Assignment 1
Gaurav Singh, 2957104
Mice Protein Expression Data Set

The data set consists of the expression levels of 77 proteins/protein modifications that produced detectable signals in the nuclear fraction of cortex. There are 38 control mice and 34 trisomic mice (Down syndrome), for a total of 72 mice.
In the experiments, 15 measurements were registered of each protein per sample/mouse.
Therefore, for control mice, there are 38x15, or 570 measurements, and for trisomic mice, there are 34x15, or 510 measurements. The dataset contains a total of 1080 measurements per protein. Each measurement can be considered as an independent sample/mouse.
The eight classes of mice are described based on features such as genotype, behaviour and treatment. According to genotype, mice can be control or trisomic. According to behaviour, some mice have been stimulated to learn (context-shock) and others have not (shock-context) and to assess the effect of the drug memantine in recovering the ability to learn in trisomic mice, some mice have been injected with the drug and others have not.

Classes:
c-CS-s: control mice, stimulated to learn, injected with saline (9 mice)
c-CS-m: control mice, stimulated to learn, injected with memantine (10 mice)
c-SC-s: control mice, not stimulated to learn, injected with saline (9 mice)
c-SC-m: control mice, not stimulated to learn, injected with memantine (10 mice)

t-CS-s: trisomy mice, stimulated to learn, injected with saline (7 mice)
t-CS-m: trisomy mice, stimulated to learn, injected with memantine (9 mice)
t-SC-s: trisomy mice, not stimulated to learn, injected with saline (9 mice)
t-SC-m: trisomy mice, not stimulated to learn, injected with memantine (9 mice)

Objective

The aim is to identify subsets of proteins that are discriminant between the classes.

The Attribute types of the data
Mouse ID
78 Values of expression levels of 77 proteins; the names of proteins are followed by indicating that they were measured in the nuclear fraction.
79 Genotype: control (c) or trisomy (t)
80 Treatment type: memantine (m) or saline (s)
81 Behaviour: context-shock (CS) or shock-context (SC)
82 Class: c-CS-s, c-CS-m, c-SC-s, c-SC-m, t-CS-s, t-CS-m, t-SC-s, t-SC-m

Source:

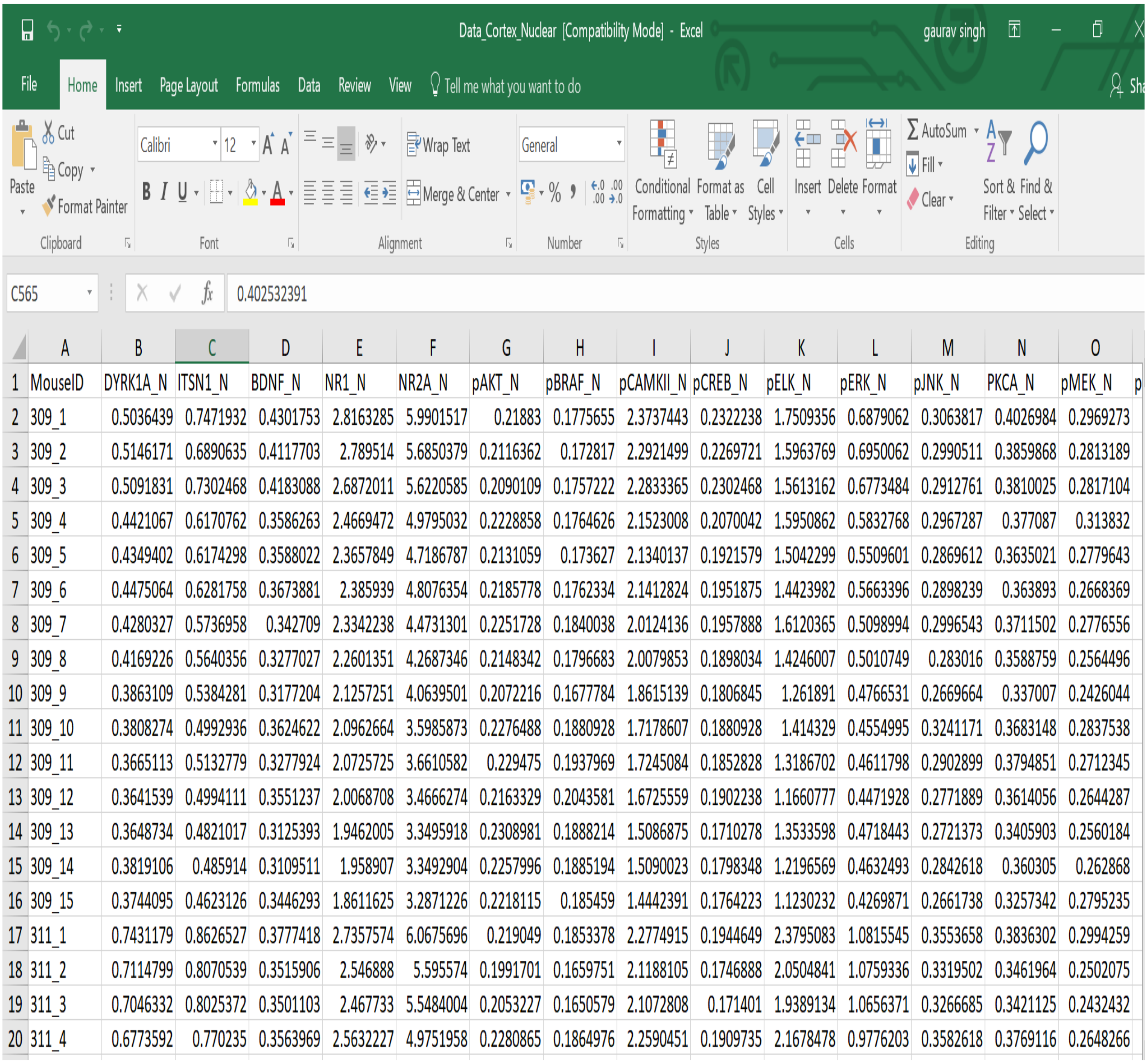
Clara Higuera Department of Software Engineering and Artificial Intelligence, Faculty of Informatics and the Department of Biochemistry and Molecular Biology, Faculty of Chemistry, University Complutense, Madrid, Spain.
Email: clarahiguera@cm.es

Katheleen J. Gardiner, creator and owner of the protein expression data, is currently with the Linda Crnic Institute for Down Syndrome, Department of Pediatrics, Department of Biochemistry and Molecular Genetics, Human Medical Genetics and Genomics, and Neuroscience Programs, University of Colorado, School of Medicine, Aurora, Colorado, USA.
Email: katheleen.gardiner@ucdenver.edu

Krzysztof J. Cios is currently with the Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, USA, and IITiS Polish Academy of Sciences, Poland.
Email: kcios@vcu.edu

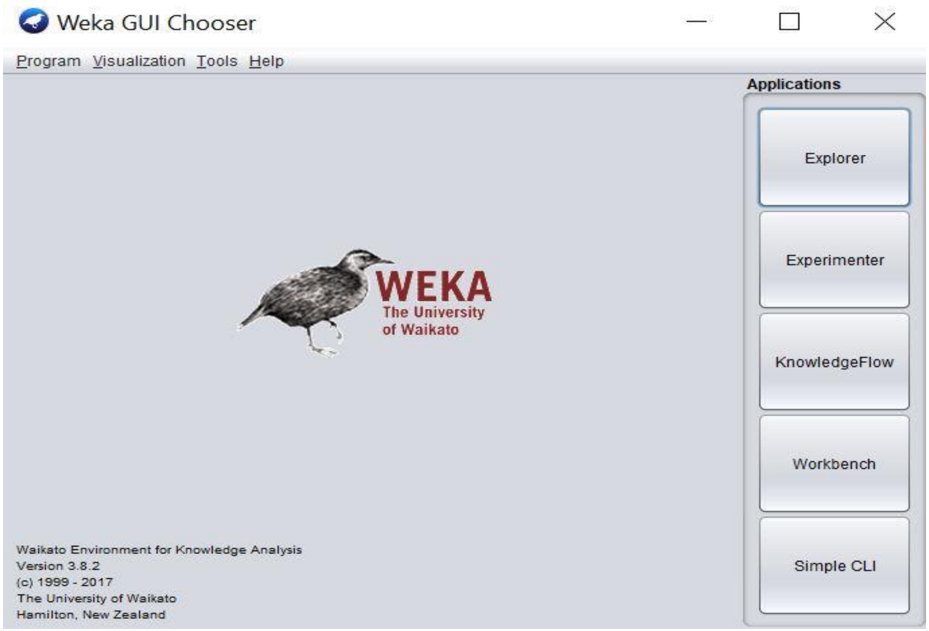
Relevant Papers:

- Higuera C, Gardiner KJ, Cios KJ (2015) Self-Organizing Feature Maps Identify Proteins Critical to Learning in a Mouse Model of Down Syndrome.
- Ahmed MM, Dhanasekaran AR, Block A, Tong S, Costa ACS, Stasko M, et al. (2015) Protein Dynamics Associated with Failed and Rescued Learning in the Ts65Dn Mouse Model of Down Syndrome.



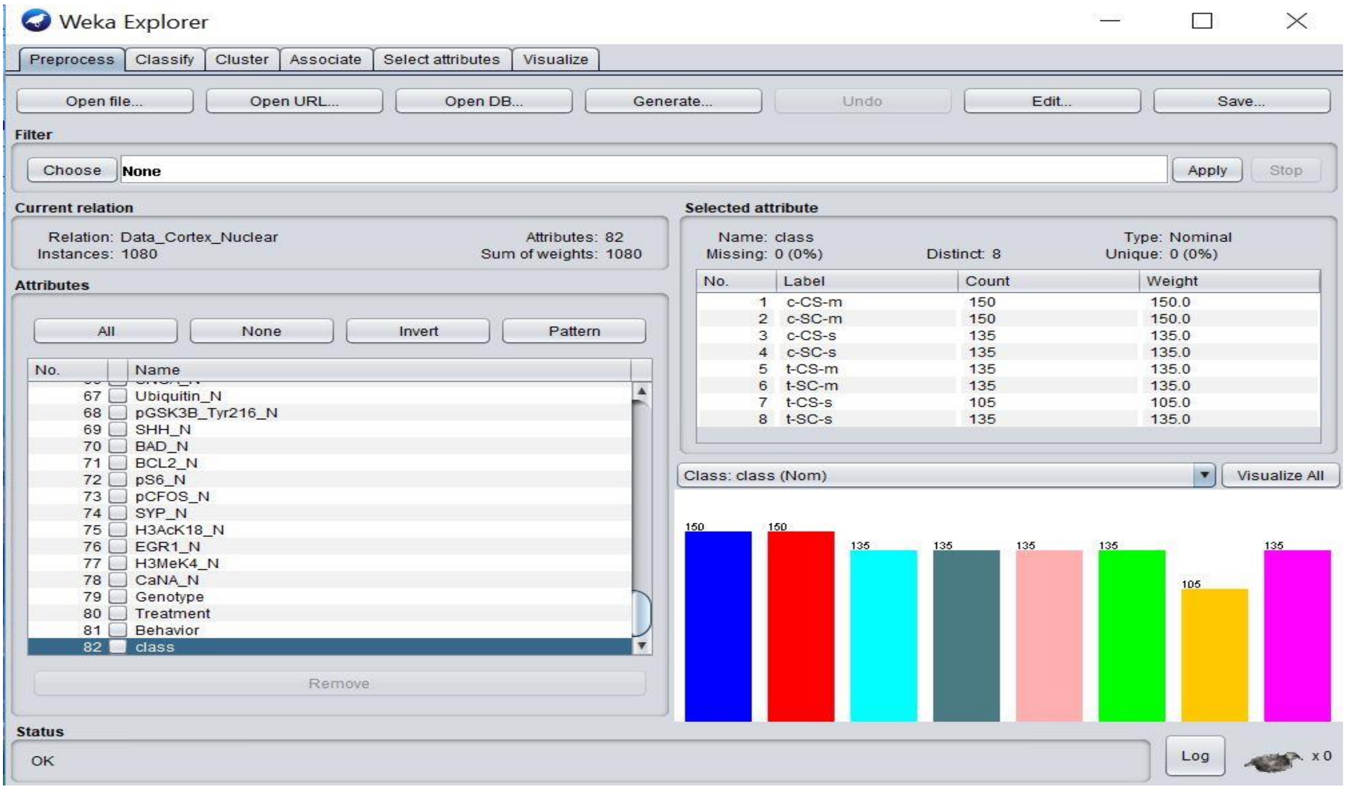
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	MouseID	DYRK1A_N	ITSN1_N	BDNF_N	NR1_N	NR2A_N	pAKT_N	pBRAF_N	pCAMKII_N	pCREB_N	pELK_N	pERK_N	pJNK_N	PKCA_N	pMEK_N
2	309_1	0.5036439	0.7471932	0.4301753	2.8163285	5.9901517	0.21883	0.1775655	2.3737443	0.2322238	1.7509356	0.6879062	0.3063817	0.4026984	0.2969273
3	309_2	0.5146171	0.6890635	0.4117703	2.789514	5.6850379	0.2116362	0.172817	2.2921499	0.2269721	1.5963769	0.6950062	0.2990511	0.3859868	0.2813189
4	309_3	0.5091831	0.7302468	0.4183088	2.6872011	5.6220585	0.2090109	0.1757222	2.2833365	0.2302468	1.5613162	0.6773484	0.2912761	0.3810025	0.2817104
5	309_4	0.4421067	0.6170762	0.3586263	2.4669472	4.9795032	0.2228858	0.1764626	2.1523008	0.2070042	1.5950862	0.5832768	0.2967287	0.377087	0.313832
6	309_5	0.4349402	0.6174298	0.3588022	2.3657849	4.7186787	0.2131059	0.173627	2.1340137	0.1921579	1.5042299	0.5509601	0.2869612	0.3635021	0.2779643
7	309_6	0.4475064	0.6281758	0.3673881	2.385939	4.8076354	0.2185778	0.1762334	2.1412824	0.1951875	1.4423982	0.5663396	0.2898239	0.363893	0.2668369
8	309_7	0.4280327	0.5736958	0.342709	2.3342238	4.4731301	0.2251728	0.1840038	2.0124136	0.1957888	1.6120365	0.5098994	0.2996543	0.3711502	0.2776556
9	309_8	0.4169226	0.5640356	0.3277027	2.2601351	4.2687346	0.2148342	0.1796683	2.0079853	0.1898034	1.4246007	0.5010749	0.283016	0.3588759	0.2564496
10	309_9	0.3863109	0.5384281	0.3177204	2.1257251	4.0639501	0.2072216	0.1677784	1.8615139	0.1806845	1.261891	0.4766531	0.2669664	0.337007	0.2426044
11	309_10	0.3808274	0.4992936	0.3624622	2.0962664	3.5985873	0.2276488	0.1880928	1.7178607	0.1880928	1.414329	0.4554995	0.3241171	0.3683148	0.2837538
12	309_11	0.3665113	0.5132779	0.3277924	2.0725725	3.6610582	0.229475	0.1937969	1.7245084	0.1852828	1.3186702	0.4611798	0.2902899	0.3794851	0.2712345
13	309_12	0.3641539	0.4994111	0.3551237	2.0068708	3.4666274	0.2163329	0.2043581	1.6725559	0.1902238	1.1660777	0.4471928	0.2771889	0.3614056	0.2644287
14	309_13	0.3648734	0.4821017	0.3125393	1.9462005	3.3495918	0.2308981	0.1888214	1.5086875	0.1710278	1.3533598	0.4718443	0.2721373	0.3405903	0.2560184
15	309_14	0.3819106	0.485914	0.3109511	1.958907	3.3492904	0.2257996	0.1885194	1.5090023	0.1798348	1.2196569	0.4632493	0.2842618	0.360305	0.262868
16	309_15	0.3744095	0.4623126	0.3446293	1.8611625	3.2871226	0.2218115	0.185459	1.4442391	0.1764223	1.1230232	0.4269871	0.2661738	0.3257342	0.2795235
17	311_1	0.7431179	0.8626527	0.3777418	2.7357574	6.0675696	0.219049	0.1853378	2.2774915	0.1944649	2.3795083	1.0815545	0.3553658	0.3836302	0.2994259
18	311_2	0.7114799	0.8070539	0.3515906	2.546888	5.595574	0.1991701	0.1659751	2.1188105	0.1746888	2.0504841	1.0759336	0.3319502	0.3461964	0.2502075
19	311_3	0.7046332	0.8025372	0.3501103	2.467733	5.5484004	0.2053227	0.1650579	2.1072808	0.171401	1.9389134	1.0656371	0.3266685	0.3421125	0.2432432
20	311_4	0.6773592	0.770235	0.3563969	2.5632227	4.9751958	0.2280865	0.1864976	2.2590451	0.1909735	2.1678478	0.9776203	0.3582618	0.3769116	0.2648266

Launch WEKA application and click on the *'Explorer'* button.

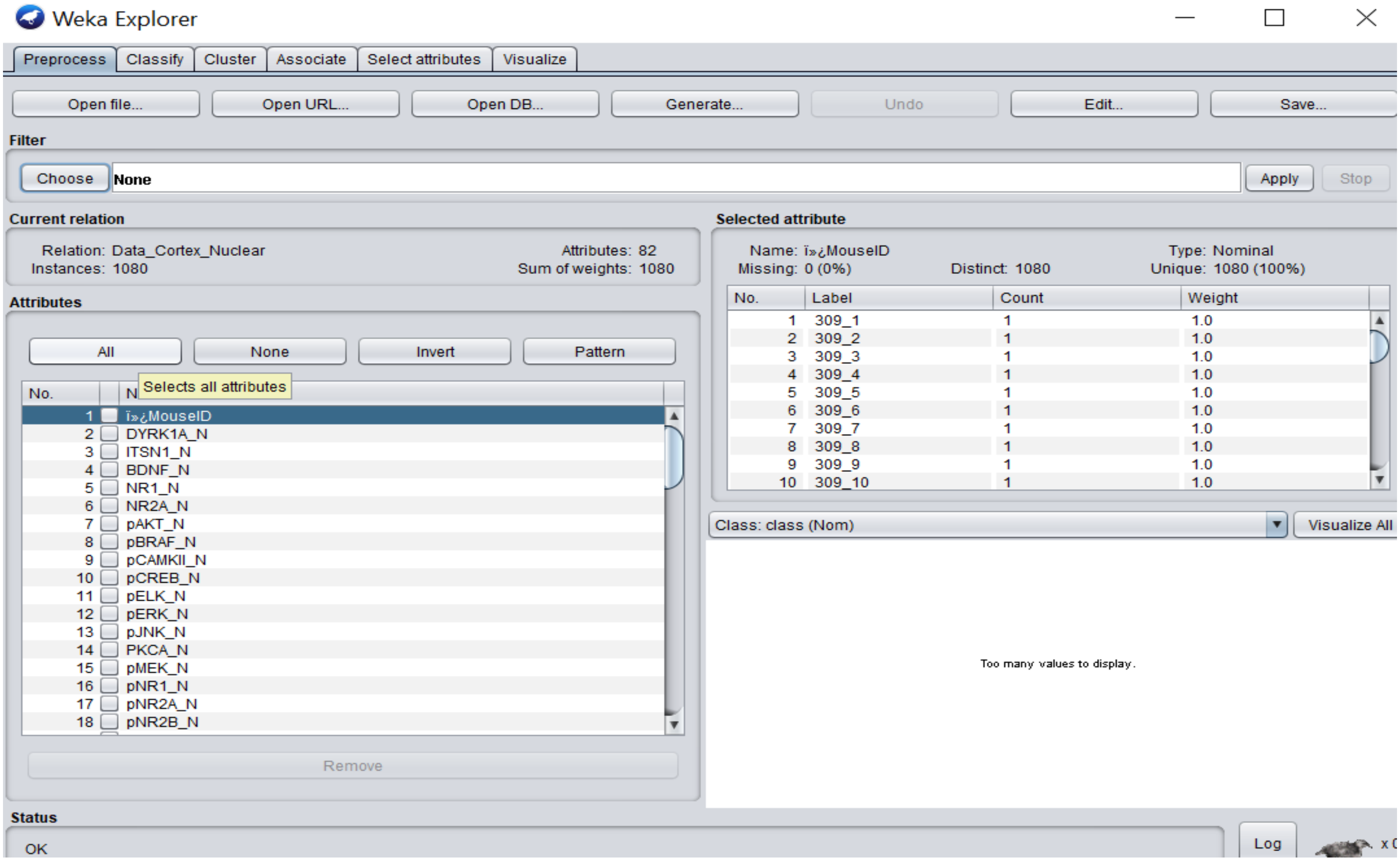


Then load the dataset into WEKA with Open file button and navigate to the directory containing the dataset file and perform the Pre-processing filters.

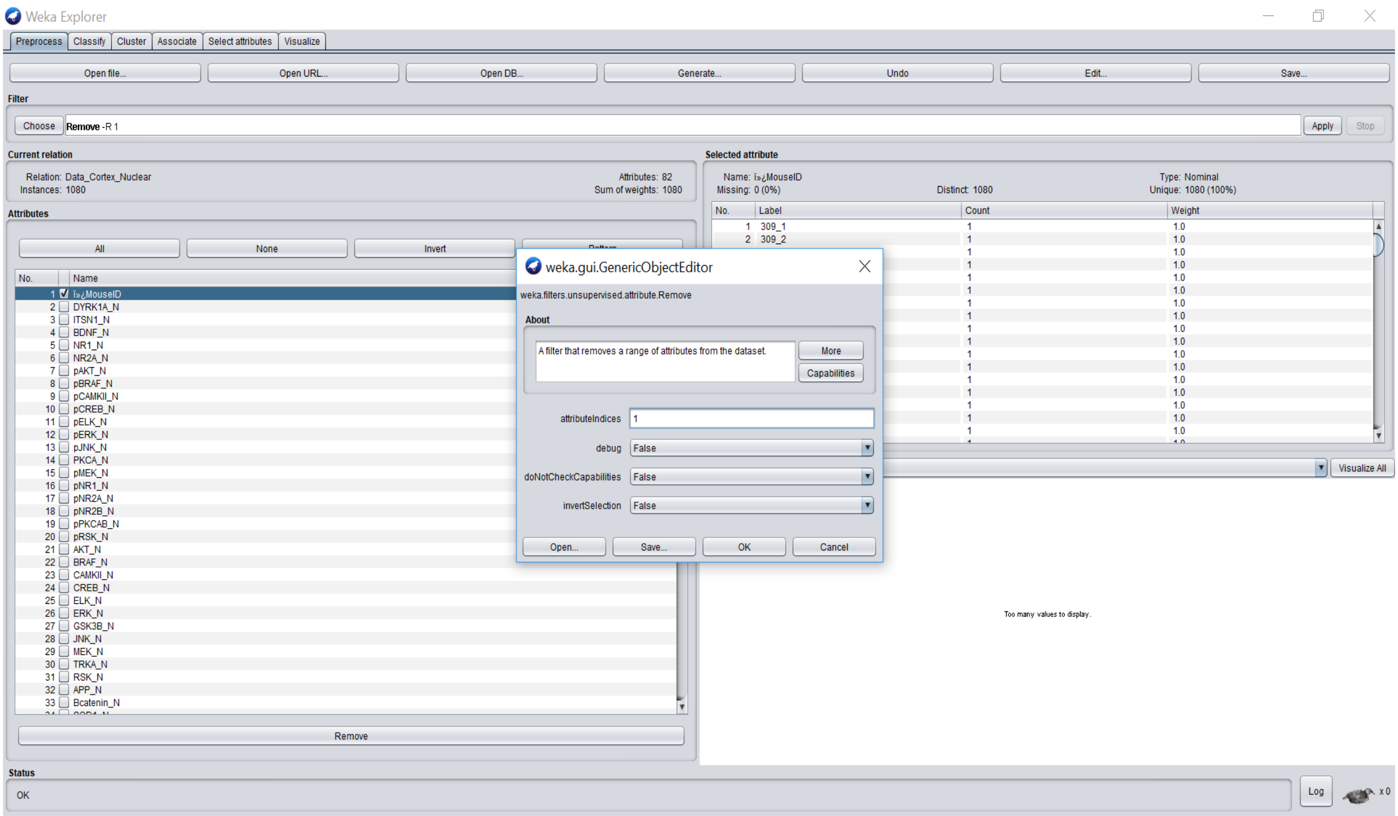
Pre-processing:

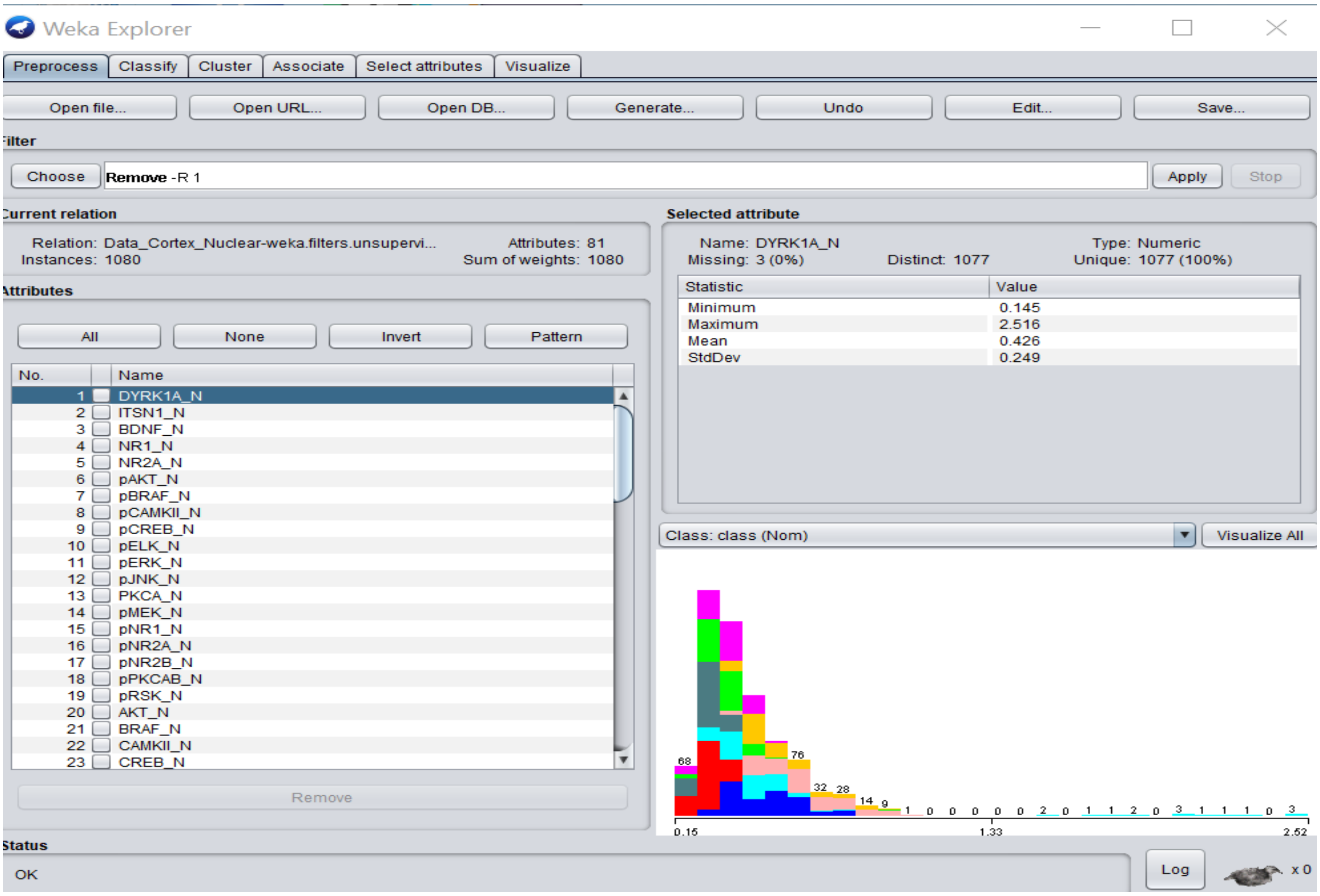


As we can see below the 1st attribute as mouseID need to remove for data mining step by using the Attribute filters in WEKA, which can be done by in the "Filter" panel, click on the "Choose" button and select "weka.filters.unsupervised.attribute.Remove" filter.

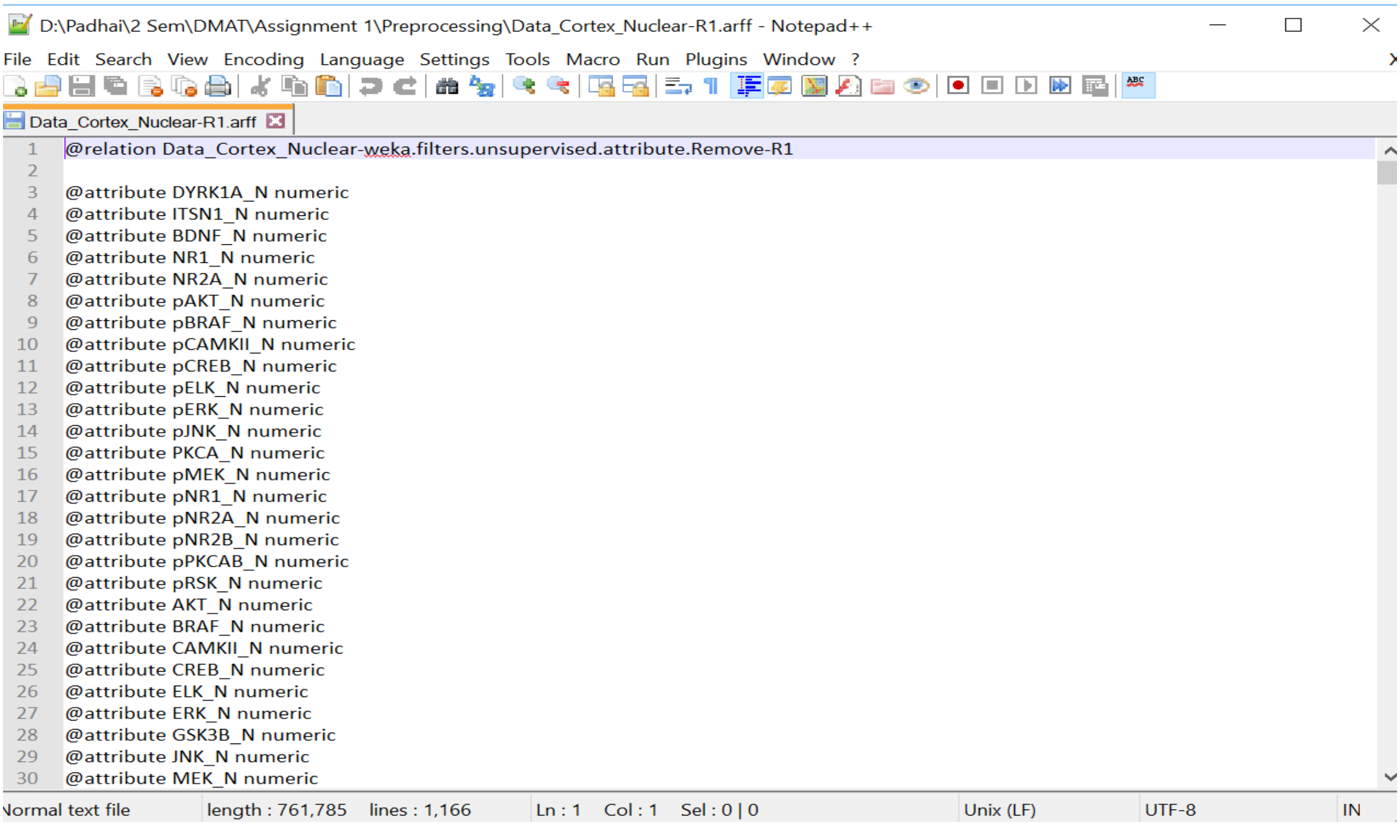


Click on text box immediately to the right of the *"Choose"* button. In the resulting dialog box enter the index of the attribute to be filtered out. In this case, we enter 1 which is the index of the *"id"* attribute. Make sure that the *"invertSelection"* option is set to false. Then click *"OK"*. Now, in the filter box you will see *"Remove -R 1"* and then click on Apply.





Then save as ***Data_Cortex_Nuclear-R1.arff*** and open the file in text editor.

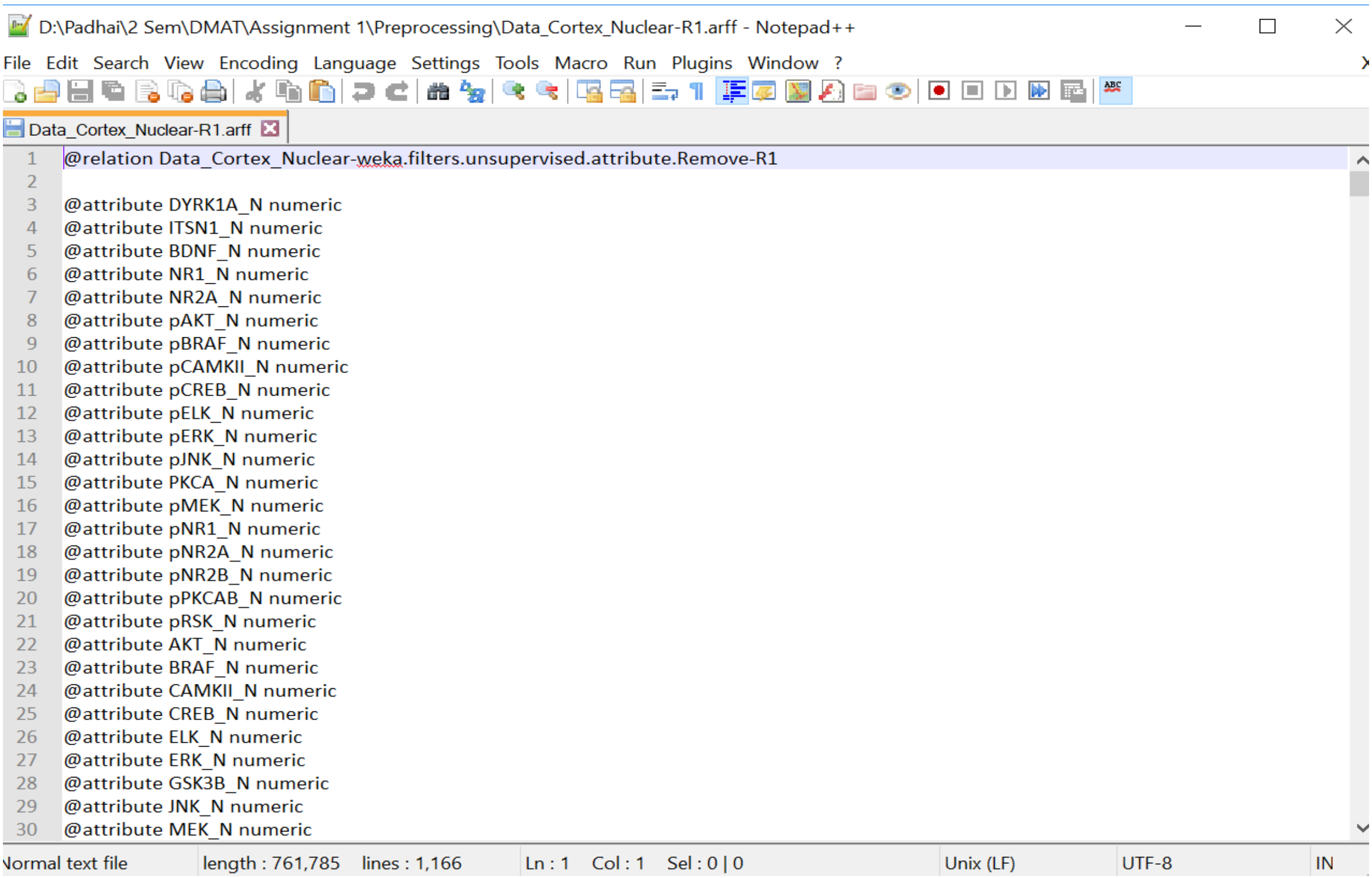



```
@relation Data_Cortex_Nuclear-weka.filters.unsupervised.attribute.Remove-R1

@attribute DYRK1A_N numeric
@attribute ITSN1_N numeric
@attribute BDNF_N numeric
@attribute NR1_N numeric
@attribute NR2A_N numeric
@attribute pAKT_N numeric
@attribute pBRAF_N numeric
@attribute pCAMKII_N numeric
@attribute pCREB_N numeric
@attribute pELK_N numeric
@attribute pERK_N numeric
@attribute pJNK_N numeric
@attribute PKCA_N numeric
@attribute pMEK_N numeric
@attribute pNR1_N numeric
@attribute pNR2A_N numeric
@attribute pNR2B_N numeric
@attribute pPKCAB_N numeric
@attribute pRSK_N numeric
@attribute AKT_N numeric
@attribute BRAF_N numeric
@attribute CAMKII_N numeric
@attribute CREB_N numeric
@attribute ELK_N numeric
@attribute ERK_N numeric
@attribute GSK3B_N numeric
@attribute JNK_N numeric
@attribute MEK_N numeric
```

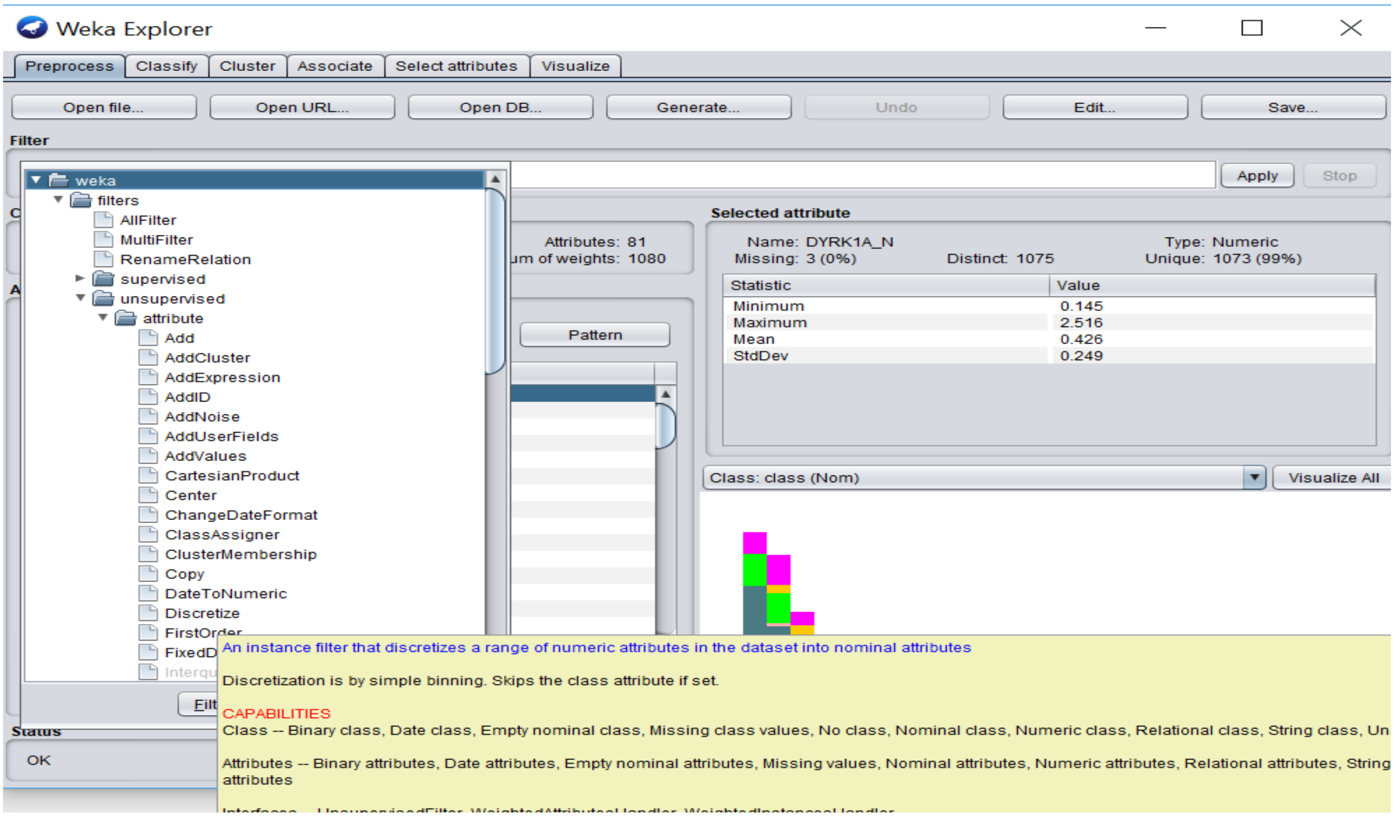
As in the new data set, the ***"id"*** attribute and all the corresponding values in the records have been removed.

Discretization

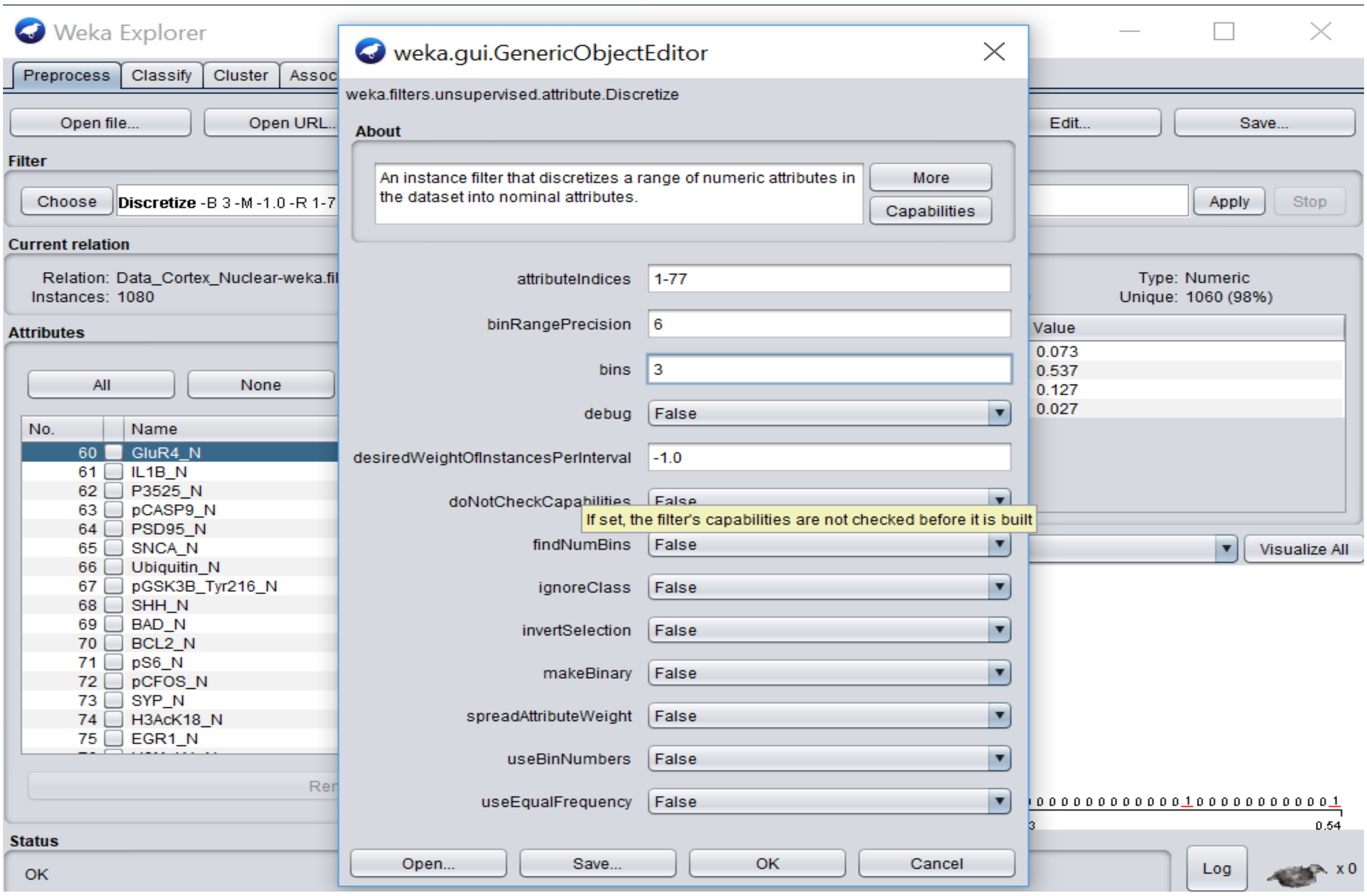
To perform advance techniques such as association rule mining data must be discretized, load the filtered data set into WEKA. There are 77 such attributes in this dataset.



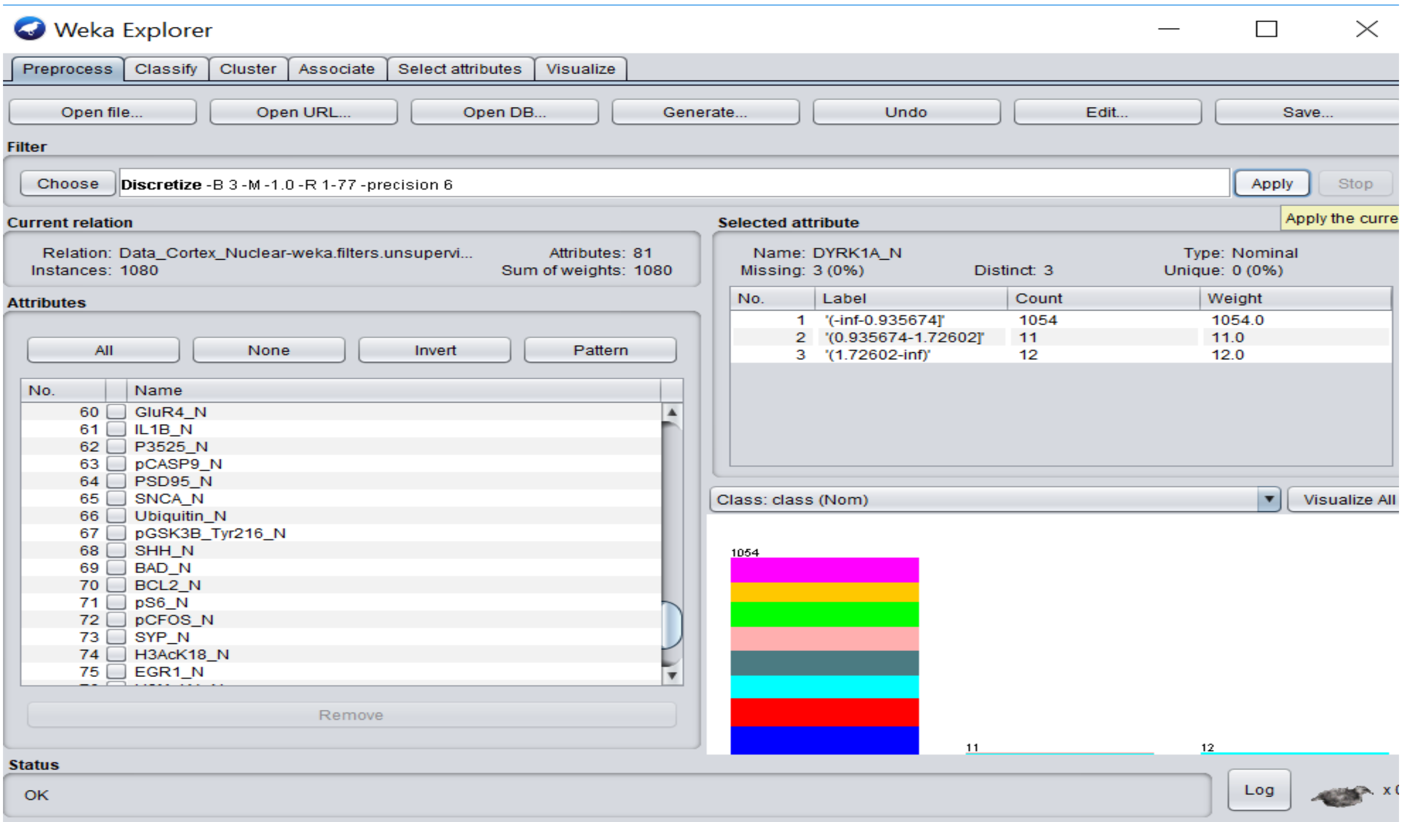
Now to perform Discretization in WEKA, open the filtered data into Weka and select Discretize in Filter.



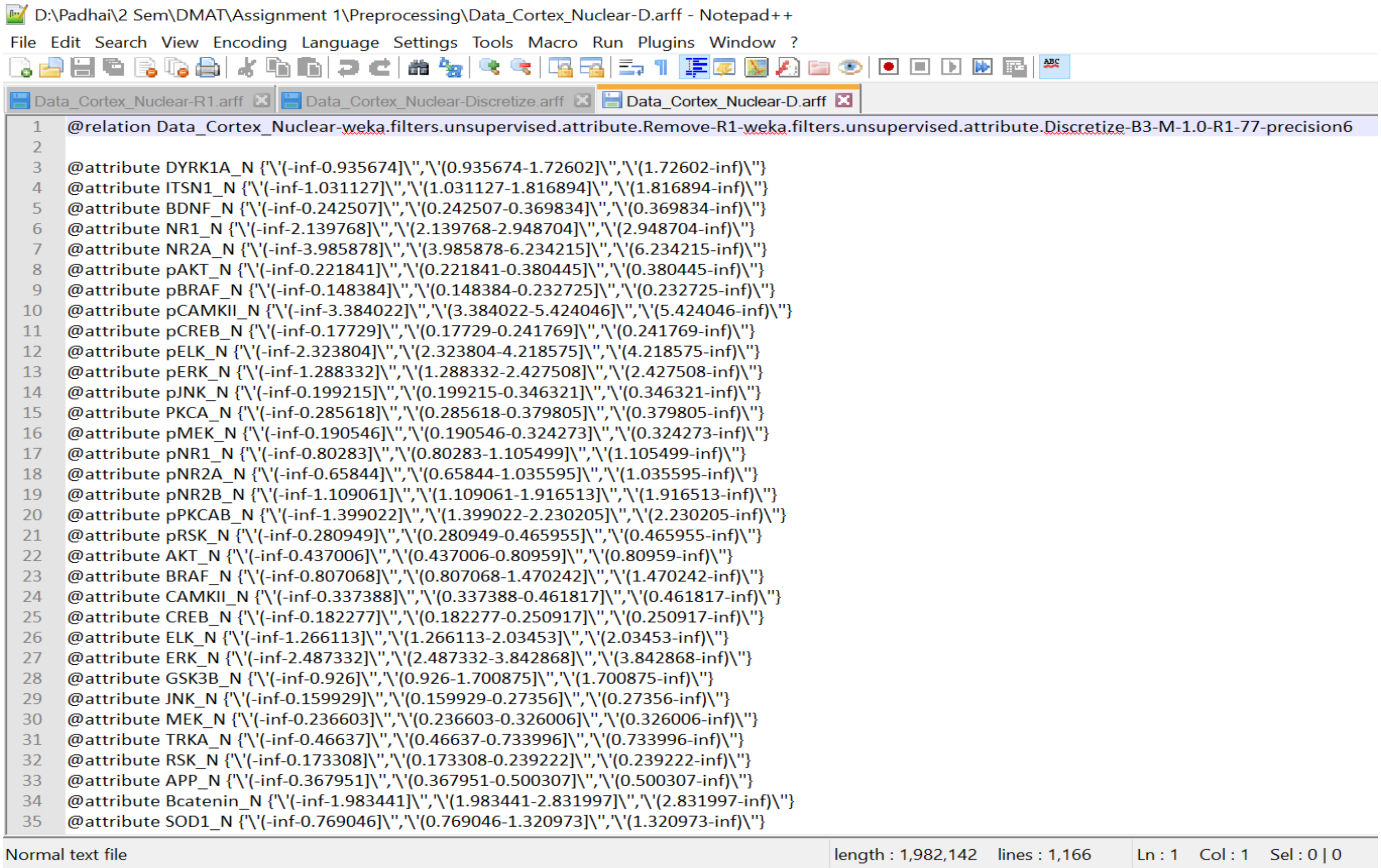
Next, click on the textbox right to the *"Choose"* button and add 1 to 77 and click on OK button.



Then click on Apply button.



And then save the file and open into the text editor.



D:\Padhai\2 Sem\DMAT\Assignment 1\Preprocessing\Data_Cortex_Nuclear-D.arff - Notepad++

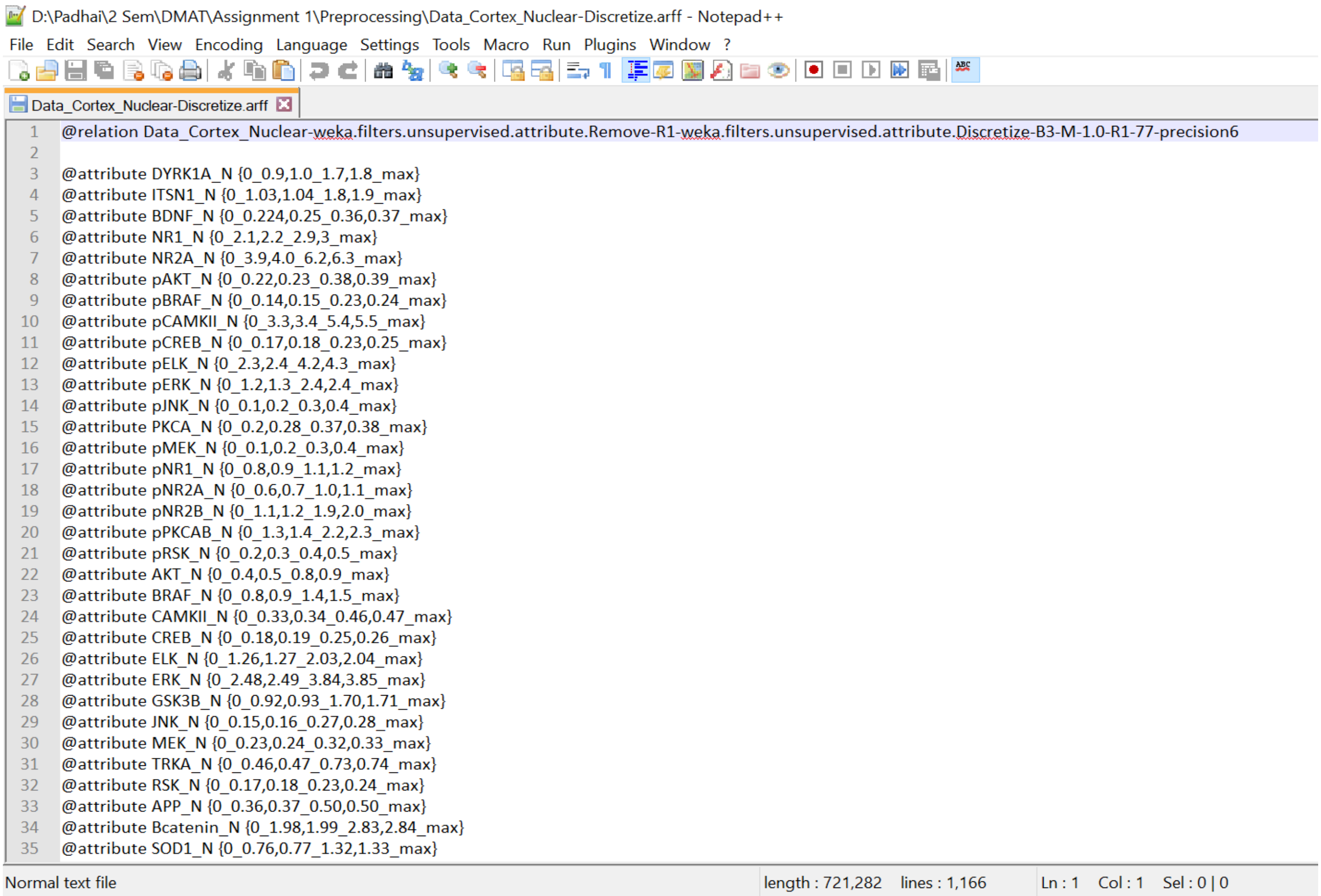
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

Data_Cortex_Nuclear-R1.arff | Data_Cortex_Nuclear-Discretize.arff | Data_Cortex_Nuclear-D.arff

```
@relation Data_Cortex_Nuclear-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1-77-precision6

@attribute DYRK1A_N {'\'(-inf-0.935674]\'','\'(0.935674-1.72602]\'','\'(1.72602-inf)\''}
@attribute ITSN1_N {'\'(-inf-1.031127]\'','\'(1.031127-1.816894]\'','\'(1.816894-inf)\''}
@attribute BDNF_N {'\'(-inf-0.242507]\'','\'(0.242507-0.369834]\'','\'(0.369834-inf)\''}
@attribute NR1_N {'\'(-inf-2.139768]\'','\'(2.139768-2.948704]\'','\'(2.948704-inf)\''}
@attribute NR2A_N {'\'(-inf-3.985878]\'','\'(3.985878-6.234215]\'','\'(6.234215-inf)\''}
@attribute pAKT_N {'\'(-inf-0.221841]\'','\'(0.221841-0.380445]\'','\'(0.380445-inf)\''}
@attribute pBRAF_N {'\'(-inf-0.148384]\'','\'(0.148384-0.232725]\'','\'(0.232725-inf)\''}
@attribute pCAMKII_N {'\'(-inf-3.384022]\'','\'(3.384022-5.424046]\'','\'(5.424046-inf)\''}
@attribute pCREB_N {'\'(-inf-0.17729]\'','\'(0.17729-0.241769]\'','\'(0.241769-inf)\''}
@attribute pELK_N {'\'(-inf-2.323804]\'','\'(2.323804-4.218575]\'','\'(4.218575-inf)\''}
@attribute pERK_N {'\'(-inf-1.288332]\'','\'(1.288332-2.427508]\'','\'(2.427508-inf)\''}
@attribute pJNK_N {'\'(-inf-0.199215]\'','\'(0.199215-0.346321]\'','\'(0.346321-inf)\''}
@attribute PKCA_N {'\'(-inf-0.285618]\'','\'(0.285618-0.379805]\'','\'(0.379805-inf)\''}
@attribute pMEK_N {'\'(-inf-0.190546]\'','\'(0.190546-0.324273]\'','\'(0.324273-inf)\''}
@attribute pNR1_N {'\'(-inf-0.80283]\'','\'(0.80283-1.105499]\'','\'(1.105499-inf)\''}
@attribute pNR2A_N {'\'(-inf-0.65844]\'','\'(0.65844-1.035595]\'','\'(1.035595-inf)\''}
@attribute pNR2B_N {'\'(-inf-1.109061]\'','\'(1.109061-1.916513]\'','\'(1.916513-inf)\''}
@attribute pPKCAB_N {'\'(-inf-1.399022]\'','\'(1.399022-2.230205]\'','\'(2.230205-inf)\''}
@attribute pRSK_N {'\'(-inf-0.280949]\'','\'(0.280949-0.465955]\'','\'(0.465955-inf)\''}
@attribute AKT_N {'\'(-inf-0.437006]\'','\'(0.437006-0.80959]\'','\'(0.80959-inf)\''}
@attribute BRAF_N {'\'(-inf-0.807068]\'','\'(0.807068-1.470242]\'','\'(1.470242-inf)\''}
@attribute CAMKII_N {'\'(-inf-0.337388]\'','\'(0.337388-0.461817]\'','\'(0.461817-inf)\''}
@attribute CREB_N {'\'(-inf-0.182277]\'','\'(0.182277-0.250917]\'','\'(0.250917-inf)\''}
@attribute ELK_N {'\'(-inf-1.266113]\'','\'(1.266113-2.03453]\'','\'(2.03453-inf)\''}
@attribute ERK_N {'\'(-inf-2.487332]\'','\'(2.487332-3.842868]\'','\'(3.842868-inf)\''}
@attribute GSK3B_N {'\'(-inf-0.926]\'','\'(0.926-1.700875]\'','\'(1.700875-inf)\''}
@attribute JNK_N {'\'(-inf-0.159929]\'','\'(0.159929-0.27356]\'','\'(0.27356-inf)\''}
@attribute MEK_N {'\'(-inf-0.236603]\'','\'(0.236603-0.326006]\'','\'(0.326006-inf)\''}
@attribute TRKA_N {'\'(-inf-0.46637]\'','\'(0.46637-0.733996]\'','\'(0.733996-inf)\''}
@attribute RSK_N {'\'(-inf-0.173308]\'','\'(0.173308-0.239222]\'','\'(0.239222-inf)\''}
@attribute APP_N {'\'(-inf-0.367951]\'','\'(0.367951-0.500307]\'','\'(0.500307-inf)\''}
@attribute Bcatenin_N {'\'(-inf-1.983441]\'','\'(1.983441-2.831997]\'','\'(2.831997-inf)\''}
@attribute SOD1_N {'\'(-inf-0.769046]\'','\'(0.769046-1.320973]\'','\'(1.320973-inf)\''}
```

Normal text file length : 1,982,142 lines : 1,166 Ln : 1 Col : 1 Sel : 0 | 0

Now filter the data for example, the lower range "(-inf-0.935674)" and so on need to replace manually with more succinct and readable ones.



D:\Padhai\2 Sem\DMAT\Assignment 1\Preprocessing\Data_Cortex_Nuclear-Discretize.arff - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

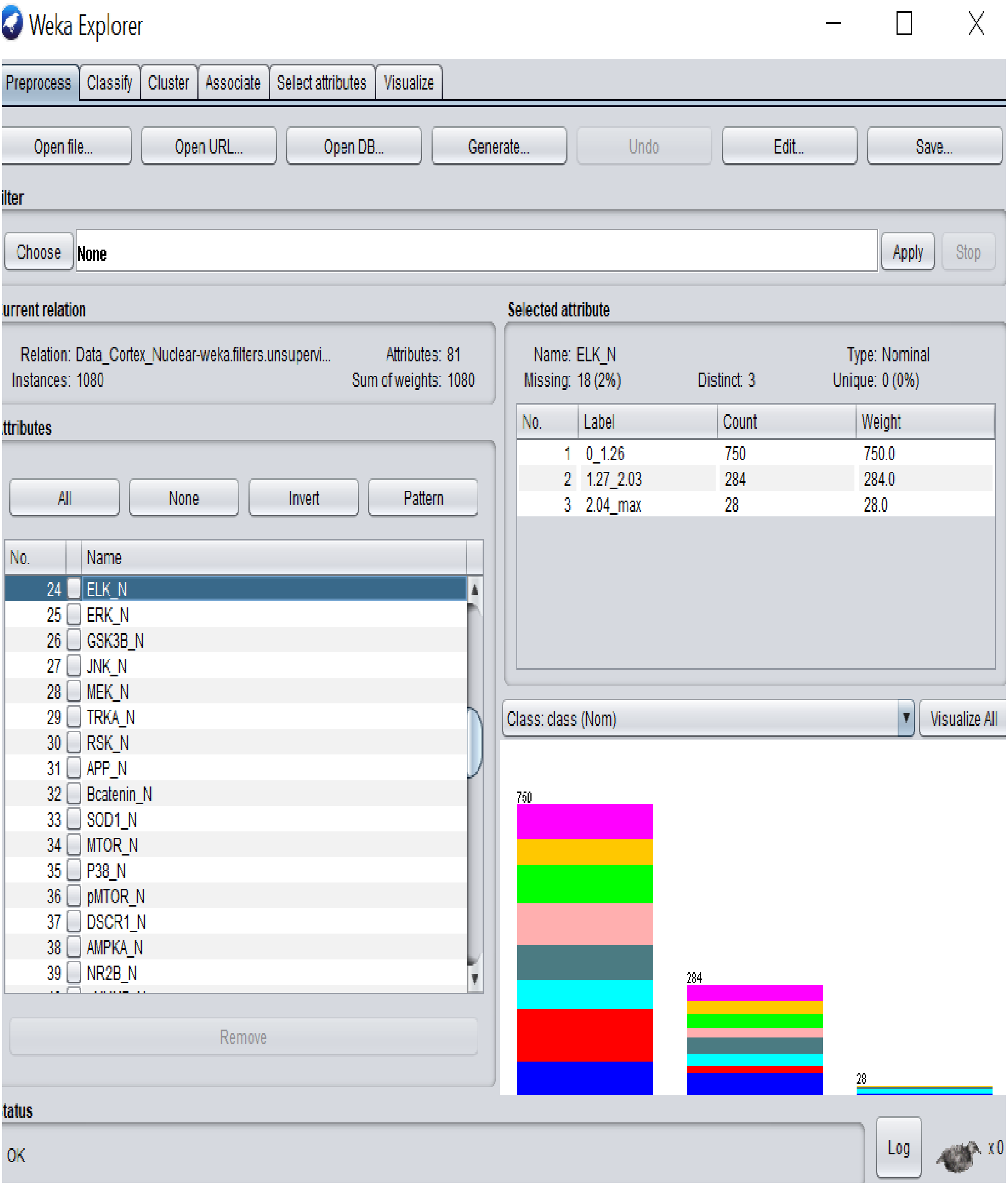
Data_Cortex_Nuclear-Discretize.arff

```
@relation Data_Cortex_Nuclear-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1-77-precision6

@attribute DYRK1A_N {0_0.9,1.0_1.7,1.8_max}
@attribute ITSN1_N {0_1.03,1.04_1.8,1.9_max}
@attribute BDNF_N {0_0.224,0.25_0.36,0.37_max}
@attribute NR1_N {0_2.1,2.2_2.9,3_max}
@attribute NR2A_N {0_3.9,4.0_6.2,6.3_max}
@attribute pAKT_N {0_0.22,0.23_0.38,0.39_max}
@attribute pBRAF_N {0_0.14,0.15_0.23,0.24_max}
@attribute pCAMKII_N {0_3.3,3.4_5.4,5.5_max}
@attribute pCREB_N {0_0.17,0.18_0.23,0.25_max}
@attribute pELK_N {0_2.3,2.4_4.2,4.3_max}
@attribute pERK_N {0_1.2,1.3_2.4,2.4_max}
@attribute pJNK_N {0_0.1,0.2_0.3,0.4_max}
@attribute PKCA_N {0_0.2,0.28_0.37,0.38_max}
@attribute pMEK_N {0_0.1,0.2_0.3,0.4_max}
@attribute pNR1_N {0_0.8,0.9_1.1,1.2_max}
@attribute pNR2A_N {0_0.6,0.7_1.0,1.1_max}
@attribute pNR2B_N {0_1.1,1.2_1.9,2.0_max}
@attribute pPKCAB_N {0_1.3,1.4_2.2,2.3_max}
@attribute pRSK_N {0_0.2,0.3_0.4,0.5_max}
@attribute AKT_N {0_0.4,0.5_0.8,0.9_max}
@attribute BRAF_N {0_0.8,0.9_1.4,1.5_max}
@attribute CAMKII_N {0_0.33,0.34_0.46,0.47_max}
@attribute CREB_N {0_0.18,0.19_0.25,0.26_max}
@attribute ELK_N {0_1.26,1.27_2.03,2.04_max}
@attribute ERK_N {0_2.48,2.49_3.84,3.85_max}
@attribute GSK3B_N {0_0.92,0.93_1.70,1.71_max}
@attribute JNK_N {0_0.15,0.16_0.27,0.28_max}
@attribute MEK_N {0_0.23,0.24_0.32,0.33_max}
@attribute TRKA_N {0_0.46,0.47_0.73,0.74_max}
@attribute RSK_N {0_0.17,0.18_0.23,0.24_max}
@attribute APP_N {0_0.36,0.37_0.50,0.50_max}
@attribute Bcatenin_N {0_1.98,1.99_2.83,2.84_max}
@attribute SOD1_N {0_0.76,0.77_1.32,1.33_max}
```

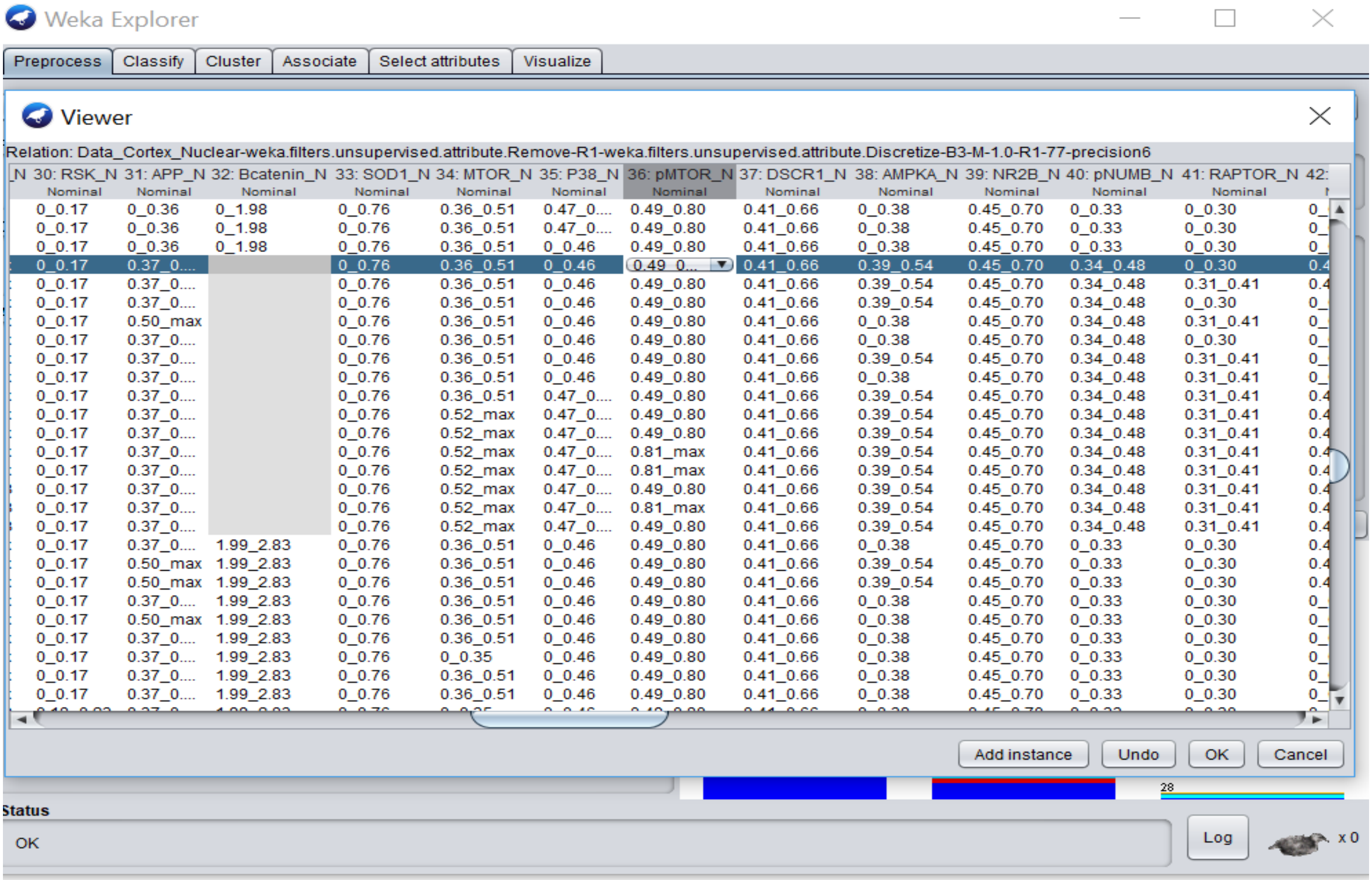
Normal text file length : 721,282 lines : 1,166 Ln : 1 Col : 1 Sel : 0 | 0

Missing Values

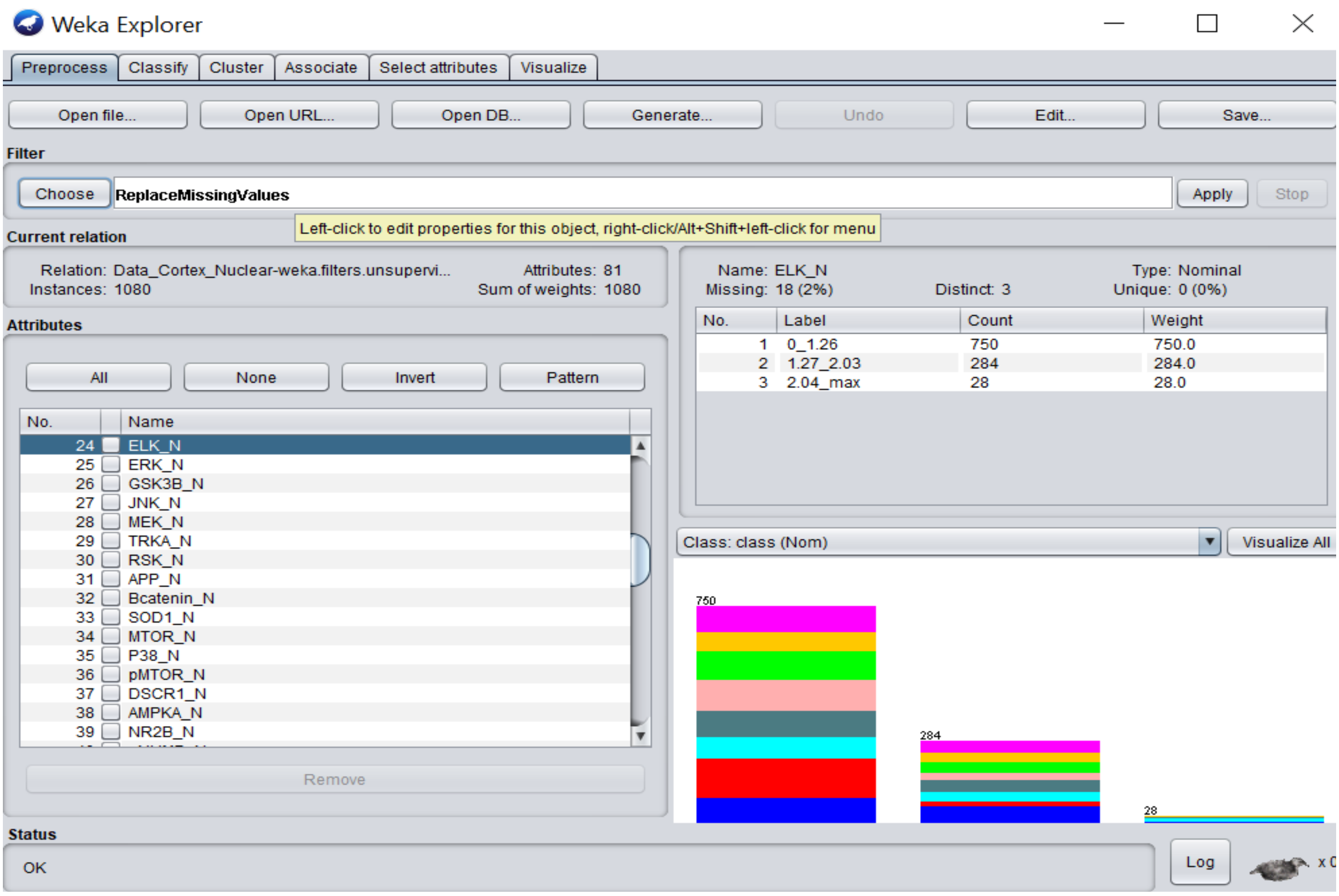
Open the filtered file into WEKA and check if any missing values in any attributes.



Edit data to make some missing values to view the missing values.



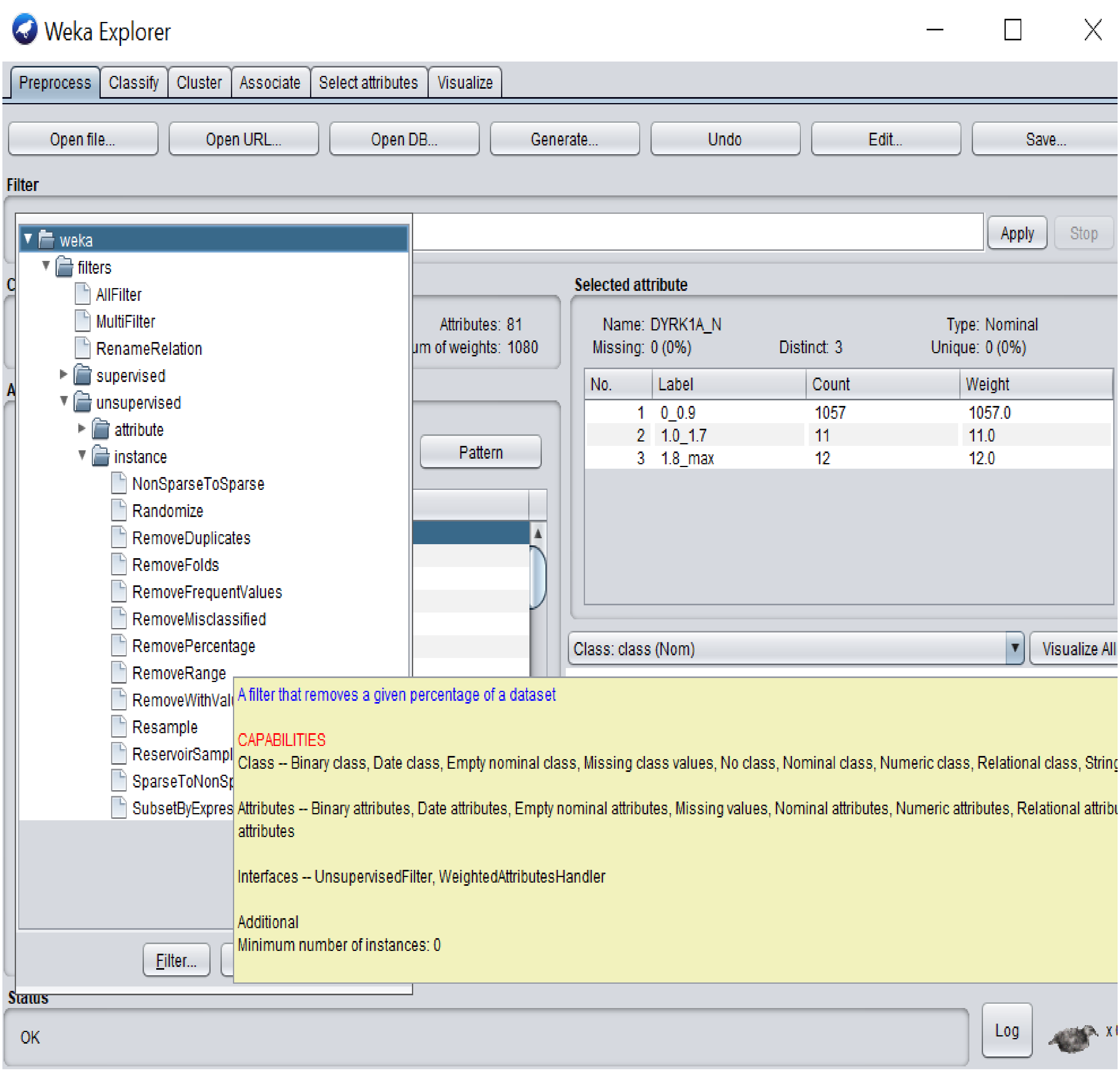
To fill the missing values, choose “ReplaceMissingValues” filter. Then click on Apply button.



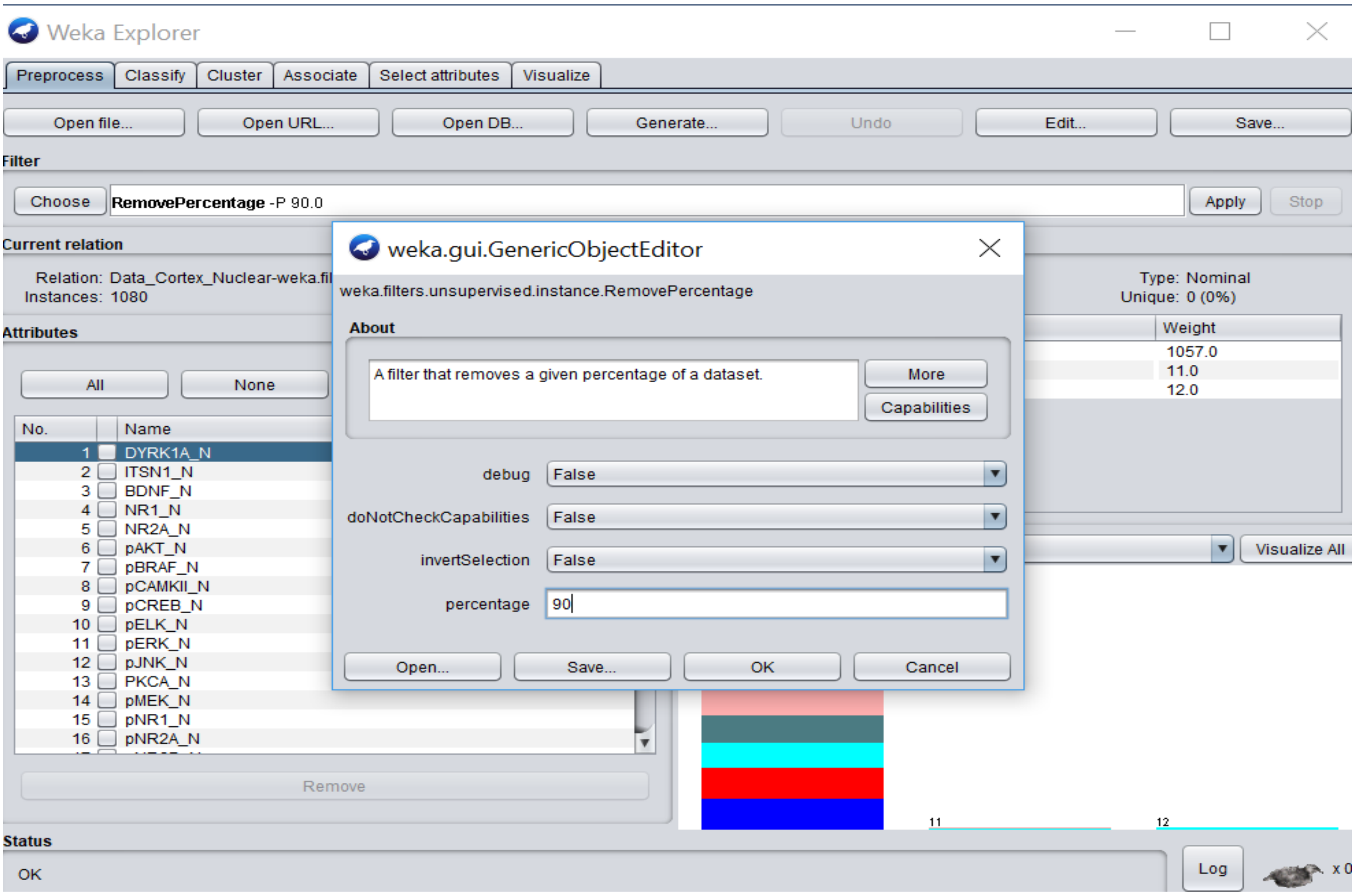
And then save the dataset into dataset.arff.

Divide the dataset into training and test set

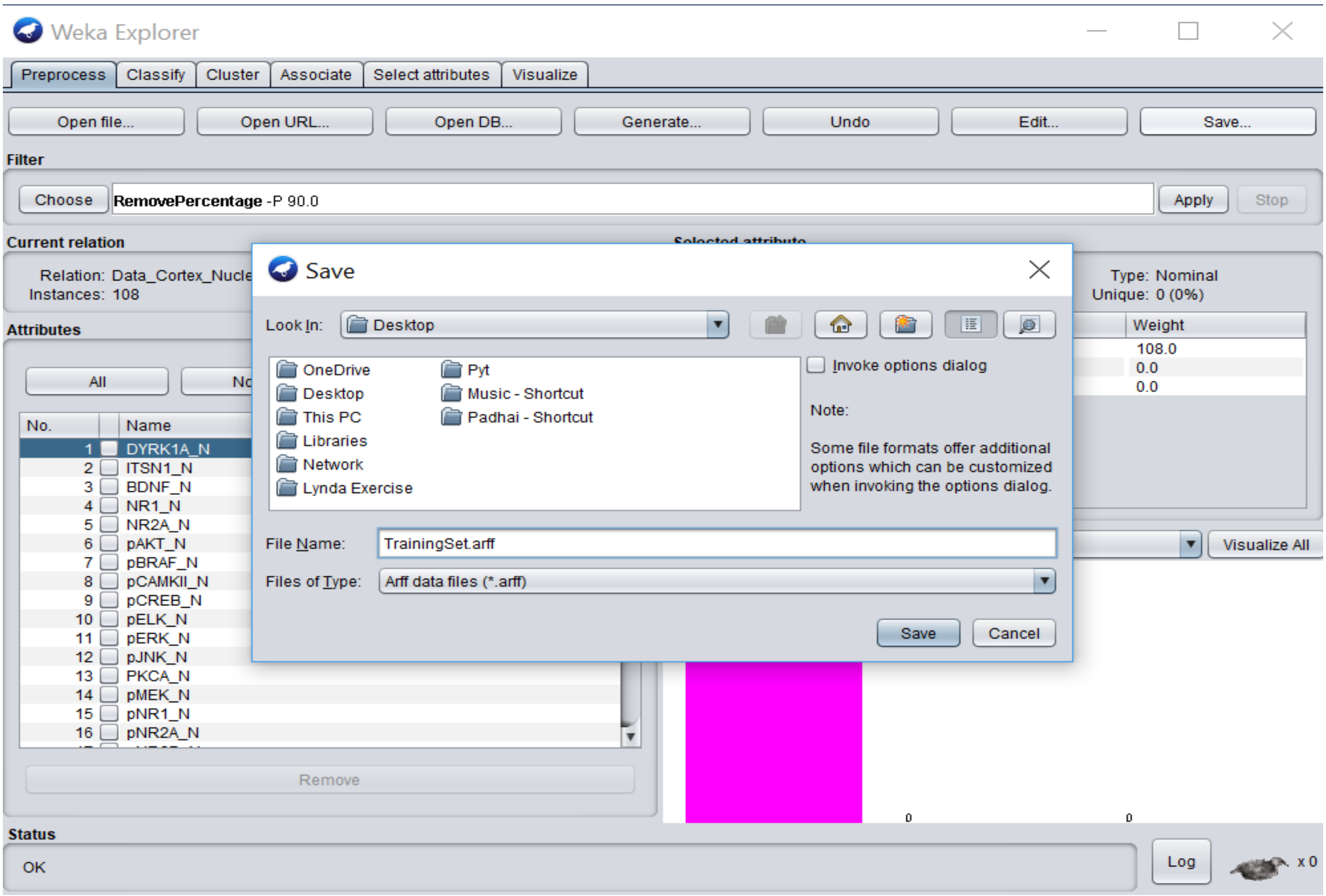
To divide the data set into training and testing, load the dataset into WEKA and choose RemovePercentage.



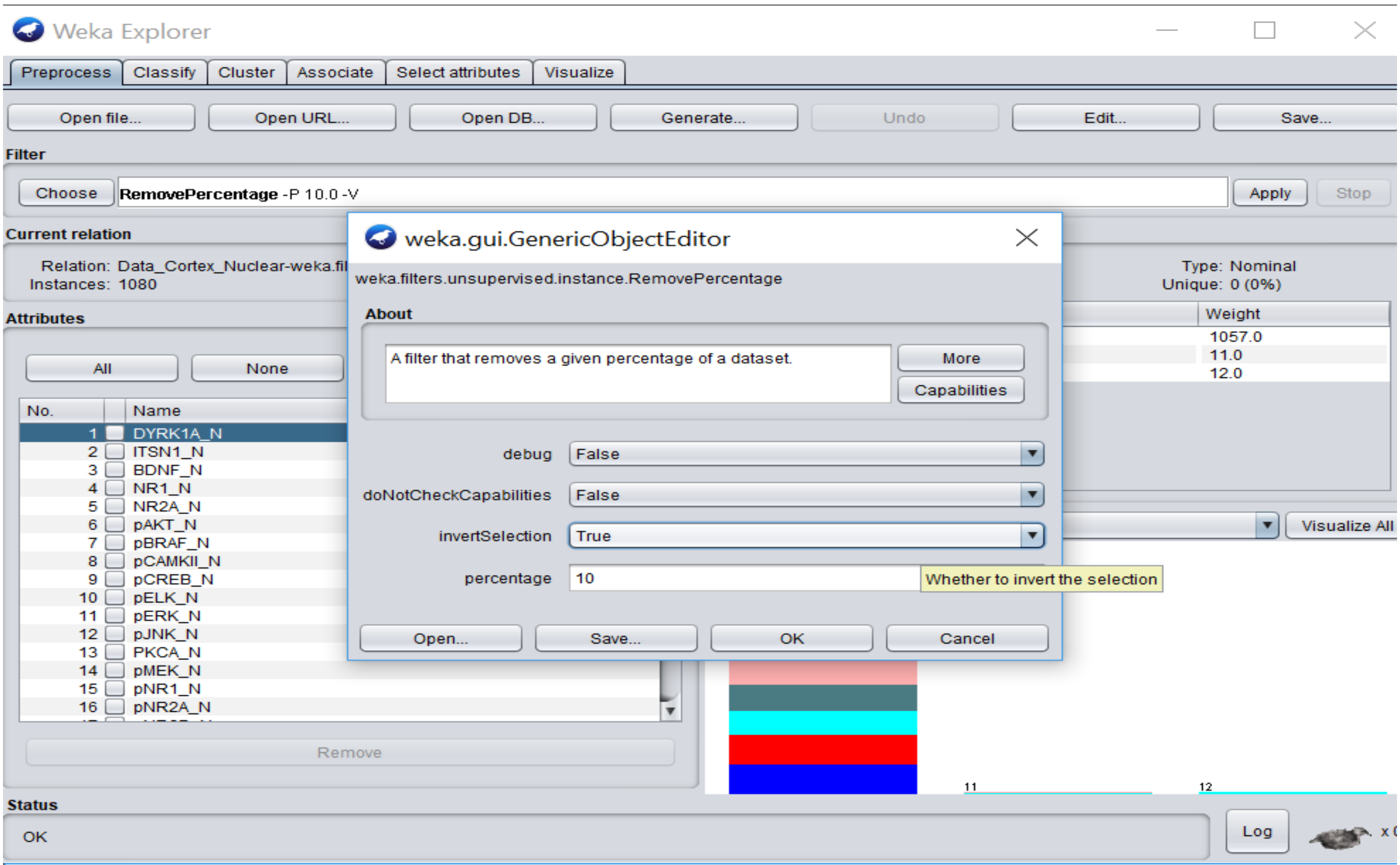
Then click on the text box next to the Choose button and set the correct percentage to 90% then click on OK button and apply.

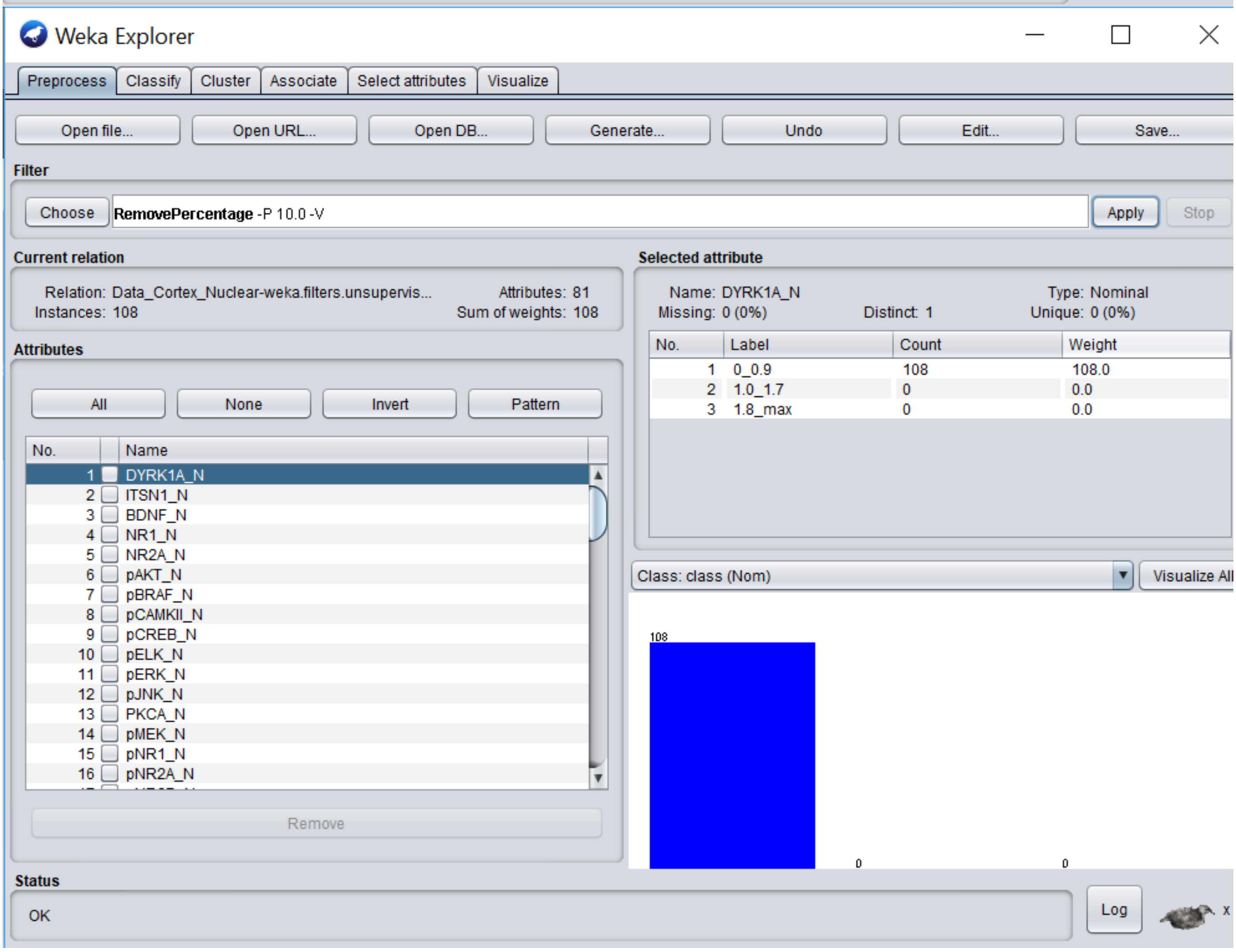


Next, save the file with TrainingSet.arff name.

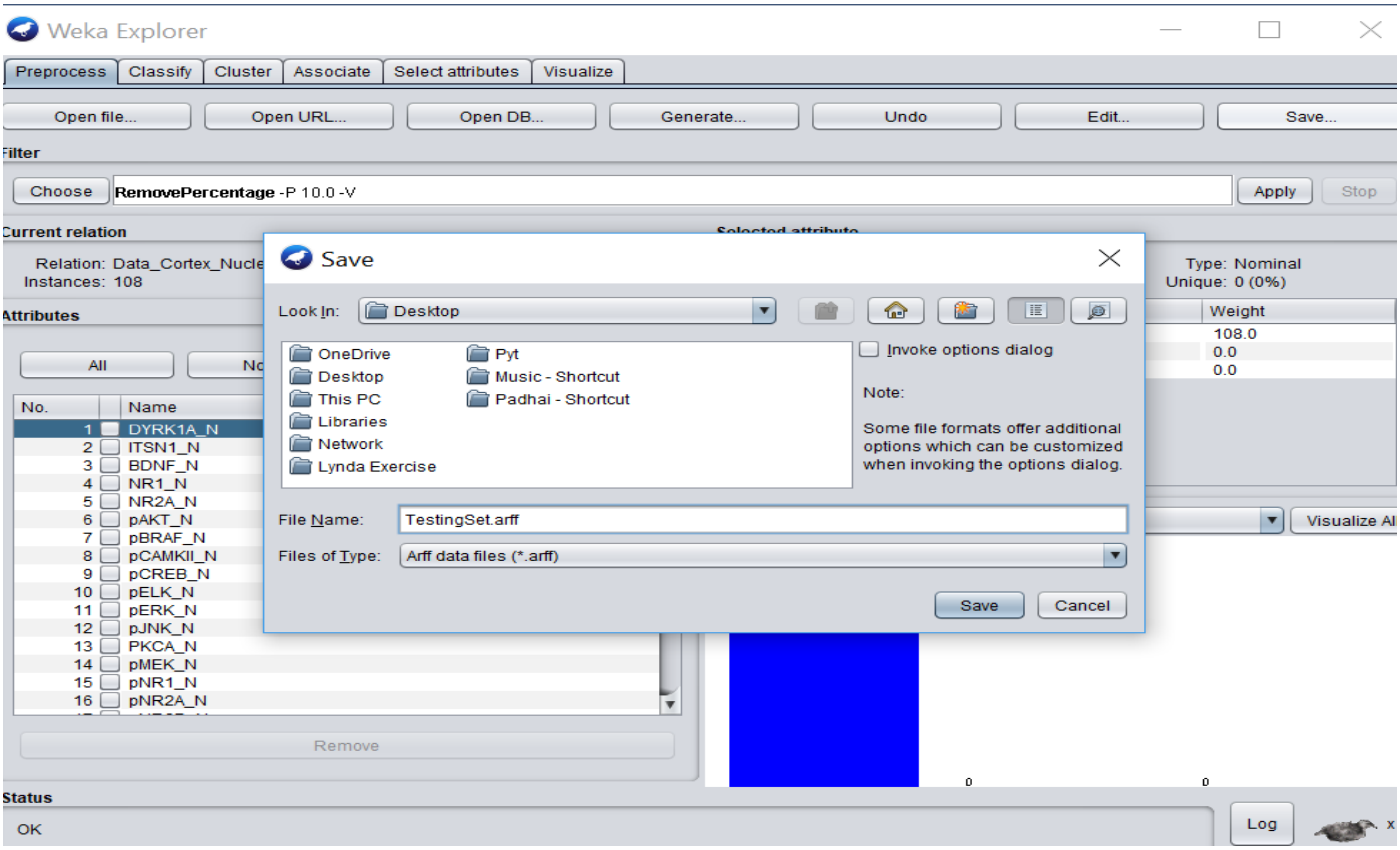


Then click on Undo to revert the changes and select the RemovePercentage filter if not yet selected set the invertSelection property to true and set the 10% and then click on apply.



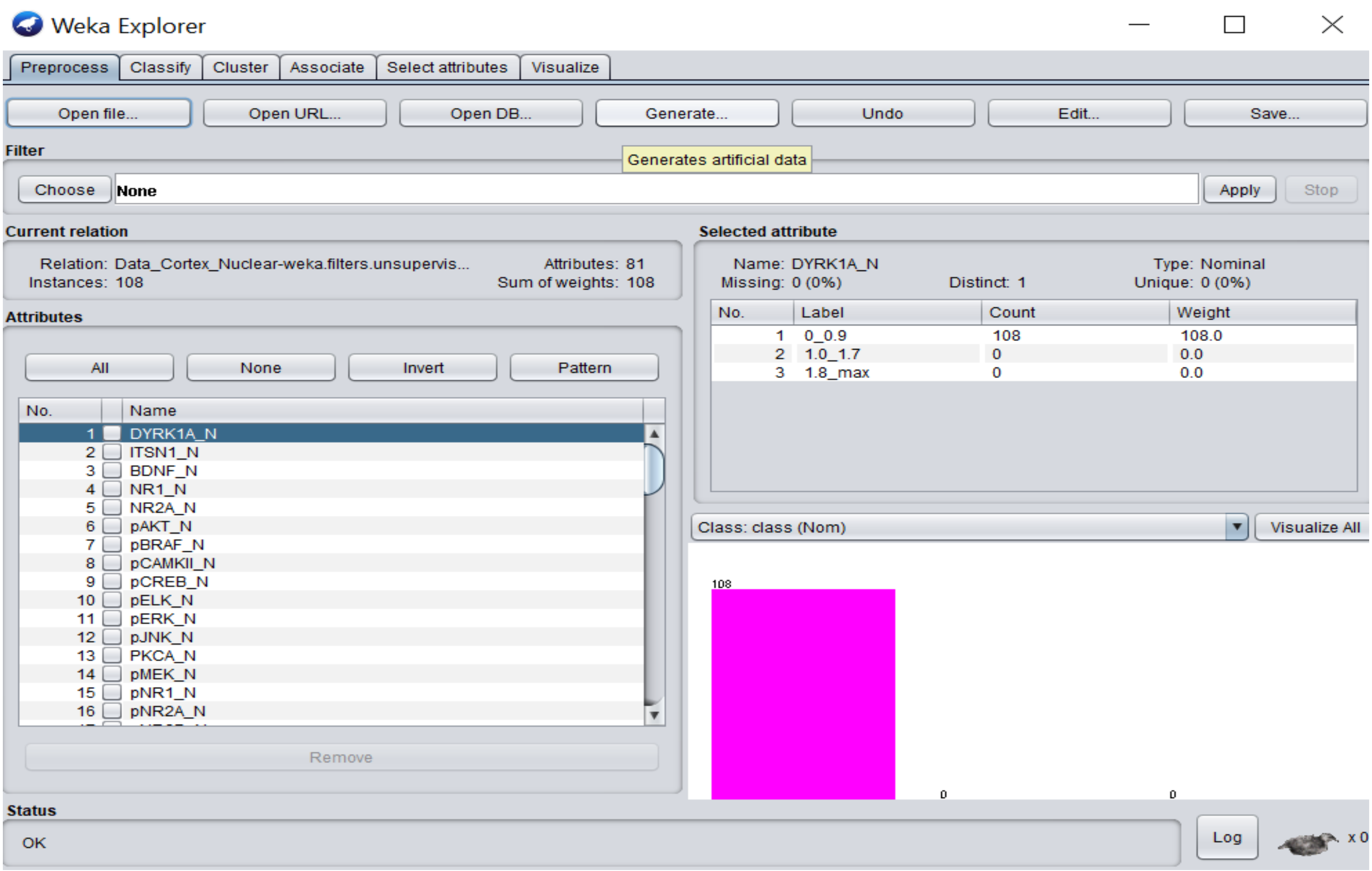


Now save as TestingSet.arff

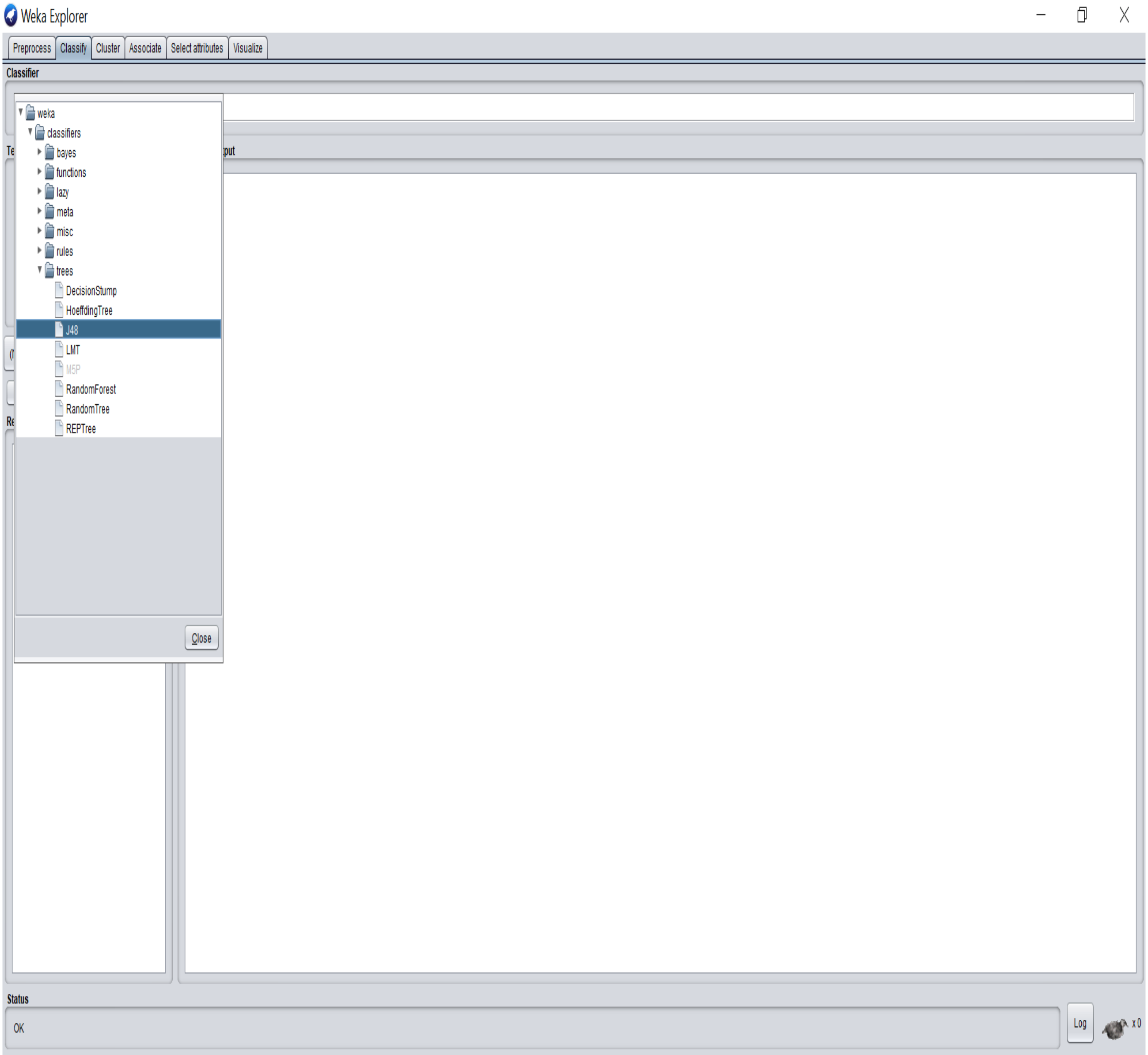


Classification

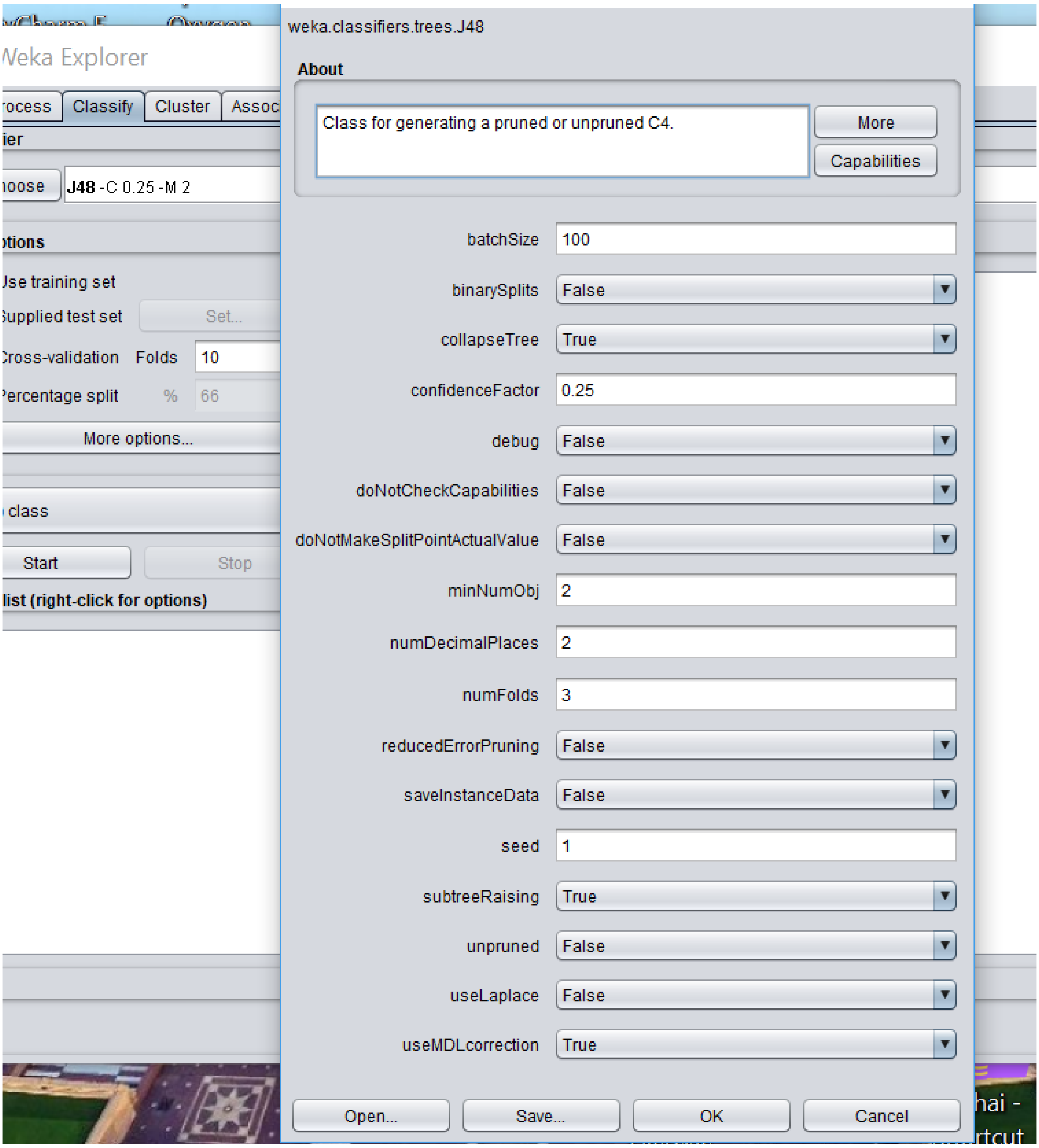
Load the data into WEKA with TrainingSet.arff



Next, we select the "Classify" tab and click the "Choose" button to select the J48 classifier.

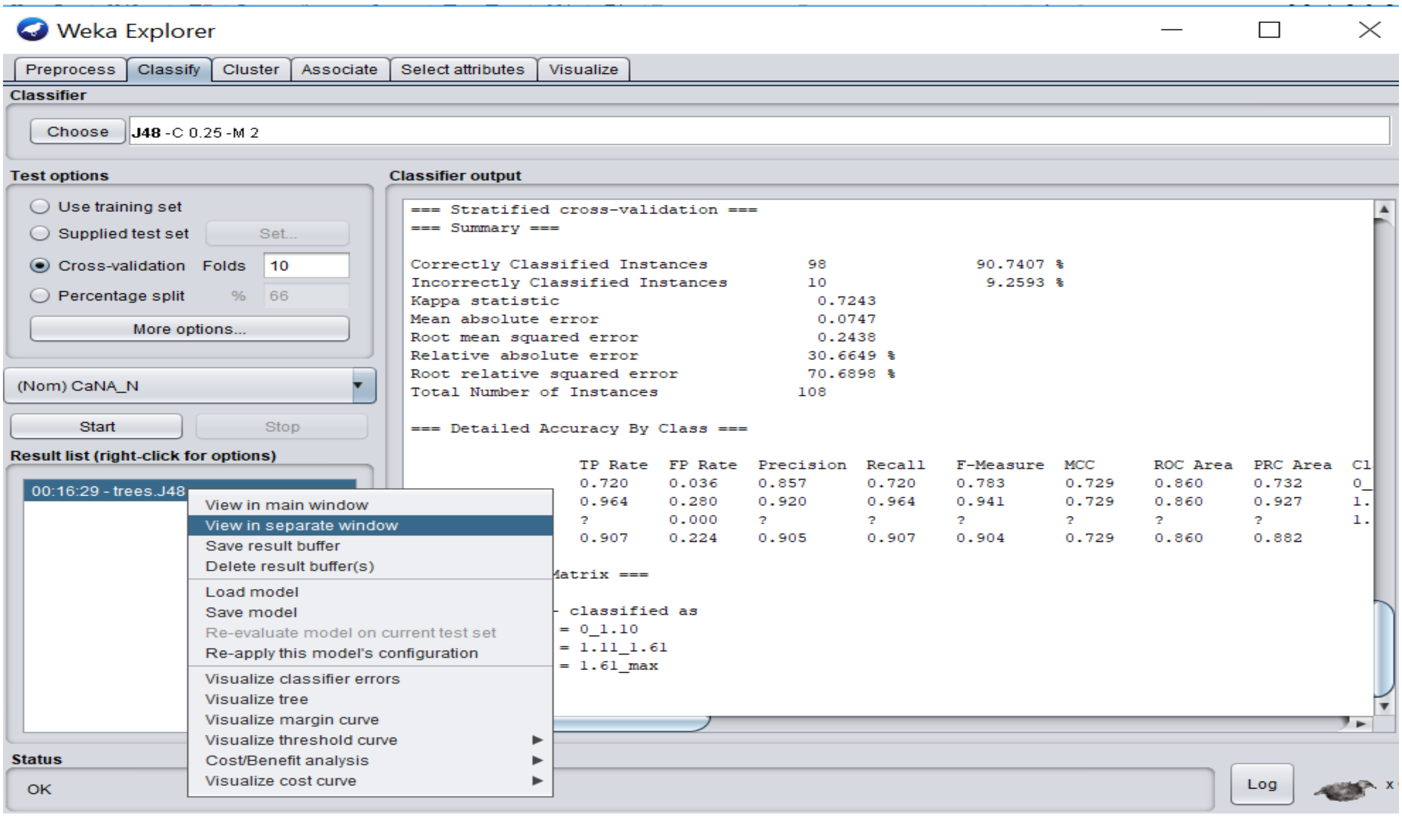


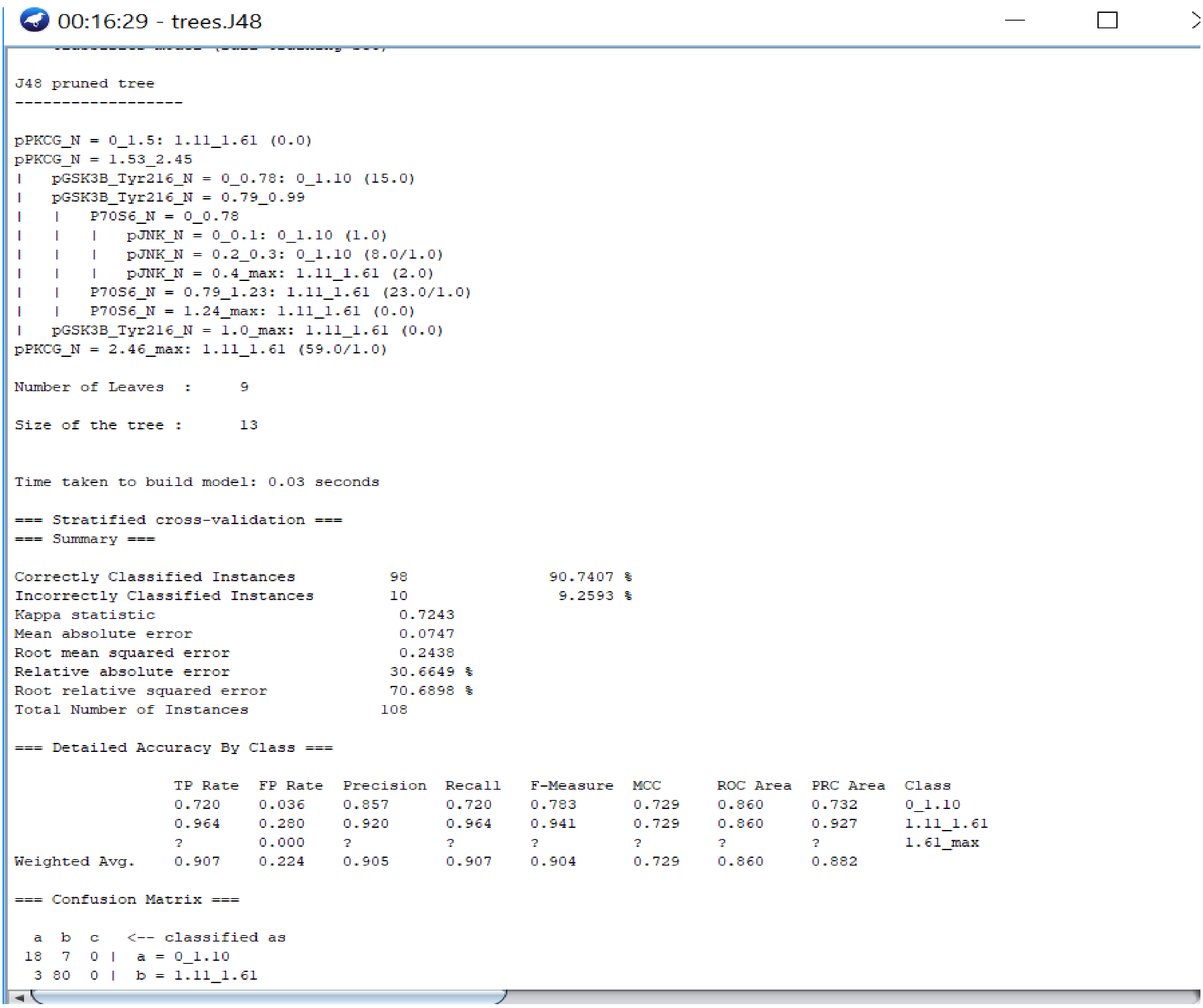
Now, we can specify the various parameters. These can be specified by clicking in the text box to the right of the "Choose" button. In this dataset we accept the default values. The default version does perform some pruning, but does not perform error pruning. Click on OK.



Under the "Test options" in the main panel as by default 10-fold cross-validation as our evaluation approach, then click "Start" to generate the model.

To view the result, right click on the result and select separate window.

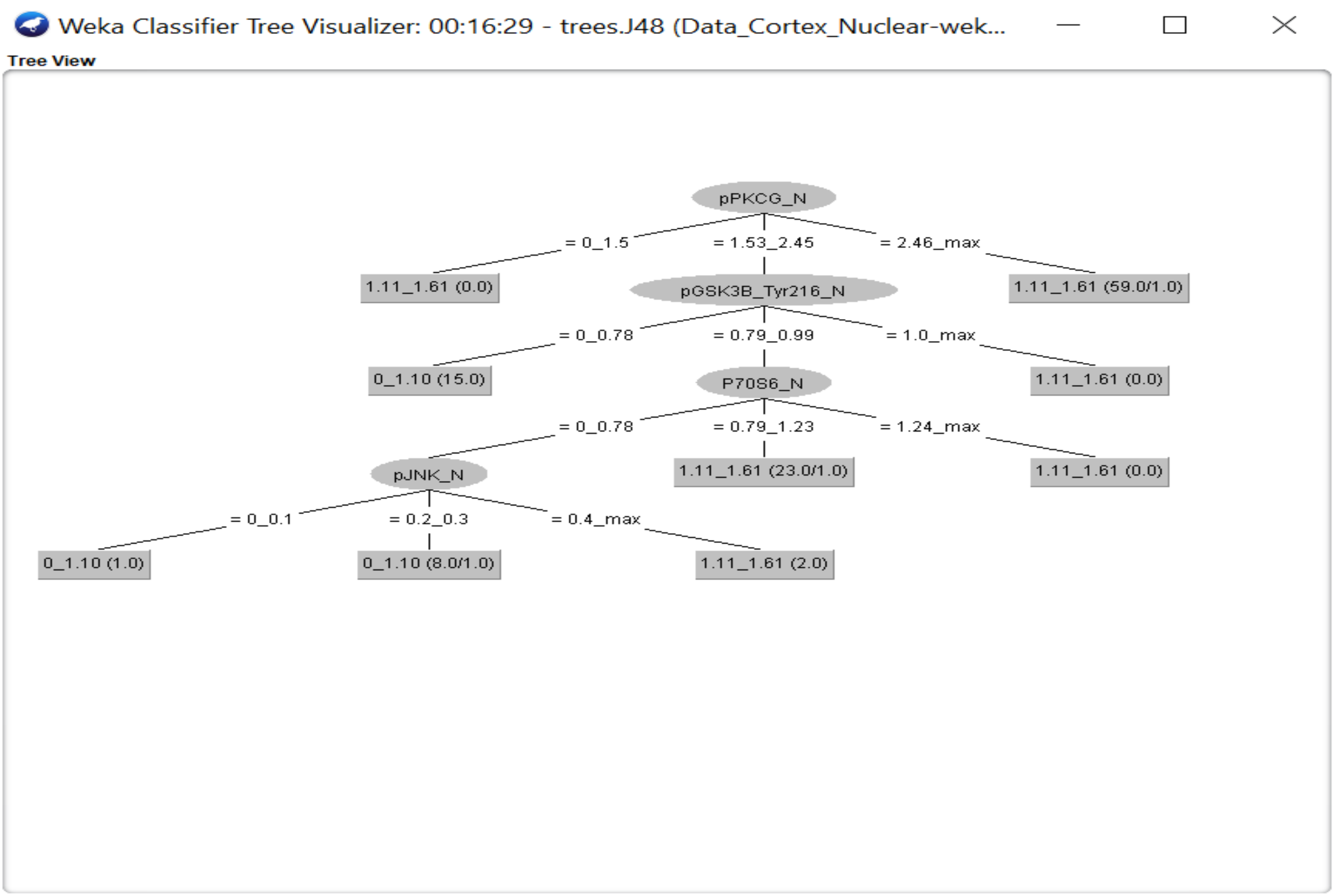




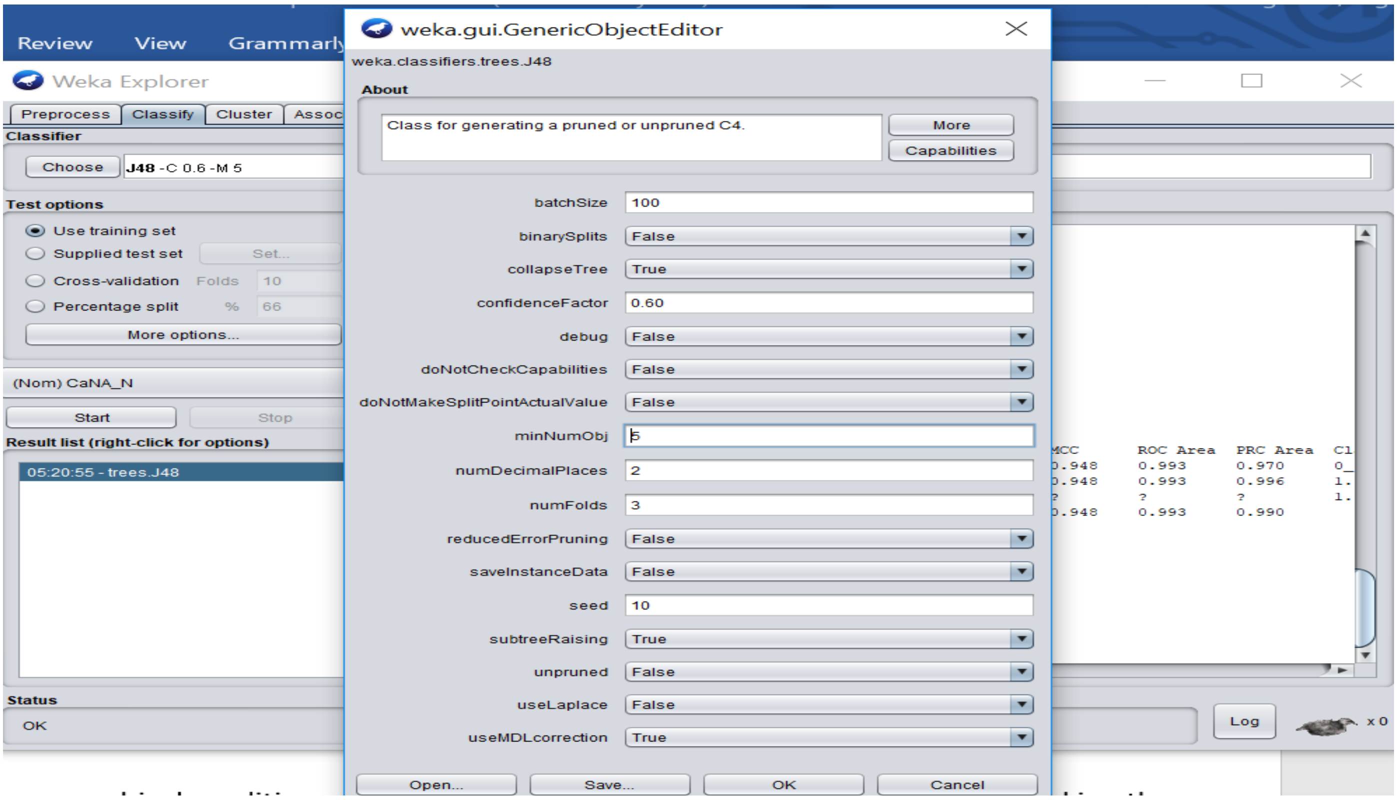
00:16:29 - trees.J48

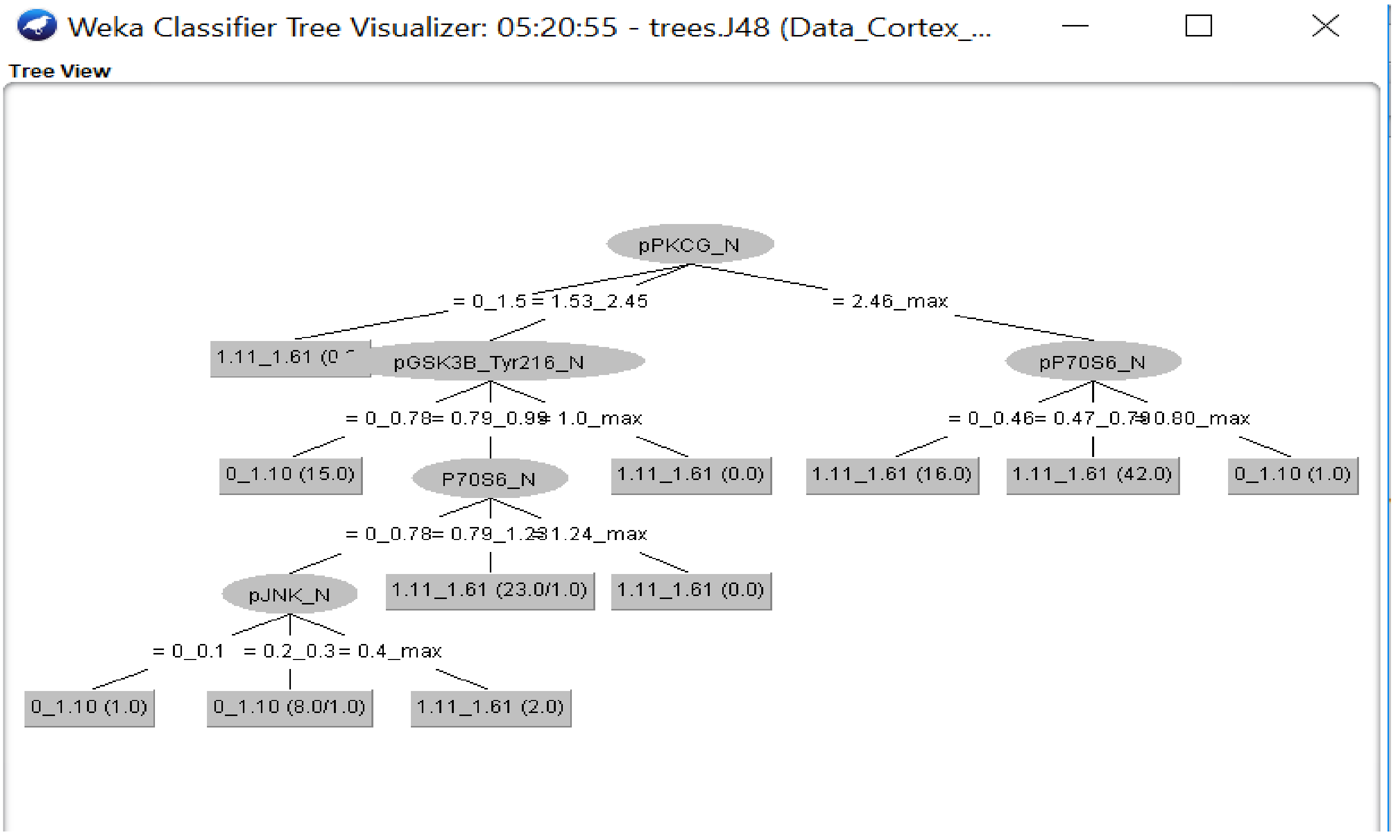
```
J48 pruned tree
------------------

pPKCG_N = 0_1.5: 1.11_1.61 (0.0)
pPKCG_N = 1.53_2.45
|   pGSK3B_Tyr216_N = 0_0.78: 0_1.10 (15.0)
|   pGSK3B_Tyr216_N = 0.79_0.99
|   |   P70S6_N = 0_0.78
|   |   |   pJNK_N = 0_0.1: 0_1.10 (1.0)
|   |   |   pJNK_N = 0.2_0.3: 0_1.10 (8.0/1.0)
|   |   |   pJNK_N = 0.4_max: 1.11_1.61 (2.0)
|   |   P70S6_N = 0.79_1.23: 1.11_1.61 (23.0/1.0)
|   |   P70S6_N = 1.24_max: 1.11_1.61 (0.0)
|   pGSK3B_Tyr216_N = 1.0_max: 1.11_1.61 (0.0)
pPKCG_N = 2.46_max: 1.11_1.61 (59.0/1.0)

Number of Leaves  :     9

Size of the tree :      13


Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          98               90.7407 %
Incorrectly Classified Instances        10                9.2593 %
Kappa statistic                          0.7243
Mean absolute error                      0.0747
Root mean squared error                  0.2438
Relative absolute error                 30.6649 %
Root relative squared error             70.6898 %
Total Number of Instances              108

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.720    0.036    0.857      0.720    0.783      0.729    0.860     0.732     0_1.10
                 0.964    0.280    0.920      0.964    0.941      0.729    0.860     0.927     1.11_1.61
                 ?        0.000    ?          ?        ?          ?        ?         ?         1.61_max
Weighted Avg.    0.907    0.224    0.905      0.907    0.904      0.729    0.860     0.882

=== Confusion Matrix ===

  a  b  c   <-- classified as
 18  7  0 |  a = 0_1.10
  3 80  0 |  b = 1.11_1.61
```

WEKA also lets us view a graphical rendition of the classification tree. This can be done by right clicking the last result set (as before) and selecting "Visualize tree" from the pop-up menu.

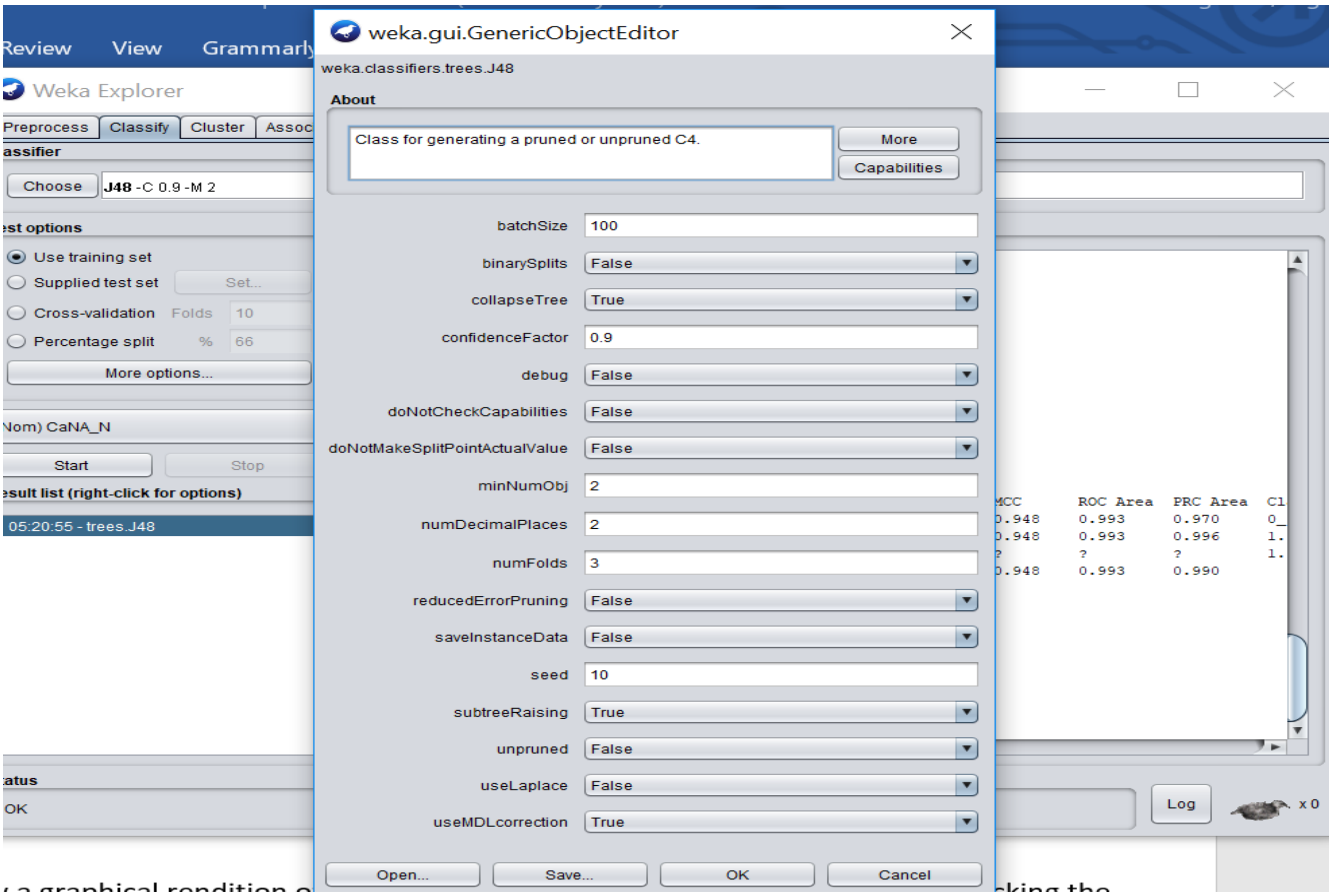


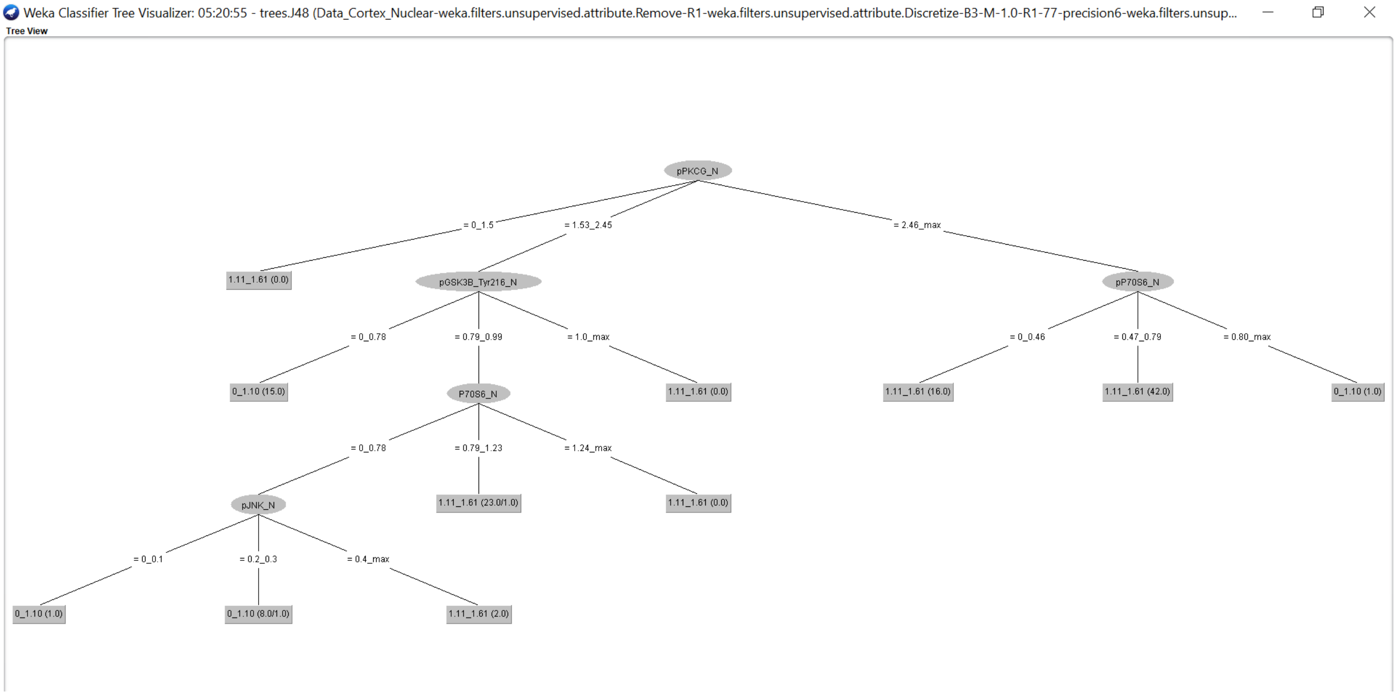
Case 2: with confidenceFactore 60% and minNumberObj =5





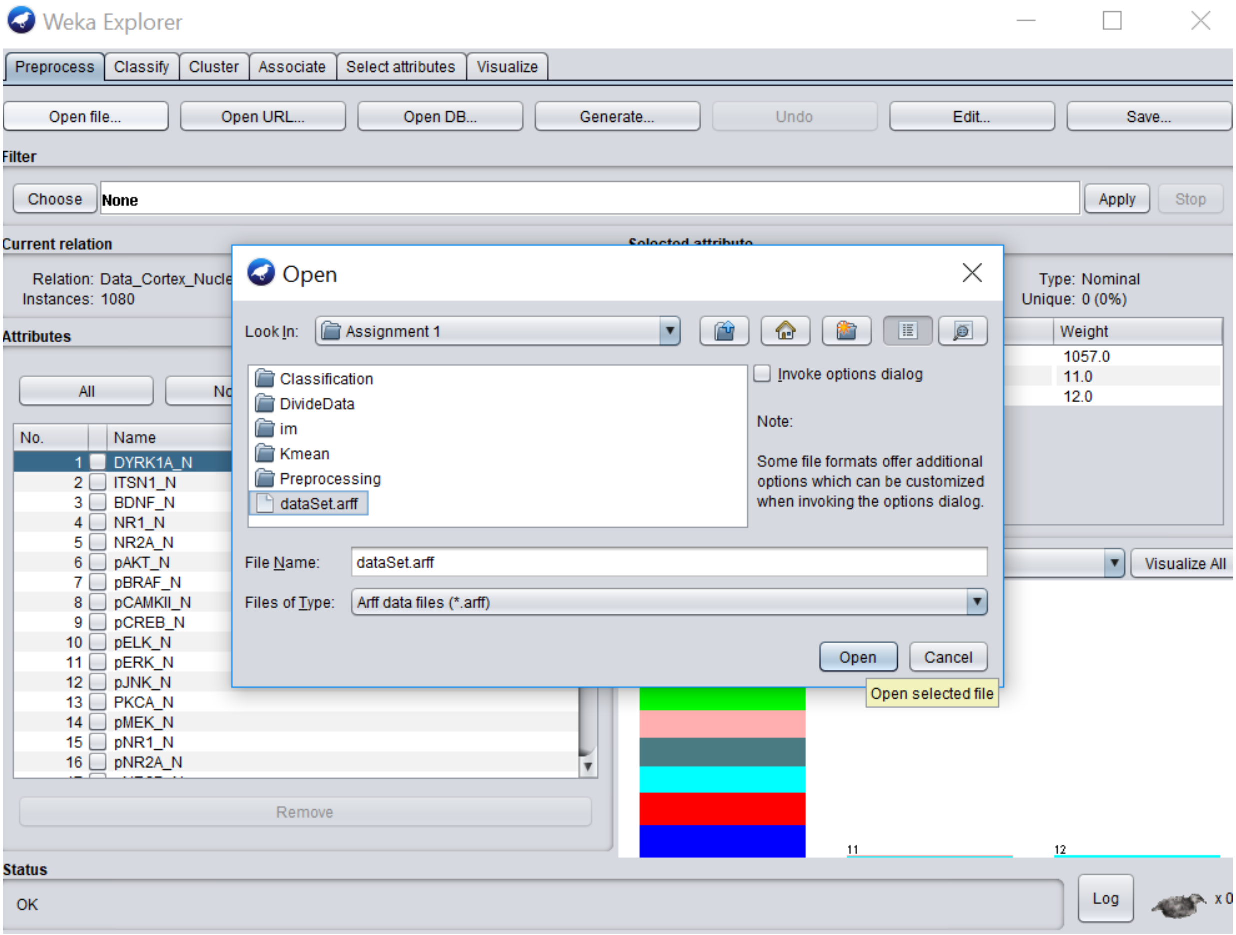
Case 3: with confidenceFactore 90% and minNumberObj =2



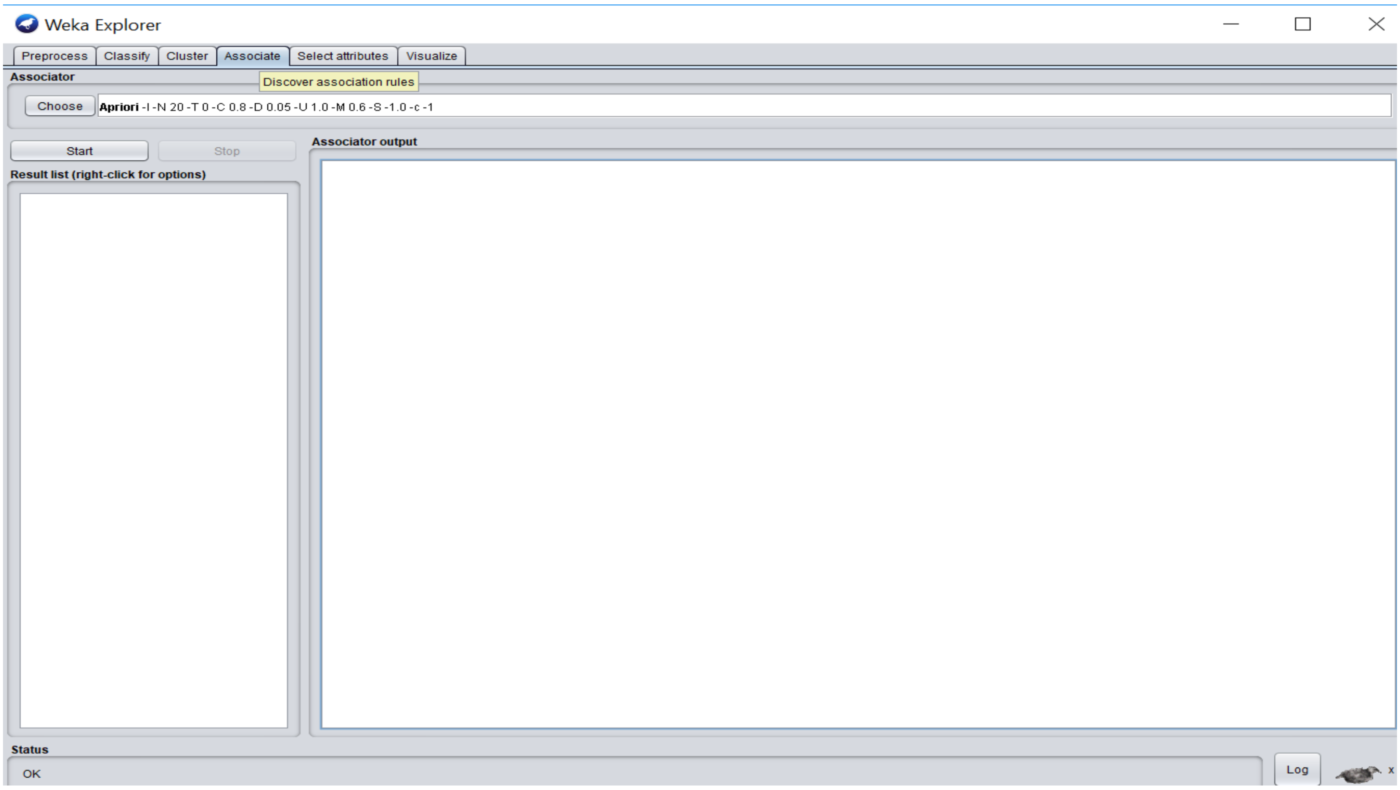


Association Rules:

Load the dataset in WEKA



In Weka, then select Associate tab.



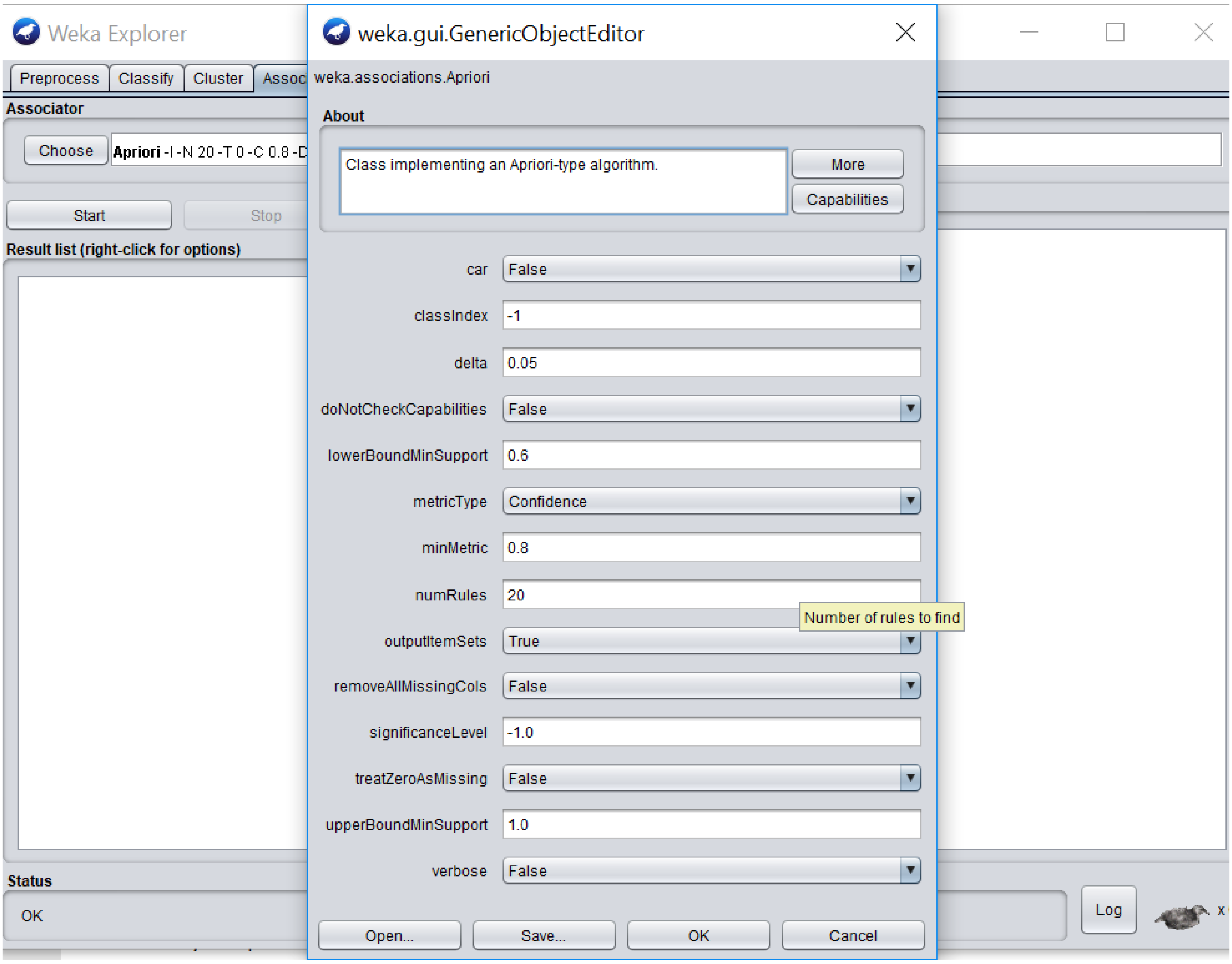
Then make sure Apriori is selected. If not then need to Choose Apriori algorithm. Then click on the text box right to the Choose button. To test the dataset with values, need to set the Apriori editor.

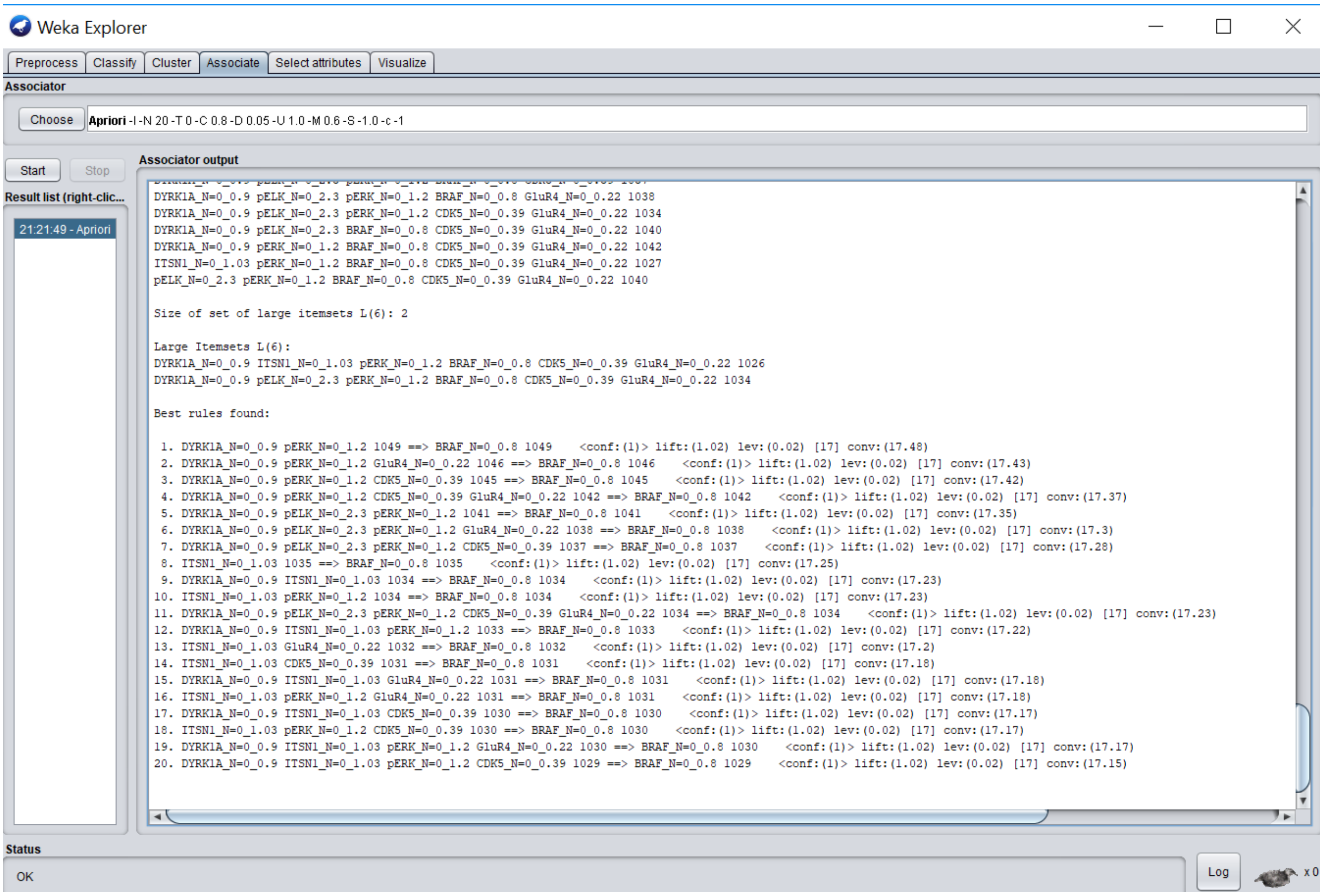
To test the dataset

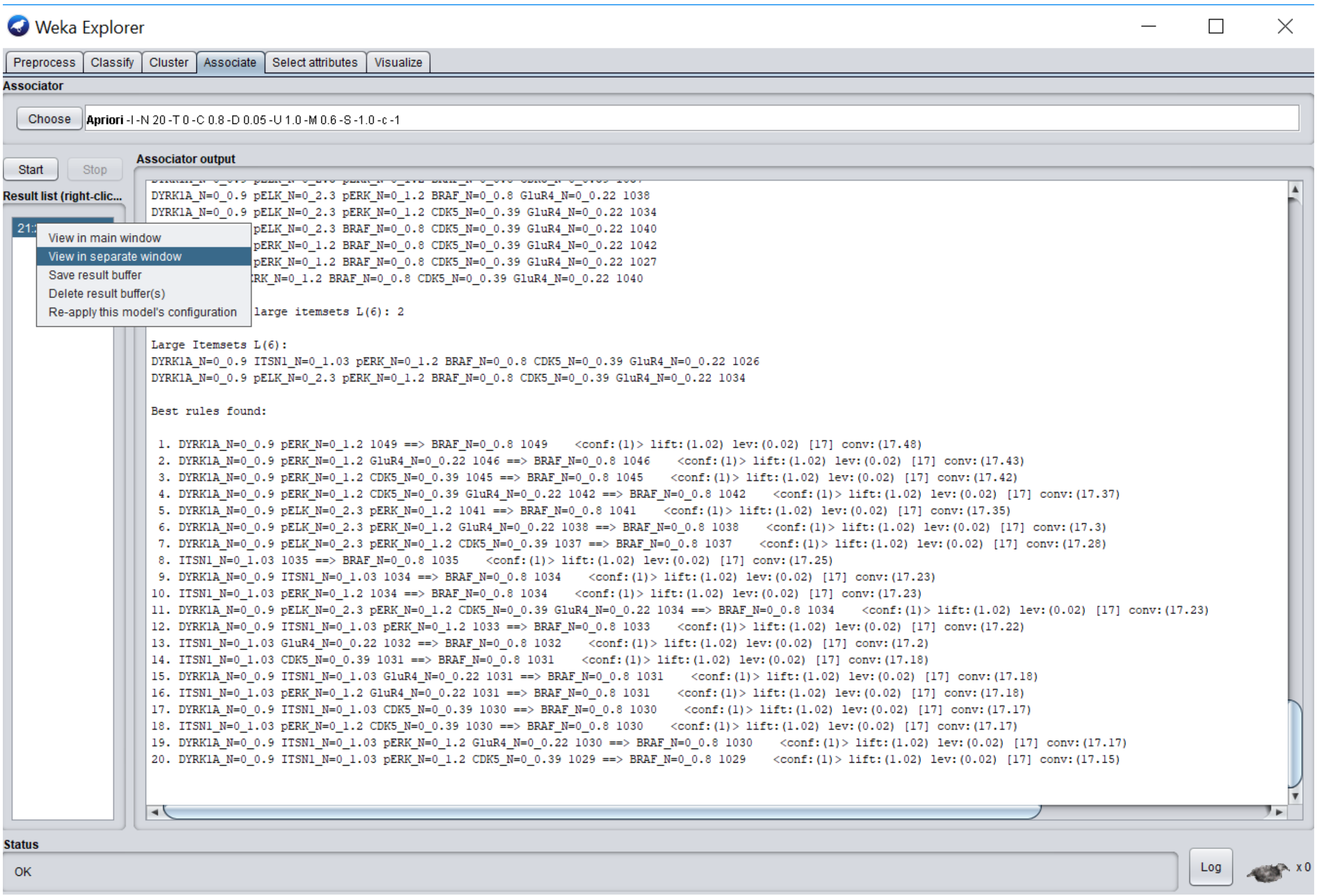
Case 1: 60% minimum support threshold and with 80% minimum confidence threshold

Therefore, set the lowerBoundMinSupport: 0.6 and minMetric: 0.80 with metricType: Confidence

And set the outputItemSets: True


Weka Explorer
weka.gui.GenericObjectEditor
Preprocess
Classify
Cluster
Assoc
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.8 -D
Start
Stop
Result list (right-click for options)
Status
OK
Log
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car False
classIndex -1
delta 0.05
doNotCheckCapabilities False
lowerBoundMinSupport 0.6
metricType Confidence
minMetric 0.8
numRules 20
Number of rules to find
outputItemSets True
removeAllMissingCols False
significanceLevel -1.0
treatZeroAsMissing False
upperBoundMinSupport 1.0
verbose False
Open...
Save...
OK
Cancel

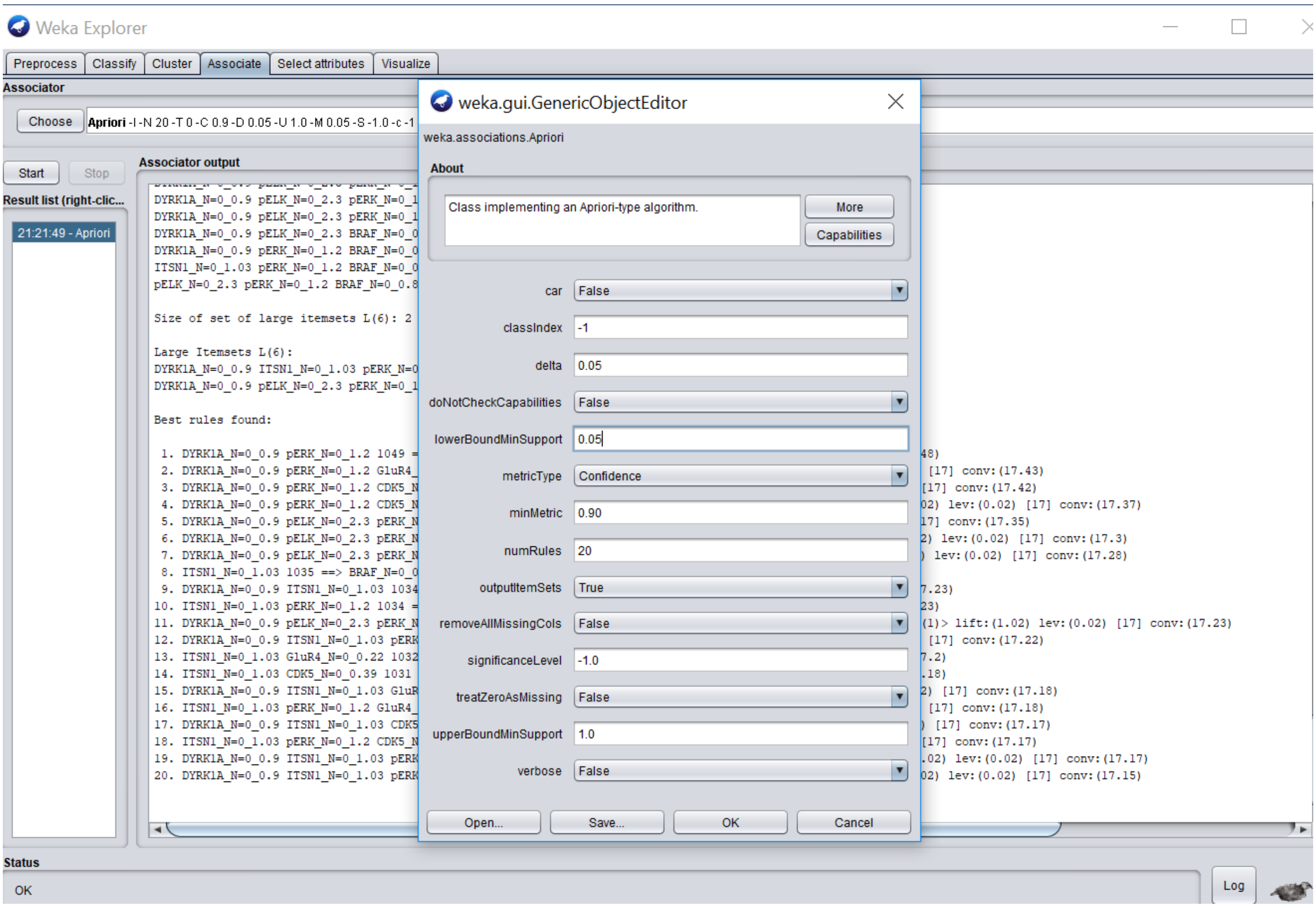

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.6 -S -1.0 -c -1
Start
Stop
Result list (right-clic...
21:21:49 - Apriori
Associator output
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1038
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034
DYRK1A_N=0_0.9 pELK_N=0_2.3 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1040
DYRK1A_N=0_0.9 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1042
ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1027
pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1040
Size of set of large itemsets L(6): 2
Large Itemsets L(6):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1026
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1034
Best rules found:
1. DYRK1A_N=0_0.9 pERK_N=0_1.2 1049 ==> BRAF_N=0_0.8 1049 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.48)
2. DYRK1A_N=0_0.9 pERK_N=0_1.2 GluR4_N=0_0.22 1046 ==> BRAF_N=0_0.8 1046 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.43)
3. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 1045 ==> BRAF_N=0_0.8 1045 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.42)
4. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1042 ==> BRAF_N=0_0.8 1042 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.37)
5. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 1041 ==> BRAF_N=0_0.8 1041 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.35)
6. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 GluR4_N=0_0.22 1038 ==> BRAF_N=0_0.8 1038 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.3)
7. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 1037 ==> BRAF_N=0_0.8 1037 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.28)
8. ITSN1_N=0_1.03 1035 ==> BRAF_N=0_0.8 1035 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.25)
9. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
10. ITSN1_N=0_1.03 pERK_N=0_1.2 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
11. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
12. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 1033 ==> BRAF_N=0_0.8 1033 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.22)
13. ITSN1_N=0_1.03 GluR4_N=0_0.22 1032 ==> BRAF_N=0_0.8 1032 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.2)
14. ITSN1_N=0_1.03 CDK5_N=0_0.39 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
15. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
16. ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
17. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
18. ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
19. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
20. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1029 ==> BRAF_N=0_0.8 1029 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.15)
Status
OK
Log
x 0

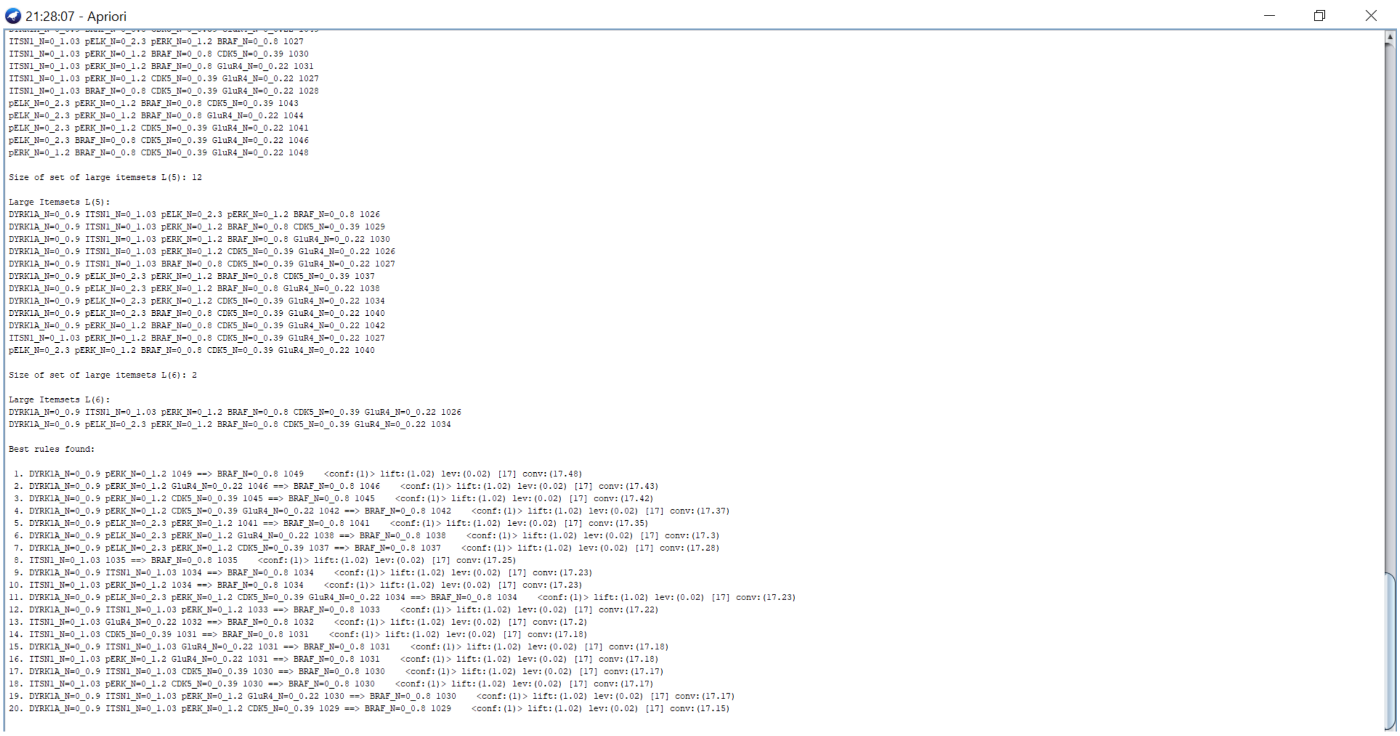

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.6 -S -1.0 -c -1
Start
Stop
Result list (right-clic...
View in main window
View in separate window
Save result buffer
Delete result buffer(s)
Re-apply this model's configuration
Associator output
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1038
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034
Large Itemsets L(6):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1026
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1034
Best rules found:
1. DYRK1A_N=0_0.9 pERK_N=0_1.2 1049 ==> BRAF_N=0_0.8 1049 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.48)
2. DYRK1A_N=0_0.9 pERK_N=0_1.2 GluR4_N=0_0.22 1046 ==> BRAF_N=0_0.8 1046 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.43)
3. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 1045 ==> BRAF_N=0_0.8 1045 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.42)
4. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1042 ==> BRAF_N=0_0.8 1042 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.37)
5. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 1041 ==> BRAF_N=0_0.8 1041 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.35)
6. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 GluR4_N=0_0.22 1038 ==> BRAF_N=0_0.8 1038 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.3)
7. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 1037 ==> BRAF_N=0_0.8 1037 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.28)
8. ITSN1_N=0_1.03 1035 ==> BRAF_N=0_0.8 1035 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.25)
9. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
10. ITSN1_N=0_1.03 pERK_N=0_1.2 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
11. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034 ==> BRAF_N=0_0.8 1034 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
12. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 1033 ==> BRAF_N=0_0.8 1033 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.22)
13. ITSN1_N=0_1.03 GluR4_N=0_0.22 1032 ==> BRAF_N=0_0.8 1032 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.2)
14. ITSN1_N=0_1.03 CDK5_N=0_0.39 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
15. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
16. ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
17. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
18. ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
19. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1030 ==> BRAF_N=0_0.8 1030 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
20. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1029 ==> BRAF_N=0_0.8 1029 <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.15)
Status
OK
Log
x 0

21:21:49 - Apriori

```
ITSN1_N=0_1.03 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 1027
ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 1030
ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1031
ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1027
ITSN1_N=0_1.03 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1028
pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 1043
pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1044
pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1041
pELK_N=0_2.3 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1046
pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1048

Size of set of large itemsets L(5): 12

Large Itemsets L(5):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 1026
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 1029
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1030
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1026
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1027
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 1037
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 GluR4_N=0_0.22 1038
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034
DYRK1A_N=0_0.9 pELK_N=0_2.3 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1040
DYRK1A_N=0_0.9 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1042
ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1027
pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1040

Size of set of large itemsets L(6): 2

Large Itemsets L(6):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1026
DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 BRAF_N=0_0.8 CDK5_N=0_0.39 GluR4_N=0_0.22 1034

Best rules found:

 1. DYRK1A_N=0_0.9 pERK_N=0_1.2 1049 ==> BRAF_N=0_0.8 1049    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.48)
 2. DYRK1A_N=0_0.9 pERK_N=0_1.2 GluR4_N=0_0.22 1046 ==> BRAF_N=0_0.8 1046    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.43)
 3. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 1045 ==> BRAF_N=0_0.8 1045    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.42)
 4. DYRK1A_N=0_0.9 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1042 ==> BRAF_N=0_0.8 1042    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.37)
 5. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 1041 ==> BRAF_N=0_0.8 1041    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.35)
 6. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 GluR4_N=0_0.22 1038 ==> BRAF_N=0_0.8 1038    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.3)
 7. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 1037 ==> BRAF_N=0_0.8 1037    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.28)
 8. ITSN1_N=0_1.03 1035 ==> BRAF_N=0_0.8 1035    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.25)
 9. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 1034 ==> BRAF_N=0_0.8 1034    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
10. ITSN1_N=0_1.03 pERK_N=0_1.2 1034 ==> BRAF_N=0_0.8 1034    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
11. DYRK1A_N=0_0.9 pELK_N=0_2.3 pERK_N=0_1.2 CDK5_N=0_0.39 GluR4_N=0_0.22 1034 ==> BRAF_N=0_0.8 1034    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.23)
12. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 1033 ==> BRAF_N=0_0.8 1033    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.22)
13. ITSN1_N=0_1.03 GluR4_N=0_0.22 1032 ==> BRAF_N=0_0.8 1032    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.2)
14. ITSN1_N=0_1.03 CDK5_N=0_0.39 1031 ==> BRAF_N=0_0.8 1031    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
15. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
16. ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1031 ==> BRAF_N=0_0.8 1031    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.18)
17. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
18. ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1030 ==> BRAF_N=0_0.8 1030    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
19. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 GluR4_N=0_0.22 1030 ==> BRAF_N=0_0.8 1030    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.17)
20. DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 CDK5_N=0_0.39 1029 ==> BRAF_N=0_0.8 1029    <conf:(1)> lift:(1.02) lev:(0.02) [17] conv:(17.15)
```

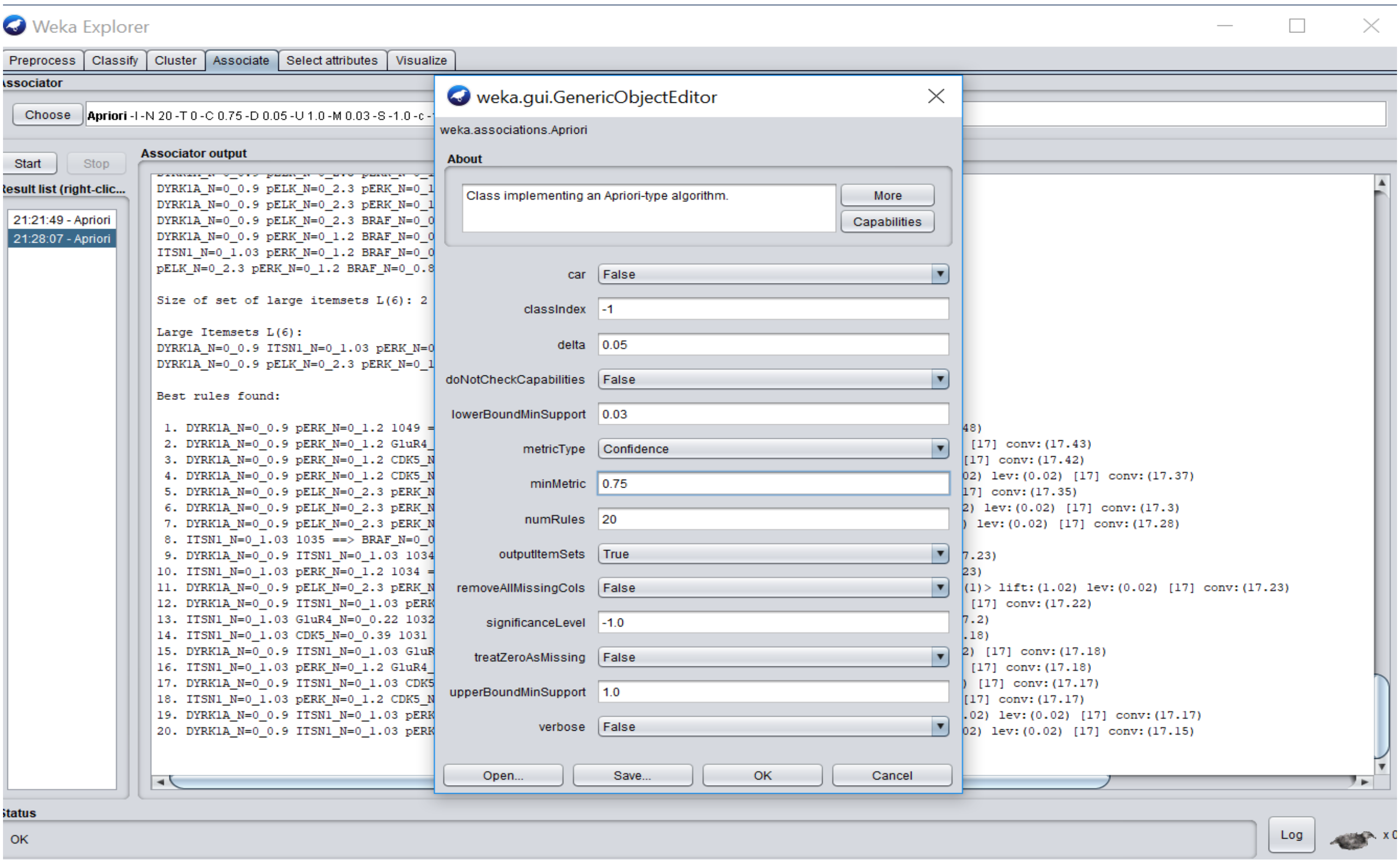
Case 2: With lowerBoundMinSupport: 0.05 and minMetric: 0.90





Case 3:

minMetricwithConfidence: 0.75, lowerBoundMinSupport: 0.03

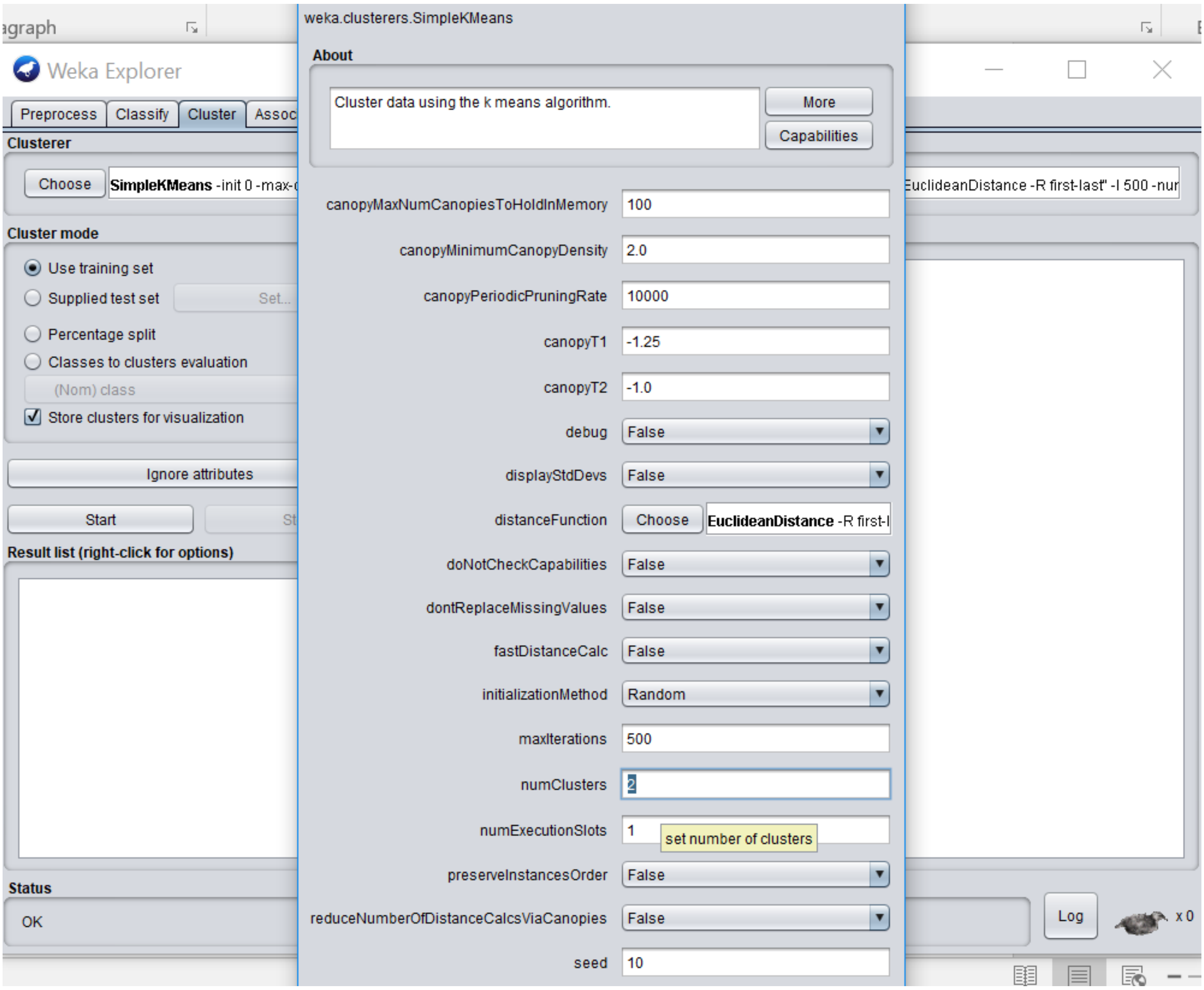

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 20 -T 0 -C 0.75 -D 0.05 -U 1.0 -M 0.03 -S -1.0 -c -
Start
Stop
Result list (right-clic...
21:21:49 - Apriori
21:28:07 - Apriori
Associator output
Size of set of large itemsets L(6): 2
Large Itemsets L(6):
Best rules found:
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
-1
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.03
metricType
Confidence
minMetric
0.75
numRules
20
outputItemSets
True
removeAllMissingCols
False
significanceLevel
-1.0
treatZeroAsMissing
False
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel
Status
OK
Log

00:14:15 - Apriori

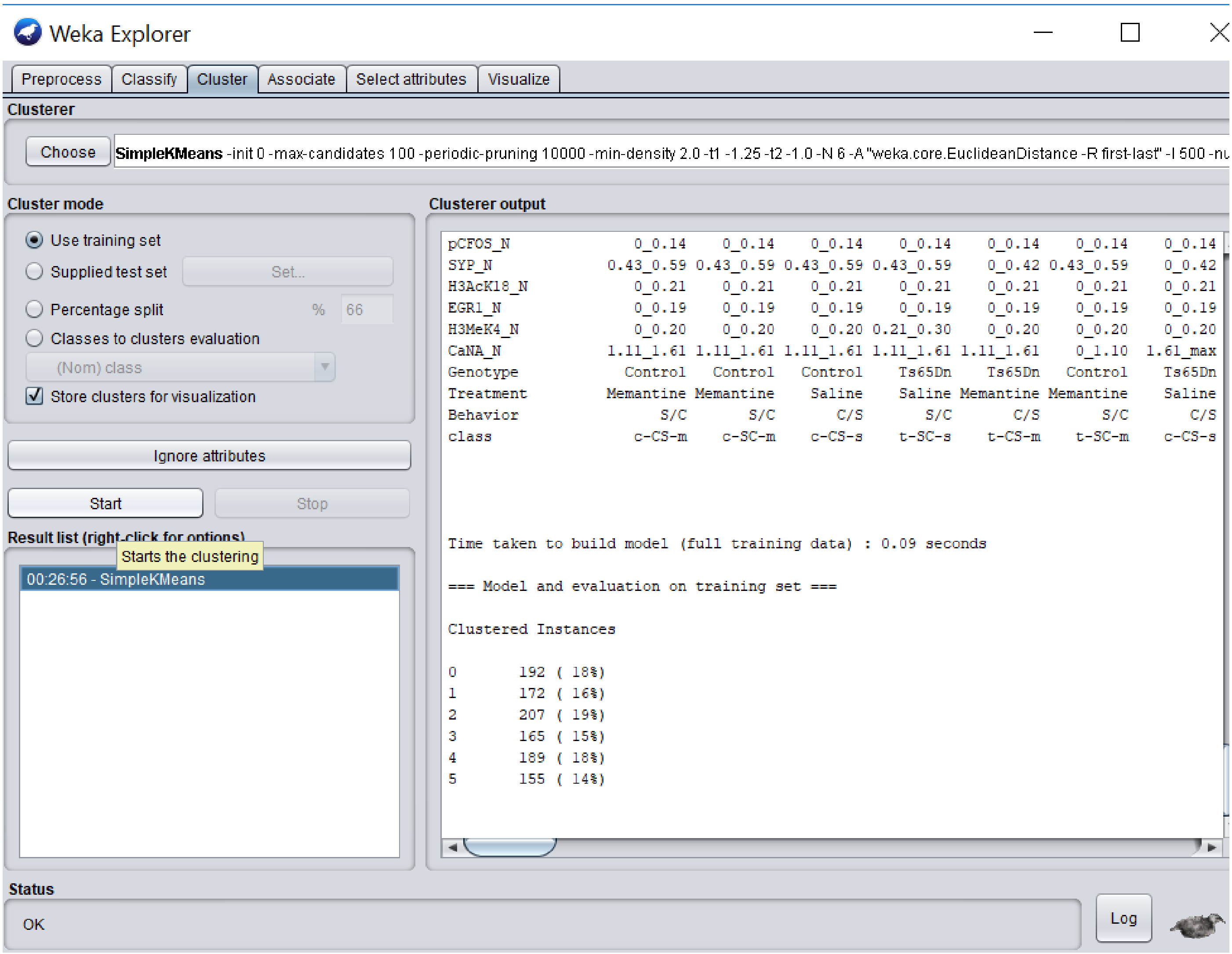
```
              Treatment
              Behavior
              class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.95 (1026 instances)
Minimum metric <confidence>: 0.75
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Large Itemsets L(1):
DYRK1A_N=0_0.9 1057
ITSN1_N=0_1.03 1035
pELK_N=0_2.3 1056
pERK_N=0_1.2 1057
BRAF_N=0_0.8 1062
CDK5_N=0_0.39 1076
GluR4_N=0_0.22 1077

Size of set of large itemsets L(2): 21

Large Itemsets L(2):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 1034
DYRK1A_N=0_0.9 pELK_N=0_2.3 1048
DYRK1A_N=0_0.9 pERK_N=0_1.2 1049
DYRK1A_N=0_0.9 BRAF_N=0_0.8 1056
DYRK1A_N=0_0.9 CDK5_N=0_0.39 1053
DYRK1A_N=0_0.9 GluR4_N=0_0.22 1054
ITSN1_N=0_1.03 pELK_N=0_2.3 1028
ITSN1_N=0_1.03 pERK_N=0_1.2 1034
ITSN1_N=0_1.03 BRAF_N=0_0.8 1035
ITSN1_N=0_1.03 CDK5_N=0_0.39 1031
ITSN1_N=0_1.03 GluR4_N=0_0.22 1032
pELK_N=0_2.3 pERK_N=0_1.2 1048
pELK_N=0_2.3 BRAF_N=0_0.8 1053
pELK_N=0_2.3 CDK5_N=0_0.39 1052
pELK_N=0_2.3 GluR4_N=0_0.22 1053
pERK_N=0_1.2 BRAF_N=0_0.8 1055
pERK_N=0_1.2 CDK5_N=0_0.39 1053
pERK_N=0_1.2 GluR4_N=0_0.22 1054
BRAF_N=0_0.8 CDK5_N=0_0.39 1058
BRAF_N=0_0.8 GluR4_N=0_0.22 1059
CDK5_N=0_0.39 GluR4_N=0_0.22 1073

Size of set of large itemsets L(3): 33

Large Itemsets L(3):
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pELK_N=0_2.3 1027
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 pERK_N=0_1.2 1033
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 BRAF_N=0_0.8 1034
DYRK1A_N=0_0.9 ITSN1_N=0_1.03 CDK5_N=0_0.39 1030
```

K-Mean

To perform clustering, select the "Cluster" tab in the Explorer and click on the "Choose" button. This results in a drop-down list of available clustering algorithms. In this case we select "SimpleKMeans". Next, click on the text box to the right of the "Choose" button to get the pop-up window.



In the pop-up window we enter 6 as the number of clusters (instead of the default values of 2) and we leave the value of "seed" default. Then click OK and Start the test.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 6 -A "weka.core.EuclideanDistance -R first-last" -I 500 -nu
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
Starts the clustering
00:26:56 - SimpleKMeans
Clusterer output
pCFOS_N 0_0.14 0_0.14 0_0.14 0_0.14 0_0.14 0_0.14 0_0.14
SYP_N 0.43_0.59 0.43_0.59 0.43_0.59 0.43_0.59 0_0.42 0.43_0.59 0_0.42
H3AcK18_N 0_0.21 0_0.21 0_0.21 0_0.21 0_0.21 0_0.21 0_0.21
EGR1_N 0_0.19 0_0.19 0_0.19 0_0.19 0_0.19 0_0.19 0_0.19
H3MeK4_N 0_0.20 0_0.20 0_0.20 0.21_0.30 0_0.20 0_0.20 0_0.20
CaNA_N 1.11_1.61 1.11_1.61 1.11_1.61 1.11_1.61 1.11_1.61 0_1.10 1.61_max
Genotype Control Control Control Ts65Dn Ts65Dn Control Ts65Dn
Treatment Memantine Memantine Saline Saline Memantine Memantine Saline
Behavior S/C S/C C/S S/C C/S S/C C/S
class c-CS-m c-SC-m c-CS-s t-SC-s t-CS-m t-SC-m c-CS-s
Time taken to build model (full training data) : 0.09 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 192 (18%)
1 172 (16%)
2 207 (19%)
3 165 (15%)
4 189 (18%)
5 155 (14%)
Status
OK
Log


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Save...
Filter
Choose
Current relation
Relation: Data
Instances: 1080
Attributes
All
No.
Weka Clusterer Visualize: 01:11:20 - SimpleKMeans (Data_Cortex_...
X: Cluster (Nom)
Y: DYRK1A_N (Nom)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: Data_Cortex_Nuclear-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1...
cluster0
cluster1
cluster2
cluster3
cluster4
cluster5
Class colour
cluster0
cluster1
cluster2
cluster3
cluster4
cluster5
Apply
Stop
Visualize All
Status
OK
Log
x 0

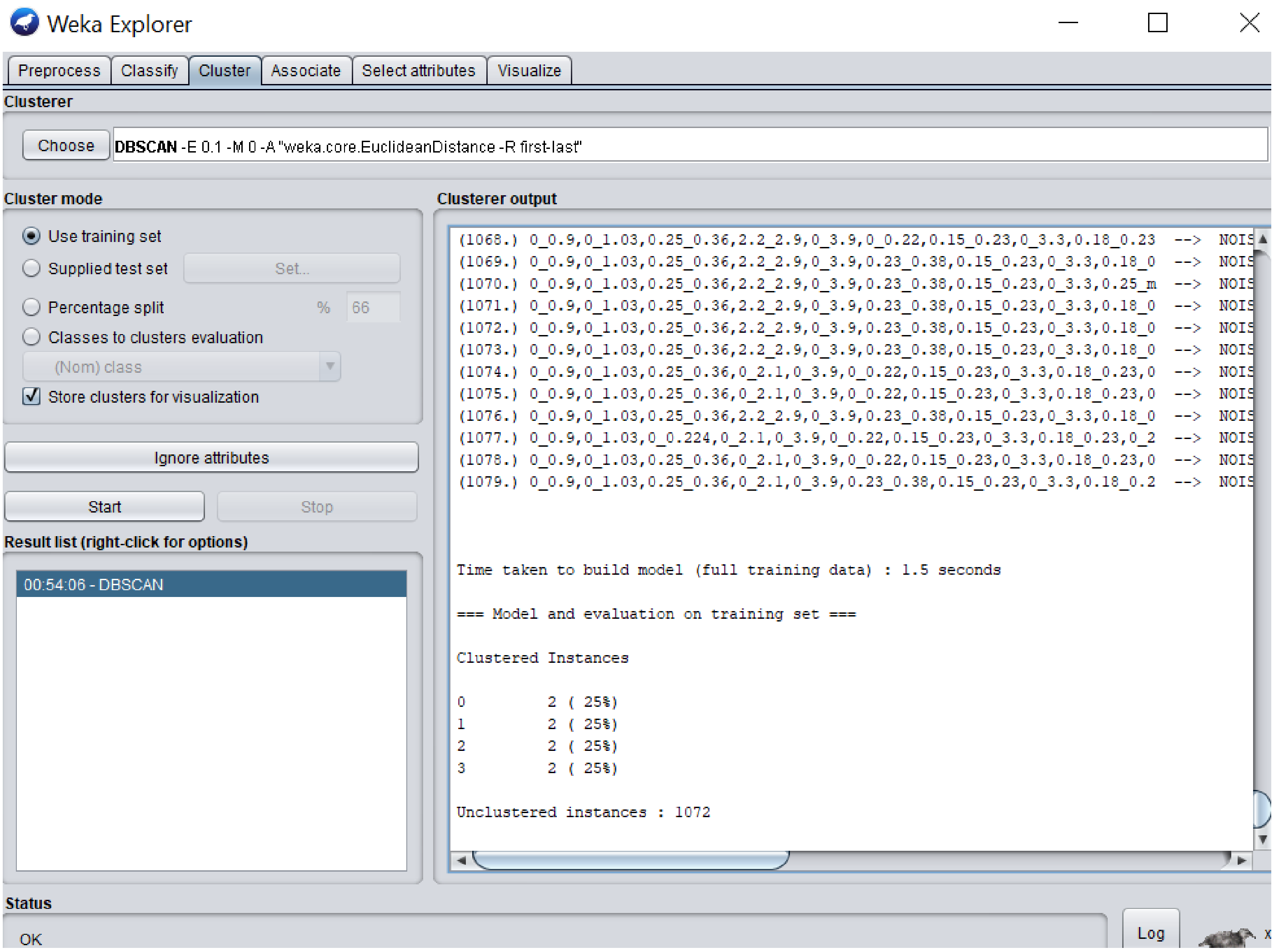
```
D:\Padhai\2 Sem\DMAT\Assignment 1\Kmean\Kmeanset.arff - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
Kmeanset.arff
@attribute PSD95_N {0_1.76,1.76_2.32,2.32_max}
@attribute SNCA_N {0_0.15,0.16_0.20,0.21_max}
@attribute Ubiquitin_N {0_1.13,1.14_1.51,1.52_max}
@attribute pGSK3B_Tyr216_N {0_0.78,0.79_0.99,1.0_max}
@attribute SHH_N {0_0.22,0.23_0.29,0.30_max}
@attribute BAD_N {0_0.15,0.16_0.21,0.22_max}
@attribute BCL2_N {0_0.14,0.15_0.20,0.21_max}
@attribute pS6_N {0_0.09,0.10_0.12,0.13_max}
@attribute pCFOS_N {0_0.14,0.15_0.19,0.20_max}
@attribute SYP_N {0_0.42,0.43_0.59,0.60_max}
@attribute H3AcK18_N {0_0.21,0.22_0.34,0.35_max}
@attribute EGR1_N {0_0.19,0.20_0.27,0.28_max}
@attribute H3MeK4_N {0_0.20,0.21_0.30,0.31_max}
@attribute CaNA_N {0_1.10,1.11_1.61,1.61_max}
@attribute Genotype {Control,Ts65Dn}
@attribute Treatment {Memantine,Saline}
@attribute Behavior {C/S,S/C}
@attribute class {c-CS-m,c-SC-m,c-CS-s,c-SC-s,t-CS-m,t-SC-m,t-CS-s,t-SC-s}
@attribute Cluster {cluster0,cluster1,cluster2,cluster3,cluster4,cluster5}

@data
0,0_0.9,0_1.03,0.37_max,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.38_max,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,2.3_max,0.3_0.4,0.9_max,0_0.8,0.34_0.46,0_0.18,1.27_2.03,2.49_3.84,0.93_
1,0_0.9,0_1.03,0.37_max,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.38_max,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.3_0.4,0.9_max,0_0.8,0.34_0.46,0_0.18,1.27_2.03,2.49_3.84,0.93_1
2,0_0.9,0_1.03,0.37_max,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.38_max,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.9_max,0_0.8,0.34_0.46,0_0.18,1.27_2.03,2.49_3.84,0.93_
3,0_0.9,0_1.03,0.25_0.36,2.2_2.9,4.0_6.2,0.23_0.38,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0_0.18,1.27_2.03,2.49_3.84,0.9
4,0_0.9,0_1.03,0.25_0.36,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0_0.18,1.27_2.03,2.49_3.84,0.93_1
5,0_0.9,0_1.03,0.25_0.36,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0_0.18,0_1.26,2.49_3.84,0.93_1.70
6,0_0.9,0_1.03,0.25_0.36,2.2_2.9,4.0_6.2,0.23_0.38,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0.19_0.25,0_1.26,2.49_3.84,0.9
7,0_0.9,0_1.03,0.25_0.36,2.2_2.9,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0.9_1.1,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0_0.18,0_1.26,2.49_3.84,0.93_1.70
8,0_0.9,0_1.03,0.25_0.36,0_2.1,4.0_6.2,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0_0.8,0_0.6,1.2_1.9,1.4_2.2,0.3_0.4,0.5_0.8,0_0.8,0_0.33,0_0.18,0_1.26,0_2.48,0.93_1.70,0.16_0.27
9,0_0.9,0_1.03,0.25_0.36,0_2.1,0_3.9,0.23_0.38,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0_0.8,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0.19_0.25,0_1.26,0_2.48,0.93_1.70,0
10,0_0.9,0_1.03,0.25_0.36,0_2.1,0_3.9,0.23_0.38,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0_0.8,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0.19_0.25,0_1.26,0_2.48,0.93_1.70,
11,0_0.9,0_1.03,0.25_0.36,0_2.1,0_3.9,0_0.22,0.15_0.23,0_3.3,0.18_0.23,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0_0.8,0_0.6,1.2_1.9,1.4_2.2,0.5_max,0.5_0.8,0_0.8,0.34_0.46,0.19_0.25,0_1.26,0_2.48,0.93_1.70,0.1
12,0_0.9,0_1.03,0.25_0.36,0_2.1,0_3.9,0.23_0.38,0.15_0.23,0_3.3,0_0.17,0_2.3,0_1.2,0.2_0.3,0.28_0.37,0.2_0.3,0_0.8,0_0.6,1.2_1.9,1.4_2.2,0.3_0.4,0.5_0.8,0_0.8,0.34_0.46,0.19_0.25,0_1.26,0_2.48,0.93_1.70,0.28
```

In Kmean, set the number of cluster 6 in which are very small in size, cluster 2 is the biggest among them with 19%. With visualization, it is difficult to determine graphically as they are very small in size. In the text editor, we can see that each instance have now its cluster.

```
6_2.32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster1
_2.32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster1
4_2.0,1.76_2.32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,
1.76_2.32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0_0.42,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster1
.32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster5
15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster5
32,0_0.15,0_1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0_0.42,0_0.21,0_0.19,0_0.20,1.11_1.61,Control,Memantine,C/S,c-CS-m,cluster5
1.13,0.79_0.99,0_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0_0.42,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster5
_0.22,0_0.15,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0_0.19,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster5
_1.13,0.79_0.99,0_0.22,0.16_0.21,0_0.14,0_0.09,0_0.14,0_0.42,0_0.21,0_0.19,0_0.20,1.11_1.61,Control,Memantine,C/S,c-CS-m,cluster2
.13,0.79_0.99,0_0.22,0.16_0.21,0_0.14,0_0.09,0_0.14,0_0.42,0_0.21,0_0.19,0_0.20,1.11_1.61,Control,Memantine,C/S,c-CS-m,cluster2
79_0.99,0_0.22,0.16_0.21,0_0.14,0.10_0.12,0_0.14,0.43_0.59,0_0.21,0.20_0.27,0_0.20,1.61_max,Control,Memantine,C/S,c-CS-m,cluster5
.99,0_0.22,0.16_0.21,0_0.14,0_0.09,0_0.14,0.43_0.59,0_0.21,0.20_0.27,0.21_0.30,1.11_1.61,Control,Memantine,C/S,c-CS-m,cluster2
```

DBSCAN

Epsilon: 0.02
minPoint: 2
Number of Clusters: 4
Unclustered: 1072



With DBscan the dataset generates only 4 clusters with equal size as 2. Tried with other values it generates 0 clusters.