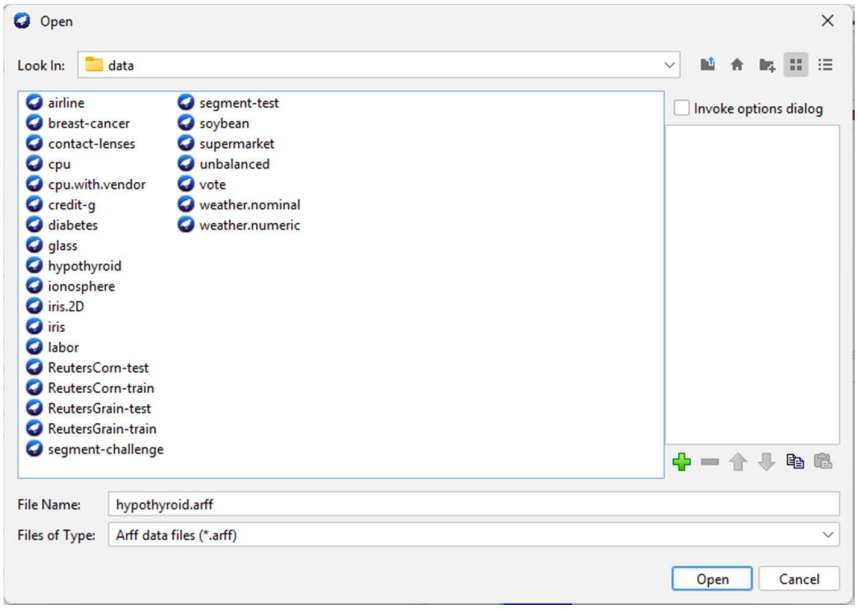
WEEK-10

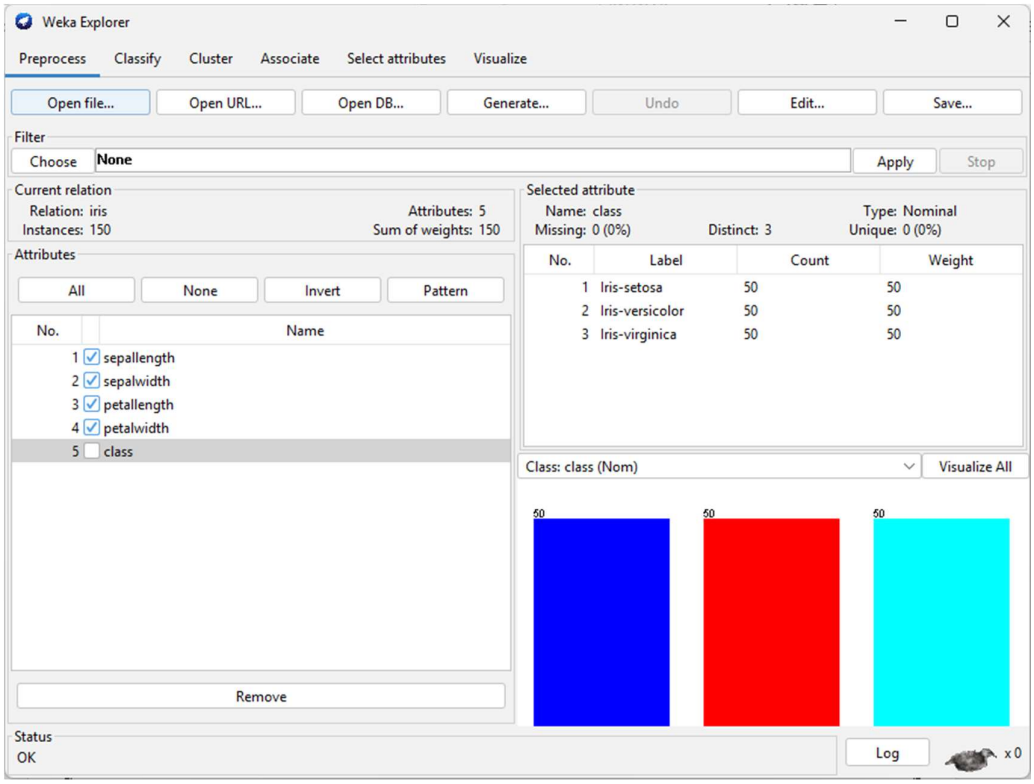
Demonstration of Clustering algorithm using K-means clustering algorithm.

Procedure for applying K-means for iris.arff

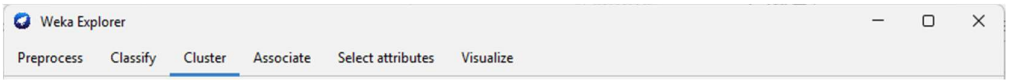
Step 1: Load the **iris.arff** data file



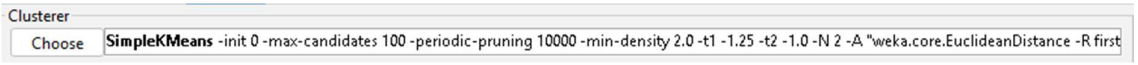
Step 2: Select all the attributes except ***class***



Step 3: Go to **Cluster** tab



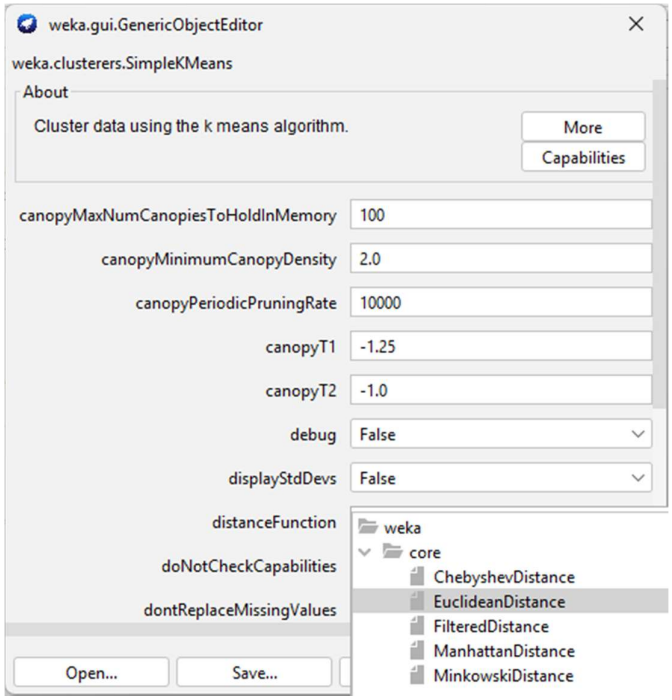
Then click on choose, under the classifier, and select **SimpleKMeans**



Click on the start. **(Output for SimpleKMeans – 3 Clusters - Euclidean)**

```
kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 7.817456892309574

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0              1              2
                          (150.0)          (50.0)         (50.0)         (50.0)
================================================================================
sepallength                5.8433           5.936          5.006          6.588
sepalwidth                  3.054            2.77          3.418          2.974
petallength                3.7587            4.26          1.464          5.552
petalwidth                 1.1987           1.326          0.244          2.026
class                 Iris-setosa Iris-versicolor     Iris-setosa  Iris-virginica




Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       50 ( 33%)
```

Step 3: Go to **SimpleKMeans → Choose the ManhattanDistance**



Output for Manhattan Distance

```
kMeans
======

Number of iterations: 3
Sum of within cluster distances: 49.87499999999999

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0             1             2
                          (150.0)          (50.0)        (50.0)        (50.0)
================================================================================
sepallength                   5.8             5.9             5           6.5
sepalwidth                      3             2.8           3.4             3
petallength                  4.35            4.35           1.5          5.55
petalwidth                    1.3             1.3           0.2             2
class                 Iris-setosa Iris-versicolor    Iris-setosa  Iris-virginica



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       50 ( 33%)
```

Output for SimpleKMeans – 4 Clusters - Euclidean Distance

```
kMeans
======

Number of iterations: 4
Within cluster sum of squared errors: 6.613823274690356

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica
Cluster 3: 5.5,4.2,1.4,0.2,Iris-setosa

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0              1              2              3
                          (150.0)          (24.0)         (26.0)         (50.0)         (50.0)
=============================================================================================
sepallength                5.8433          6.3292         5.5731          6.588          5.006
sepalwidth                  3.054          2.9792         2.5769          2.974          3.418
petallength                3.7587             4.6         3.9462          5.552          1.464
petalwidth                 1.1987          1.4625            1.2          2.026          0.244
class                 Iris-setosa Iris-versicolor Iris-versicolor  Iris-virginica     Iris-setosa



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       24 ( 16%)
1       26 ( 17%)
2       50 ( 33%)
3       50 ( 33%)
```

Output for SimpleKMeans – 4 Clusters - Manhattan Distance

```
kMeans
======

Number of iterations: 4
Sum of within cluster distances: 44.35852165725046

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica
Cluster 3: 5.5,4.2,1.4,0.2,Iris-setosa

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0              1              2              3
                          (150.0)          (26.0)         (24.0)         (50.0)         (50.0)
=============================================================================================
sepallength                  5.8             6.3            5.6            6.5              5
sepalwidth                     3               3            2.6              3            3.4
petallength                 4.35             4.6              4           5.55            1.5
petalwidth                   1.3             1.5            1.2              2            0.2
class                 Iris-setosa Iris-versicolor Iris-versicolor  Iris-virginica     Iris-setosa



Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       26 ( 17%)
1       24 ( 16%)
2       50 ( 33%)
3       50 ( 33%)
```

Step 4: Go to **HierarchialCluster → Choose the EuclideanDistance**

Output for HierarchialCluster – 3 Clusters - Euclidean Distance

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

Cluster 0
((((((((((((((((((((0.0:0.03254,0.0:0.03254):0.00913,(0.0:0.03254,0.0:0.03254):0.00913):0.00332,((0.0:0.02778,0.0:0.02778):0.00476,0.0:0.03254):0.01244

Cluster 1
((((((((((((((((1.0:0.07344,(((1.0:0.06508,1.0:0.06508):0.00066,(1.0:0.05008,1.0:0.05008):0.01566):0.00224,1.0:0.06798):0.00546):0.00188,(1.0:0.07137,(

Cluster 2
((((((((((((2.0:0.08983,(2.0:0.06047,2.0:0.06047):0.02935):0.01175,2.0:0.10158):0.01245,(2.0:0.10743,((((2.0:0.07148,(2.0:0.05008,2.0:0.05008):0.02141

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       50 ( 33%)
```

Output for HierarchialCluster – 3 Clusters - Manhattan Distance

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

Cluster 0
((((((((((((((((((((((((0.0:0.04167,0.0:0.04167):0.00306,0.0:0.04473):0,0.0:0.04473):0.01389,((0.0:0.02778,0.0:0.02778):0.01695,0.0:0.04473):0.01389):

Cluster 1
((((((((((((1.0:0.14501,(((((((((1.0:0.08945,1.0:0.08945):0.00306,(1.0:0.06944,1.0:0.06944):0.02307):0.02472,(1.0:0.1064,(1.0:0.05556,1.0:0.05556):0.0

Cluster 2
((((((((((((((((2.0:0.15113,(2.0:0.10334,2.0:0.10334):0.04779):0.00471,2.0:0.15584):0.00612,(2.0:0.15113,((((2.0:0.08639,2.0:0.08639):0.04779,(2.0:0.11

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       50 ( 33%)
```

Output for HierarchialCluster – 4 Clusters - Euclidean Distance

```
Cluster 0
((((((((((((((((((((((0.0:0.03254,0.0:0.03254):0.00913,(0.0:0.03254,0.0:0.03254):0.00913):0.00332,((0.0:0.02778,0.0:0.02778):0.00476,0.0:0.03254):0.012

Cluster 1
((((((((((((((((((1.0:0.07344,(((1.0:0.06508,1.0:0.06508):0.00066,(1.0:0.05008,1.0:0.05008):0.01566):0.00224,1.0:0.06798):0.00546):0.00188,(1.0:0.07137

Cluster 2
((((((((((((2.0:0.08983,(2.0:0.06047,2.0:0.06047):0.02935):0.01175,2.0:0.10158):0.01245,(2.0:0.10743,((((2.0:0.07148,(2.0:0.05008,2.0:0.05008):0.0214

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       49 ( 33%)
3        1 (  1%)
```

Output for HierarchialCluster – 4 Clusters - Manhattan Distance

```
Cluster 0
((((((((((((((((((((((((0.0:0.04167,0.0:0.04167):0.00306,0.0:0.04473):0,0.0:0.04473):0.01389,((0.0:0.02778,0.0:0.02778):0.01695,0.0:0.04473):0.01389)

Cluster 1
((((((((((((1.0:0.14501,(((((((((1.0:0.08945,1.0:0.08945):0.00306,(1.0:0.06944,1.0:0.06944):0.02307):0.02472,(1.0:0.1064,(1.0:0.05556,1.0:0.05556):0.

Cluster 2
((((((((((((((((2.0:0.15113,(2.0:0.10334,2.0:0.10334):0.04779):0.00471,2.0:0.15584):0.00612,(2.0:0.15113,((((2.0:0.08639,2.0:0.08639):0.04779,(2.0:0.11

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       49 ( 33%)
3        1 (  1%)
```