732A75 Advanced Data Mining laboratory 2 report

Yuki Washio and Nicolas Taba

March 4, 2021

1 Introduction

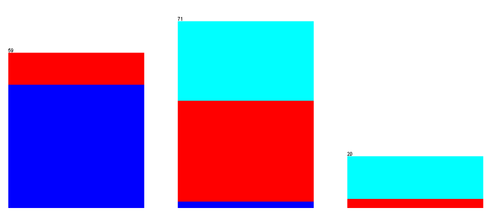
The aim of this laboratory exercise is to use assosciation analysis to describe clusters obtained from a mined dataset.

In this laboratory, we use the iris dataset. This dataset features 50 instances of each 3 classes of flowers (Iris Setosa, Iris Versicolor and Iris Virginica). Each of these data points has data pertaining to the sepal and petal length and width given in cm.

We will be using the seed value 10 whenever prompted.

2 Using assosciation analysis with optimal parameters

In this section of the laboratory, we know the make up of this dataset and want to familiarize ourselves with the assosciation analysis tool of Weka. We begin by discretizing the dataset in 3 bins of data since continuous attributes cannot be processed by the software. We perform the simple Kmeans algorithm on the data like in the previous laboratory and crosstabulate the clusters with the class labels.



(a) data separated into 3 bins according to petal-width

```
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      %
```

(b) kmeans clustering using k=3

Figure 1: clustering using Kmeans with k=3 and n=3 bins

Here we see that a few elements of versicolor and virginica are wrongly clustered and account for a 68% error in the clustering. We now perform the apriori algorithm on this clustering to try and ascertain if the rules used to make this clusters have high confidence.

We use the apriori algorithm to find rules that have the clusters as consequent.

```
 1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50    conf:(1)
 2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 3. class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 4. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 5. petallength='(-inf-2.966667]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 6. petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 7. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 8. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 48 ==> cluster=cluster1 48    conf:(1)
 9. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' 47 ==> cluster=cluster3 47    conf:(1)
10. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
11. sepallength='(-inf-5.5]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
12. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
13. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
14. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
15. petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 47 ==> cluster=cluster1 47    conf:(1)
16. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
17. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40    conf:(1)
18. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 40 ==> cluster=cluster2 40    conf:(1)
19. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
20. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
21. sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
22. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
23. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
24. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
25. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
26. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
27. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
28. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
29. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
30. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
31. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
32. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
33. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33    conf:(1)
34. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 33 ==> cluster=cluster1 33    conf:(1)
35. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31    conf:(1)
36. petalwidth='(0.9-1.7]' class=Iris-versicolor 49 ==> cluster=cluster1 48    conf:(0.98)
```

Figure 2: Apriori rules for k=3 and n=3 bins.

In Fig. 2, we have the 36 best rules for this clustering solution. All 35 first rules point to all three clusters with a confidence of 100%, with most rules indicating high confidence in identifying cluster 3 and few rules for clusters 2 and 1.

This is a good indication that our clustering is correct as several rules can be found to cluster the data to one class using one or many determinant with very high support. The 5 first rules in Fig. 2 concern cluster 3 with a support of 50. This indicates that cluster 3 is the setosa class and that this particular class can easily be identified with respect to the others.

The documentation of the dataset indicates that one of the classes is linearly separable from the others whereas the other two are not linearly separable from each other. The result we obtained is thus the expected optimal parameters when we use a very well known dataset. (documentation: https://www.ida.liu.se/~732A75/lab/iris.arff)

3 Inverstigation of change of bins and cluster numbers

We now wish to investigate how the change in number of bins and clusters affects the clustering algorithm and the apriori algorithm. We know from the previous laboratory that the assignment of the correct number of clusters to the number of classes that we believe compose the set is an important factor in how the Kmeans algorithm performs, but we do not know yet how the number of bins affects this performance. We will use the same process as previously with the apriori algorithm to assess these effects.

3.1 Changing the number of bins

We will keep the number of clusters fixed to the optimal value while changing the number of bins. We will use the previous result with n=3 and repeat the experiment with n=5 and 10.

```
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 %
```

(a) Kmeans with 5 bins and k=3

```
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 %
```

(b) Kmeans with 10 bins and k=3

```
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      %
```

(c) Kmeans with 3 bins and k=3

Figure 3: Kmeans algorithm output using different number of bins

From Fig. 3, we can see that increasing the number of bins has decreased the performance of the Kmeans algorithm. When the number of bins is much larger than the optimal (n=10 bins) we find that the performance goes as low as 52% error rate in this case. Most of the errors occur for the versicolor and virginica in the case of n= 5.

We now try to find rules assosciated to each clusters for 5 and 10 bins.

```
 1. petallength='(-inf-2.18]' 50 ==> cluster=cluster1 50    conf:(1)
 2. class=Iris-setosa 50 ==> cluster=cluster1 50    conf:(1)
 3. petallength='(-inf-2.18]' class=Iris-setosa 50 ==> cluster=cluster1 50    conf:(1)
 4. petalwidth='(-inf-0.58]' 49 ==> cluster=cluster1 49    conf:(1)
 5. petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 49 ==> cluster=cluster1 49    conf:(1)
 6. petalwidth='(-inf-0.58]' class=Iris-setosa 49 ==> cluster=cluster1 49    conf:(1)
 7. petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' class=Iris-setosa 49 ==> cluster=cluster1 49    conf:(1)
 8. sepallength='(-inf-5.02]' 32 ==> cluster=cluster1 32    conf:(1)
 9. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' 28 ==> cluster=cluster1 28    conf:(1)
10. sepallength='(-inf-5.02]' class=Iris-setosa 28 ==> cluster=cluster1 28    conf:(1)
11. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' class=Iris-setosa 28 ==> cluster=cluster1 28    conf:(1)
12. sepallength='(-inf-5.02]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster1 27    conf:(1)
13. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' 27 ==> cluster=cluster1 27    conf:(1)
14. sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster1 27    conf:(1)
15. sepalwidth='(2.96-3.44]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
16. petallength='(3.36-4.54]' petalwidth='(1.06-1.54]' 27 ==> cluster=cluster2 27    conf:(1)
17. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster1 27    conf:(1)
18. sepallength='(-inf-5.02]' petalwidth='(-inf-0.58]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
19. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 27 ==> cluster=cluster1 27    conf:(1)
20. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
21. sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
22. petallength='(3.36-4.54]' petalwidth='(1.06-1.54]' class=Iris-versicolor 27 ==> cluster=cluster2 27    conf:(1)
23. sepallength='(-inf-5.02]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
24. sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' class=Iris-setosa 27 ==> cluster=cluster1 27    conf:(1)
25. petalwidth='(2.02-inf)' 23 ==> cluster=cluster3 23    conf:(1)
26. petalwidth='(2.02-inf)' class=Iris-virginica 23 ==> cluster=cluster3 23    conf:(1)
27. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' 22 ==> cluster=cluster1 22    conf:(1)
28. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' 22 ==> cluster=cluster1 22    conf:(1)
29. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' 22 ==> cluster=cluster1 22    conf:(1)
30. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' class=Iris-setosa 22 ==> cluster=cluster1 22    conf:(1)
31. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' 22 ==> cluster=cluster1 22    conf:(1)
32. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' class=Iris-setosa 22 ==> cluster=cluster1 22    conf:(1)
33. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petalwidth='(-inf-0.58]' class=Iris-setosa 22 ==> cluster=cluster1 22    conf:(1)
34. sepallength='(-inf-5.02]' sepalwidth='(2.96-3.44]' petallength='(-inf-2.18]' petalwidth='(-inf-0.58]' class=Iris-setosa 22 ==> cluster=cluster1 22    conf:(1)
35. sepallength='(5.02-5.74]' petallength='(-inf-2.18]' 21 ==> cluster=cluster1 21    conf:(1)
36. sepallength='(5.02-5.74]' petalwidth='(-inf-0.58]' 21 ==> cluster=cluster1 21    conf:(1)
37. sepallength='(5.02-5.74]' class=Iris-setosa 21 ==> cluster=cluster1 21    conf:(1)
```

Figure 4: Apriori rules for k=3 and n=5 bins.

```
Best rules found:

 1. class=Iris-setosa 50 ==> cluster=cluster1 50    conf:(1)
 2. petalwidth='(-inf-0.34]' 41 ==> cluster=cluster1 41    conf:(1)
 3. petalwidth='(-inf-0.34]' class=Iris-setosa 41 ==> cluster=cluster1 41    conf:(1)
 4. petallength='(-inf-1.59]' 37 ==> cluster=cluster1 37    conf:(1)
 5. petallength='(-inf-1.59]' class=Iris-setosa 37 ==> cluster=cluster1 37    conf:(1)
 6. petallength='(-inf-1.59]' petalwidth='(-inf-0.34]' 33 ==> cluster=cluster1 33    conf:(1)
 7. petallength='(-inf-1.59]' petalwidth='(-inf-0.34]' class=Iris-setosa 33 ==> cluster=cluster1 33    conf:(1)
 8. petallength='(3.95-4.54]' class=Iris-versicolor 25 ==> cluster=cluster2 25    conf:(1)
 9. sepallength='(4.66-5.02]' 23 ==> cluster=cluster1 23    conf:(1)
10. petalwidth='(1.06-1.3]' 21 ==> cluster=cluster2 21    conf:(1)
11. petalwidth='(1.06-1.3]' class=Iris-versicolor 21 ==> cluster=cluster2 21    conf:(1)
12. sepallength='(4.66-5.02]' class=Iris-setosa 19 ==> cluster=cluster1 19    conf:(1)
13. sepallength='(4.66-5.02]' petalwidth='(-inf-0.34]' 17 ==> cluster=cluster1 17    conf:(1)
14. sepallength='(4.66-5.02]' petalwidth='(-inf-0.34]' class=Iris-setosa 17 ==> cluster=cluster1 17    conf:(1)
15. sepalwidth='(2.96-3.2]' petalwidth='(-inf-0.34]' 16 ==> cluster=cluster1 16    conf:(1)
16. sepalwidth='(2.96-3.2]' class=Iris-setosa 16 ==> cluster=cluster1 16    conf:(1)
17. sepalwidth='(2.96-3.2]' petalwidth='(-inf-0.34]' class=Iris-setosa 16 ==> cluster=cluster1 16    conf:(1)
18. petallength='(3.95-4.54]' 26 ==> cluster=cluster2 25    conf:(0.96)
19. sepallength='(6.1-6.46]' 20 ==> cluster=cluster2 17    conf:(0.85)
20. sepallength='(6.46-6.82]' 18 ==> cluster=cluster3 15    conf:(0.83)
```

Figure 5: Apriori rules for k=3 and n=10 bins.

In Fig. 4 we see that the rules with highest confidence mostly point to cluster 1. There are some instances of cluster 3 being a consequent (rule 26 and 27). We can think from this that only 1 class can be clearly said to be dissimilar to the others.

In Fig 5, we find only 20 rules with high confidence (> 0.8) with rules that have all 3 clusters as consequent. However, we will notice that some of these rules are not "good" rules such as rule 1 that has "belonging to class iris-setosa" yielding cluster 1. This is not a good rule. We can ascertain from these rules that the bin number is not the correct one since the rules that we get are not good with some having low support 15 instances for rule 20 for example.

3.2 Changing the number of clusters

We now will investigate how the number of clusters affects the clustering algorithm and how the effects identified in the previous lab can be also identified using the rules produced by the apriori algorithm.

We will keep n=3 bins constant while varying the number of clusters that are investigated.

```
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 %
```

(a) Kmeans with 3 bins and k=4

```
=== 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      %
```

(b) Kmeans with 3 bins and k=5

Figure 6: Kmeans algorithm output using different number of clusters and the same number of bins

In Fig 6, we observe that the number of incorrectly clustered instances increases with a rate going from 6% in the optimal case to 18% for 5 clusters. In the case of 4 clusters, all instances of setosa are correctly identified and cluster 3 has very few elements. This could lead us to think already at this point that some of these elements could be clustered elsewhere or are maybe a group of outliers belonging to another class.

We now investigate the rules that are yielded by the apriori algorithm for the previous two cases.

```
 1. petallength='(-inf-2.966667]' 50 ==> cluster=cluster3 50    conf:(1)
 2. petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 3. class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 4. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 50 ==> cluster=cluster3 50    conf:(1)
 5. petallength='(-inf-2.966667]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 6. petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 7. petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 50 ==> cluster=cluster3 50    conf:(1)
 8. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' 47 ==> cluster=cluster3 47    conf:(1)
 9. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
10. sepallength='(-inf-5.5]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
11. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 47 ==> cluster=cluster3 47    conf:(1)
12. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
13. sepallength='(-inf-5.5]' petalwidth='(-inf-0.9]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
14. sepallength='(-inf-5.5]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 47 ==> cluster=cluster3 47    conf:(1)
15. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40    conf:(1)
16. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 40 ==> cluster=cluster2 40    conf:(1)
17. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
18. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
19. sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
20. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster3 36    conf:(1)
21. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
22. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
23. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
24. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
25. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
26. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster3 36    conf:(1)
27. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
28. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
29. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
30. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster3 36    conf:(1)
31. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33    conf:(1)
32. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 33 ==> cluster=cluster1 33    conf:(1)
33. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31    conf:(1)
34. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 30 ==> cluster=cluster1 30    conf:(1)
35. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 29 ==> cluster=cluster2 29    conf:(1)
36. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' class=Iris-virginica 28 ==> cluster=cluster2 28    conf:(1)
37. sepalwidth='(-inf-2.8]' class=Iris-versicolor 27 ==> cluster=cluster1 27    conf:(1)
38. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 27 ==> cluster=cluster1 27    conf:(1)
39. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' class=Iris-versicolor 27 ==> cluster=cluster1 27    conf:(1)
40. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' class=Iris-versicolor 26 ==> cluster=cluster1 26    conf:(1)
41. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 26 ==> cluster=cluster2 26    conf:(1)
42. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 26 ==> cluster=cluster1 26    conf:(1)
43. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 26 ==> cluster=cluster2 26    conf:(1)
44. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 24 ==> cluster=cluster2 24    conf:(1)
45. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 24 ==> cluster=cluster2 24    conf:(1)
46. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 19    conf:(1)
47. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 18 ==> cluster=cluster1 18    conf:(1)
48. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 18 ==> cluster=cluster2 18    conf:(1)
49. 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)
50. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 18 ==> cluster=cluster1 18    conf:(1)
51. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' class=Iris-virginica 17 ==> cluster=cluster2 17    conf:(1)
52. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' class=Iris-virginica 17 ==> cluster=cluster2 17    conf:(1)
53. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
54. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' class=Iris-versicolor 16 ==> cluster=cluster1 16    conf:(1)
55. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 16 ==> cluster=cluster2 16    conf:(1)
56. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
57. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 16 ==> cluster=cluster2 16    conf:(1)
58. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' class=Iris-versicolor 16 ==> cluster=cluster1 16    conf:(1)
59. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 16 ==> cluster=cluster2 16    conf:(1)
60. 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)
61. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' class=Iris-versicolor 15 ==> cluster=cluster1 15    conf:(1)
62. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 15 ==> cluster=cluster2 15    conf:(1)
63. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' class=Iris-virginica 15 ==> cluster=cluster2 15    conf:(1)
64. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 15 ==> cluster=cluster1 15    conf:(1)
65. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 15 ==> cluster=cluster2 15    conf:(1)
```

Figure 7: Apriori rules for k=4 and n=3 bins.

```
 1. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 40 ==> cluster=cluster2 40    conf:(1)
 2. petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 40 ==> cluster=cluster2 40    conf:(1)
 3. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster5 36    conf:(1)
 4. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36    conf:(1)
 5. sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
 6. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' 36 ==> cluster=cluster5 36    conf:(1)
 7. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36    conf:(1)
 8. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
 9. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36    conf:(1)
10. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
11. sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
12. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' 36 ==> cluster=cluster5 36    conf:(1)
13. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
14. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
15. sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
16. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' petallength='(-inf-2.966667]' petalwidth='(-inf-0.9]' class=Iris-setosa 36 ==> cluster=cluster5 36    conf:(1)
17. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 33 ==> cluster=cluster1 33    conf:(1)
18. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 33 ==> cluster=cluster1 33    conf:(1)
19. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 31 ==> cluster=cluster1 31    conf:(1)
20. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 30 ==> cluster=cluster1 30    conf:(1)
21. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 29 ==> cluster=cluster2 29    conf:(1)
22. sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' class=Iris-virginica 28 ==> cluster=cluster2 28    conf:(1)
23. sepalwidth='(-inf-2.8]' class=Iris-versicolor 27 ==> cluster=cluster1 27    conf:(1)
24. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' 27 ==> cluster=cluster1 27    conf:(1)
25. sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' class=Iris-versicolor 27 ==> cluster=cluster1 27    conf:(1)
26. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' class=Iris-versicolor 26 ==> cluster=cluster1 26    conf:(1)
27. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 26 ==> cluster=cluster2 26    conf:(1)
28. sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 26 ==> cluster=cluster1 26    conf:(1)
29. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 26 ==> cluster=cluster2 26    conf:(1)
30. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 24 ==> cluster=cluster2 24    conf:(1)
31. sepallength='(5.5-6.7]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 24 ==> cluster=cluster2 24    conf:(1)
32. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 19    conf:(1)
33. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' 18 ==> cluster=cluster1 18    conf:(1)
34. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' 18 ==> cluster=cluster2 18    conf:(1)
35. 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)
36. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 18 ==> cluster=cluster1 18    conf:(1)
37. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' class=Iris-virginica 17 ==> cluster=cluster2 17    conf:(1)
38. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(1.7-inf)' class=Iris-virginica 17 ==> cluster=cluster2 17    conf:(1)
39. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
40. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' class=Iris-versicolor 16 ==> cluster=cluster1 16    conf:(1)
41. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 16 ==> cluster=cluster2 16    conf:(1)
42. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 16 ==> cluster=cluster2 16    conf:(1)
43. sepallength='(6.7-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 16 ==> cluster=cluster2 16    conf:(1)
44. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petalwidth='(0.9-1.7]' class=Iris-versicolor 16 ==> cluster=cluster1 16    conf:(1)
45. sepallength='(6.7-inf)' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 16 ==> cluster=cluster2 16    conf:(1)
46. 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)
47. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' class=Iris-versicolor 15 ==> cluster=cluster1 15    conf:(1)
48. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' 15 ==> cluster=cluster2 15    conf:(1)
49. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' class=Iris-virginica 15 ==> cluster=cluster2 15    conf:(1)
50. sepallength='(5.5-6.7]' sepalwidth='(-inf-2.8]' petallength='(2.966667-4.933333]' petalwidth='(0.9-1.7]' class=Iris-versicolor 15 ==> cluster=cluster1 15    conf:(1)
51. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' petalwidth='(1.7-inf)' class=Iris-virginica 15 ==> cluster=cluster2 15    conf:(1)
52. sepallength='(5.5-6.7]' petalwidth='(0.9-1.7]' 38 ==> cluster=cluster1 37    conf:(0.97)
53. sepallength='(-inf-5.5]' sepalwidth='(2.8-3.6]' 37 ==> cluster=cluster5 36    conf:(0.97)
54. sepallength='(5.5-6.7]' petalwidth='(0.9-1.7]' class=Iris-versicolor 35 ==> cluster=cluster1 34    conf:(0.97)
55. sepallength='(5.5-6.7]' petallength='(2.966667-4.933333]' class=Iris-versicolor 34 ==> cluster=cluster1 33    conf:(0.97)
56. sepalwidth='(2.8-3.6]' class=Iris-virginica 29 ==> cluster=cluster2 28    conf:(0.97)
57. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' 28 ==> cluster=cluster2 27    conf:(0.96)
58. sepalwidth='(2.8-3.6]' petallength='(4.933333-inf)' class=Iris-virginica 27 ==> cluster=cluster2 26    conf:(0.96)
59. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' 19 ==> cluster=cluster1 18    conf:(0.95)
60. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petallength='(2.966667-4.933333]' class=Iris-versicolor 19 ==> cluster=cluster1 18    conf:(0.95)
61. sepallength='(5.5-6.7]' sepalwidth='(2.8-3.6]' petalwidth='(0.9-1.7]' class=Iris-versicolor 19 ==> cluster=cluster1 18    conf:(0.95)
62. sepallength='(5.5-6.7]' class=Iris-versicolor 36 ==> cluster=cluster1 34    conf:(0.94)
63. sepallength='(6.7-inf)' petallength='(4.933333-inf)' 17 ==> cluster=cluster2 16    conf:(0.94)
64. sepallength='(6.7-inf)' class=Iris-virginica 17 ==> cluster=cluster2 16    conf:(0.94)
65. sepallength='(6.7-inf)' petallength='(4.933333-inf)' class=Iris-virginica 17 ==> cluster=cluster2 16    conf:(0.94)
```

Figure 8: Apriori rules for k=5 and n=3 bins.

As stated in the previous section, the rules produced here are not good rules. They contain determinants that have the class attribute to establish a rule to a cluster consequent. Although they have high confidence, the support for some of these rules is low with many values below 20. Due to these rules being of poor quality, we are lead to think that the clustering is erroneous.

4 Discussion

In this work, we established that the apriori algorithm could be used to ascertain the quality of a clustering algorithm by analyzing the rules it yields. A primary glance at the Kmeans algorithm can indicate if there are issues in the number of bins or clusters. Too many clusters or bins will yield poor performance and many errors of clustering.