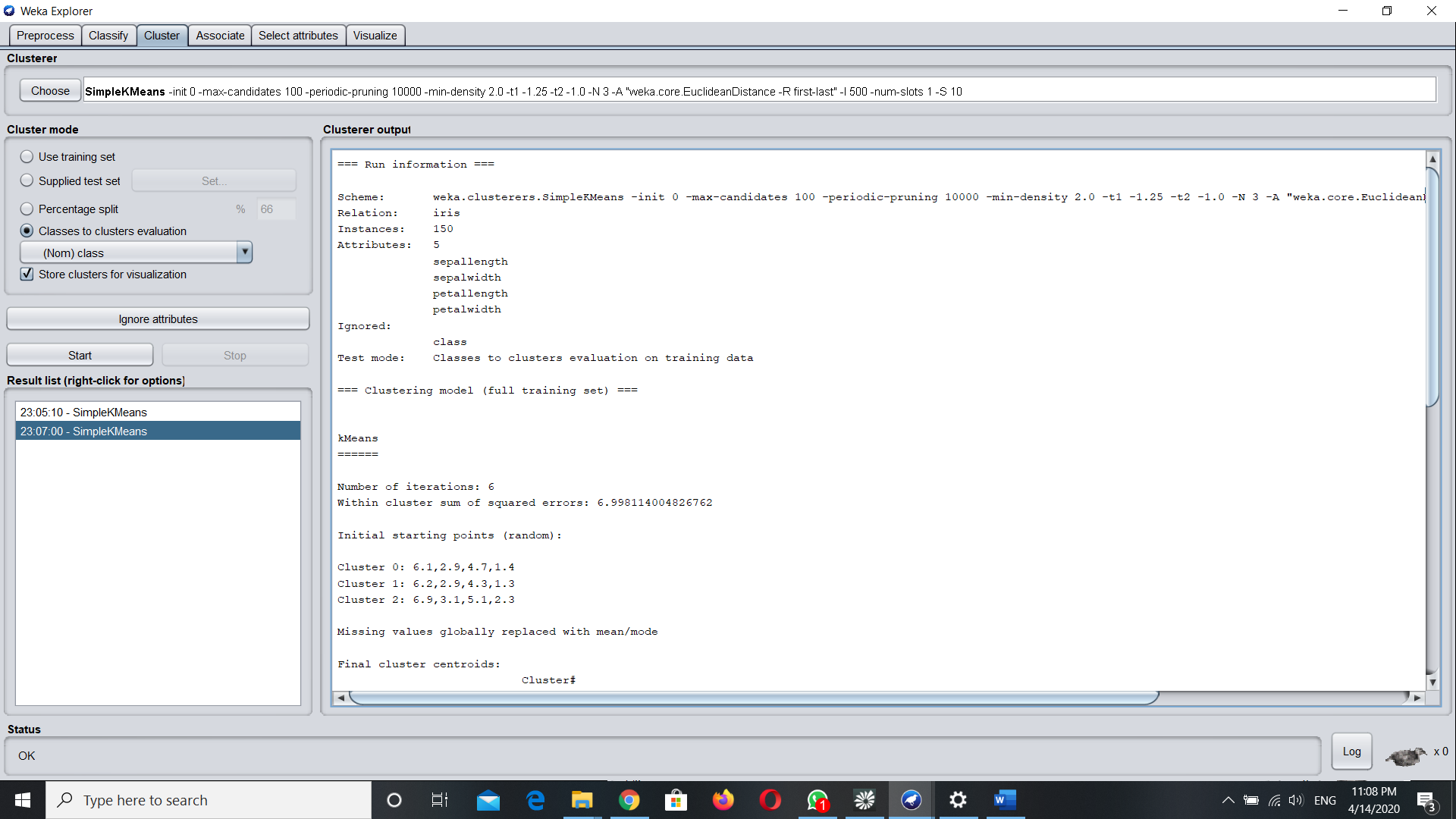
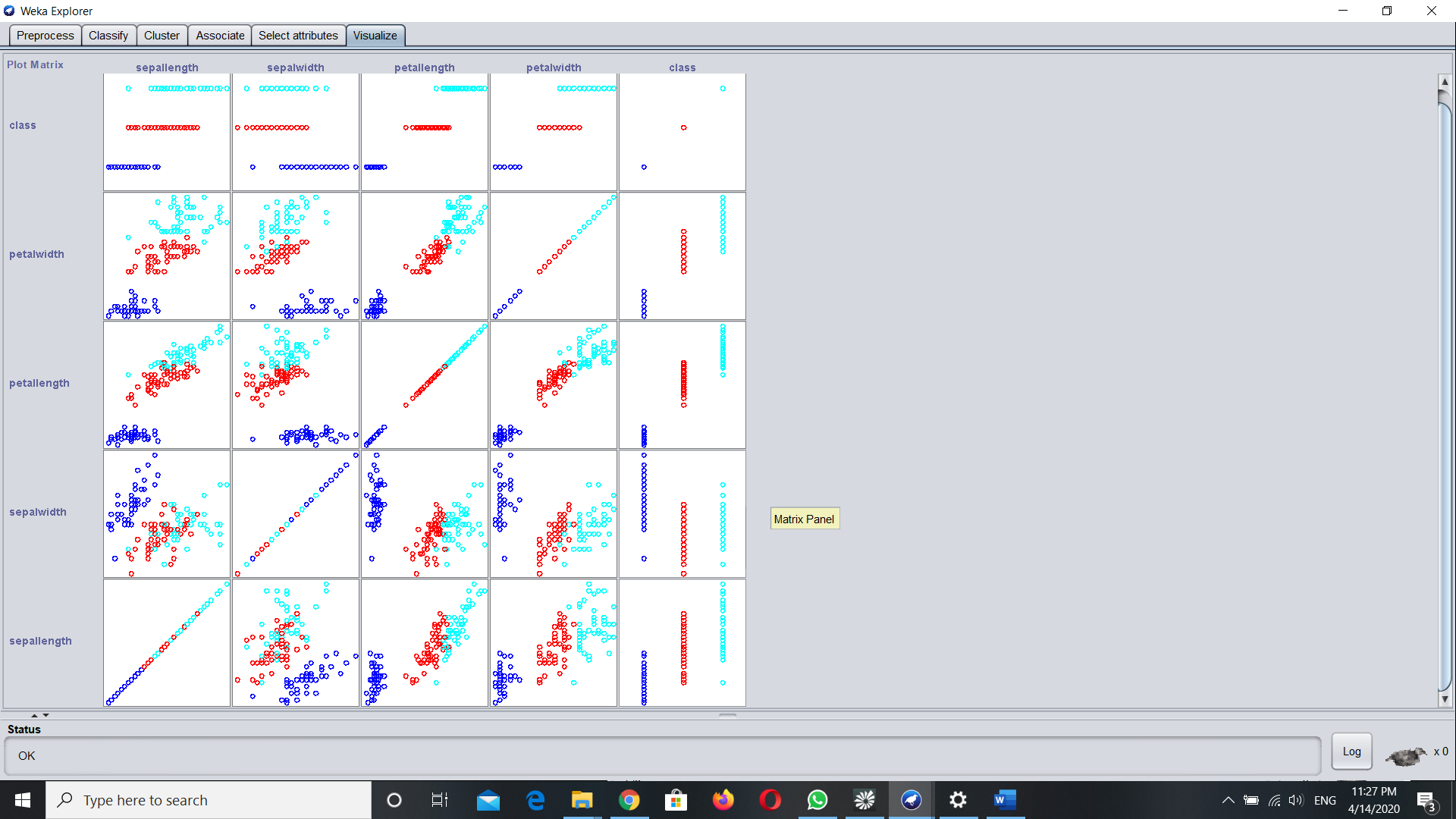
Cloud Computing Group Activity- Group 9

WEKA- Clustering by Kmeans on Iris dataset





=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: iris

Instances: 150

Attributes: 5

sepallength

sepalwidth

petallength

petalwidth

Ignored:

class

Test mode: Classes to clusters evaluation on training data

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

kMeans

======

Number of iterations: 6

Within cluster sum of squared errors: 6.998114004826762

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4

Cluster 1: 6.2,2.9,4.3,1.3

Cluster 2: 6.9,3.1,5.1,2.3

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1 2

(150.0) (61.0) (50.0) (39.0)

=========================================================

sepallength 5.8433 5.8885 5.006 6.8462

sepalwidth 3.054 2.7377 3.418 3.0821

petallength 3.7587 4.3967 1.464 5.7026

petalwidth 1.1987 1.418 0.244 2.0795

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

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

Clustered Instances

0 61 ( 41%)

1 50 ( 33%)

2 39 ( 26%)

Class attribute: class

Classes to Clusters:

0 1 2 <-- assigned to cluster

0 50 0 | Iris-setosa

47 0 3 | Iris-versicolor

14 0 36 | Iris-virginica

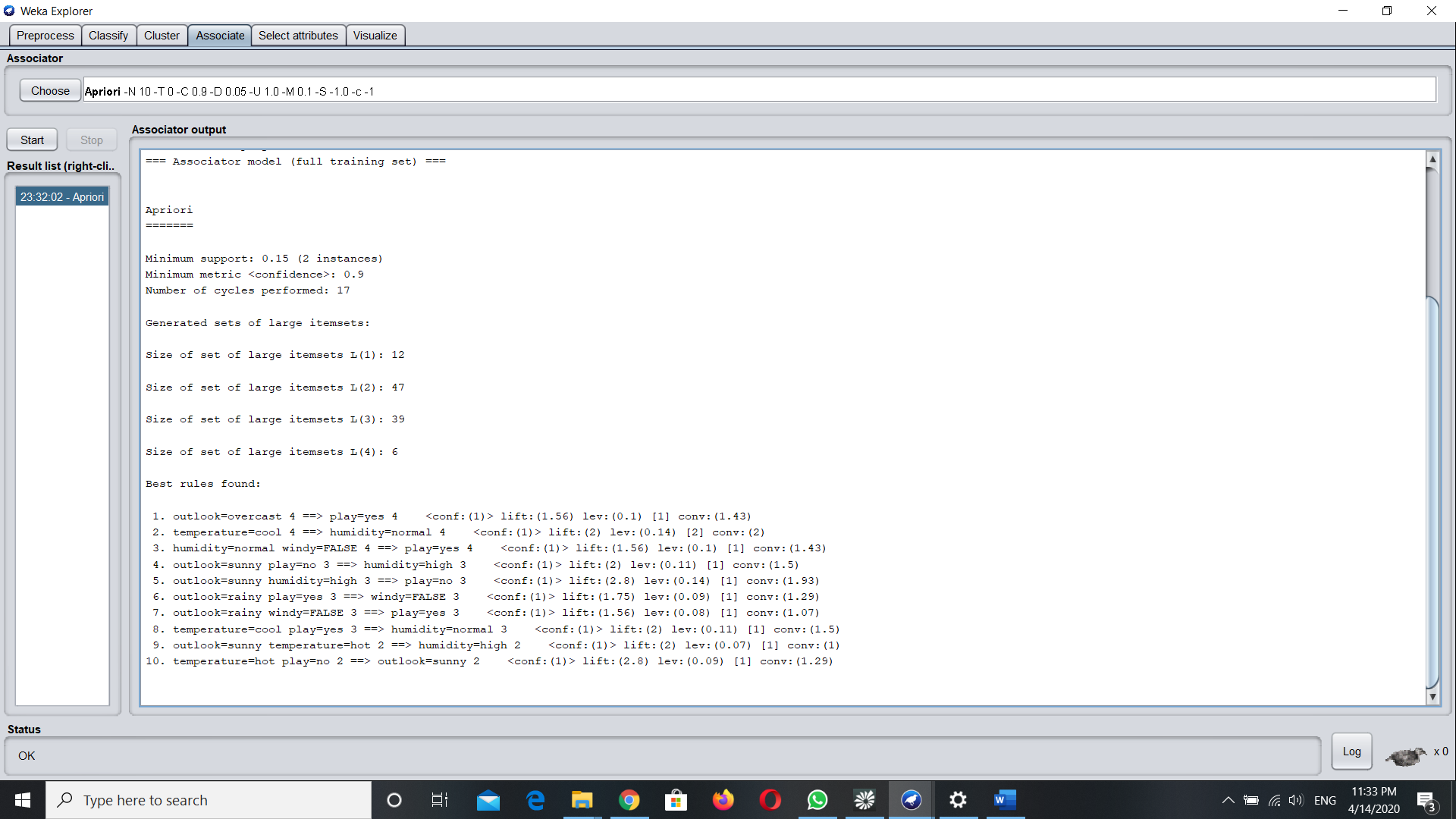
Cluster 0 <-- Iris-versicolor

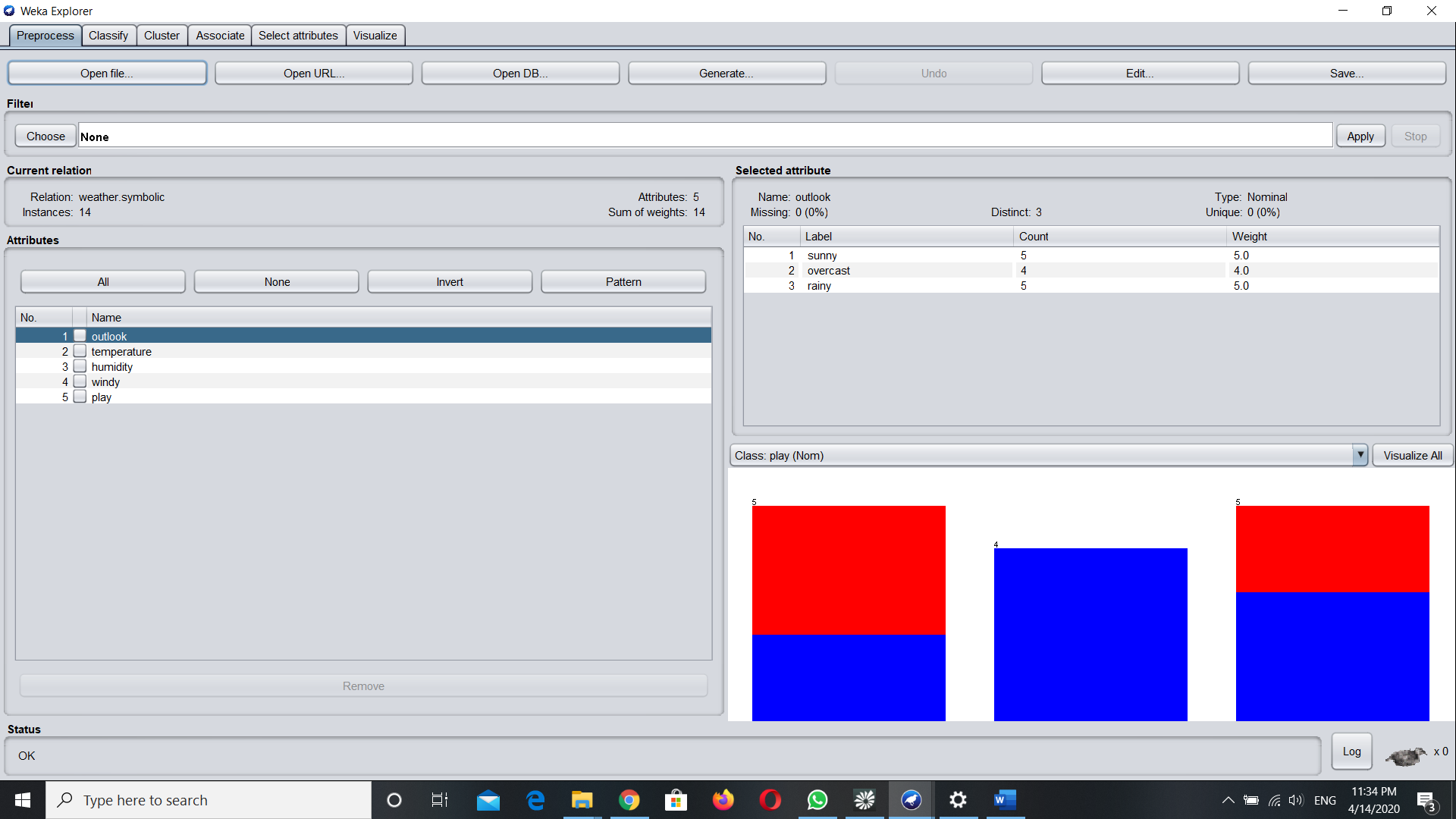
Cluster 1 <-- Iris-setosa

Cluster 2 <-- Iris-virginica

Incorrectly clustered instances : 17.0 11.3333 %

WEKA- Association Rule Mining on Weather Data





=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.15 (2 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)

2. temperature=cool 4 ==> humidity=normal 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)

3. humidity=normal windy=FALSE 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)

4. outlook=sunny play=no 3 ==> humidity=high 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)

5. outlook=sunny humidity=high 3 ==> play=no 3 <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)

6. outlook=rainy play=yes 3 ==> windy=FALSE 3 <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)

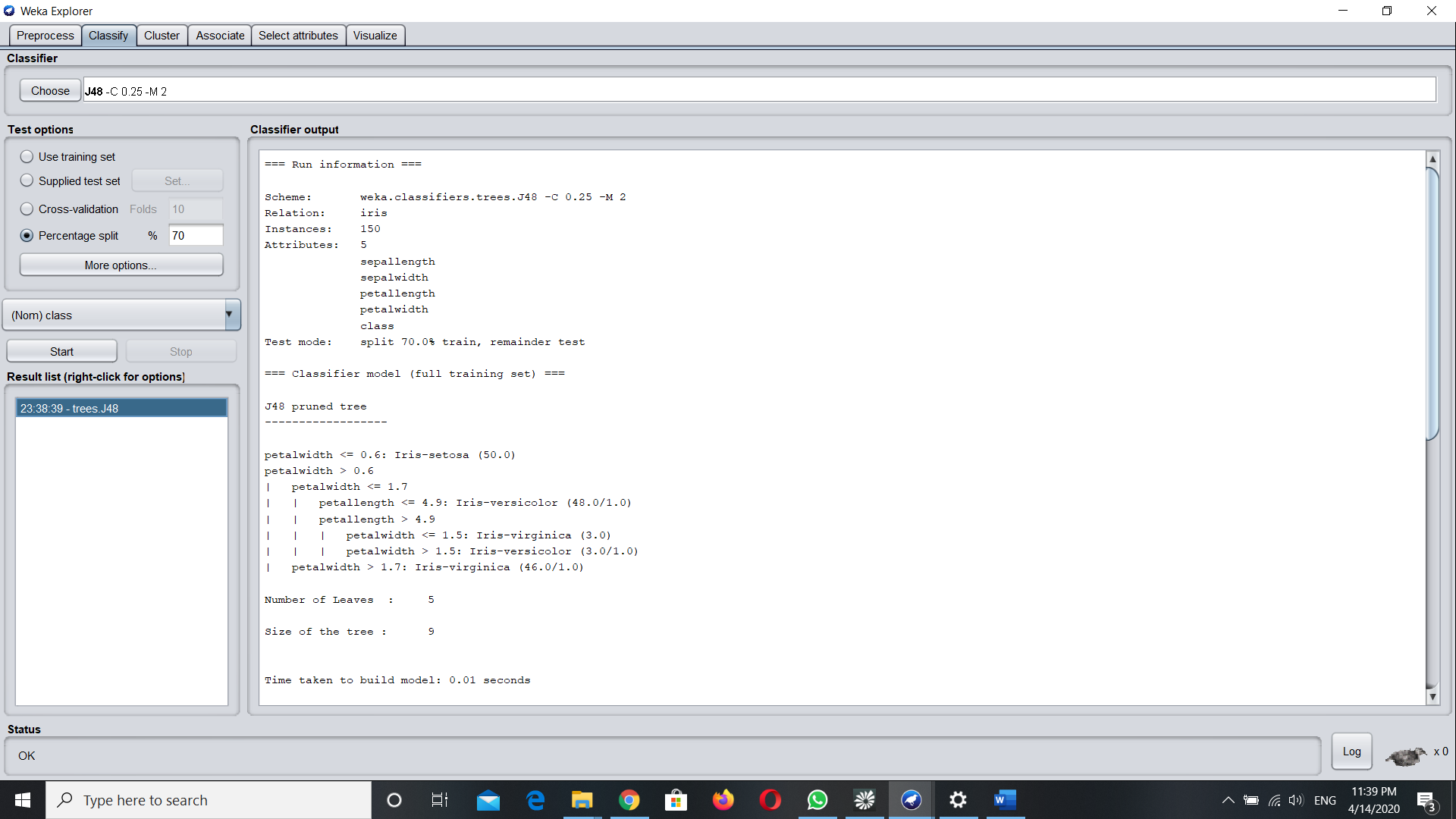
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)

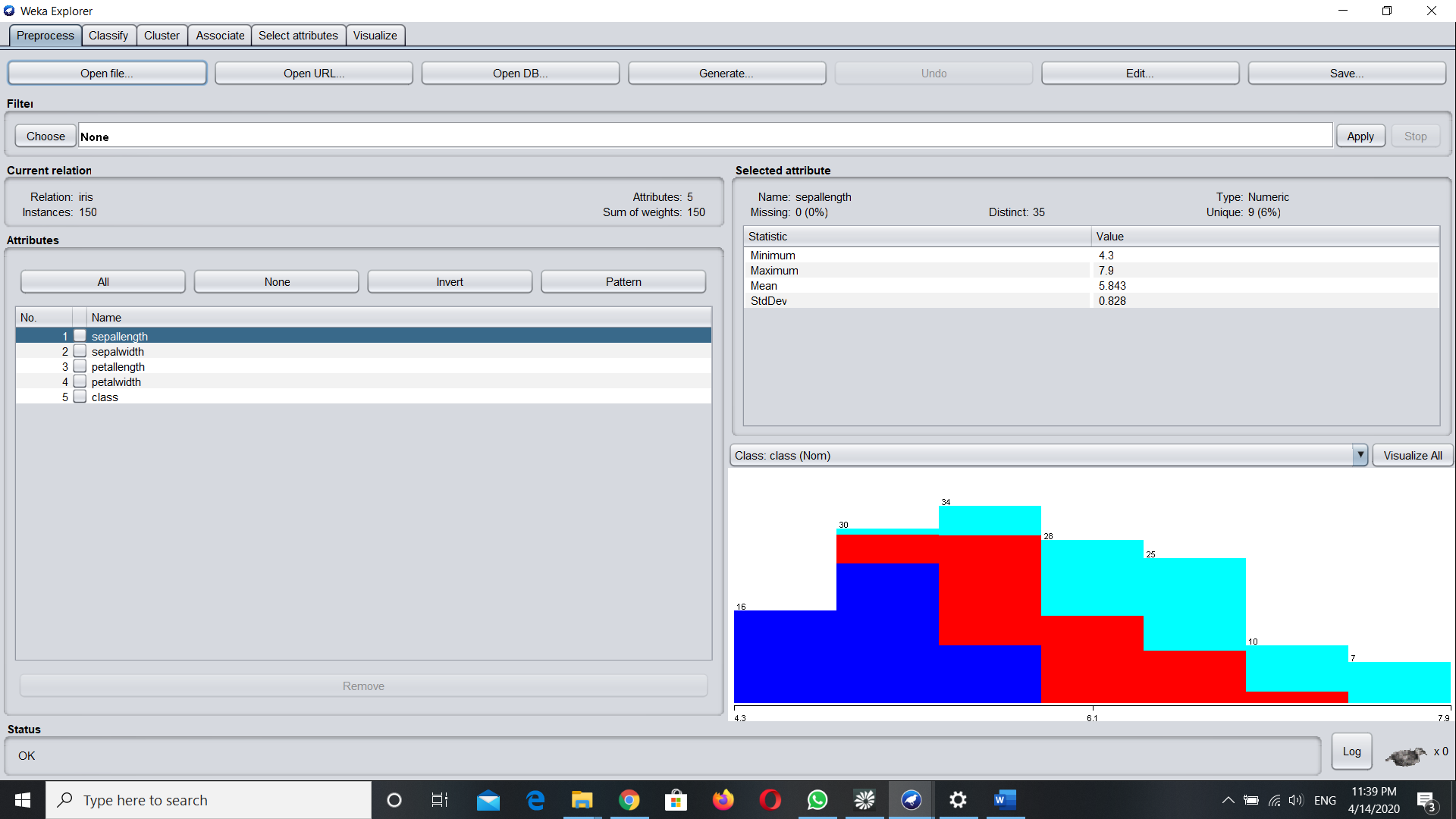
8. temperature=cool play=yes 3 ==> humidity=normal 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)

9. outlook=sunny temperature=hot 2 ==> humidity=high 2 <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)

10. temperature=hot play=no 2 ==> outlook=sunny 2 <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)

WEKA – Decision Tree on Iris Dataset





=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: iris

Instances: 150

Attributes: 5

sepallength

sepalwidth

petallength

petalwidth

class

Test mode: split 70.0% train, remainder test

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

J48 pruned tree

------------------

petalwidth <= 0.6: Iris-setosa (50.0)

petalwidth > 0.6

| petalwidth <= 1.7

| | petallength <= 4.9: Iris-versicolor (48.0/1.0)

| | petallength > 4.9

| | | petalwidth <= 1.5: Iris-virginica (3.0)

| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)

| petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 5

Size of the tree : 9

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances 43 95.5556 %

Incorrectly Classified Instances 2 4.4444 %

Kappa statistic 0.9331

Mean absolute error 0.0416

Root mean squared error 0.1682

Relative absolute error 9.3466 %

Root relative squared error 35.6559 %

Total Number of Instances 45

=== 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

1.000 0.069 0.889 1.000 0.941 0.910 0.966 0.889 Iris-versicolor

0.867 0.000 1.000 0.867 0.929 0.901 0.964 0.931 Iris-virginica

Weighted Avg. 0.956 0.025 0.960 0.956 0.955 0.935 0.976 0.938

=== Confusion Matrix ===

a b c <-- classified as

14 0 0 | a = Iris-setosa

0 16 0 | b = Iris-versicolor

0 2 13 | c = Iris-virginica