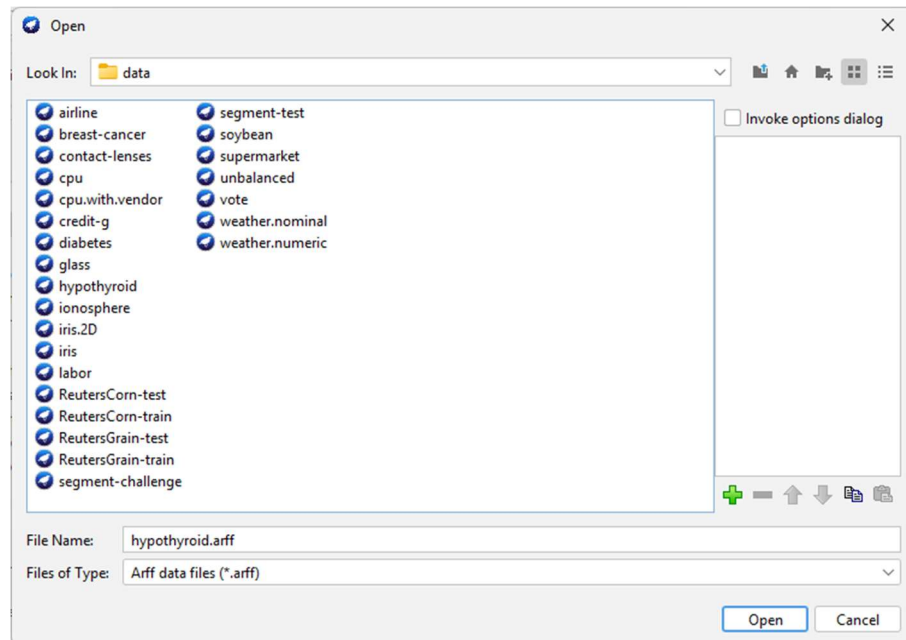# WEEK-9
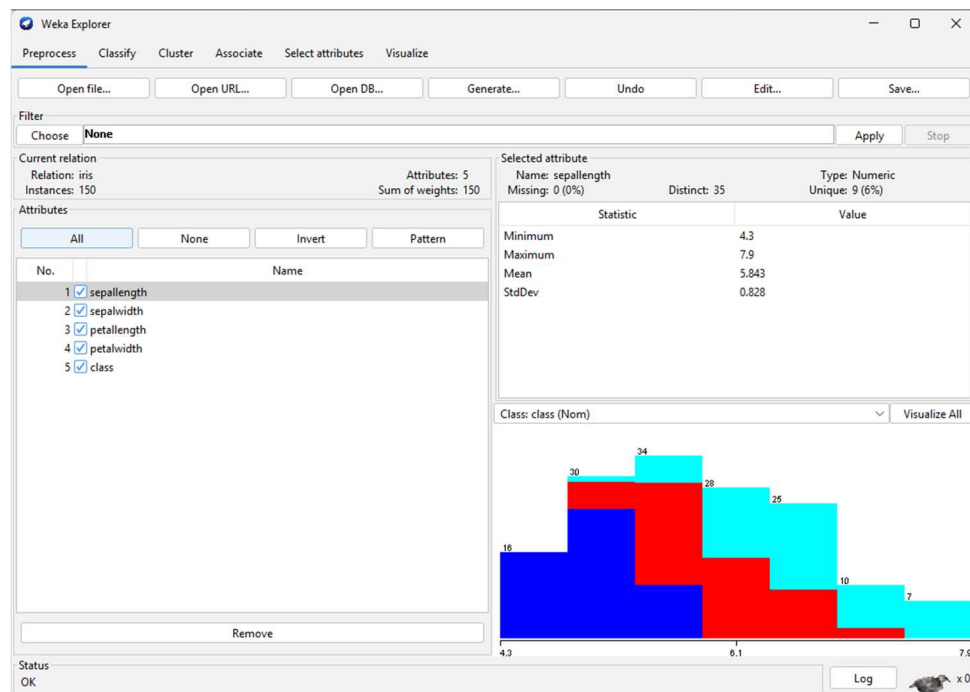
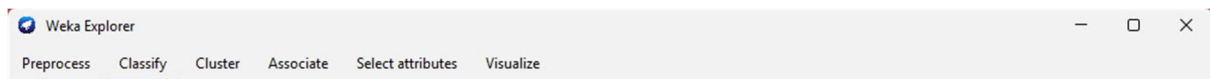## Demonstration of Classification algorithm using KNN approach.

Procedure for applying KNN approach for iris.arff

Step 1: Load the **iris.arff** data file
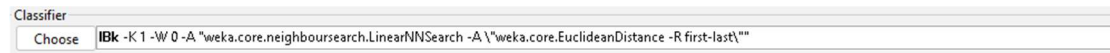

Step 2: Select all the attributes

Step 3: Go to classify tab

Weka Explorer — □ ×

Preprocess Classify Cluster Associate Select attributes Visualize

Then click on choose, under the classifier, and select **IBk**

Classifier

Choose | IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Click on the start. **(Output for Euclidean Distance)**

```
Correctly Classified Instances       143               95.3333 %
Incorrectly Classified Instances       7                4.6667 %
Kappa statistic                        0.93
Mean absolute error                    0.0399
Root mean squared error                0.1747
Relative absolute error                8.9763 %
Root relative squared error           37.0695 %
Total Number of Instances            150

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     Iris-setosa
                0.940    0.040    0.922      0.940   0.931      0.896  0.952     0.887     Iris-versicolor
                0.920    0.030    0.939      0.920   0.929      0.895  0.947     0.894     Iris-virginica
Weighted Avg.   0.953    0.023    0.953      0.953   0.953      0.930  0.966     0.927

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 47  3 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```
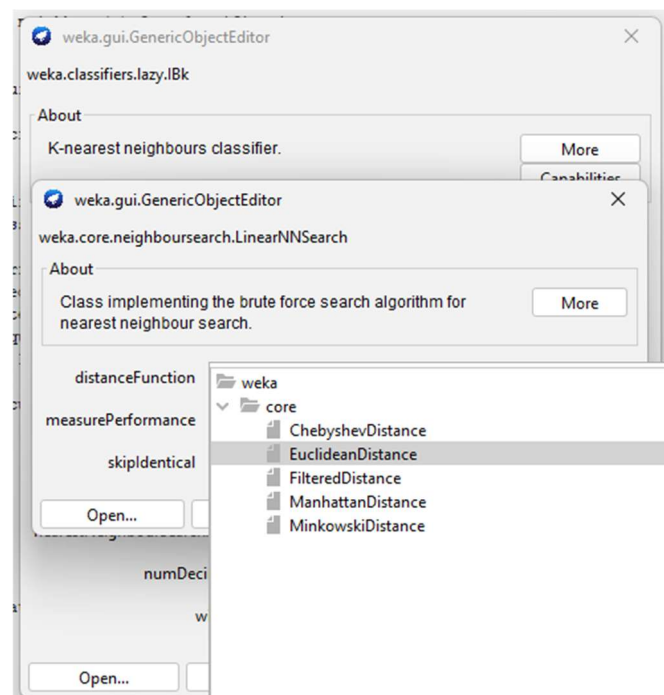
Step 3: Go to **IBk → LinearNNSearch → Choose all the left over 4 distances**

weka.gui.GenericObjectEditor   ×

weka.classifiers.lazy.IBk

About

K-nearest neighbours classifier.   More

                                                    Capabilities

weka.gui.GenericObjectEditor   ×

weka.core.neighboursearch.LinearNNSearch

About

Class implementing the brute force search algorithm for   More
nearest neighbour search.

distanceFunction   📁 weka
                        ∨ 📁 core
measurePerformance   🗎 ChebyshevDistance
                            🗎 EuclideanDistance
skipIdentical   🗎 FilteredDistance
                            🗎 ManhattanDistance
Open...   🗎 MinkowskiDistance

numDeci

w

Open...

Output for ChebyShev Distance

```
Correctly Classified Instances        144             96      %
Incorrectly Classified Instances        6              4      %
Kappa statistic                       0.94
Mean absolute error                   0.0337
Root mean squared error               0.1559
Relative absolute error               7.5898 %
Root relative squared error          33.0814 %
Total Number of Instances             150

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     Iris-setosa
              0.960    0.040    0.923      0.960   0.941      0.911  0.975     0.926     Iris-versicolor
              0.920    0.020    0.958      0.920   0.939      0.910  0.963     0.934     Iris-virginica
Weighted Avg. 0.960    0.020    0.960      0.960   0.960      0.940  0.979     0.953

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 48  2 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```

Output for Filtered Distance

```
Correctly Classified Instances        144             96      %
Incorrectly Classified Instances        6              4      %
Kappa statistic                       0.94
Mean absolute error                   0.0356
Root mean squared error               0.1618
Relative absolute error               8.0076 %
Root relative squared error          34.3296 %
Total Number of Instances             150

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     Iris-setosa
              0.960    0.040    0.923      0.960   0.941      0.911  0.961     0.900     Iris-versicolor
              0.920    0.020    0.958      0.920   0.939      0.910  0.952     0.911     Iris-virginica
Weighted Avg. 0.960    0.020    0.960      0.960   0.960      0.940  0.971     0.937

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 48  2 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```

Output for Manhattan Distance

```
Correctly Classified Instances        141             94      %
Incorrectly Classified Instances        9              6      %
Kappa statistic                       0.91
Mean absolute error                   0.0486
Root mean squared error               0.1981
Relative absolute error              10.9353 %
Root relative squared error          42.0173 %
Total Number of Instances             150

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              1.000    0.000    1.000      1.000   1.000      1.000  1.000     1.000     Iris-setosa
              0.900    0.040    0.918      0.900   0.909      0.864  0.933     0.861     Iris-versicolor
              0.920    0.050    0.902      0.920   0.911      0.866  0.938     0.863     Iris-virginica
Weighted Avg. 0.940    0.030    0.940      0.940   0.940      0.910  0.957     0.908

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 45  5 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```

Output for Minkowski Distance

```
Correctly Classified Instances        143              95.3333 %
Incorrectly Classified Instances        7               4.6667 %
Kappa statistic                       0.93
Mean absolute error                   0.0399
Root mean squared error               0.1747
Relative absolute error               8.9763 %
Root relative squared error          37.0695 %
Total Number of Instances             150

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
              1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     Iris-setosa
              0.940    0.040    0.922      0.940   0.931      0.896   0.952     0.887     Iris-versicolor
              0.920    0.030    0.939      0.920   0.929      0.895   0.947     0.894     Iris-virginica
Weighted Avg. 0.953    0.023    0.953      0.953   0.953      0.930   0.966     0.927

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 47  3 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```
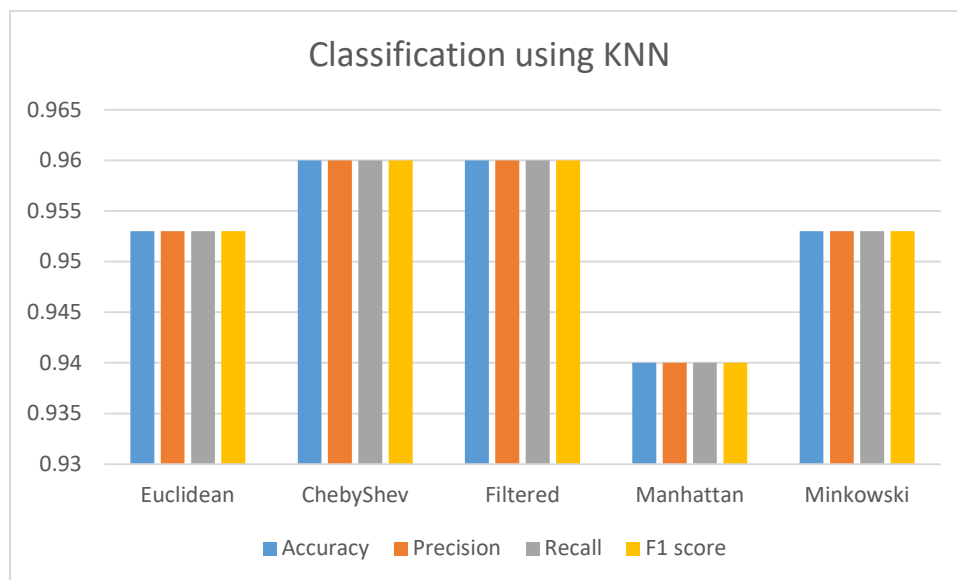
Visualization

Classification using KNN

Algo (IBk-{distance})	Accuracy	Precision	Recall	F1 score
Euclidean	0.953	0.953	0.953	0.953
ChebyShev	0.960	0.960	0.960	0.960
Filtered	0.960	0.960	0.960	0.960
Manhattan	0.940	0.940	0.940	0.940
Minkowski	0.953	0.953	0.953	0.953