

Practical 8

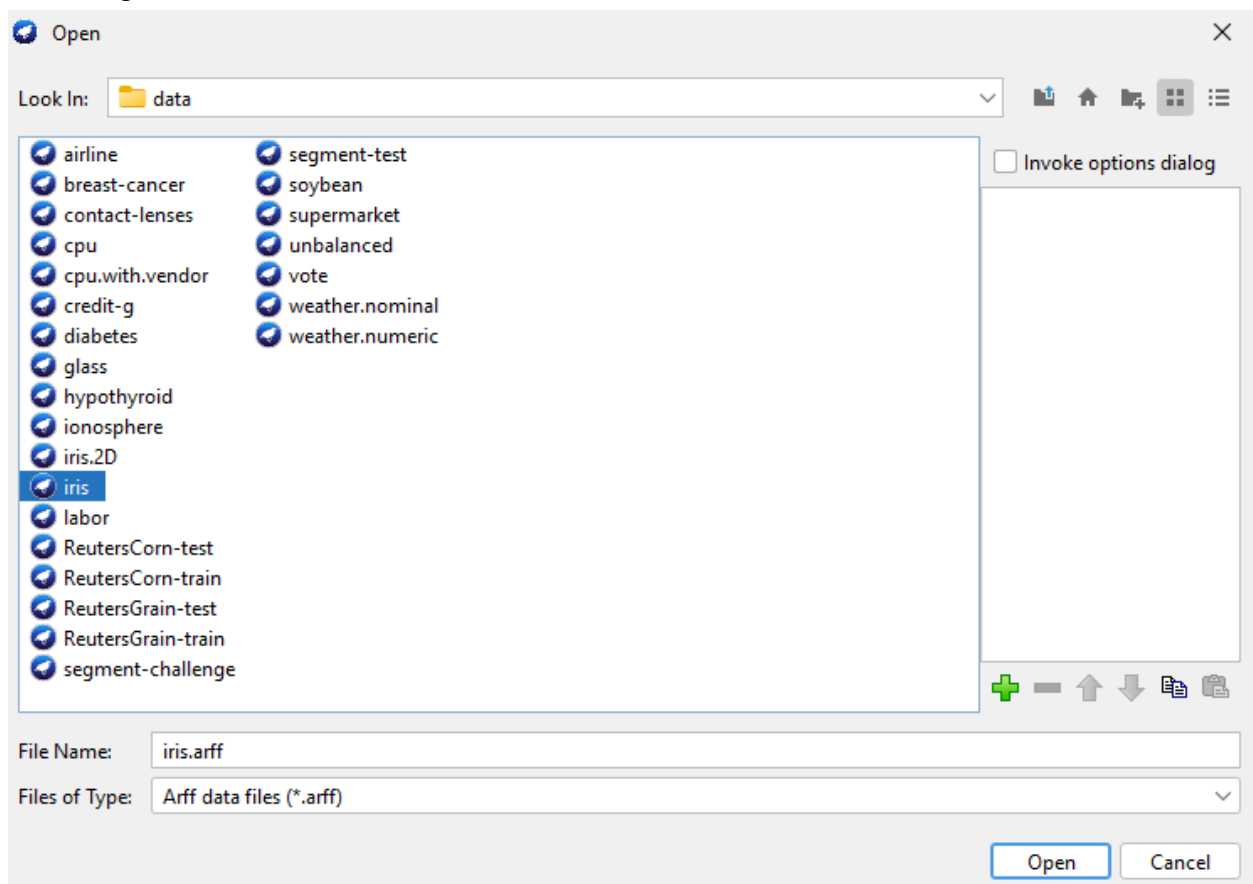
Name: Heet Dhanuka

Roll No.: 34

Batch: B2

Aim: Perform data mining using K-Means Clustering Technique in WEKA DATA MINING TOOL.

Selecting the Dataset:



Preprocessing the Dataset:

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

ChooseNone

Apply

Stop

Current relation

Relation: iris

Instances: 150

Attributes: 5

Sum of weights: 150

Attributes

AllNoneInvertPattern

No.	Name
1	<input checked="" type="checkbox"/> sepallength
2	<input checked="" type="checkbox"/> sepalwidth
3	<input checked="" type="checkbox"/> petallength
4	<input checked="" type="checkbox"/> petalwidth
5	<input checked="" type="checkbox"/> class

Remove

Selected attribute

Name: petallength

Missing: 0 (0%)

Distinct: 43

Type: Numeric

Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Class: class (Nom)

Visualize All

Class	Count
Setosa	50
Versicolour	34
Virginica	47

Applying K means clustering:

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set Set...

☐ Percentage split % 66

☐ Classes to clusters evaluation
(Nom) class v

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

11:06:51 - SimpleKMeans

Clusterer output

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:

Attribute	Full Data (150.0)	Cluster# 0 (24.0)	1 (26.0)	2 (50.0)	3 (50.0)
sepalength	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-virginica	Iris-setosa

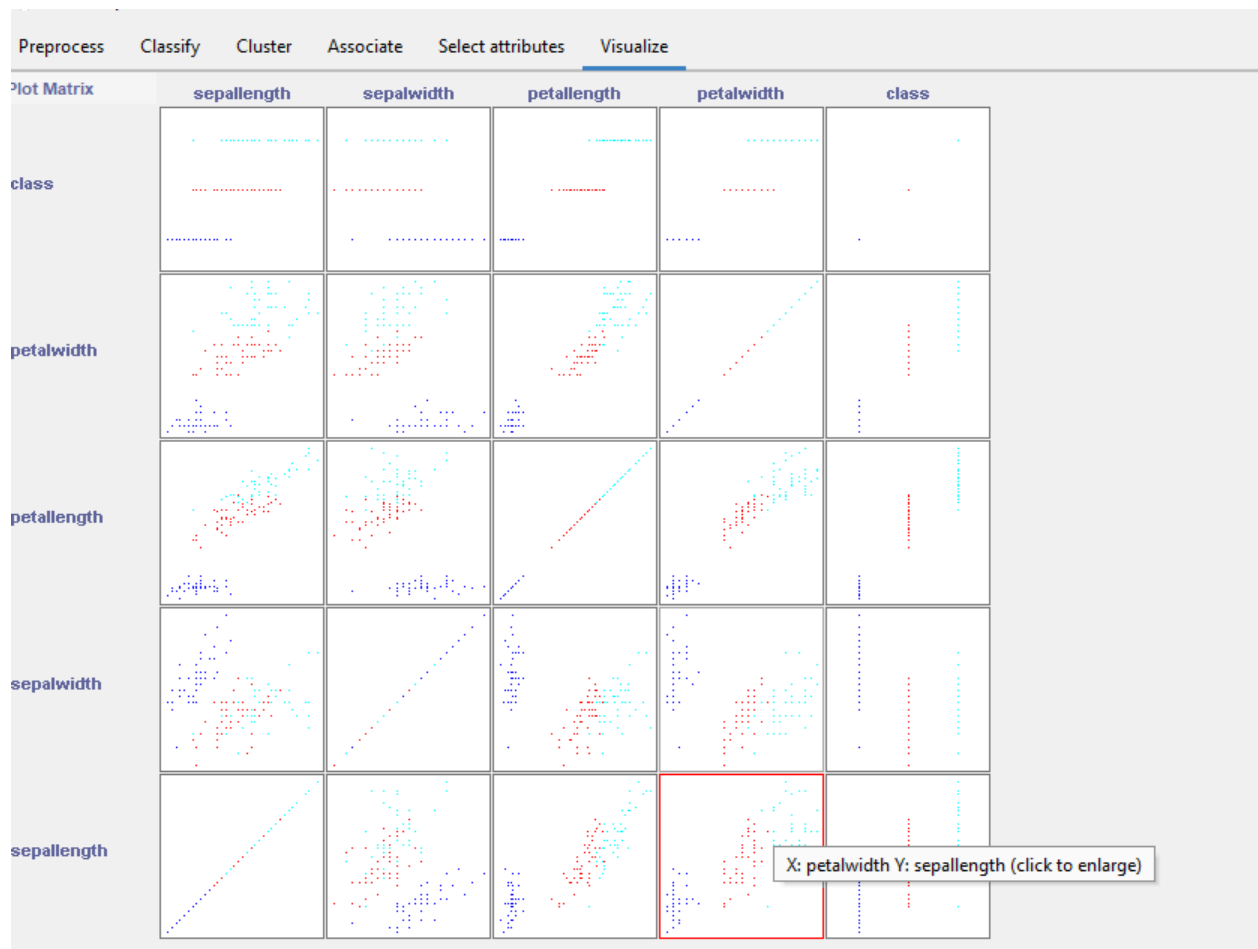
Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

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

Visualizing K means clustering:



Output:

