# CSE 5243 HW5 (Lab 4)

*Arathi Mani & Roee Ebenstein*


- Work Split
- Algorithm 1
  - Step 0:
  - Step 1 + 2:
- Algorithm 2
  - In Summary
- Aggregated Results
  - Running Times
  - Error Rate/F-measure
  - Skew Characteristic (for Clusters)
- Appendix A - Output of no clustering, 0.1% support and 90% confidence
- Appendix B - Output of no clusters, 0.08% support and 90% confidence
- Appendix C - Output of 8 clusters, 0.1% support and 90% confidence
- Appendix D - Output of 8 clusters, 0.08% support and 90% confidence
- Appendix E - Output of 16 clusters, 0.1% support and 90% confidence
- Appendix F - Output of 16 clusters, 0.08 support and 90% confidence


## Work Split

Roee: Setup the open source code and did algorithm 1.
Arathi: Setup the file transformations for clustering files and did algorithm 2.
We both worked on the report.

## Algorithm 1

We decided to use the LUCS-KDD implementation of the CBA algorithm given here: http://cgi.csc.liv.ac.uk/~frans/KDD/Software/CBA/cba.html. We assumed the tool works correctly, and therefore we did not manually confirm the results on the big dataset. We did run experiments on 10 records that we engineered for testing, for which the tool supplied the right results (part of those files are included with the code too).

### Step 0:

In order to the use open source software we had to transform the output of the second homework that listed each transaction vector and enumerate each word. For example, in the original data set, a transaction was listed as <ID, Attribute1, Attribute2, Attribute3, Attribute4, ClassLabel>. This was transformed by dropping the ID and enumerating each attribute to an integer. The final transaction vector had a form <1, 2, 3, 4, 5> and each number represented the same word throughout all documents. This transformation was a requirement in order to successfully use the open source software.

In order to complete the transformation, we built a hashmap between a word and its enumerated integer value. We ran the entire list of topics and classes through the hashmap to extract unique words so that there were no duplicates in the hashmap. Then, we reconstructed the file by first ignoring the first attribute (which was the article ID), and then mapping each word to its integer by performing a lookup in the hashmap we had just generated. We also sorted each transaction so that the numbers appeared in numerical order (on each line, including the classifications); this was another requirement in order to successfully use the open source software.

Each line has one topic at the end, and the enumerated integer value of the topic was always bigger than those of the attributes (another requirement of the tool).

### Step 1 + 2:

The tool we're using automatically determine which are test and which are validation sets, and therefore we did not have to randomly chose those ourselves. The tool split the set into 50% for training and 50% for testing.

The tool has a default of confidence 80%, support 20%. These defaults are too high for our dataset (no rules are being generated), but the rule generation algorithm ran very quickly.

We started exploring the values around 0.01% to 1% for support. It looks like the optimal support for this data-set is around 0.1% based on the running time of the algorithm as well as based on the number of rules that were generated.

We noticed that a lower support has a longer run time. We found this to be true because lowering the support means that more rules will pass through therefore extending the run time. For this data set at least, we did not notice big difference in run time for different confidence within the same support. It also looked like the run time difference between 0.1% to 0.08% of support is exponential (0.1% ran in a few minutes while .08% took hours to run), while the difference is much lower for 0.1% and 1%, and 1% to 20%.

We also noticed that reducing the confidence level increased the number of rules that were generated/added to the set, as expected. Regardless, the rules with the highest confidence remained at the top of the list which meant that we could build a mechanism to reuse this data.

Step two is being done as part of this section, in the "startClassification" method.

We supply the code of the apriori open source we used. In this code you can clearly see that the library prunes the rules supplied in the same approach required in class.

## Algorithm 2

The clusters for the second part of this lab were constructed by reusing code from our previous lab (lab4). We chose to use the K-means algorithm and Manhattan distance metric to construct

the clusters because K-means had a much faster running time than our implemented Hierarchical clustering algorithm and Manhattan distance metric because our previous lab reported that the final entropy did not differ much between Manhattan and Euclidean and the running time of Manhattan was faster than Euclidean.

For 16 clusters, we used a third party software to construct the clusters because our implementation led to some clusters with 0 elements. We used the same distance metric, Manhattan distance, to form the clusters. The third party clustering software was downloaded from http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm and was developed at Stanford University. It provided a nice GUI for us to load the data and a wide variety of options to output the clusters. The output from this software was transformed into the clustering file format that we used in our implementation through a function in the Lab5 code called fromOpenSourceToNormal which took the generated clustering file as a parameter. This function basically transposed the clusters; the generated clustering file mapped each document to a cluster and the output of the function mapped each cluster to an array of document ids.

From this file, the function createKTransactionFiles was run. This function takes a filename as a parameter and creates n number of cluster files. Each file is formatted exactly the same way as described for Algorithm 1. These files are constructed by first creating a hashmap that maps a cluster number to an array of document ids (just like the input file), then creating another hashmap that maps the document id to its transaction vector (inputted from the file used for algorithm 1) and creating k files that only contain the transaction vectors associated with each cluster. These files are iteratively sent through the open source software which, again, enumerates the topics/classes, and generates the output. By sending each file separately through the open source software, we were able to generate separate classifiers for each cluster and split each cluster 50:50 for training and testing to generate accurate error metrics. Also, the running time was greatly reduced by sending smaller clusters of files through the code which can be seen in the table below.

We noticed that due to the skew of the clusters, there was not too much difference between changing the support from 0.1% to 0.08%. For many of the clusters, there were less than 10 documents associated and when there is such few documents, it becomes difficult to generate accurate rules. However, the accuracy (as reported by the f-measure) did improve slightly from 8 clusters to 16 clusters as shown in the tables below. We deduce that generating more even clusters, perhaps through a more rigorous clustering algorithm like hierarchical clustering, will reduce the standard deviation of the the clusters and therefore will also reduce the error rate of the rules generated. This will eventually lead to a lower error rate than classifying the documents without clustering.

## In Summary

Overall, the time to model and train the data was lower for clustering than for not clustering the set. Overall, the error rate was lower for more clusters and even though the error rate was lower

for the set without any clusters, we expect that with a more rigorous clustering algorithm, that the error will eventually be lower than that of not clustering. Also, the standard deviation was lower based on increasing the number of clusters and we suspect that there is a correlation between lowering the standard deviation between the size of the clusters and lowering the error rate of the classifying based on association rules.

# Aggregated Results

## Running Times

The running time is the addition of the time to create the model from 50% of the set plus the time to classify 50% of the rest of the set.

| | 0.1% Support; 90% Confidence | 0.08% Support; 90% Confidence |
|---|---|---|
| No clustering | 297.32 seconds | fails, out of memory |
| 8 Clusters | 2.99 seconds | 3.12 seconds |
| 16 Clusters | 2.83 seconds | 2.74 seconds |

## Error Rate/F-measure

The error rate was found by subtracting the generated accuracy rate (the f-measure) from 100%. The error rates/f-measures for the clusters was calculated by doing a weighted average of the results outputted by each of the clusters.

| | 0.1% Support; 90% Confidence | 0.08% Support; 90% Confidence |
|---|---|---|
| No clustering | 74.96% | fails, out of memory |
| 8 Clusters | 77.01% | 77.0% |
| 16 Clusters | 75.33% | 75.29% |

## Skew Characteristic (for Clusters)

We choose to quantify the skew by using standard deviation for the clusters. The standard deviation calculation was done based on the number of documents that was assigned to each cluster.

| | Calculated Standard Deviation |
|---|---|
| 8 Clusters | 3285.1 |

| | |
|---|---|
| 16 Clusters | 2307.5 |

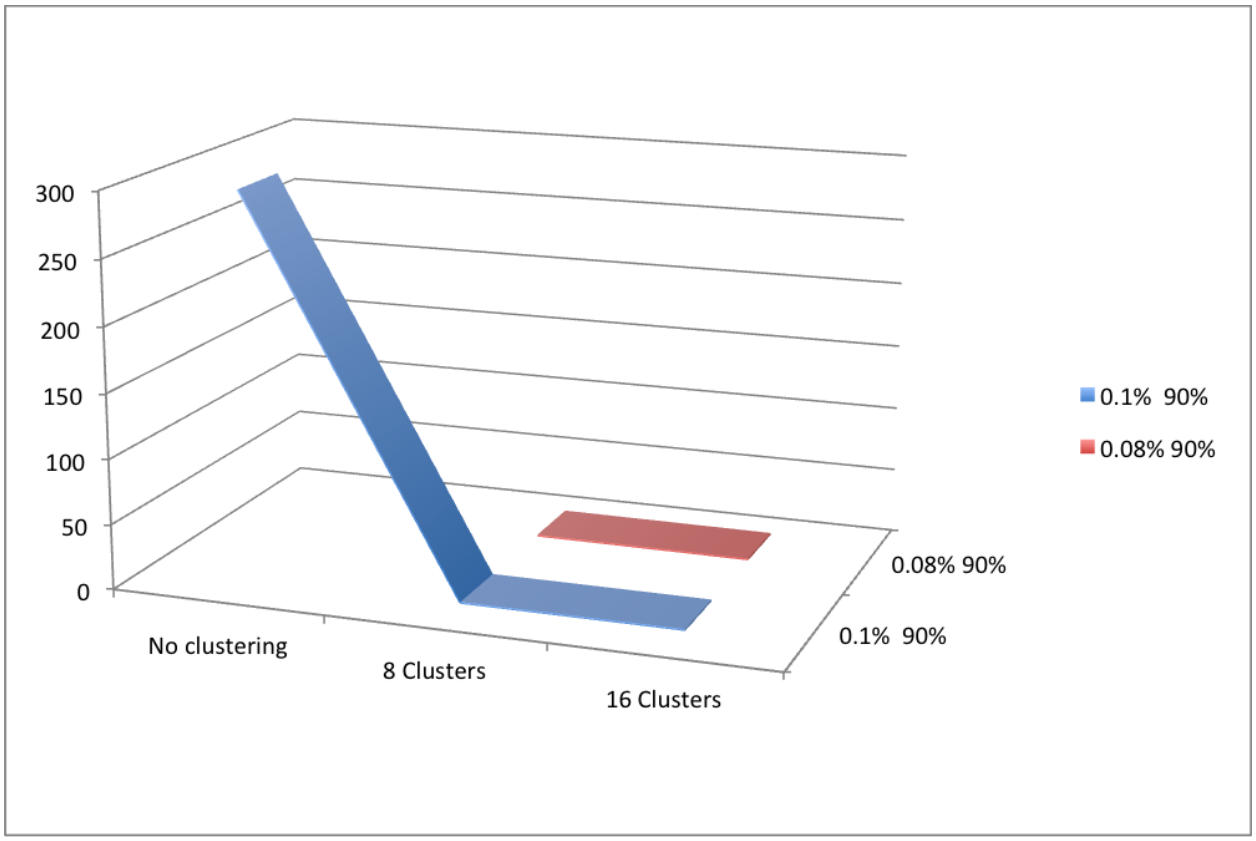


**Figure 1:** Running time of building the model and classifying documents.

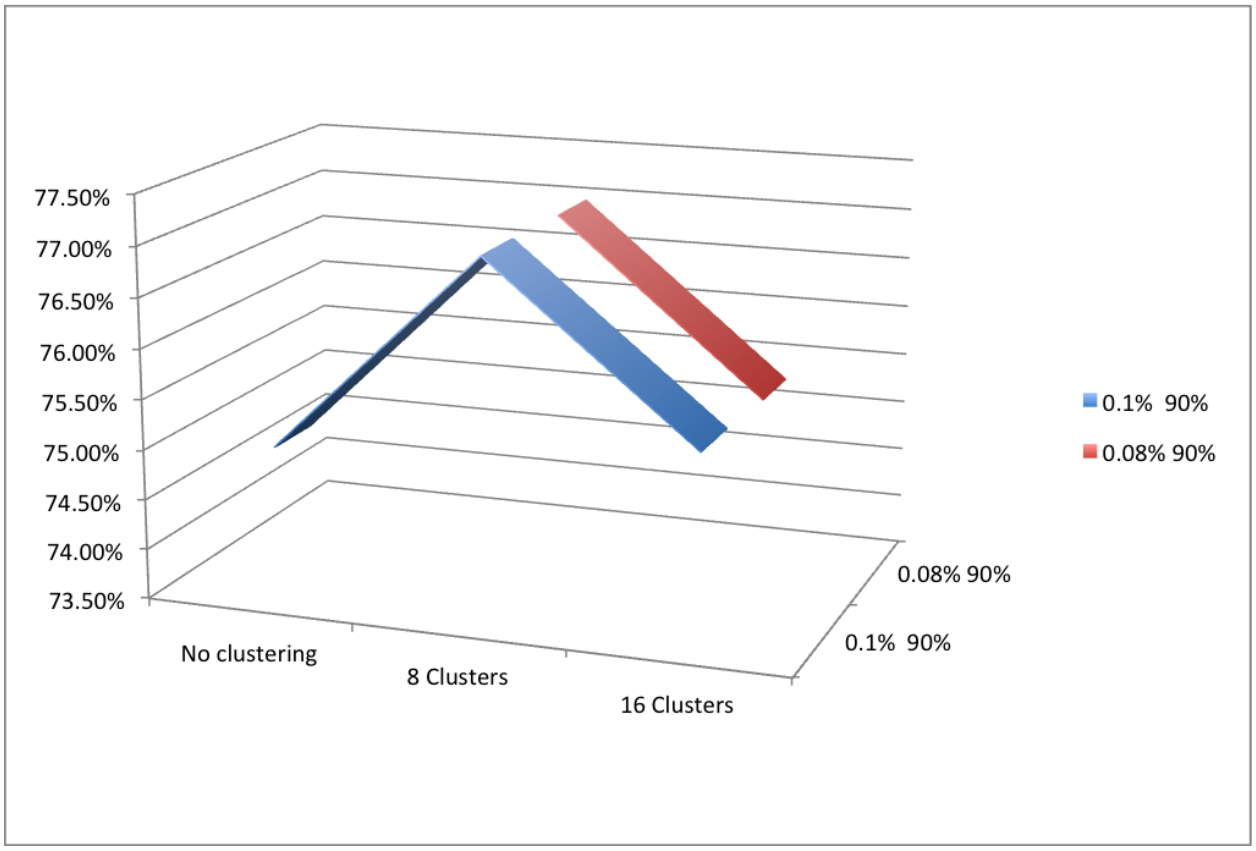


**Figure 2:** Error rate of the classified documents.

# Appendix A - Output of no clustering, 0.1% support and 90% confidence

SETTINGS
--------
Training file name = ../output/WordList.csv.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 119

Reading input file: ../output/WordList.csv.fix
Number of records = 12116
Number of columns = 6971
Min support = 12.12 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 6.06 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 6058
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (6.06 records)
Generation time = 297.32 seconds (4.96 mins)
Number of frequent sets = 34155441
Number of T-tree nodes created = 35542464
Number of T-tree Updates    = 67949945
Number of rules         = 54
Accuracy = 25.04
AUC value = 0.0121
(1)  {390 799}  ->  {6870}  100.0
(2)  {1438}  ->  {6861}  100.0
(3)  {618 799}  ->  {6870}  100.0
(4)  {1 390}  ->  {6870}  100.0
(5)  {390 618}  ->  {6870}  100.0
(6)  {521}  ->  {6853}  100.0
(7)  {1191}  ->  {6861}  100.0
(8)  {1 799}  ->  {6870}  100.0
(9)  {390 619}  ->  {6870}  100.0
(10)  {1 619}  ->  {6870}  100.0
(11)  {1541 1578}  ->  {6881}  100.0
(12)  {44}  ->  {6853}  100.0
(13)  {2502}  ->  {6853}  100.0
(14)  {1 267}  ->  {6853}  100.0
(15)  {284}  ->  {6853}  100.0
(16)  {160}  ->  {6853}  100.0
(17)  {994}  ->  {6870}  100.0
(18)  {996}  ->  {6870}  100.0
(19)  {864 1321}  ->  {6853}  100.0
(20)  {1465 835}  ->  {6853}  100.0
(21)  {278 282}  ->  {6853}  100.0
(22)  {388 1578}  ->  {6881}  100.0
(23)  {2867}  ->  {6853}  100.0
(24)  {1 296}  ->  {6853}  100.0
(25)  {1 1537}  ->  {6853}  100.0
(26)  {1 280}  ->  {6853}  100.0
(27)  {618 619}  ->  {6870}  100.0

(28) {929 1127} -> {6870} 100.0
(29) {953 524} -> {6880} 100.0
(30) {762 760} -> {6887} 100.0
(31) {1210 2834} -> {6909} 100.0
(32) {1486} -> {6853} 100.0
(33) {2846} -> {6853} 100.0
(34) {1444} -> {6853} 100.0
(35) {1491} -> {6853} 100.0
(36) {1398} -> {6853} 100.0
(37) {1949} -> {6853} 100.0
(38) {1854} -> {6870} 100.0
(39) {2078} -> {6880} 100.0
(40) {2905} -> {6909} 100.0
(41) {483} -> {6853} 95.23
(42) {475} -> {6870} 94.44
(43) {1094} -> {6853} 93.75
(44) {286} -> {6853} 92.85
(45) {1 1465} -> {6853} 91.66
(46) {1252} -> {6853} 90.9
(47) {1809} -> {6853} 90.9
(48) {1 134} -> {6853} 90.0
(49) {24 718} -> {6870} 90.0
(50) {88 967} -> {6915} 90.0
(51) {2030} -> {6853} 90.0
(52) {2312} -> {6853} 90.0
(53) {2551} -> {6853} 90.0
(54) null -> {6861} 0.0

## Appendix B - Output of no clusters, 0.08% support and 90% confidence

## Appendix C - Output of 8 clusters, 0.1% support and 90% confidence

SETTINGS
---
Training file name = ../output/KMeansCluster0.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 2

Reading input file: ../output/KMeansCluster0.txt.fix

```
Number of records = 6
Number of columns = 2
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name            = ../output/KMeansCluster1.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster1.txt.fix
Number of records = 8
Number of columns = 2
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------
```

```
Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name            = ../output/KMeansCluster2.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 7

Reading input file: ../output/KMeansCluster2.txt.fix
Number of records = 9
Number of columns = 7
Min support       = 0.01 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
```

```
Minimum support = 0.01 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 5
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.01 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 5
Number of T-tree nodes created = 17
Number of T-tree Updates       = 5
T-tree Storage          = 72 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name            = ../output/KMeansCluster3.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster3.txt.fix
Number of records = 2
Number of columns = 6
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
```

```
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 7
Number of T-tree nodes created = 10
Number of T-tree Updates       = 7
T-tree Storage          = 120 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {5}  100.0
(2)  null  ->  {5}  0.0

SETTINGS
--------
Training file name            = ../output/KMeansCluster4.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 3

Reading input file: ../output/KMeansCluster4.txt.fix
Number of records = 5
Number of columns = 7
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 8
Number of T-tree nodes created = 14
```

Number of T-tree Updates = 9
T-tree Storage = 132 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {5} 100.0
(2) null -> {5} 0.0

SETTINGS
---
Training file name = ../output/KMeansCluster5.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 114

Reading input file: ../output/KMeansCluster5.txt.fix
Number of records = 9298
Number of columns = 288
Min support = 9.3 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 4.65 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4649
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (4.65 records)
Generation time = 2.99 seconds (0.05 mins)
Number of frequent sets = 265
Number of T-tree nodes created = 5618
Number of T-tree Updates = 1864
T-tree Storage = 30584 (Bytes)
Number of rules = 3
Accuracy = 23.1
AUC value = 0.0115
(1) {23} -> {203} 100.0

(2) {24} -> {203} 94.44
(3) null -> {175} 0.0

SETTINGS
---
Training file name = ../output/KMeansCluster6.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 4

Reading input file: ../output/KMeansCluster6.txt.fix
Number of records = 5
Number of columns = 8
Min support = 0.01 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 8
Number of T-tree nodes created = 15
Number of T-tree Updates = 10
T-tree Storage = 136 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {5} 100.0
(2) null -> {6} 0.0

SETTINGS
---
Training file name = ../output/KMeansCluster7.txt.fix

```
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster7.txt.fix
Number of records = 9
Number of columns = 2
Min support       = 0.01 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.01 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 5
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.01 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
```

# Appendix D - Output of 8 clusters, 0.08% support and 90% confidence

```
Generated Rules for 8 Clusters with Support of 0.08% and Confidence of 90%
*********************************
SETTINGS
--------
Training file name            = ../output/KMeansCluster0.txt.fix
```

```
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster0.txt.fix
Number of records = 6
Number of columns = 2
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
*******************************
SETTINGS
--------
Training file name            = ../output/KMeansCluster1.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster1.txt.fix
```

```
Number of records = 8
Number of columns = 2
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
**********
SETTINGS
--------
Training file name            = ../output/KMeansCluster2.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 7

Reading input file: ../output/KMeansCluster2.txt.fix
Number of records = 9
Number of columns = 7
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------
```

```
Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 5
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 5
Number of T-tree nodes created = 17
Number of T-tree Updates       = 5
T-tree Storage          = 72 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
*********************************
SETTINGS
--------
Training file name            = ../output/KMeansCluster3.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster3.txt.fix
Number of records = 2
Number of columns = 6
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
```

```
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 7
Number of T-tree nodes created = 10
Number of T-tree Updates       = 7
T-tree Storage          = 120 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {5}  100.0
(2)  null  ->  {5}  0.0
*********************************
SETTINGS
--------
Training file name           = ../output/KMeansCluster4.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 3

Reading input file: ../output/KMeansCluster4.txt.fix
Number of records = 5
Number of columns = 7
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
```

```
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 8
Number of T-tree nodes created = 14
Number of T-tree Updates       = 9
T-tree Storage          = 132 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {5}  100.0
(2)  null  ->  {5}  0.0
*********************************
SETTINGS
--------
Training file name           = ../output/KMeansCluster5.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 114

Reading input file: ../output/KMeansCluster5.txt.fix
Number of records = 9298
Number of columns = 288
Min support       = 7.44 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 3.72 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4649
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (3.72 records)
Generation time = 3.12 seconds (0.05 mins)
Number of frequent sets = 345
Number of T-tree nodes created = 6628
Number of T-tree Updates       = 2112
```

T-tree Storage = 36988 (Bytes)
Number of rules = 3
Accuracy = 23.1
AUC value = 0.0115
(1) {23} -> {203} 100.0
(2) {24} -> {203} 94.44
(3) null -> {175} 0.0
**********
SETTINGS
---
Training file name = ../output/KMeansCluster6.txt.fix
Support (default 20%) = 0.08
Confidence (default 80%) = 90.0
Number of classes = 4

Reading input file: ../output/KMeansCluster6.txt.fix
Number of records = 5
Number of columns = 8
Min support = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 8
Number of T-tree nodes created = 15
Number of T-tree Updates = 10
T-tree Storage = 136 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {5} 100.0

```
(2)  null  ->  {6}  0.0
*********************************
SETTINGS
--------
Training file name            = ../output/KMeansCluster7.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/KMeansCluster7.txt.fix
Number of records = 9
Number of columns = 2
Min support       = 0.01 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 5
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
```

# Appendix E - Output of 16 clusters, 0.1% support and 90% confidence

```
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster0.txt.fix
Support (default 20%)        = 0.1
Confidence (default 80%)     = 90.0
Number of classes            = 3

Reading input file: ../output/2ndKMeansCluster0.txt.fix
Number of records = 3
Number of columns = 21
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.05 seconds (0.0 mins)
Number of frequent sets = 1024
Number of T-tree nodes created = 1044
Number of T-tree Updates       = 1024
T-tree Storage          = 14412 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {19}  100.0
(2)  null  ->  {20}  0.0

SETTINGS
--------
```

```
Training file name            = ../output/2ndKMeansCluster1.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 10

Reading input file: ../output/2ndKMeansCluster1.txt.fix
Number of records = 14
Number of columns = 10
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.01 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 7
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.01 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 6
Number of T-tree nodes created = 25
Number of T-tree Updates       = 6
T-tree Storage          = 92 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster2.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 5
```

```
Reading input file: ../output/2ndKMeansCluster2.txt.fix
Number of records = 6
Number of columns = 11
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 11
Number of T-tree nodes created = 27
Number of T-tree Updates       = 11
T-tree Storage          = 228 (Bytes)
Number of rules         = 3
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {8}  100.0
(2)  {3}  ->  {9}  100.0
(3)  null  ->  {7}  0.0

SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster3.txt.fix
Support (default 20%)        = 0.1
Confidence (default 80%)     = 90.0
Number of classes            = 4

Reading input file: ../output/2ndKMeansCluster3.txt.fix
Number of records = 12
Number of columns = 4
Min support       = 0.01 (records)
START APRIORI-TFP CBA
```

```
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.01 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 6
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.01 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 3
Number of T-tree nodes created = 7
Number of T-tree Updates       = 3
T-tree Storage          = 44 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster4.txt.fix
Support (default 20%)        = 0.1
Confidence (default 80%)     = 90.0
Number of classes            = 3

Reading input file: ../output/2ndKMeansCluster4.txt.fix
Number of records = 4
Number of columns = 9
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647
```

```
Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 16
Number of T-tree nodes created = 24
Number of T-tree Updates       = 16
T-tree Storage          = 256 (Bytes)
Number of rules         = 2
Accuracy = 50.0
AUC value = 0.1667
(1)  {1}  ->  {8}  100.0
(2)  null  ->  {7}  0.0

SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster5.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 1

Reading input file: ../output/2ndKMeansCluster5.txt.fix
Number of records = 1
Number of columns = 5
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
```

```
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 7
Number of T-tree nodes created = 9
Number of T-tree Updates       = 7
T-tree Storage          = 116 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {5}  100.0
(2)  null  ->  {5}  0.0

SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster6.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 1

Reading input file: ../output/2ndKMeansCluster6.txt.fix
Number of records = 1
Number of columns = 3
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 3
Number of T-tree nodes created = 4
```

Number of T-tree Updates = 3
T-tree Storage = 52 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {3} 100.0
(2) null -> {3} 0.0

SETTINGS
---
Training file name = ../output/2ndKMeansCluster7.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 113

Reading input file: ../output/2ndKMeansCluster7.txt.fix
Number of records = 9237
Number of columns = 281
Min support = 9.24 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 4.62 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4619
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (4.62 records)
Generation time = 2.77 seconds (0.05 mins)
Number of frequent sets = 256
Number of T-tree nodes created = 4931
Number of T-tree Updates = 1584
T-tree Storage = 27400 (Bytes)
Number of rules = 3
Accuracy = 24.88
AUC value = 0.0188
(1) {48} -> {233} 100.0

(2) {43} -> {224} 90.9
(3) null -> {172} 0.0

SETTINGS
---
Training file name = ../output/2ndKMeansCluster8.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 23

Reading input file: ../output/2ndKMeansCluster8.txt.fix
Number of records = 44
Number of columns = 65
Min support = 0.04 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.02 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 22
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.02 records)
Generation time = 0.01 seconds (0.0 mins)
Number of frequent sets = 79
Number of T-tree nodes created = 308
Number of T-tree Updates = 84
T-tree Storage = 2600 (Bytes)
Number of rules = 5
Accuracy = 9.09
AUC value = 0.0445
(1) {4} -> {45} 100.0
(2) {2} -> {47} 100.0
(3) {5} -> {50} 100.0
(4) {9} -> {54} 100.0
(5) null -> {43} 0.0

```
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster9.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/2ndKMeansCluster9.txt.fix
Number of records = 3
Number of columns = 2
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster10.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
```

```
Number of classes            = 3

Reading input file: ../output/2ndKMeansCluster10.txt.fix
Number of records = 5
Number of columns = 5
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 4
Number of T-tree nodes created = 8
Number of T-tree Updates       = 5
T-tree Storage          = 72 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {4}  100.0
(2)  null  ->  {3}  0.0

SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster11.txt.fix
Support (default 20%)        = 0.1
Confidence (default 80%)     = 90.0
Number of classes            = 1

Reading input file: ../output/2ndKMeansCluster11.txt.fix
Number of records = 2
Number of columns = 1
Min support       = 0.0 (records)
```

START APRIORI-TFP CBA
---

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 1
Number of T-tree nodes created = 1
Number of T-tree Updates       = 1
T-tree Storage          = 16 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
---
Training file name            = ../output/2ndKMeansCluster12.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/2ndKMeansCluster12.txt.fix
Number of records = 4
Number of columns = 2
Min support       = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS   = 80000
Max size antecedent  = 2147483647

```
Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED

SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster13.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 1

Reading input file: ../output/2ndKMeansCluster13.txt.fix
Number of records = 1
Number of columns = 1
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
```

```
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 1
Number of T-tree nodes created = 1
Number of T-tree Updates       = 1
T-tree Storage          = 16 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster14.txt.fix
Support (default 20%)         = 0.1
Confidence (default 80%)      = 90.0
Number of classes             = 1

Reading input file: ../output/2ndKMeansCluster14.txt.fix
Number of records = 1
Number of columns = 7
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
```

Number of frequent sets = 15
Number of T-tree nodes created = 18
Number of T-tree Updates = 15
T-tree Storage = 236 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {7} 100.0
(2) null -> {7} 0.0

SETTINGS
---
Training file name = ../output/2ndKMeansCluster15.txt.fix
Support (default 20%) = 0.1
Confidence (default 80%) = 90.0
Number of classes = 3

Reading input file: ../output/2ndKMeansCluster15.txt.fix
Number of records = 3
Number of columns = 7
Min support = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.1, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.1% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 6
Number of T-tree nodes created = 13
Number of T-tree Updates = 6
T-tree Storage = 132 (Bytes)
Number of rules = 2

Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {5} 100.0
(2) null -> {6} 0.0

## Appendix F - Output of 16 clusters, 0.08 support and 90% confidence

SETTINGS
---
Training file name = ../output/2ndKMeansCluster0.txt.fix
Support (default 20%) = 0.08
Confidence (default 80%) = 90.0
Number of classes = 3

Reading input file: ../output/2ndKMeansCluster0.txt.fix
Number of records = 3
Number of columns = 21
Min support = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.03 seconds (0.0 mins)
Number of frequent sets = 1024
Number of T-tree nodes created = 1044
Number of T-tree Updates = 1024
T-tree Storage = 14412 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {19} 100.0

```
(2)  null  ->  {20}  0.0
*********************************
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster1.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 10

Reading input file: ../output/2ndKMeansCluster1.txt.fix
Number of records = 14
Number of columns = 10
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.01 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 7
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.01 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 6
Number of T-tree nodes created = 25
Number of T-tree Updates       = 6
T-tree Storage          = 92 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
*********************************
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster2.txt.fix
```

```
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 5

Reading input file: ../output/2ndKMeansCluster2.txt.fix
Number of records = 6
Number of columns = 11
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 11
Number of T-tree nodes created = 27
Number of T-tree Updates       = 11
T-tree Storage          = 228 (Bytes)
Number of rules         = 3
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {8}  100.0
(2)  {3}  ->  {9}  100.0
(3)  null  ->  {7}  0.0
********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster3.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 4

Reading input file: ../output/2ndKMeansCluster3.txt.fix
```

```
Number of records = 12
Number of columns = 4
Min support       = 0.01 (records)
START APRIORI-TFP CBA
------------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 6
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 3
Number of T-tree nodes created = 7
Number of T-tree Updates       = 3
T-tree Storage          = 44 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
*********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster4.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 3

Reading input file: ../output/2ndKMeansCluster4.txt.fix
Number of records = 4
Number of columns = 9
Min support       = 0.0 (records)
START APRIORI-TFP CBA
------------------------------
```

```
Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 16
Number of T-tree nodes created = 24
Number of T-tree Updates       = 16
T-tree Storage          = 256 (Bytes)
Number of rules         = 2
Accuracy = 50.0
AUC value = 0.1667
(1)  {1}  ->  {8}  100.0
(2)  null  ->  {7}  0.0
*********************************
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster5.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 1

Reading input file: ../output/2ndKMeansCluster5.txt.fix
Number of records = 1
Number of columns = 5
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
```

```
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 7
Number of T-tree nodes created = 9
Number of T-tree Updates       = 7
T-tree Storage          = 116 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {5}  100.0
(2)  null  ->  {5}  0.0
*********************************
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster6.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 1

Reading input file: ../output/2ndKMeansCluster6.txt.fix
Number of records = 1
Number of columns = 3
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
```

Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 3
Number of T-tree nodes created = 4
Number of T-tree Updates = 3
T-tree Storage = 52 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {3} 100.0
(2) null -> {3} 0.0
**********
SETTINGS
---
Training file name = ../output/2ndKMeansCluster7.txt.fix
Support (default 20%) = 0.08
Confidence (default 80%) = 90.0
Number of classes = 113

Reading input file: ../output/2ndKMeansCluster7.txt.fix
Number of records = 9237
Number of columns = 281
Min support = 7.39 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 3.7 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 4619
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (3.7 records)
Generation time = 2.7 seconds (0.05 mins)
Number of frequent sets = 324
Number of T-tree nodes created = 5464
Number of T-tree Updates = 1775
T-tree Storage = 32572 (Bytes)

```
Number of rules         = 4
Accuracy = 24.92
AUC value = 0.0193
(1)  {48}  ->  {233}  100.0
(2)  {1 13 24}  ->  {195}  100.0
(3)  {43}  ->  {224}  90.9
(4)  null  ->  {172}  0.0
********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster8.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 23

Reading input file: ../output/2ndKMeansCluster8.txt.fix
Number of records = 44
Number of columns = 65
Min support       = 0.04 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.02 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 22
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.02 records)
Generation time = 0.01 seconds (0.0 mins)
Number of frequent sets = 79
Number of T-tree nodes created = 308
Number of T-tree Updates       = 84
T-tree Storage          = 2600 (Bytes)
Number of rules         = 5
Accuracy = 9.09
AUC value = 0.0445
(1)  {4}  ->  {45}  100.0
```

(2) {2} -> {47} 100.0
(3) {5} -> {50} 100.0
(4) {9} -> {54} 100.0
(5) null -> {43} 0.0
**********
SETTINGS
---
Training file name = ../output/2ndKMeansCluster9.txt.fix
Support (default 20%) = 0.08
Confidence (default 80%) = 90.0
Number of classes = 2

Reading input file: ../output/2ndKMeansCluster9.txt.fix
Number of records = 3
Number of columns = 2
Min support = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS = 80000
Max size antecedent = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates = 2
T-tree Storage = 28 (Bytes)
Number of rules = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
**********

```
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster10.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 3

Reading input file: ../output/2ndKMeansCluster10.txt.fix
Number of records = 5
Number of columns = 5
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 3
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 4
Number of T-tree nodes created = 8
Number of T-tree Updates       = 5
T-tree Storage          = 72 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {4}  100.0
(2)  null  ->  {3}  0.0
*********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster11.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 1
```

```
Reading input file: ../output/2ndKMeansCluster11.txt.fix
Number of records = 2
Number of columns = 1
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 1
Number of T-tree nodes created = 1
Number of T-tree Updates       = 1
T-tree Storage          = 16 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster12.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 2

Reading input file: ../output/2ndKMeansCluster12.txt.fix
Number of records = 4
Number of columns = 2
Min support       = 0.0 (records)
```

START APRIORI-TFP CBA
---

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 2
Number of T-tree nodes created = 3
Number of T-tree Updates       = 2
T-tree Storage          = 28 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
**********
SETTINGS
---
Training file name           = ../output/2ndKMeansCluster13.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 1

Reading input file: ../output/2ndKMeansCluster13.txt.fix
Number of records = 1
Number of columns = 1
Min support       = 0.0 (records)
START APRIORI-TFP CBA
---

Max number of CARS   = 80000
Max size antecedent  = 2147483647

```
Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
NO RULES GENERATED
No classification rules generated!
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 1
Number of T-tree nodes created = 1
Number of T-tree Updates       = 1
T-tree Storage          = 16 (Bytes)
Number of rules         = 0
Accuracy = 0.0
AUC value = 0.0
NO RULES GENERATED
**********
SETTINGS
--------
Training file name           = ../output/2ndKMeansCluster14.txt.fix
Support (default 20%)        = 0.08
Confidence (default 80%)     = 90.0
Number of classes            = 1

Reading input file: ../output/2ndKMeansCluster14.txt.fix
Number of records = 1
Number of columns = 7
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
```

```
Number of records in training set = 1
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
Number of frequent sets = 15
Number of T-tree nodes created = 18
Number of T-tree Updates       = 15
T-tree Storage          = 236 (Bytes)
Number of rules         = 2
Accuracy = 0.0
AUC value = 0.0
(1)  {1}  ->  {7}  100.0
(2)  null  ->  {7}  0.0
*********************************
SETTINGS
--------
Training file name            = ../output/2ndKMeansCluster15.txt.fix
Support (default 20%)         = 0.08
Confidence (default 80%)      = 90.0
Number of classes             = 3

Reading input file: ../output/2ndKMeansCluster15.txt.fix
Number of records = 3
Number of columns = 7
Min support       = 0.0 (records)
START APRIORI-TFP CBA
-----------------------------

Max number of CARS   = 80000
Max size antecedent  = 2147483647

Support = 0.08, Confidence = 90.0
Minimum support = 0.0 (Records)
Max num frequent sets  = 2147483647
Max size of antecedent = 2147483647
Number of records in training set = 2
NOTE: Data set reordered
Creating P-tree table
Apriori-TFP with X-Checking
Minimum support threshold = 0.08% (0.0 records)
Generation time = 0.0 seconds (0.0 mins)
```

Number of frequent sets = 6
Number of T-tree nodes created = 13
Number of T-tree Updates = 6
T-tree Storage = 132 (Bytes)
Number of rules = 2
Accuracy = 0.0
AUC value = 0.0
(1) {1} -> {5} 100.0
(2) null -> {6} 0.0