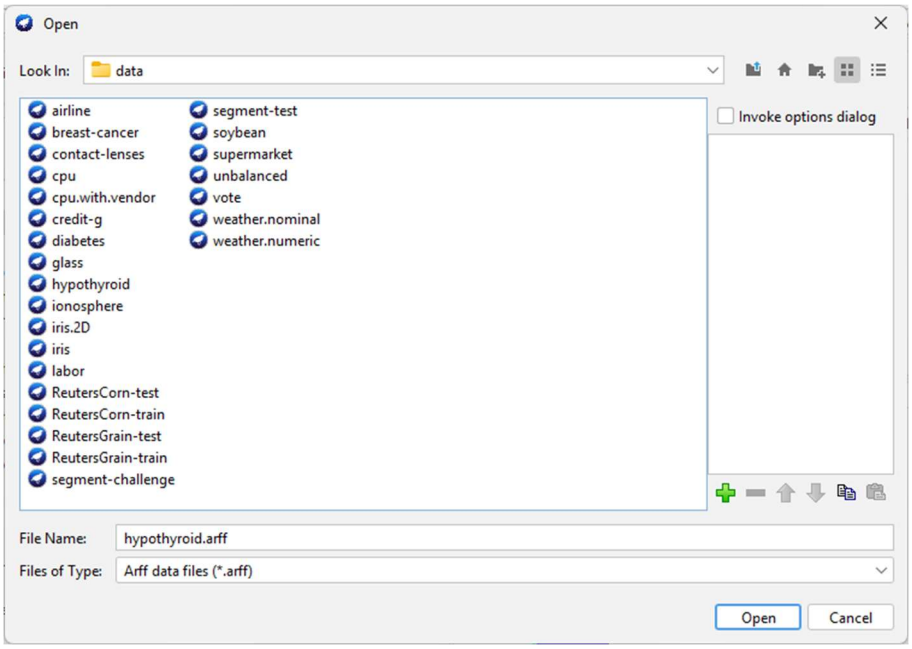
WEEK-9

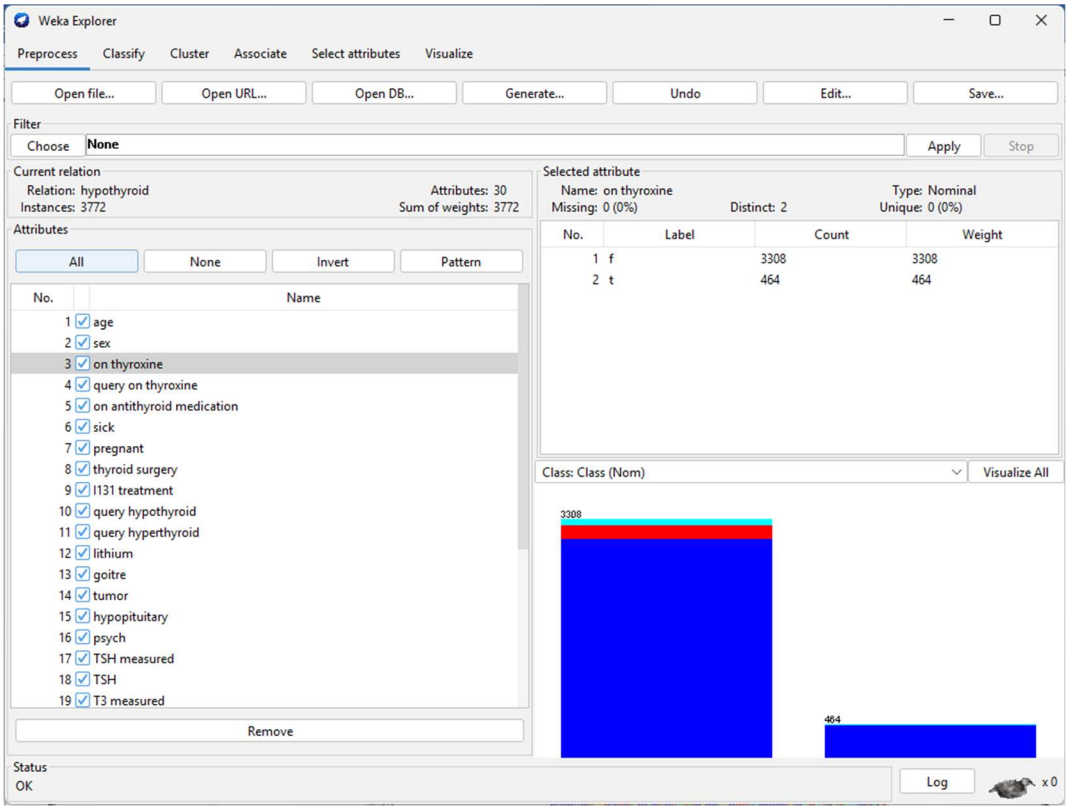
Demonstration of Classification algorithm using KNN approach.

Procedure for applying KNN approach for hypothyroid.arff

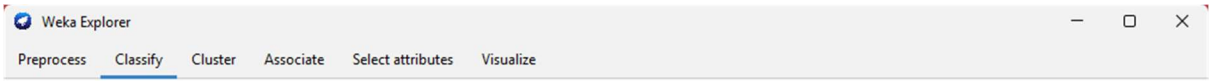
Step 1: Load the **hypothyroid.arff** data file



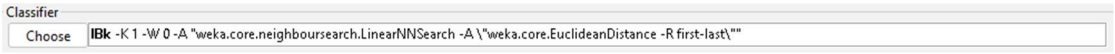
Step 2: Select all the attributes



Step 3: Go to classify tab



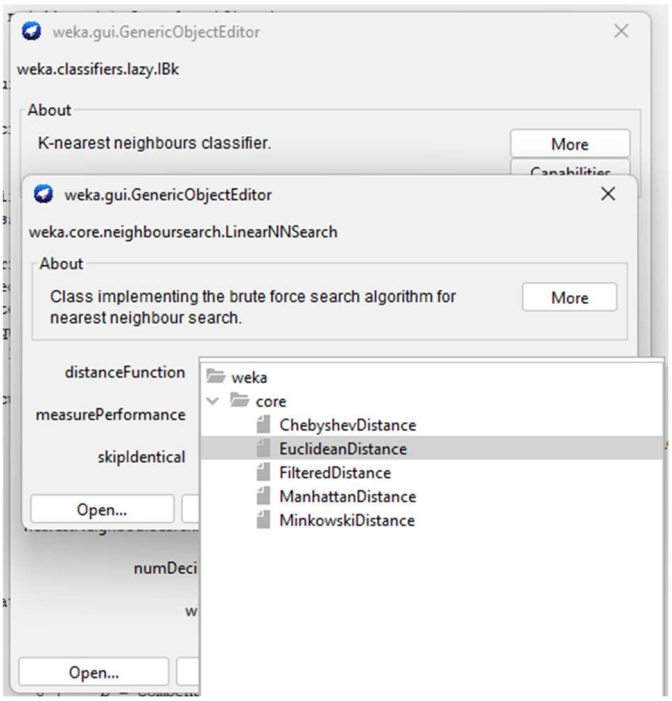
Then click on choose, under the classifier, and select **IBk**



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

```
Correctly Classified Instances        3452               91.5164 %
Incorrectly Classified Instances       320                8.4836 %
Kappa statistic                          0.3605
Mean absolute error                      0.0428
Root mean squared error                  0.2058
Relative absolute error                 58.7244 %
Root relative squared error            108.0856 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.964    0.619    0.949      0.964    0.956      0.379    0.682     0.950     negative
               0.191    0.035    0.227      0.191    0.207      0.169    0.587     0.090     compensated_hypothyroid
               0.642    0.004    0.813      0.642    0.718      0.716    0.829     0.534     primary_hypothyroid
               0.000    0.000    ?          0.000    ?          ?        0.902     0.003     secondary_hypothyroid
Weighted Avg.  0.915    0.573    ?          0.915    ?          ?        0.681     0.895

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3354  118    9    0 |    a = negative
  152   37    5    0 |    b = compensated_hypothyroid
   26    8   61    0 |    c = primary_hypothyroid
    2    0    0    0 |    d = secondary_hypothyroid
```

Step 3: Go to **IBk → LinearNNSearch → Choose all the left over 4 distances**



Output for ChebyShev Distance

```
Correctly Classified Instances        3481               92.2853 %
Incorrectly Classified Instances       291                7.7147 %
Kappa statistic                          0
Mean absolute error                      0.0725
Root mean squared error                  0.1904
Relative absolute error                 99.5118 %
Root relative squared error             99.9997 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.923      1.000    0.960      ?        0.498     0.923     negative
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.051     compensated_hypothyroid
                 0.000    0.000    ?          0.000    ?          ?        0.486     0.025     primary_hypothyroid
                 0.000    0.000    ?          0.000    ?          ?        0.100     0.001     secondary_hypothyroid
Weighted Avg.    0.923    0.923    ?          0.923    ?          ?        0.498     0.855

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3481    0    0    0 |    a = negative
  194    0    0    0 |    b = compensated_hypothyroid
   95    0    0    0 |    c = primary_hypothyroid
    2    0    0    0 |    d = secondary_hypothyroid
```

Output for Filtered Distance

```
Correctly Classified Instances        3492               92.5769 %
Incorrectly Classified Instances       280                7.4231 %
Kappa statistic                          0.4123
Mean absolute error                      0.0375
Root mean squared error                  0.1925
Relative absolute error                 51.4598 %
Root relative squared error            101.1049 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.973    0.598    0.951      0.973    0.962      0.437    0.700     0.953     negative
                 0.206    0.023    0.325      0.206    0.252      0.228    0.601     0.114     compensated_hypothyroid
                 0.674    0.005    0.780      0.674    0.723      0.719    0.858     0.559     primary_hypothyroid
                 0.000    0.001    0.000      0.000    0.000      -0.001   0.901     0.003     secondary_hypothyroid
Weighted Avg.    0.926    0.553    0.914      0.926    0.919      0.433    0.699     0.899

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3388   76   13    4 |    a = negative
  150   40    4    0 |    b = compensated_hypothyroid
   23    7   64    1 |    c = primary_hypothyroid
    1    0    1    0 |    d = secondary_hypothyroid
```

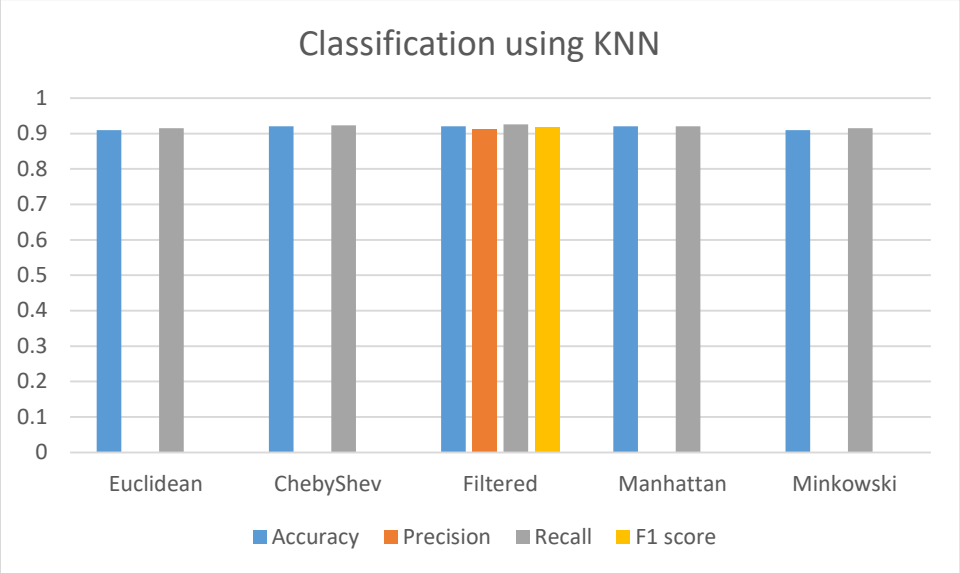
Output for Manhattan Distance

```
Correctly Classified Instances        3472               92.0467 %
Incorrectly Classified Instances       300                7.9533 %
Kappa statistic                          0.3821
Mean absolute error                      0.0402
Root mean squared error                  0.1993
Relative absolute error                 55.0915 %
Root relative squared error            104.6535 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.969    0.615    0.950      0.969    0.959      0.402    0.688     0.951     negative
                 0.201    0.030    0.265      0.201    0.229      0.195    0.595     0.099     compensated_hypothyroid
                 0.642    0.004    0.824      0.642    0.722      0.721    0.835     0.545     primary_hypothyroid
                 0.000    0.000    ?          0.000    ?          ?        0.902     0.003     secondary_hypothyroid
Weighted Avg.    0.920    0.569    ?          0.920    ?          ?        0.687     0.896

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3372  100    9    0 |    a = negative
  151   39    4    0 |    b = compensated_hypothyroid
   26    8   61    0 |    c = primary_hypothyroid
    2    0    0    0 |    d = secondary_hypothyroid
```

Output for Minkowski Distance

```
Correctly Classified Instances        3452               91.5164 %
Incorrectly Classified Instances       320                8.4836 %
Kappa statistic                          0.3605
Mean absolute error                      0.0428
Root mean squared error                  0.2058
Relative absolute error                 58.7244 %
Root relative squared error            108.0856 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.964    0.619    0.949      0.964    0.956      0.379    0.682     0.950     negative
                 0.191    0.035    0.227      0.191    0.207      0.169    0.587     0.090     compensated_hypothyroid
                 0.642    0.004    0.813      0.642    0.718      0.716    0.829     0.534     primary_hypothyroid
                 0.000    0.000    ?          0.000    ?          ?        0.902     0.003     secondary_hypothyroid
Weighted Avg.    0.915    0.573    ?          0.915    ?          ?        0.681     0.895

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3354  118    9    0 |    a = negative
  152   37    5    0 |    b = compensated_hypothyroid
   26    8   61    0 |    c = primary_hypothyroid
    2    0    0    0 |    d = secondary_hypothyroid
```

Visualization



Algo (IBk-{distance})	Accuracy	Precision	Recall	F1 score
Euclidean	0.91	?	0.915	?
ChebyShev	0.92	?	0.923	?
Filtered	0.92	0.914	0.926	0.919
Manhattan	0.92	?	0.920	?
Minkowski	0.91	?	0.915	?