

# Advanced Data Mining Association Analysis-1

*Nahid Farazmand (nahfa911) and Anubhav Dikshit(anudi287)*

*February 27, 2019*

In this exercise we want to cluster a given dataset and use association analysis to describe the clusters obtained. for this purpose, we will work with one of the most well-known datasets in the data mining literature, namely the Iris dataset. The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample, they are the length and the width of sepal and petal.

## First step: Clustering (by usin simple K-mean algorithm)

Here we know there are 3 types of Iris flowers; so K is 3 and we have chosen the seed = 10 (from the previous exercise we know that changing the rate of seed will change the initial starting points of the K-mean process). By using K-mean the result will be as follow:

```
##          class  cluster
## 51  Iris-versicolor cluster1
## 52  Iris-versicolor cluster1
## 53  Iris-versicolor cluster1
## 54  Iris-versicolor cluster1
## 55  Iris-versicolor cluster1
## 56  Iris-versicolor cluster1
## 57  Iris-versicolor cluster1
## 58  Iris-versicolor cluster1
## 59  Iris-versicolor cluster1
## 60  Iris-versicolor cluster1
## 61  Iris-versicolor cluster1
## 62  Iris-versicolor cluster1
## 63  Iris-versicolor cluster1
## 64  Iris-versicolor cluster1
## 65  Iris-versicolor cluster1
## 66  Iris-versicolor cluster1
## 67  Iris-versicolor cluster1
## 68  Iris-versicolor cluster1
## 69  Iris-versicolor cluster1
## 70  Iris-versicolor cluster1
## 72  Iris-versicolor cluster1
## 73  Iris-versicolor cluster1
## 74  Iris-versicolor cluster1
## 75  Iris-versicolor cluster1
## 76  Iris-versicolor cluster1
## 77  Iris-versicolor cluster1
## 79  Iris-versicolor cluster1
## 80  Iris-versicolor cluster1
## 81  Iris-versicolor cluster1
## 82  Iris-versicolor cluster1
## 83  Iris-versicolor cluster1
## 84  Iris-versicolor cluster1
## 85  Iris-versicolor cluster1
## 86  Iris-versicolor cluster1
## 87  Iris-versicolor cluster1
```

```
## 88 Iris-versicolor cluster1
## 89 Iris-versicolor cluster1
## 90 Iris-versicolor cluster1
## 91 Iris-versicolor cluster1
## 92 Iris-versicolor cluster1
## 93 Iris-versicolor cluster1
## 94 Iris-versicolor cluster1
## 95 Iris-versicolor cluster1
## 96 Iris-versicolor cluster1
## 97 Iris-versicolor cluster1
## 98 Iris-versicolor cluster1
## 99 Iris-versicolor cluster1
## 100 Iris-versicolor cluster1
## 107 Iris-virginica cluster1
## 120 Iris-virginica cluster1
## 122 Iris-virginica cluster1
## 124 Iris-virginica cluster1
## 127 Iris-virginica cluster1
## 134 Iris-virginica cluster1
## 135 Iris-virginica cluster1
## 71 Iris-versicolor cluster2
## 78 Iris-versicolor cluster2
## 101 Iris-virginica cluster2
## 102 Iris-virginica cluster2
## 103 Iris-virginica cluster2
## 104 Iris-virginica cluster2
## 105 Iris-virginica cluster2
## 106 Iris-virginica cluster2
## 108 Iris-virginica cluster2
## 109 Iris-virginica cluster2
## 110 Iris-virginica cluster2
## 111 Iris-virginica cluster2
## 112 Iris-virginica cluster2
## 113 Iris-virginica cluster2
## 114 Iris-virginica cluster2
## 115 Iris-virginica cluster2
## 116 Iris-virginica cluster2
## 117 Iris-virginica cluster2
## 118 Iris-virginica cluster2
## 119 Iris-virginica cluster2
## 121 Iris-virginica cluster2
## 123 Iris-virginica cluster2
## 125 Iris-virginica cluster2
## 126 Iris-virginica cluster2
## 128 Iris-virginica cluster2
## 129 Iris-virginica cluster2
## 130 Iris-virginica cluster2
## 131 Iris-virginica cluster2
## 132 Iris-virginica cluster2
## 133 Iris-virginica cluster2
## 136 Iris-virginica cluster2
## 137 Iris-virginica cluster2
## 138 Iris-virginica cluster2
## 139 Iris-virginica cluster2
```

```
## 140 Iris-virginica cluster2
## 141 Iris-virginica cluster2
## 142 Iris-virginica cluster2
## 143 Iris-virginica cluster2
## 144 Iris-virginica cluster2
## 145 Iris-virginica cluster2
## 146 Iris-virginica cluster2
## 147 Iris-virginica cluster2
## 148 Iris-virginica cluster2
## 149 Iris-virginica cluster2
## 150 Iris-virginica cluster2
## 1      Iris-setosa cluster3
## 2      Iris-setosa cluster3
## 3      Iris-setosa cluster3
## 4      Iris-setosa cluster3
## 5      Iris-setosa cluster3
## 6      Iris-setosa cluster3
## 7      Iris-setosa cluster3
## 8      Iris-setosa cluster3
## 9      Iris-setosa cluster3
## 10     Iris-setosa cluster3
## 11     Iris-setosa cluster3
## 12     Iris-setosa cluster3
## 13     Iris-setosa cluster3
## 14     Iris-setosa cluster3
## 15     Iris-setosa cluster3
## 16     Iris-setosa cluster3
## 17     Iris-setosa cluster3
## 18     Iris-setosa cluster3
## 19     Iris-setosa cluster3
## 20     Iris-setosa cluster3
## 21     Iris-setosa cluster3
## 22     Iris-setosa cluster3
## 23     Iris-setosa cluster3
## 24     Iris-setosa cluster3
## 25     Iris-setosa cluster3
## 26     Iris-setosa cluster3
## 27     Iris-setosa cluster3
## 28     Iris-setosa cluster3
## 29     Iris-setosa cluster3
## 30     Iris-setosa cluster3
## 31     Iris-setosa cluster3
## 32     Iris-setosa cluster3
## 33     Iris-setosa cluster3
## 34     Iris-setosa cluster3
## 35     Iris-setosa cluster3
## 36     Iris-setosa cluster3
## 37     Iris-setosa cluster3
## 38     Iris-setosa cluster3
## 39     Iris-setosa cluster3
## 40     Iris-setosa cluster3
## 41     Iris-setosa cluster3
## 42     Iris-setosa cluster3
## 43     Iris-setosa cluster3
```

```
## 44      Iris-setosa cluster3
## 45      Iris-setosa cluster3
## 46      Iris-setosa cluster3
## 47      Iris-setosa cluster3
## 48      Iris-setosa cluster3
## 49      Iris-setosa cluster3
## 50      Iris-setosa cluster3
```

```
kMeans
=====
```

```
Number of iterations: 3
Within cluster sum of squared errors: 96.0
```

```
Initial starting points (random):
```

```
Cluster 0: '\ '(5.5-6.7]'\ '\ '(2.8-3.6]'\ '\ '(2.966667-4.933333]'\ '\ '(0.9-1.7]'\ '\ '
Cluster 1: '\ '(6.7-inf]'\ '\ '(2.8-3.6]'\ '\ '(4.933333-inf]'\ '\ '(1.7-inf]'\ '\ '
Cluster 2: '\ '(-inf-5.5]'\ '\ '(3.6-inf]'\ '\ '(-inf-2.966667]'\ '\ '(-inf-0.9]'\ '\ '
```

```
Missing values globally replaced with mean/mode
```

```
Final cluster centroids:
```

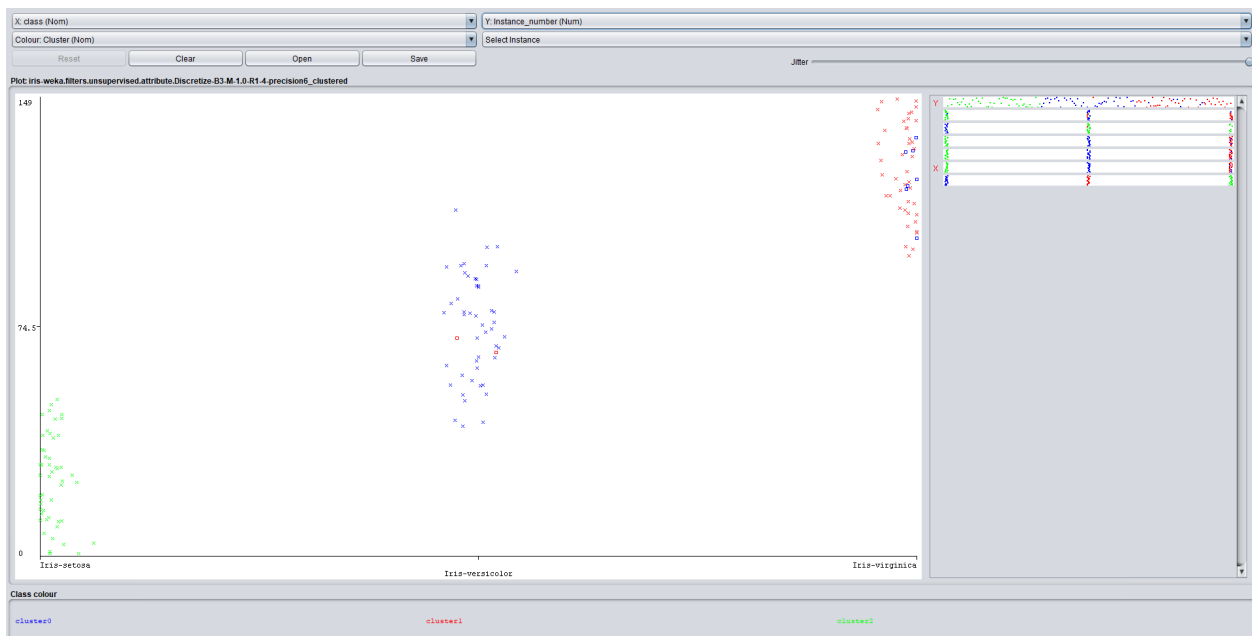
Attribute	Full Data (150.0)	Cluster# 0 (55.0)	1 (45.0)	2 (50.0)
sepal.length	' (5.5-6.7] '	' (5.5-6.7] '	' (5.5-6.7] '	' (-inf-5.5] '
sepal.width	' (2.8-3.6] '	' (-inf-2.8] '	' (2.8-3.6] '	' (2.8-3.6] '
petal.length	' (2.966667-4.933333] '	' (2.966667-4.933333] '	' (4.933333-inf] '	' (-inf-2.966667] '
petal.width	' (0.9-1.7] '	' (0.9-1.7] '	' (1.7-inf] '	' (-inf-0.9] '

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

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

```
Clustered Instances
```

```
0      55 ( 37%)
1      45 ( 30%)
2      50 ( 33%)
```



As we can see, cluster 1 is exactly same as class 1, cluster 1 contains two observations from class 3 and cluster 2 contains 7 observations from class 2.

## Second step: Association analysis (Apriori algorithm)

Here, at first we create a new attribute that represents the cluster label assigned to each instance.

After that we will implement Apriori algorithm and then we try to Find rules that are accurate and such that the antecedent does not contain the class attribute and the consequent only contains the cluster attribute. Such rules for the 3 clusters should help us to describe the instances grouped in each cluster.

**1. Number of clusters = 3 (in clustering), Number of bins = 3 (In discretizing), Minimum support: 0.3 (45 instances), Minimum metric (confidence): 0.8**

Founded rules:

1. sepallength= $[-\infty, 5.5]$  petallength= $[-\infty, 2.966667]$  47 ==> cluster=cluster3 47
2. sepallength= $[-\infty, 5.5]$  petalwidth= $[-\infty, 0.9]$  47 ==> cluster=cluster3 47
3. sepallength= $[-\infty, 5.5]$  petallength= $[-\infty, 2.966667]$  petalwidth= $[-\infty, 0.9]$  47 ==> cluster=cluster3 47

**1. Number of clusters = 3 (in clustering), Number of bins = 2 (In discretizing), Minimum support: 0.4 (60 instances), Minimum metric (confidence): 0.8**

```
##           class  cluster
## 1      Iris-setosa cluster1
## 2      Iris-setosa cluster1
## 3      Iris-setosa cluster1
## 4      Iris-setosa cluster1
## 5      Iris-setosa cluster1
## 6      Iris-setosa cluster1
## 7      Iris-setosa cluster1
## 8      Iris-setosa cluster1
## 9      Iris-setosa cluster1
## 10     Iris-setosa cluster1
## 11     Iris-setosa cluster1
## 12     Iris-setosa cluster1
## 13     Iris-setosa cluster1
## 14     Iris-setosa cluster1
## 15     Iris-setosa cluster1
## 16     Iris-setosa cluster1
## 17     Iris-setosa cluster1
## 18     Iris-setosa cluster1
## 19     Iris-setosa cluster1
## 20     Iris-setosa cluster1
## 21     Iris-setosa cluster1
## 22     Iris-setosa cluster1
## 23     Iris-setosa cluster1
## 24     Iris-setosa cluster1
## 25     Iris-setosa cluster1
## 26     Iris-setosa cluster1
## 27     Iris-setosa cluster1
## 28     Iris-setosa cluster1
## 29     Iris-setosa cluster1
## 30     Iris-setosa cluster1
```

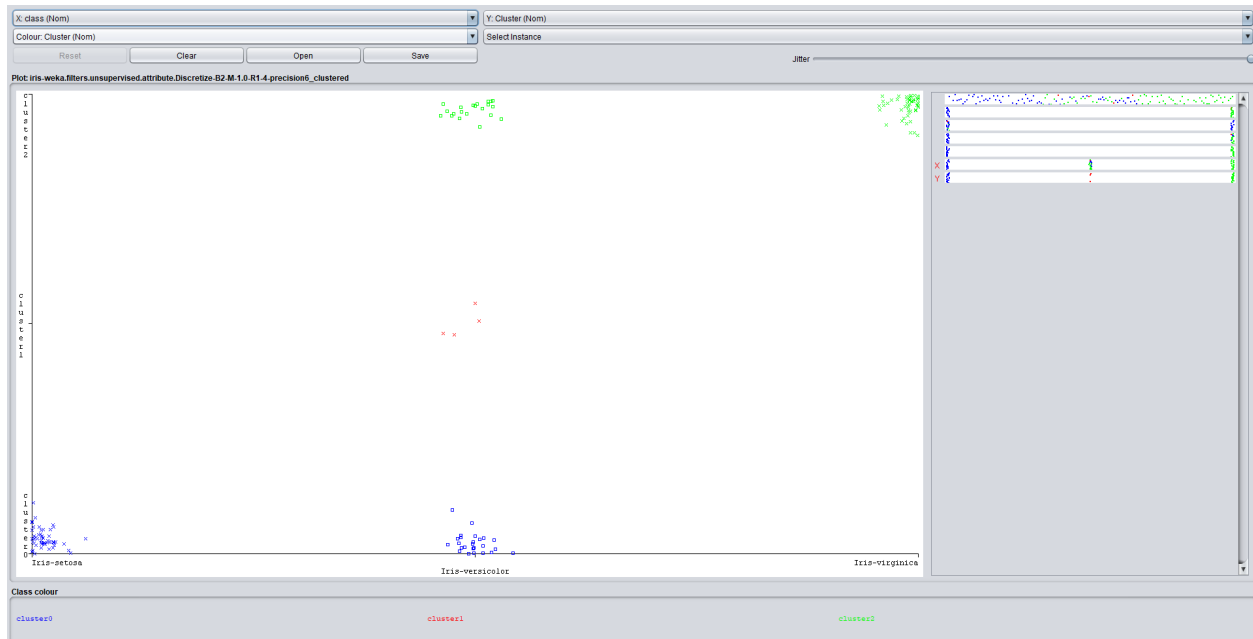
```
## 31      Iris-setosa cluster1
## 32      Iris-setosa cluster1
## 33      Iris-setosa cluster1
## 34      Iris-setosa cluster1
## 35      Iris-setosa cluster1
## 36      Iris-setosa cluster1
## 37      Iris-setosa cluster1
## 38      Iris-setosa cluster1
## 39      Iris-setosa cluster1
## 40      Iris-setosa cluster1
## 41      Iris-setosa cluster1
## 42      Iris-setosa cluster1
## 43      Iris-setosa cluster1
## 44      Iris-setosa cluster1
## 45      Iris-setosa cluster1
## 46      Iris-setosa cluster1
## 47      Iris-setosa cluster1
## 48      Iris-setosa cluster1
## 49      Iris-setosa cluster1
## 50      Iris-setosa cluster1
## 54      Iris-versicolor cluster1
## 56      Iris-versicolor cluster1
## 58      Iris-versicolor cluster1
## 60      Iris-versicolor cluster1
## 61      Iris-versicolor cluster1
## 63      Iris-versicolor cluster1
## 65      Iris-versicolor cluster1
## 68      Iris-versicolor cluster1
## 70      Iris-versicolor cluster1
## 72      Iris-versicolor cluster1
## 74      Iris-versicolor cluster1
## 80      Iris-versicolor cluster1
## 81      Iris-versicolor cluster1
## 82      Iris-versicolor cluster1
## 83      Iris-versicolor cluster1
## 89      Iris-versicolor cluster1
## 90      Iris-versicolor cluster1
## 91      Iris-versicolor cluster1
## 93      Iris-versicolor cluster1
## 94      Iris-versicolor cluster1
## 95      Iris-versicolor cluster1
## 96      Iris-versicolor cluster1
## 97      Iris-versicolor cluster1
## 99      Iris-versicolor cluster1
## 100     Iris-versicolor cluster1
## 59      Iris-versicolor cluster2
## 75      Iris-versicolor cluster2
## 88      Iris-versicolor cluster2
## 98      Iris-versicolor cluster2
## 51      Iris-versicolor cluster3
## 52      Iris-versicolor cluster3
## 53      Iris-versicolor cluster3
## 55      Iris-versicolor cluster3
## 57      Iris-versicolor cluster3
```

```
## 62 Iris-versicolor cluster3
## 64 Iris-versicolor cluster3
## 66 Iris-versicolor cluster3
## 67 Iris-versicolor cluster3
## 69 Iris-versicolor cluster3
## 71 Iris-versicolor cluster3
## 73 Iris-versicolor cluster3
## 76 Iris-versicolor cluster3
## 77 Iris-versicolor cluster3
## 78 Iris-versicolor cluster3
## 79 Iris-versicolor cluster3
## 84 Iris-versicolor cluster3
## 85 Iris-versicolor cluster3
## 86 Iris-versicolor cluster3
## 87 Iris-versicolor cluster3
## 92 Iris-versicolor cluster3
## 101 Iris-virginica cluster3
## 102 Iris-virginica cluster3
## 103 Iris-virginica cluster3
## 104 Iris-virginica cluster3
## 105 Iris-virginica cluster3
## 106 Iris-virginica cluster3
## 107 Iris-virginica cluster3
## 108 Iris-virginica cluster3
## 109 Iris-virginica cluster3
## 110 Iris-virginica cluster3
## 111 Iris-virginica cluster3
## 112 Iris-virginica cluster3
## 113 Iris-virginica cluster3
## 114 Iris-virginica cluster3
## 115 Iris-virginica cluster3
## 116 Iris-virginica cluster3
## 117 Iris-virginica cluster3
## 118 Iris-virginica cluster3
## 119 Iris-virginica cluster3
## 120 Iris-virginica cluster3
## 121 Iris-virginica cluster3
## 122 Iris-virginica cluster3
## 123 Iris-virginica cluster3
## 124 Iris-virginica cluster3
## 125 Iris-virginica cluster3
## 126 Iris-virginica cluster3
## 127 Iris-virginica cluster3
## 128 Iris-virginica cluster3
## 129 Iris-virginica cluster3
## 130 Iris-virginica cluster3
## 131 Iris-virginica cluster3
## 132 Iris-virginica cluster3
## 133 Iris-virginica cluster3
## 134 Iris-virginica cluster3
## 135 Iris-virginica cluster3
## 136 Iris-virginica cluster3
## 137 Iris-virginica cluster3
## 138 Iris-virginica cluster3
```

```

## 139 Iris-virginica cluster3
## 140 Iris-virginica cluster3
## 141 Iris-virginica cluster3
## 142 Iris-virginica cluster3
## 143 Iris-virginica cluster3
## 144 Iris-virginica cluster3
## 145 Iris-virginica cluster3
## 146 Iris-virginica cluster3
## 147 Iris-virginica cluster3
## 148 Iris-virginica cluster3
## 149 Iris-virginica cluster3
## 150 Iris-virginica cluster3

```



Founded rules:

1. sepallength= $(-\infty, 6.1]$  petalwidth= $(-\infty, 1.3]$  74  $\implies$  cluster=cluster1 74
2. petallength= $(3.95, \infty)$  petalwidth= $(1.3, \infty)$  71  $\implies$  cluster=cluster3 71
3. petallength= $(-\infty, 3.95]$  61  $\implies$  cluster=cluster1 61
4. sepallength= $(-\infty, 6.1]$  petallength= $(-\infty, 3.95]$  61  $\implies$  cluster=cluster1 61
5. sepalwidth= $(-\infty, 3.2]$  petallength= $(3.95, \infty)$  petalwidth= $(1.3, \infty)$  61  $\implies$  cluster=cluster3 61
6. petallength= $(-\infty, 3.95]$  petalwidth= $(-\infty, 1.3]$  60  $\implies$  cluster=cluster1 60
7. sepallength= $(-\infty, 6.1]$  petallength= $(-\infty, 3.95]$  petalwidth= $(-\infty, 1.3]$  60  $\implies$  cluster=cluster1 60
8. petalwidth= $(1.3, \infty)$  72  $\implies$  cluster=cluster3 71
9. sepalwidth= $(-\infty, 3.2]$  petalwidth= $(1.3, \infty)$  62  $\implies$  cluster=cluster3 61
10. petallength= $(-\infty, 3.95]$  61  $\implies$  petalwidth= $(-\infty, 1.3]$  cluster=cluster1 60
11. petalwidth= $(-\infty, 1.3]$  78  $\implies$  cluster=cluster1 74

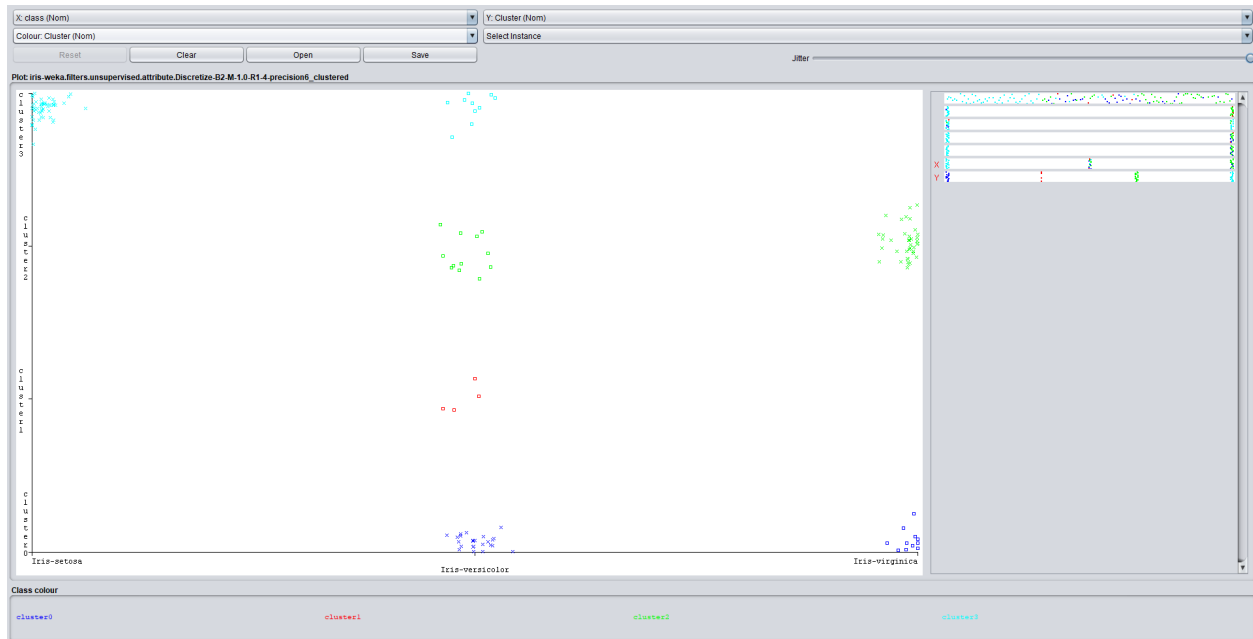


1. Number of clusters = 4 (in clustering), Number of bins = 2 (In discretizing), 0.3 (45 instances), Minimum metric (confidence): 0.8

```
##          class  cluster
## 54 Iris-versicolor cluster1
## 56 Iris-versicolor cluster1
## 60 Iris-versicolor cluster1
## 62 Iris-versicolor cluster1
## 63 Iris-versicolor cluster1
## 64 Iris-versicolor cluster1
## 67 Iris-versicolor cluster1
## 68 Iris-versicolor cluster1
## 71 Iris-versicolor cluster1
## 72 Iris-versicolor cluster1
## 74 Iris-versicolor cluster1
## 79 Iris-versicolor cluster1
## 84 Iris-versicolor cluster1
## 85 Iris-versicolor cluster1
## 86 Iris-versicolor cluster1
## 89 Iris-versicolor cluster1
## 90 Iris-versicolor cluster1
## 91 Iris-versicolor cluster1
## 92 Iris-versicolor cluster1
## 93 Iris-versicolor cluster1
## 95 Iris-versicolor cluster1
## 96 Iris-versicolor cluster1
## 97 Iris-versicolor cluster1
## 100 Iris-versicolor cluster1
## 102 Iris-virginica cluster1
## 107 Iris-virginica cluster1
## 114 Iris-virginica cluster1
## 115 Iris-virginica cluster1
## 120 Iris-virginica cluster1
## 122 Iris-virginica cluster1
## 128 Iris-virginica cluster1
## 135 Iris-virginica cluster1
## 139 Iris-virginica cluster1
## 143 Iris-virginica cluster1
## 150 Iris-virginica cluster1
## 59 Iris-versicolor cluster2
## 75 Iris-versicolor cluster2
## 88 Iris-versicolor cluster2
## 98 Iris-versicolor cluster2
## 51 Iris-versicolor cluster3
## 52 Iris-versicolor cluster3
## 53 Iris-versicolor cluster3
## 55 Iris-versicolor cluster3
## 57 Iris-versicolor cluster3
## 66 Iris-versicolor cluster3
## 69 Iris-versicolor cluster3
## 73 Iris-versicolor cluster3
## 76 Iris-versicolor cluster3
## 77 Iris-versicolor cluster3
## 78 Iris-versicolor cluster3
```

```
## 87  Iris-versicolor cluster3
## 101  Iris-virginica cluster3
## 103  Iris-virginica cluster3
## 104  Iris-virginica cluster3
## 105  Iris-virginica cluster3
## 106  Iris-virginica cluster3
## 108  Iris-virginica cluster3
## 109  Iris-virginica cluster3
## 110  Iris-virginica cluster3
## 111  Iris-virginica cluster3
## 112  Iris-virginica cluster3
## 113  Iris-virginica cluster3
## 116  Iris-virginica cluster3
## 117  Iris-virginica cluster3
## 118  Iris-virginica cluster3
## 119  Iris-virginica cluster3
## 121  Iris-virginica cluster3
## 123  Iris-virginica cluster3
## 124  Iris-virginica cluster3
## 125  Iris-virginica cluster3
## 126  Iris-virginica cluster3
## 127  Iris-virginica cluster3
## 129  Iris-virginica cluster3
## 130  Iris-virginica cluster3
## 131  Iris-virginica cluster3
## 132  Iris-virginica cluster3
## 133  Iris-virginica cluster3
## 134  Iris-virginica cluster3
## 136  Iris-virginica cluster3
## 137  Iris-virginica cluster3
## 138  Iris-virginica cluster3
## 140  Iris-virginica cluster3
## 141  Iris-virginica cluster3
## 142  Iris-virginica cluster3
## 144  Iris-virginica cluster3
## 145  Iris-virginica cluster3
## 146  Iris-virginica cluster3
## 147  Iris-virginica cluster3
## 148  Iris-virginica cluster3
## 149  Iris-virginica cluster3
## 1      Iris-setosa cluster4
## 2      Iris-setosa cluster4
## 3      Iris-setosa cluster4
## 4      Iris-setosa cluster4
## 5      Iris-setosa cluster4
## 6      Iris-setosa cluster4
## 7      Iris-setosa cluster4
## 8      Iris-setosa cluster4
## 9      Iris-setosa cluster4
## 10     Iris-setosa cluster4
## 11     Iris-setosa cluster4
## 12     Iris-setosa cluster4
## 13     Iris-setosa cluster4
## 14     Iris-setosa cluster4
```

```
## 15      Iris-setosa cluster4
## 16      Iris-setosa cluster4
## 17      Iris-setosa cluster4
## 18      Iris-setosa cluster4
## 19      Iris-setosa cluster4
## 20      Iris-setosa cluster4
## 21      Iris-setosa cluster4
## 22      Iris-setosa cluster4
## 23      Iris-setosa cluster4
## 24      Iris-setosa cluster4
## 25      Iris-setosa cluster4
## 26      Iris-setosa cluster4
## 27      Iris-setosa cluster4
## 28      Iris-setosa cluster4
## 29      Iris-setosa cluster4
## 30      Iris-setosa cluster4
## 31      Iris-setosa cluster4
## 32      Iris-setosa cluster4
## 33      Iris-setosa cluster4
## 34      Iris-setosa cluster4
## 35      Iris-setosa cluster4
## 36      Iris-setosa cluster4
## 37      Iris-setosa cluster4
## 38      Iris-setosa cluster4
## 39      Iris-setosa cluster4
## 40      Iris-setosa cluster4
## 41      Iris-setosa cluster4
## 42      Iris-setosa cluster4
## 43      Iris-setosa cluster4
## 44      Iris-setosa cluster4
## 45      Iris-setosa cluster4
## 46      Iris-setosa cluster4
## 47      Iris-setosa cluster4
## 48      Iris-setosa cluster4
## 49      Iris-setosa cluster4
## 50      Iris-setosa cluster4
## 58 Iris-versicolor cluster4
## 61 Iris-versicolor cluster4
## 65 Iris-versicolor cluster4
## 70 Iris-versicolor cluster4
## 80 Iris-versicolor cluster4
## 81 Iris-versicolor cluster4
## 82 Iris-versicolor cluster4
## 83 Iris-versicolor cluster4
## 94 Iris-versicolor cluster4
## 99 Iris-versicolor cluster4
```



Founded rules:

1. petallength= $(-\infty, 3.95]$  petalwidth= $(-\infty, 1.3]$  60  $\Rightarrow$  cluster=cluster4 60
2. sepallength= $(-\infty, 6.1]$  petallength= $(-\infty, 3.95]$  petalwidth= $(-\infty, 1.3]$  60  $\Rightarrow$  cluster=cluster4 60
3. sepallength= $(6.1, \infty)$  petalwidth= $(1.3, \infty)$  51  $\Rightarrow$  cluster=cluster3 51
4. sepallength= $(6.1, \infty)$  petallength= $(3.95, \infty)$  petalwidth= $(1.3, \infty)$  51  $\Rightarrow$  cluster=cluster3 51
5. sepallength= $(6.1, \infty)$  petalwidth= $(1.3, \infty)$  51  $\Rightarrow$  petallength= $(3.95, \infty)$  cluster=cluster3 51