.
  weight sum
  precision



Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         425                81.7308 %
Incorrectly Classified Instances        95                18.2692 %
Kappa statistic                          0.1137
Mean absolute error                      0.2521
Root mean squared error                  0.3674
Relative absolute error                 89.3538 %
Root relative squared error             97.9726 %
Total Number of Instances              520

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.125    0.042    0.379      0.125    0.188      0.136    0.704     0.330     Yes
                 0.958    0.875    0.843      0.958    0.897      0.136    0.704     0.910     No
Weighted Avg.    0.817    0.734    0.765      0.817    0.777      0.136    0.704     0.812

=== Confusion Matrix ===

   a    b   <-- classified as
  11   77 |   a = Yes
  18  414 |   b = No
```
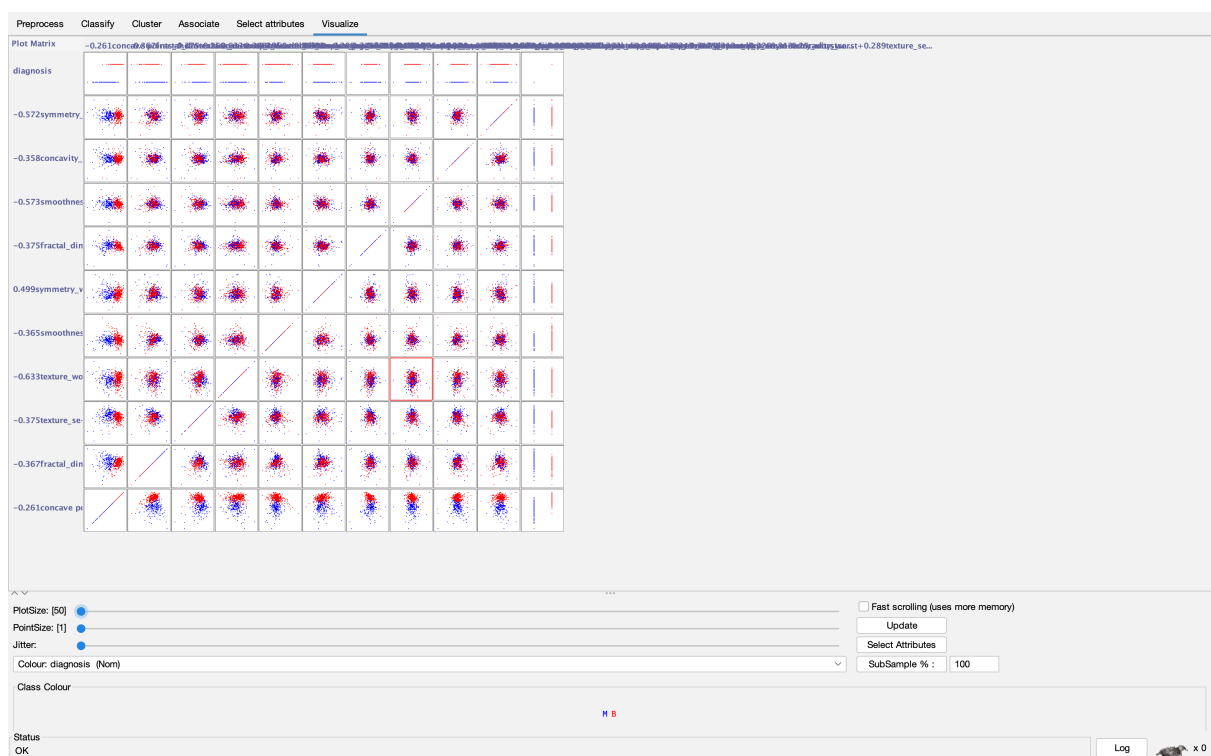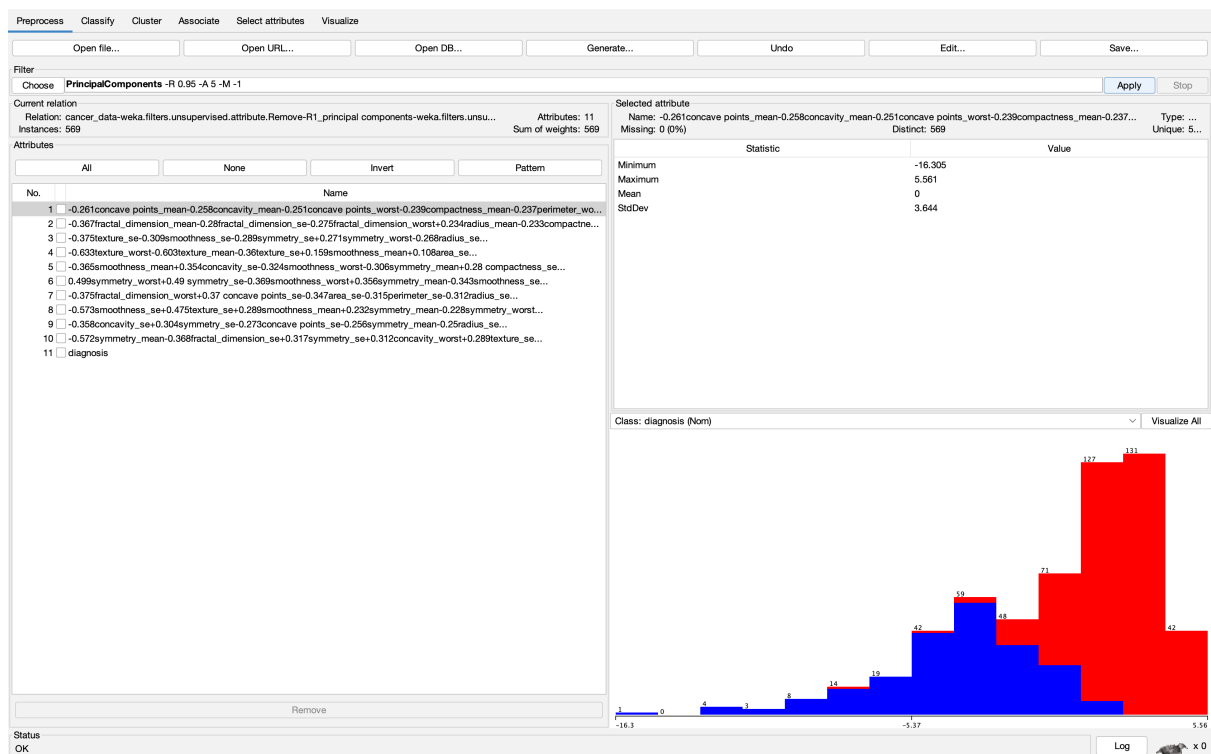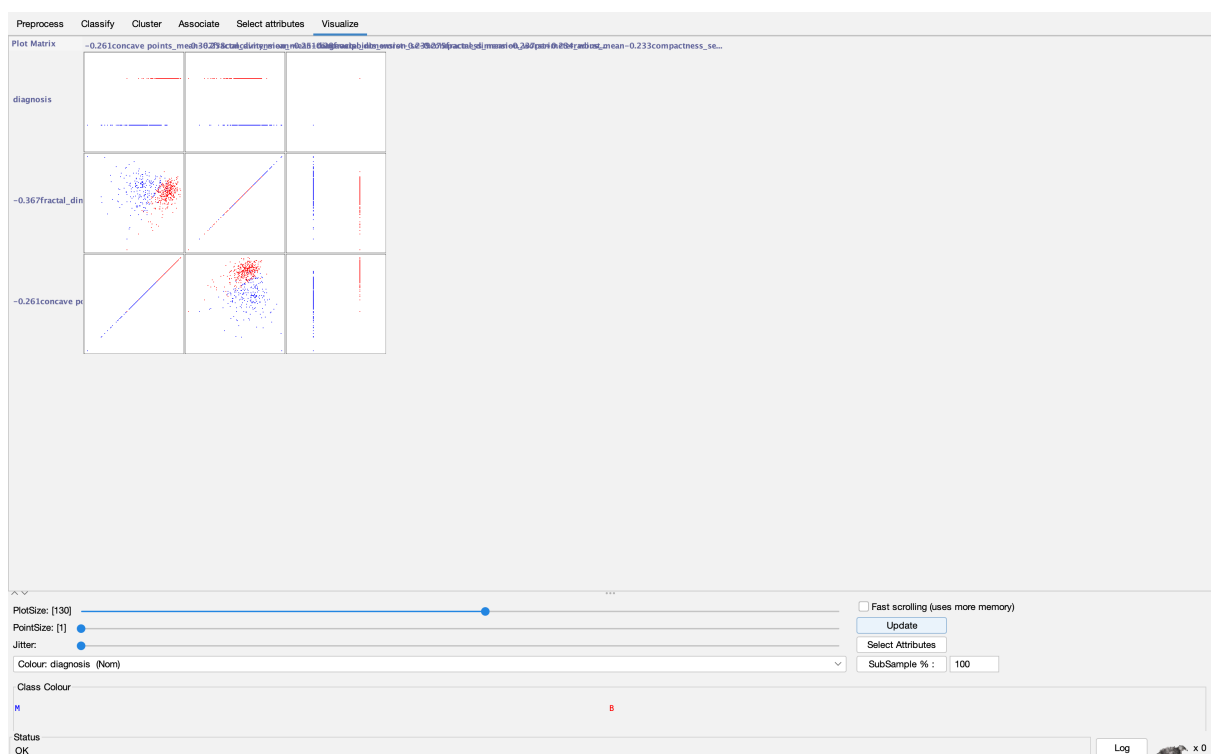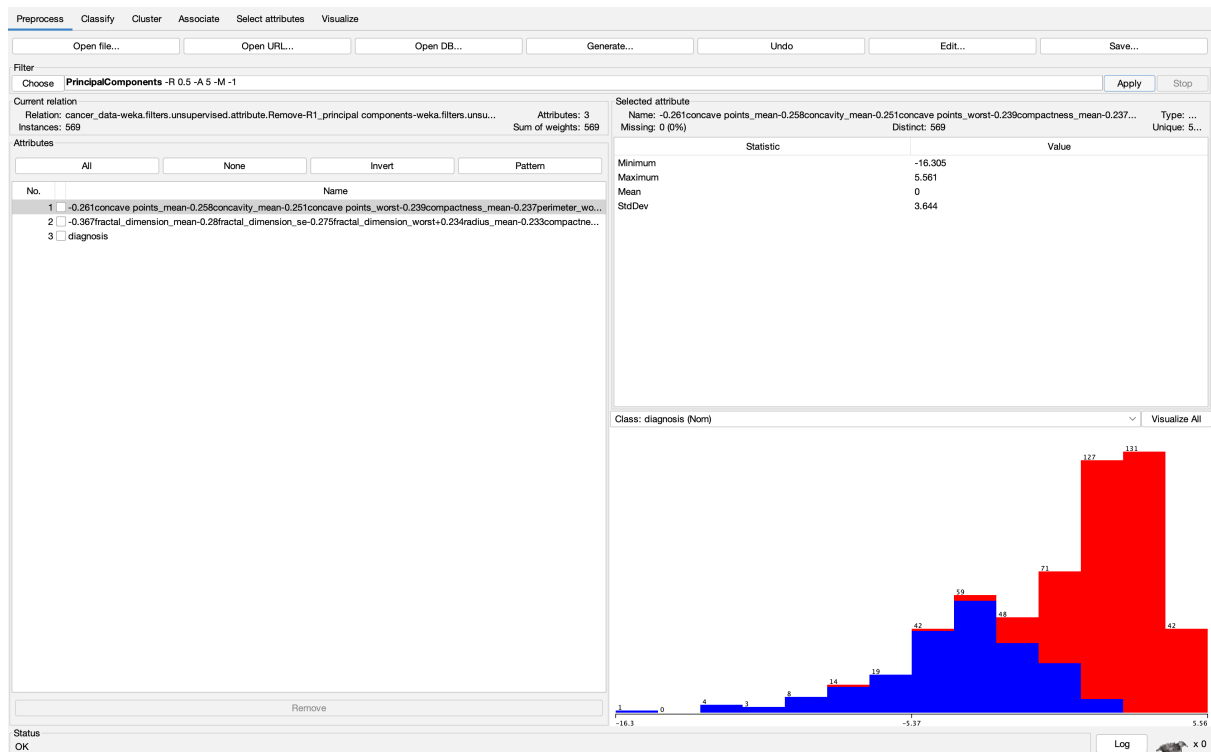
Cancer Dataset

PCA with Correlation Matrix

- variance is 0.95

- variance is 0.5

Classification Algorithms

- we remove the **last 2 attributes** as it had no effect on overall data

Decision Trees

```
|   |    -0.367fractal_dimension_mean-0.28fractal_dimension_se-0.275fractal_dimension_worst+0.234radius_mean-0.
|   |    -0.367fractal_dimension_mean-0.28fractal_dimension_se-0.275fractal_dimension_worst+0.234radius_mean-0.
|   -0.261concave points_mean-0.258concavity_mean-0.251concave points_worst-0.239compactness_mean-0.237perimet

Number of Leaves  :     6

Size of the tree :      11


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         536               94.2004 %
Incorrectly Classified Instances        33                5.7996 %
Kappa statistic                          0.8763
Mean absolute error                      0.0815
Root mean squared error                  0.2223
Relative absolute error                 17.4252 %
Root relative squared error             45.9737 %
Total Number of Instances              569

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.929    0.050    0.916      0.929    0.923      0.876    0.961     0.943     M
                 0.950    0.071    0.958      0.950    0.954      0.876    0.961     0.957     B
Weighted Avg.    0.942    0.063    0.942      0.942    0.942      0.876    0.961     0.952

=== Confusion Matrix ===

   a   b   <-- classified as
 197  15 |   a = M
  18 339 |   b = B
```

Random Forest

```
=== Summary ===

Correctly Classified Instances         534               93.8489 %
Incorrectly Classified Instances        35                6.1511 %
Kappa statistic                          0.8683
Mean absolute error                      0.0813
Root mean squared error                  0.2187
Relative absolute error                 17.388  %
Root relative squared error             45.2351 %
Total Number of Instances              569

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.915    0.048    0.919      0.915    0.917      0.868    0.977     0.966     M
                 0.952    0.085    0.950      0.952    0.951      0.868    0.977     0.979     B
Weighted Avg.    0.938    0.071    0.938      0.938    0.938      0.868    0.977     0.974

=== Confusion Matrix ===

   a   b   <-- classified as
 194  18 |   a = M
  17 340 |   b = B
```

Naive Bayes

```
-0.367fractal_dimension_mean-0.28fractal_dimension_se-0.275fractal_dimension_worst+0.234radius_mean-0.233compa
  mean
  std. dev.
  weight sum
  precision


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         519               91.2127 %
Incorrectly Classified Instances        50                8.7873 %
Kappa statistic                          0.8099
Mean absolute error                      0.1204
Root mean squared error                  0.2536
Relative absolute error                 25.7542 %
Root relative squared error             52.4474 %
Total Number of Instances              569

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.854    0.053    0.905      0.854   0.879      0.811    0.972     0.943     M
                 0.947    0.146    0.916      0.947   0.931      0.811    0.972     0.984     B
Weighted Avg.    0.912    0.112    0.912      0.912   0.912      0.811    0.972     0.969

=== Confusion Matrix ===

   a    b   <-- classified as
 181   31 |   a = M
  19  338 |   b = B
```