Advanced Data Mining - Lab 2

Stefano Toffol (steto820) || Mim Kemal Tekin (mimte666)

11 March, 2019

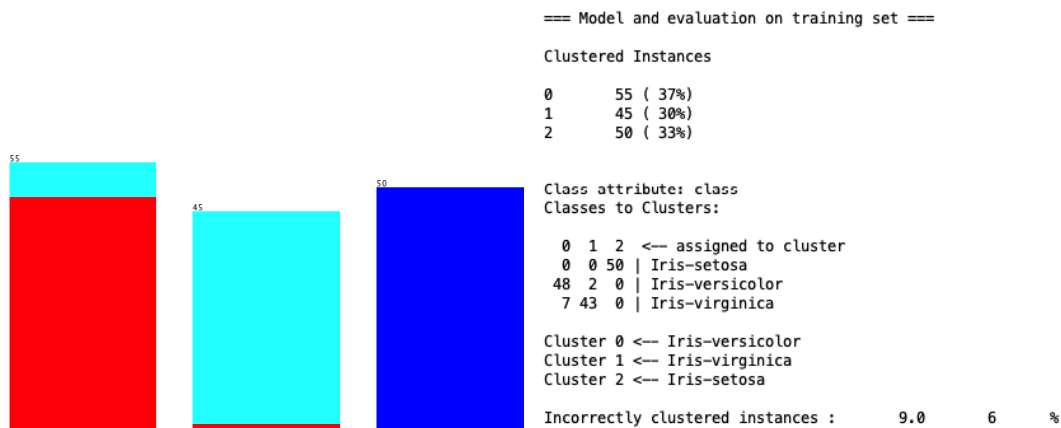## Contents

Lab 2: Association Analysis - 1

In this lab we will use Iris dataset. This dataset has 4 numeric continues features and one categorical feature which is name of the Species. We will use only first 4 numeric continues features in order to use them with a clustering algorithm (K-Means for all experiments). But with numeric continues features we cannot cluster the data, so first of all we have to split these feauteres to different count of bins. We will try different count of bins. We use 3 (recommended bin count), 5 and 10 as bin count for each features which are "sepallength", "sepalwidth", "petallength" and "petalwidth". All cluster names in Clustering sections references to association rules, not "Model and evaluation on training set" plot, so if we use cluster1 this means "Cluster 0" in that plot.

Recommended Parameters k=3; bins=3

Clustering


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

Clustered Instances

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


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0  0 50 | Iris-setosa
 48  2  0 | Iris-versicolor
  7 43  0 | Iris-virginica


Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-virginica
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances :      9.0      6      %
```

Before starting our parameter experiments we will use recommended parameters and see how the association rules are good. Recommended parameters are bins=3 and k=3. First we run k-means algorithm with k=3 and

we get the clusters below respectively cluster1, cluster2 and cluster3. Additionally we can see the performance of clustering at the right side above We have some incorrectly clustered instances, but it is not majority, just 6%. The only bad thing is that incorrect clustered instances are mostly from one class which is Iris-setosa.

Association Rules

```
Best rules found:

 1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50    conf:(1)
 2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 3. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 4. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48    conf:(1)
 5. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' 47 ==> cluster=cluster3 47    conf:(1)
 6. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
 7. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
 8. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40    conf:(1)
 9. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
10. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
11. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
12. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
13. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
14. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
15. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33    conf:(1)
16. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31    conf:(1)
17. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 30 ==> cluster=cluster1 30    conf:(1)
18. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 29 ==> cluster=cluster2 29    conf:(1)
19. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 28 ==> cluster=cluster2 28    conf:(1)
20. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 27 ==> cluster=cluster1 27    conf:(1)
21. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 26 ==> cluster=cluster2 26    conf:(1)
22. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 24 ==> cluster=cluster2 24    conf:(1)
23. sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 21 ==> cluster=cluster1 21    conf:(1)
24. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 19    conf:(1)
25. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 18 ==> cluster=cluster1 18    conf:(1)
26. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 18 ==> cluster=cluster2 18    conf:(1)
27. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 18 ==> cluster=cluster1 18    conf:(1)
28. sepallength='(6.7-inf)' petallength='(4.933333-inf)' 17 ==> cluster=cluster2 17    conf:(1)
29. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
30. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 16 ==> cluster=cluster2 16    conf:(1)
31. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
32. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 15 ==> cluster=cluster1 15    conf:(1)
33. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 15 ==> cluster=cluster2 15    conf:(1)
34. sepallength='(5.5-6.7]' petalwidth='(0.9-1.7]' 38 ==> cluster=cluster1 37    conf:(0.97)
35. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' 37 ==> cluster=cluster3 36    conf:(0.97)
36. petalwidth='(0.9-1.7]' 54 ==> cluster=cluster1 52    conf:(0.96)
37. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 18    conf:(0.95)
38. petallength='(2.966667-4.933333]' 54 ==> cluster=cluster1 51    conf:(0.94)
39. petalwidth='(1.7-inf)' 46 ==> cluster=cluster2 43    conf:(0.93)
40. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' 39 ==> cluster=cluster1 36    conf:(0.92)
41. petallength='(4.933333-inf)' 46 ==> cluster=cluster2 42    conf:(0.91)
42. sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 23 ==> cluster=cluster1 21    conf:(0.91)
43. sepallength='(5.5-6.7]' petalwidth='(1.7-inf)' 30 ==> cluster=cluster2 27    conf:(0.9)
```

After getting clusters above we can move forward to find association rules with apriori algorithm. We can see the rules that we obtain above, we have 43 rules which have more than 90% confidence. We can see our first 32 rules have 100% confidence. We captured at least one rule for each cluster as we want. Moreover, we have many rules have 100% confidence for each cluster. We have 11 rules for cluster1; 12 rules for cluster2 and cluster3 with 100% conf. We can say by looking all these foundings, this is a good clustering and binning settings for creating strong association rules, because we have almost equal count of rules for each cluster in the best rules. Moreover in total we have

After this recommended settings we will test different bin values by keeping same k value for clustering which is already recommended and the actual cluster count in the dataset. Below we will discuss affect of different bin values to association rules. We use bins=3,5 and 10.

Experiments

k=3; bins=5

Clustering

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

Clustered Instances

0      63 ( 42%)
1      35 ( 23%)
2      52 ( 35%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0  0 50 | Iris-setosa
 15 33  2 | Iris-versicolor
 48  2  0 | Iris-virginica

Cluster 0 <-- Iris-virginica
Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances :       19.0      12.6667 %
```

With this binning setting, we split all features to 5 bins. We can see that Iris-versicolor have many incorrectly clustered instances. Except this Iris-setosa is clustered perfectly again and Iris-virginica is also close to perfect, but in total we have 19 incorrectly clustered instances which means 12.66% of total data. It is slightly more than double of k=3;bins=3 settings. We can see the effects of increased bins count as increasing incorrectly clustered instances.
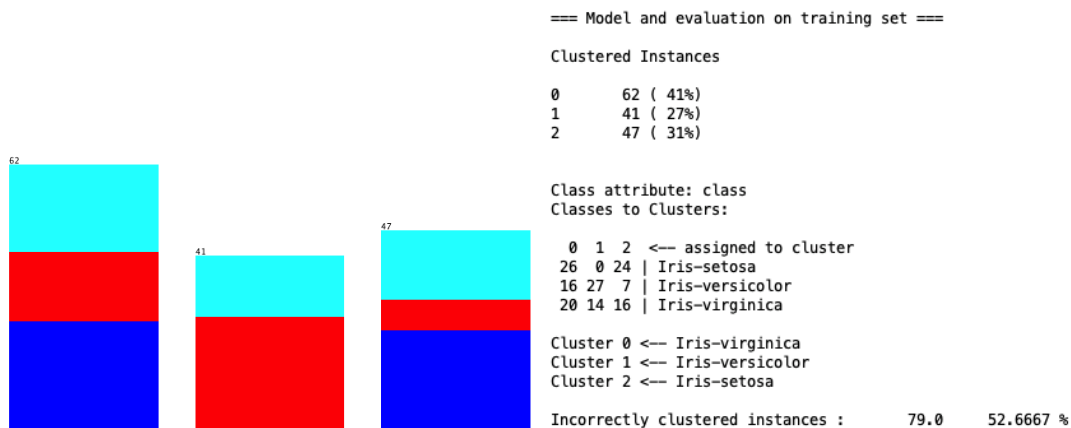
Association Rules

```
 Best rules found:

  1. petallength='(-inf-2.18]' 50 ==> cluster=cluster3 50    conf:(1)
  2. petalwidth='(-inf-0.58]' 49 ==> cluster=cluster3 49    conf:(1)
  3. petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 49 ==> cluster=cluster3 49    conf:(1)
  4. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' 28 ==> cluster=cluster3 28    conf:(1)
  5. sepallength='(-inf-5.02]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster3 27    conf:(1)
  6. sepallength='(5.74-6.46]' petallength='(4.54-5.72]' 27 ==> cluster=cluster3 27    conf:(1)
  7. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' 27 ==> cluster=cluster3 27    conf:(1)
  8. sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster3 27    conf:(1)
  9. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster3 27    conf:(1)
 10. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster3 27    conf:(1)
 11. sepalwidth='(2.96-3.44]' petallength='(4.54-5.72]' 25 ==> cluster=cluster3 25    conf:(1)
 12. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' 22 ==> cluster=cluster3 22    conf:(1)
 13. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' 22 ==> cluster=cluster3 22    conf:(1)
 14. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' 22 ==> cluster=cluster3 22    conf:(1)
 15. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 22 ==> cluster=cluster3 22    conf:(1)
 16. sepallength='(5.02-5.74]' petallength='(-inf-2.18]' 21 ==> cluster=cluster3 21    conf:(1)
 17. sepallength='(5.02-5.74]' petalwidth='(-inf-0.58]' 21 ==> cluster=cluster3 21    conf:(1)
 18. sepallength='(5.02-5.74]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 21 ==> cluster=cluster3 21    conf:(1)
 19. petallength='(4.54-5.72]' petalwidth='(1.54-2.02]' 20 ==> cluster=cluster1 20    conf:(1)
 20. sepalwidth='(2.48-2.96]' petallength='(3.36-4.54]' 18 ==> cluster=cluster2 18    conf:(1)
 21. sepallength='(5.02-5.74]' petallength='(3.36-4.54]' 17 ==> cluster=cluster2 17    conf:(1)
 22. sepalwidth='(2.96-3.44]' petalwidth='(2.02-inf)' 17 ==> cluster=cluster1 17    conf:(1)
 23. sepalwidth='(3.44-3.92]' petallength='(-inf-2.18]' 17 ==> cluster=cluster3 17    conf:(1)
 24. sepallength='(5.02-5.74]' petalwidth='(1.06-1.54]' 16 ==> cluster=cluster2 16    conf:(1)
 25. sepalwidth='(3.44-3.92]' petalwidth='(-inf-0.58]' 16 ==> cluster=cluster3 16    conf:(1)
 26. sepallength='(5.74-6.46]' sepalwidth='(2.48-2.96]' petallength='(4.54-5.72]' 16 ==> cluster=cluster1 16    conf:(1)
 27. sepalwidth='(3.44-3.92]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 16 ==> cluster=cluster3 16    conf:(1)
 28. sepallength='(5.74-6.46]' petalwidth='(1.54-2.02]' 15 ==> cluster=cluster1 15    conf:(1)
 29. sepallength='(5.02-5.74]' petallength='(3.36-4.54]' petalwidth='(1.06-1.54]' 15 ==> cluster=cluster2 15    conf:(1)
 30. sepalwidth='(6.46-7.18]' sepalwidth='(2.96-3.44]' petallength='(4.54-5.72]' 15 ==> cluster=cluster1 15    conf:(1)
 31. sepalwidth='(2.48-2.96]' petallength='(3.36-4.54]' petalwidth='(1.06-1.54]' 15 ==> cluster=cluster2 15    conf:(1)
 32. petalwidth='(1.54-2.02]' 29 ==> cluster=cluster1 28    conf:(0.97)
 33. petalwidth='(2.02-inf)' 23 ==> cluster=cluster1 22    conf:(0.96)
 34. sepallength='(-inf-5.02]' 32 ==> cluster=cluster3 30    conf:(0.94)
 35. petallength='(5.72-inf)' 16 ==> cluster=cluster1 15    conf:(0.94)
 36. petallength='(4.54-5.72]' 47 ==> cluster=cluster1 44    conf:(0.94)
 37. petallength='(3.36-4.54]' petalwidth='(1.06-1.54]' 27 ==> cluster=cluster2 25    conf:(0.93)
 38. sepallength='(6.46-7.18]' sepalwidth='(2.96-3.44]' 20 ==> cluster=cluster1 18    conf:(0.9)
```

This time although we increased the rule count to 50, we have only 38 rules which is greater than 90% confidence. This number is less than previous one. This shows us when we have more bins it is harder to create association rules. 31 of them have 100% confidence. We have rules with 100% confidence for cluster3

in majority again. We have 19 rules for cluster3, 5 rules for cluster2 and 7 rules for cluster1. As we can see that we have strong rules for each cluster and we get good rules as before.

k=3; bins=10

Clustering

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

Clustered Instances

0        62 ( 41%)
1        41 ( 27%)
2        47 ( 31%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
 26  0 24 | Iris-setosa
 16 27  7 | Iris-versicolor
 20 14 16 | Iris-virginica

Cluster 0 <-- Iris-virginica
Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances :      79.0     52.6667 %
```

This time we used 10 bins. We can see that we have 79 incorrectly clustered instances which is 52.66% of the total data. The experiments that we did before we always have Iris-setosa clustered perfectly, but this time it is incorrectly clustered almost half of it. We can see increasing bin count affects the wellness of clustering as we observed before.

Association Rules

```
Best rules found:

  1. sepallength='(4.66-5.02]' petalwidth='(-inf-0.34]' 17 ==> cluster=cluster3 17     conf:(1)
  2. sepalwidth='(2.96-3.2]' petalwidth='(-inf-0.34]' 16 ==> cluster=cluster3 16    conf:(1)
  3. sepallength='(4.66-5.02]' 23 ==> cluster=cluster3 22     conf:(0.96)
```

This time we have only 3 rules which are only for cluster3. With this settings beside having a bad clustering performance, we also get bad association rules.

Conclusion of Bin Count Experiments

Statistics of 100% Confidence

	cl1	cl2	cl3	total
k=3;bins=3	0.3125000	0.3437500	0.3750000	32
k=3;bins=5	0.2258065	0.1612903	0.6129032	31
k=3;bins=10	0.0000000	0.0000000	1.0000000	2

Statistics of Total

	cl1	cl2	cl3	total
k=3;bins=3	0.3720930	0.3255814	0.3023256	43
k=3;bins=5	0.3157895	0.1578947	0.5263158	38
k=3;bins=10	0.0000000	0.0000000	1.0000000	3

As far as we see, increasing bin count affects the clustering and association algorithms. More categories to cluster and associate means more confusing, more complex models and it is not good idea to use bins=10. In

next experiments we will not use bins=10. We can use bins=3 or bins=5. We will select bins=3, because it provides more quantity of strong association rules. k=3;bins=3 experiment has 43 rules in total. 32/43 have 100% confidence, 11/43 have 90-99.9% confidence. On the other hand, k=3;bins=5 has 38 rules in total and 31/38 have 100% confidence, 7/38 have 90-99.9% confident. But as we can see from the statistics above k=3;bins=3 has almost equally number of rules, so it is sensible to continue with bins=3. The statistics present counts of all clusters in proportional. k=3;bins=5 settings create more rules for cluster3 than for the others.

k=4; bins=3

Clustering


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

Clustered Instances

0       52 ( 35%)
1       44 ( 29%)
2       50 ( 33%)
3        4 (  3%)


Class attribute: class
Classes to Clusters:

  0  1  2  3  <-- assigned to cluster
  0  0 50  0 | Iris-setosa
 45  2  0  3 | Iris-versicolor
  7 42  0  1 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-virginica
Cluster 2 <-- Iris-setosa
Cluster 3 <-- No class

Incorrectly clustered instances :     13.0      8.6667 %
```
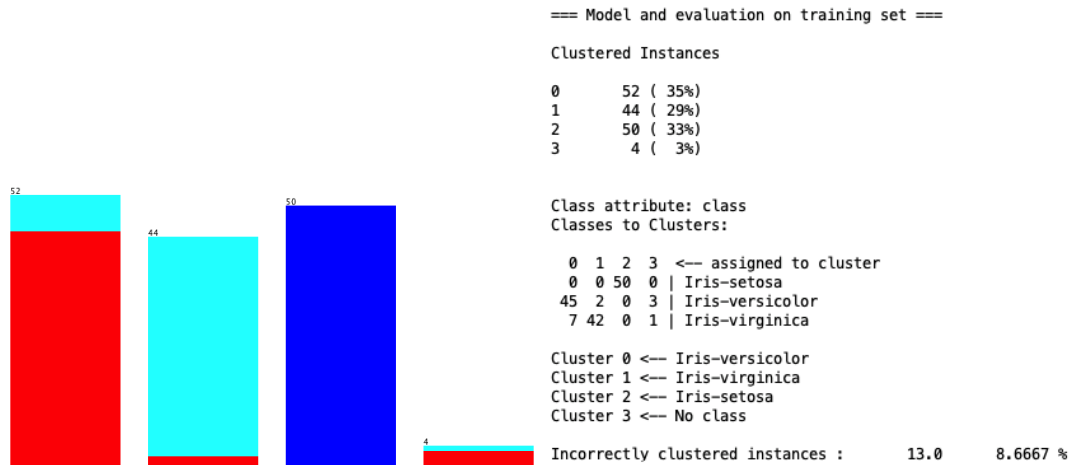
In this experiment part, we will increase cluster count. As we can see we have 3 cluster (1,2,3) which has majority inside them, but cluster4 is too small to say that it is a cluster next to others. It has only 4 element inside it. We cannot say this is a good clustering result. Strongest cluster is cluster3 here, because all Iris-setosa instances are captured in this cluster. We will see how this clustering results will affect the association rules.

Association Rules
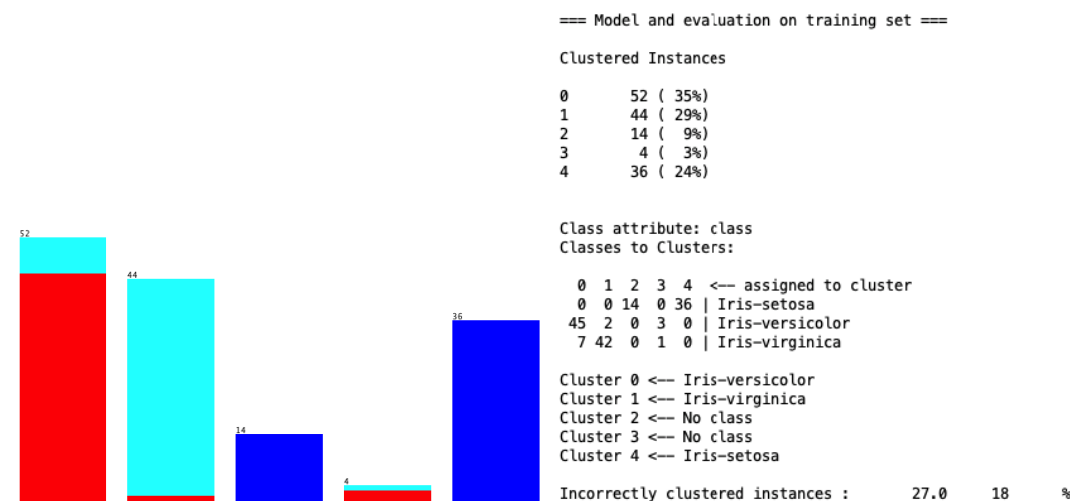
```
Best rules found:

  1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50    conf:(1)
  2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
  3. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
  4. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' 47 ==> cluster=cluster3 47    conf:(1)
  5. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
  6. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
  7. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40    conf:(1)
  8. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
  9. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
 10. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
 11. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
 12. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
 13. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
 14. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33    conf:(1)
 15. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31    conf:(1)
 16. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 30 ==> cluster=cluster1 30    conf:(1)
 17. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 29 ==> cluster=cluster2 29    conf:(1)
 18. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 27 ==> cluster=cluster1 27    conf:(1)
 19. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 26 ==> cluster=cluster2 26    conf:(1)
 20. sepallength='(5.5-6.7]' petallength='(4.333333-inf)' petalwidth='(1.7-inf)' 24 ==> cluster=cluster2 24    conf:(1)
 21. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 19    conf:(1)
 22. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 18 ==> cluster=cluster1 18    conf:(1)
 23. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 18 ==> cluster=cluster2 18    conf:(1)
 24. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 18 ==> cluster=cluster1 18    conf:(1)
 25. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
 26. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 16 ==> cluster=cluster2 16    conf:(1)
 27. sepallength='(6.7-inf)' petallength='(4.333333-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
 28. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 15 ==> cluster=cluster1 15    conf:(1)
 29. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 15 ==> cluster=cluster2 15    conf:(1)
 30. sepallength='(5.5-6.7]' petalwidth='(0.9-1.7]' 38 ==> cluster=cluster1 37    conf:(0.97)
 31. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' 37 ==> cluster=cluster3 36    conf:(0.97)
 32. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 28 ==> cluster=cluster2 27    conf:(0.96)
 33. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 18    conf:(0.95)
 34. sepallength='(6.7-inf)' petallength='(4.333333-inf)' 17 ==> cluster=cluster2 16    conf:(0.94)
 35. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 45    conf:(0.94)
 36. petalwidth='(1.7-inf)' 46 ==> cluster=cluster2 43    conf:(0.93)
 37. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' 39 ==> cluster=cluster1 36    conf:(0.92)
 38. petalwidth='(0.9-1.7]' 54 ==> cluster=cluster1 49    conf:(0.91)
 39. sepallength='(5.5-6.7]' petalwidth='(1.7-inf)' 30 ==> cluster=cluster2 27    conf:(0.9)
```

We have 29/39 rules have 100% confidence, but there is no rule for cluster4. This situation is expected after that bad clustering. Because we do not have enough instances in cluster4 to associate.

k=5; bins=3

Clustering

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

Clustered Instances

0       52 ( 35%)
1       44 ( 29%)
2       14 (  9%)
3        4 (  3%)
4       36 ( 24%)


Class attribute: class
Classes to Clusters:

   0  1  2  3  4  <-- assigned to cluster
   0  0 14  0 36 | Iris-setosa
  45  2  0  3  0 | Iris-versicolor
   7 42  0  1  0 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-virginica
Cluster 2 <-- No class
Cluster 3 <-- No class
Cluster 4 <-- Iris-setosa

Incorrectly clustered instances :     27.0     18     %
```

With this settings also we have same situation. We have the same cluster which has only 4 element (cluster4). In this time we have also splitted clusters which have only Iris-setosa instances which are really good clustered in previous experiments (cluster3 and cluster5). We can observe that increasing cluster count splits clusters to many small pieces and some of instances which have strong relation between their features resists this split.

Association Rules

```
Best rules found:

 1. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40     conf:(1)
 2. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster5 36     conf:(1)
 3. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36     conf:(1)
 4. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster5 36     conf:(1)
 5. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36     conf:(1)
 6. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36     conf:(1)
 7. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36     conf:(1)
 8. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33     conf:(1)
 9. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31     conf:(1)
10. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 30 ==> cluster=cluster1 30     conf:(1)
11. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 29 ==> cluster=cluster2 29     conf:(1)
12. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 27 ==> cluster=cluster1 27     conf:(1)
13. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 26 ==> cluster=cluster2 26     conf:(1)
14. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 24 ==> cluster=cluster2 24     conf:(1)
15. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 19     conf:(1)
16. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 18 ==> cluster=cluster1 18     conf:(1)
17. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 18 ==> cluster=cluster2 18     conf:(1)
18. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 18 ==> cluster=cluster1 18     conf:(1)
19. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16     conf:(1)
20. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 16 ==> cluster=cluster2 16     conf:(1)
21. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16     conf:(1)
22. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 15 ==> cluster=cluster1 15     conf:(1)
23. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 15 ==> cluster=cluster2 15     conf:(1)
24. sepallength='(5.5-6.7]' petalwidth='(0.9-1.7]' 38 ==> cluster=cluster1 37     conf:(0.97)
25. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' 37 ==> cluster=cluster5 36     conf:(0.97)
26. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 28 ==> cluster=cluster2 27     conf:(0.96)
27. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 18     conf:(0.95)
28. sepallength='(6.7-inf)' petallength='(4.933333-inf)' 17 ==> cluster=cluster2 16     conf:(0.94)
29. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 45     conf:(0.94)
30. petalwidth='(1.7-inf)' 46 ==> cluster=cluster2 43     conf:(0.93)
31. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' 39 ==> cluster=cluster1 36     conf:(0.92)
32. petalwidth='(0.9-1.7]' 54 ==> cluster=cluster1 49     conf:(0.91)
33. sepallength='(5.5-6.7]' petalwidth='(1.7-inf)' 30 ==> cluster=cluster2 27     conf:(0.9)
```

We have 23/33 rules have 100% confidence which does not contain rules for cluster4. This problem is very similar to previous experiment with k=4. We should create association rules for each cluster that satisfies minimum support threshold. Obviously, increasing k value is not good for our association rules.

Final Conclusion

Statistics of 100% Confidence

	cl1	cl2	cl3	cl4	cl5	total
k=3;bins=3	0.3125000	0.3437500	0.3750000	0	0.0000000	32
k=4;bins=3	0.2758621	0.3103448	0.4137931	0	0.0000000	29
k=5;bins=3	0.3478261	0.3913043	0.0000000	0	0.2608696	23

Statistics of Total

	cl1	cl2	cl3	cl4	cl5	total
k=3;bins=3	0.3720930	0.3255814	0.3023256	0	0.0000000	43
k=4;bins=3	0.3333333	0.3333333	0.3333333	0	0.0000000	39
k=5;bins=3	0.3939394	0.3939394	0.0000000	0	0.2121212	33

The statistics above, if k=3 we expect to see a value different from 0 at cl1, cl2, cl3 or same thing if k=4, we expect something at cl1, cl2, cl3 and cl4. As we can see we have missing values at k=4;bins=3 (cl4) and k=5;bins=3 (cl3, cl4). There is only one option left that we can choose which is recommended settings k=3;bins=3.