732A75: Association Analysis-1

Liu ID: vinbe289, muhha866

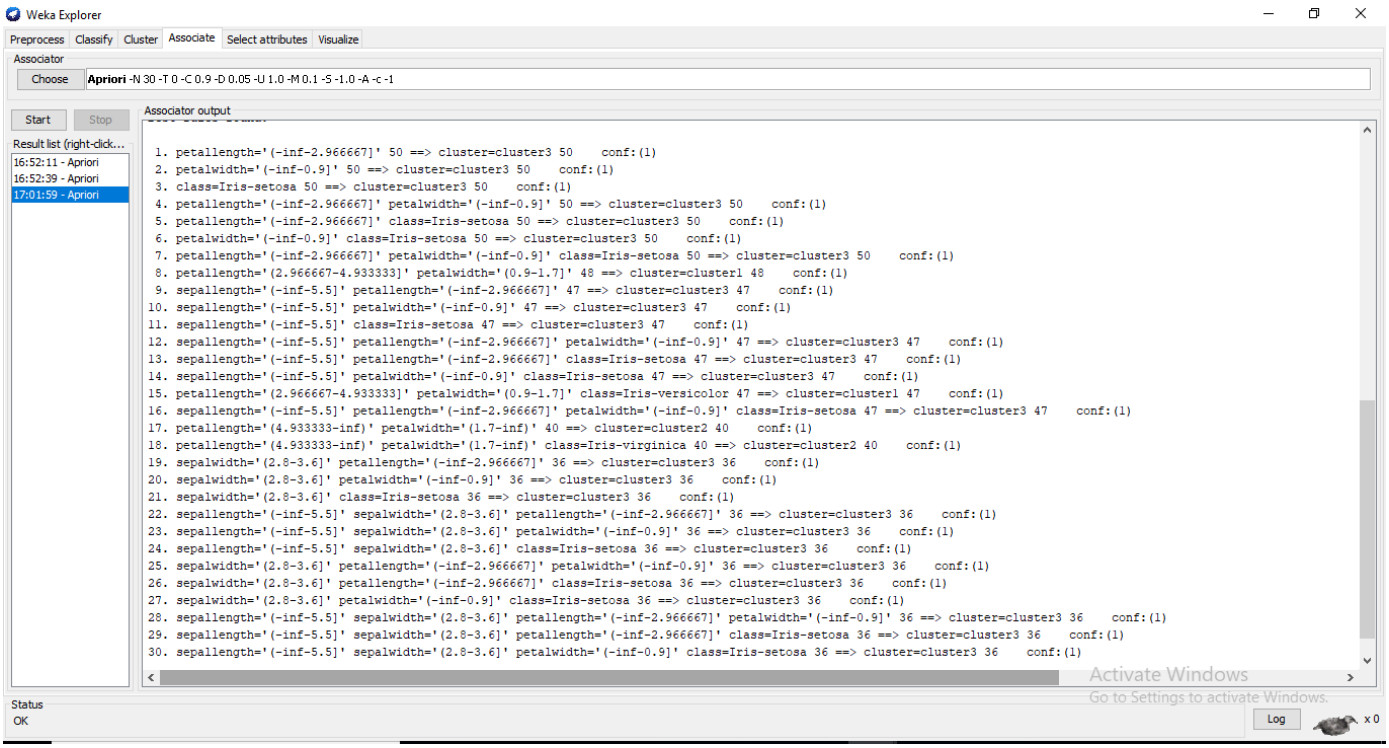
Rules that are accurate and such that the antecedent does not contain the class attribute and the consequent only contains the cluster attribute.

For SimpleKmeans,3 bin,3 cluster: car=True

petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50 conf:(1)

petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48 conf:(1)

petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40 conf:(1)

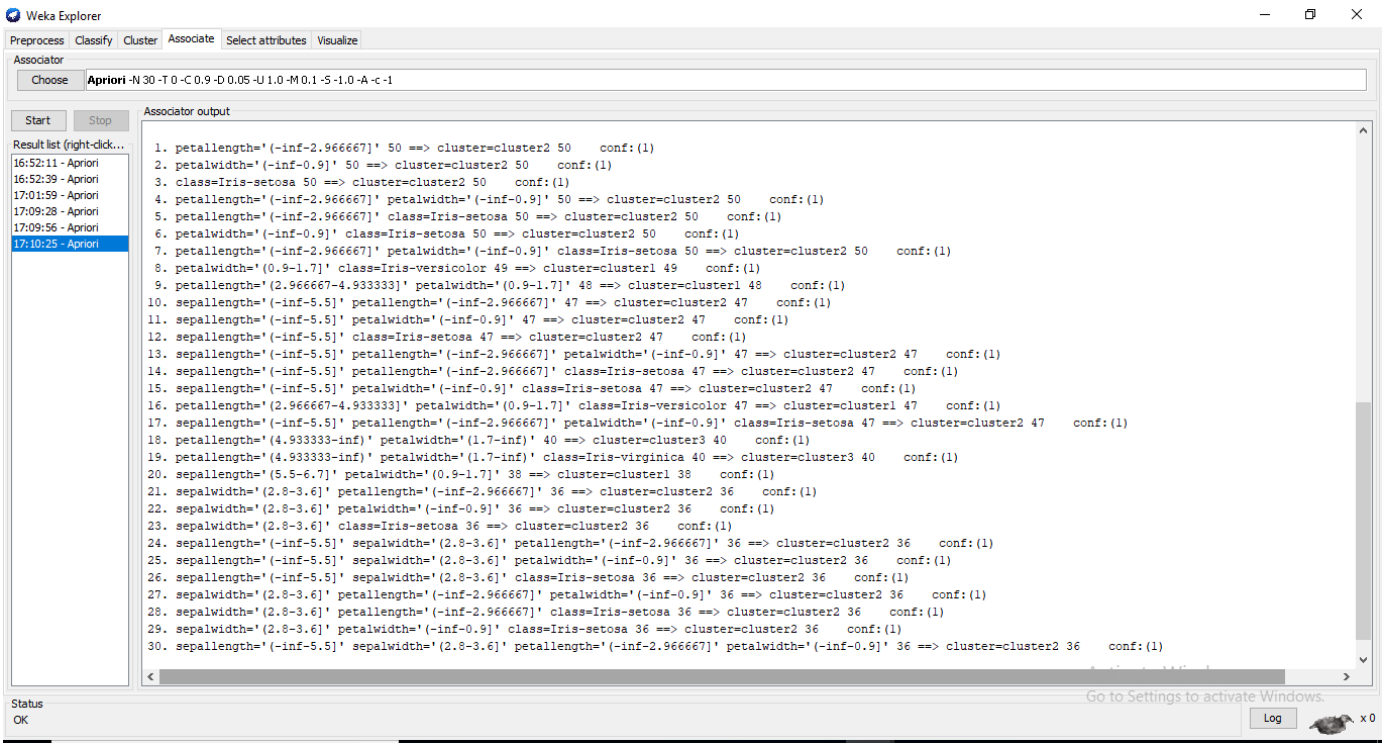


For EM,3 bin,3 cluster: car=True

petallength='(-inf-2.966667]' 50 ==> cluster=cluster2 50 conf:(1)

petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48 conf:(1)

petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster3 40 conf:(1)



<u>For HC,3 bin,3 cluster: car=True</u>

petallength='(2.966667-4.933333]' 54 ==> cluster=cluster2 54 conf:(1)

petallength='(-inf-2.966667]' 50 ==> cluster=cluster1 50 conf:(1)

No rule for cluster 3 since only 117 rules exist that satisfies the minimum confidence and there is only 1 instance of cluster 3. This rule does not satisfy the condition for considered minimum confidence of 0.9.

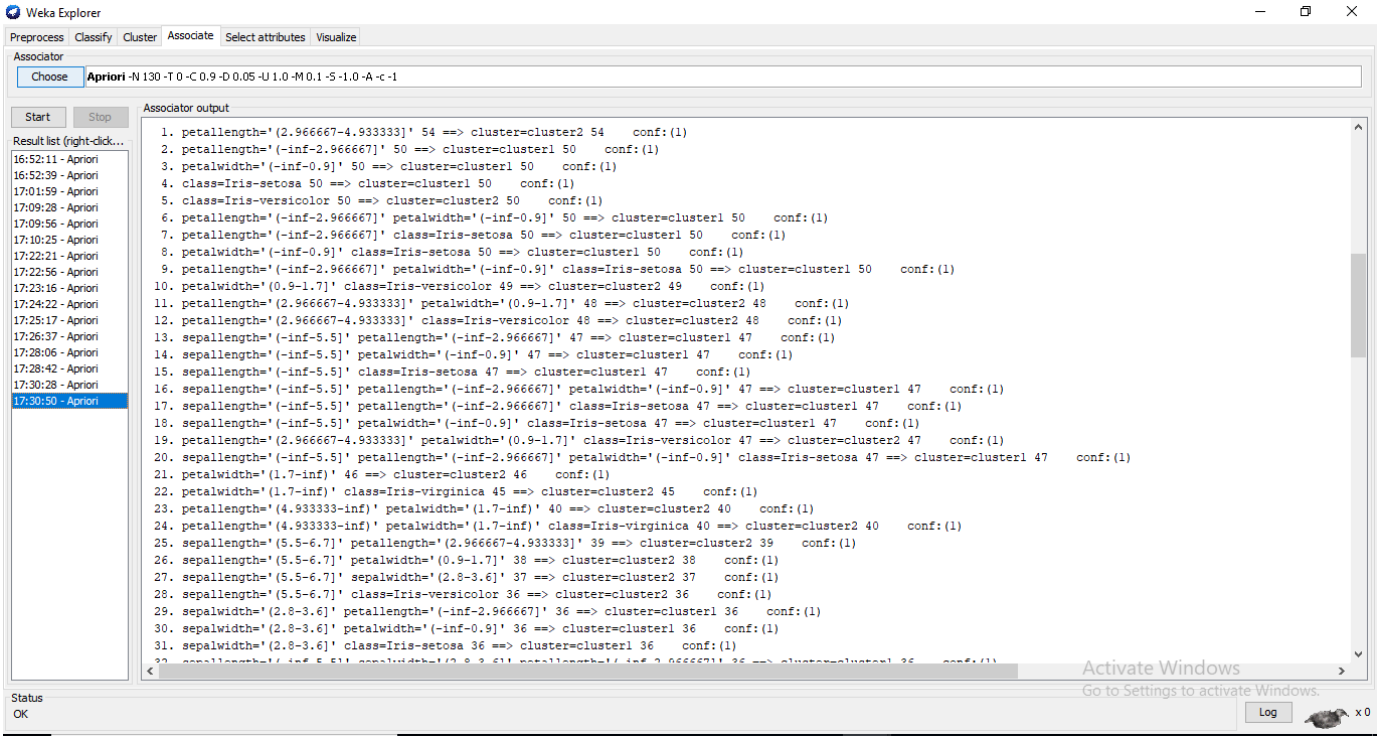
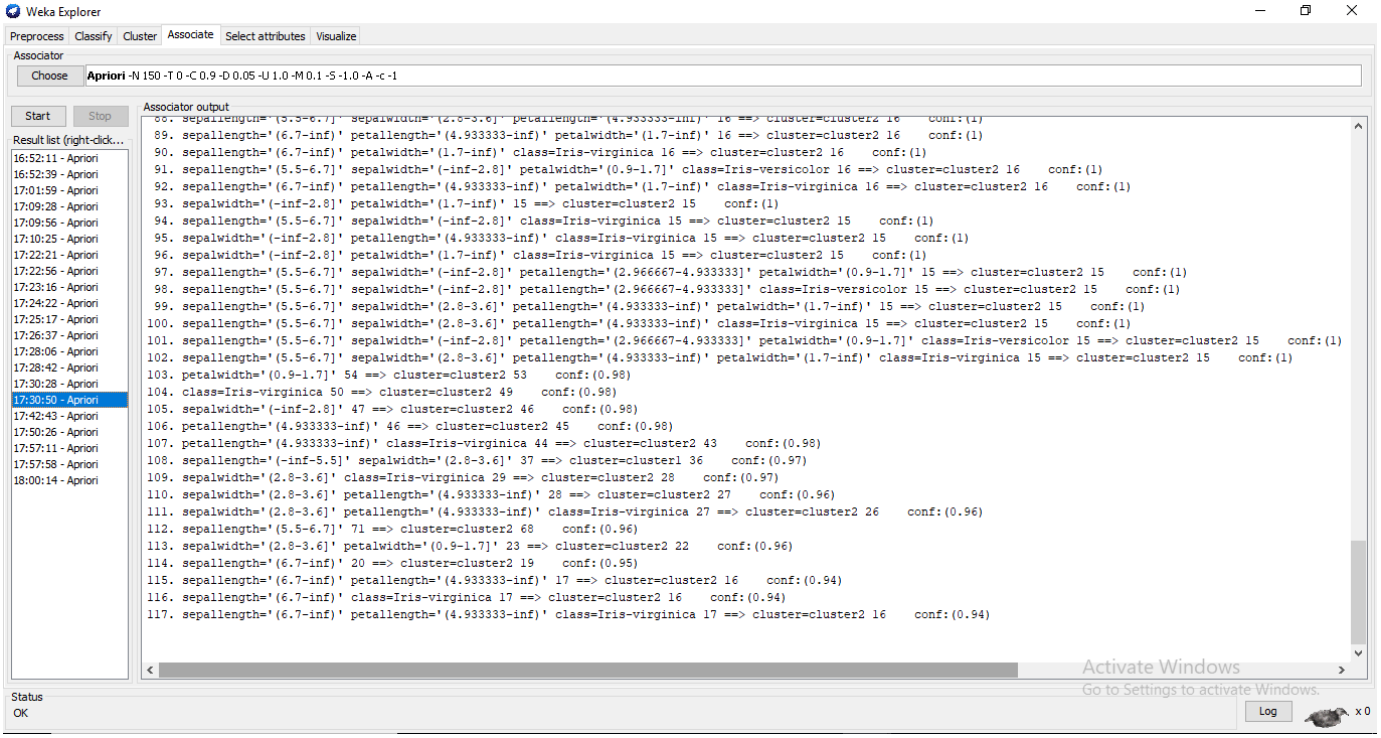


Image describing the maximum number of rules generated that satisfies the condition of minimum confidence:

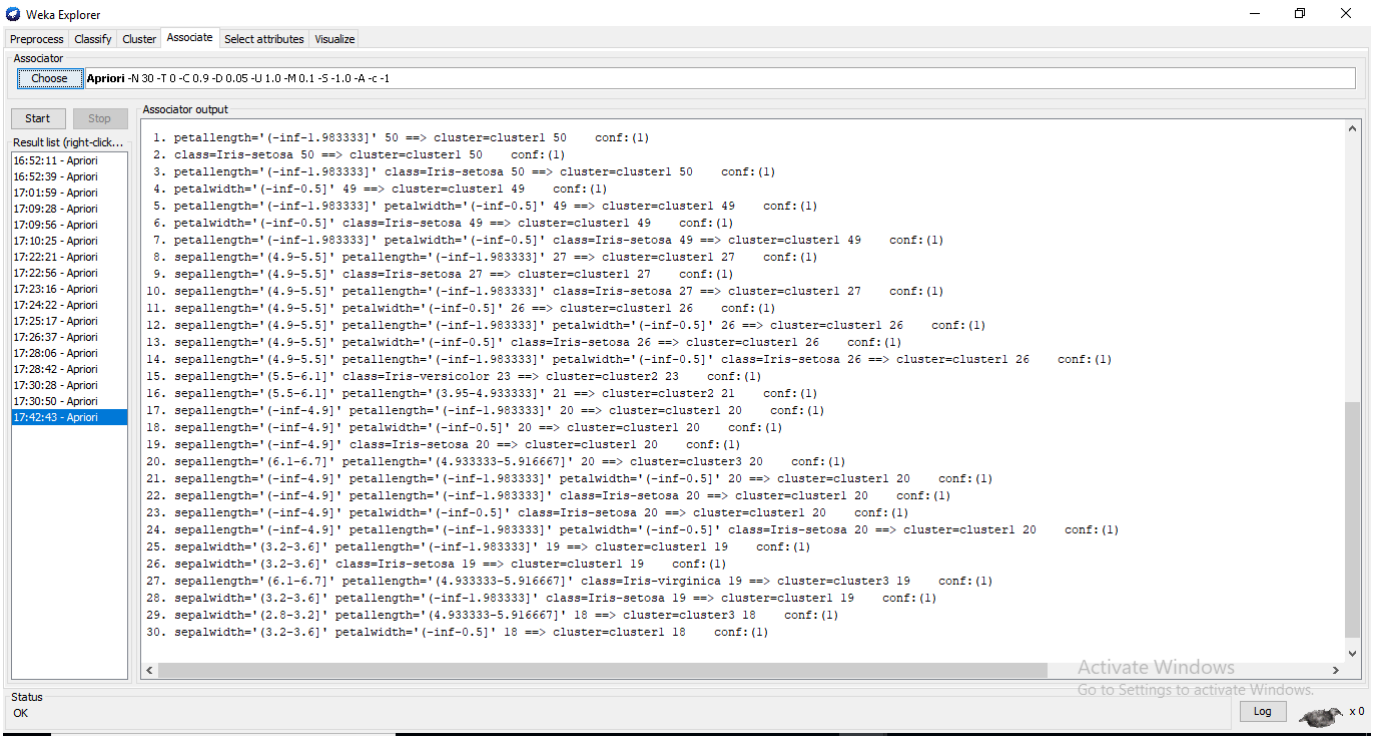


For SimpleKmeans,6 bin,3 cluster: car=True

petallength='(-inf-1.983333]' 50 ==> cluster=cluster1 50 conf:(1)

sepallength='(5.5-6.1]' petallength='(3.95-4.933333]' 21 ==> cluster=cluster2 21 conf:(1)

sepallength='(6.1-6.7]' petallength='(4.933333-5.916667]' 20 ==> cluster=cluster3 20 conf:(1)

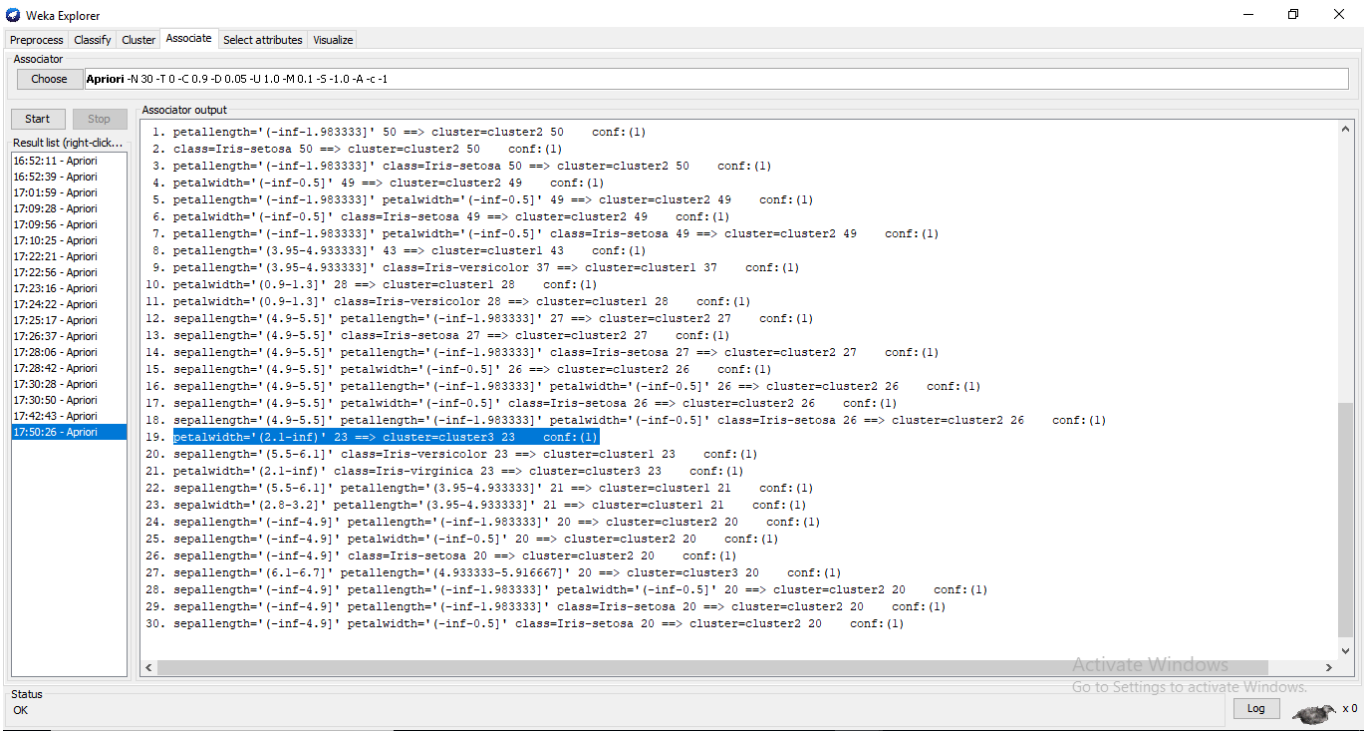


For EM,6 bin,3 cluster: car=True

petallength='(-inf-1.983333]' 50 ==> cluster=cluster2 50 conf:(1)

petallength='(3.95-4.933333]' 43 ==> cluster=cluster1 43 conf:(1)

petalwidth='(2.1-inf)' 23 ==> cluster=cluster3 23 conf:(1)



For HC,6 bin,3 cluster: car=True

petallength='(-inf-1.983333]' 50 ==> cluster=cluster1 50 conf:(1)

petallength='(3.95-4.933333]' 43 ==> cluster=cluster2 43 conf:(1)

No rule for cluster 3 since only 84 rules exist that satisfies the minimum confidence and there are only 7 instances of cluster 3. This rule does not satisfy the condition for considered minimum confidence of 0.9.

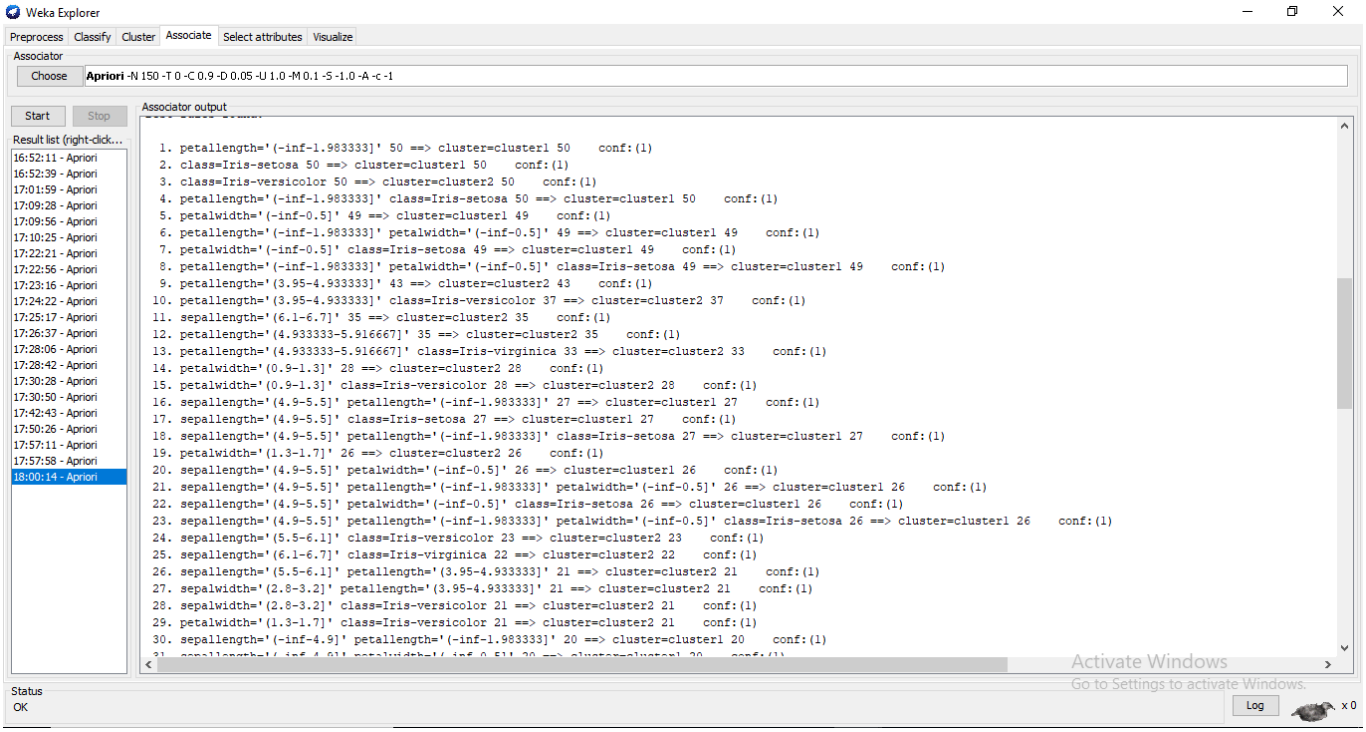
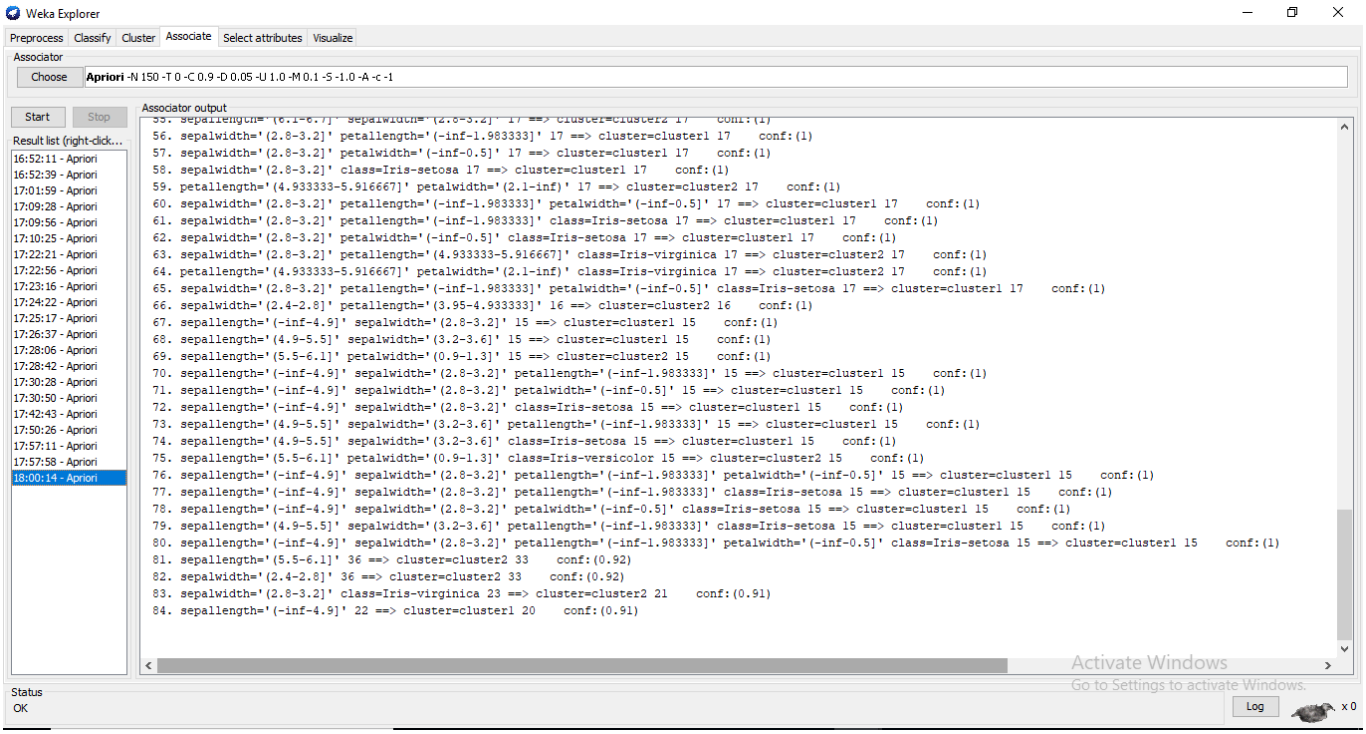


Image describing the maximum number of rules generated that satisfies the condition of minimum confidence:



Try to explain the differences between results in different experiments as well as reasons for why these differences occurred.

Algorithms used to cluster the given dataset into 3 clusters

1. SimpleKmeans
2. EM algorithm
3. Hierarchical clustering

Number of bins used

1. 3
2. 6

The number of clusters were kept constant at 3 for all the experiments.

Experiments were conducted with different combinations of the clustering algorithms and number of bins mentioned above. Six such trails are described above.

The results i.e. the association rules generated for each experiment is mentioned above .

The minimum confidence used was the default value of 0.9. Hence association rules generated are the rules which satisfy the condition of minimum confidence.

No association rules are generated where cluster 3 is the consequent since the minimum confidence considered for an association rule to be put to the output is 0.9 and none of the association rules with cluster 3 as the consequent satisfy this rule.

Hierarchical clustering seems to perform bad compared to the other 2 algorithms since it contains very less number of instances in cluster 3, 1 in the case where number of bins is 3 and 7 in the case where number of bins is 6, this in turn results in the generation of less number of association rules and the probability of one of these rules satisfying the condition of minimum confidence is low since the number of instances is itself very low.

The support of the determinant and the determinant plus the consequent decreases in all the experiments (with different clustering algorithms) with the increase in bin size from 3 to 6.