Experiment-2.3

Student Name: Ashish Kumar
UID: 23MAI10008
Branch: CSE AIML
Section/Group: 23MAI-1
Semester: 01
Date of Performance: 16/10/2023
Subject Name: Artificial Intelligence Lab
Subject Code: 23CSH-621

Aim of the Experiment :

Aim of the Experiment is to apply K means clustering to sample dataset from UCI Machine Learning Repository. Analyze how parameter tuning affects the results of K means clustering algorithm.

Objective of the Experiment :

Task to be done for this experiment is that we have to perform following tasks:

a) Apply K means clustering to sample dataset from UCI Machine Learning Repository.

b) Analyze the performance of K means clustering algorithm.

Algorithm/ Steps for Experiment :

Step 1: Download the **Cancer dataset** from UCI Machine Learning repository.

Step 2: Open the WEKA Tool and open the **'Explorer'** tab.

Home | breast-cancer.csv

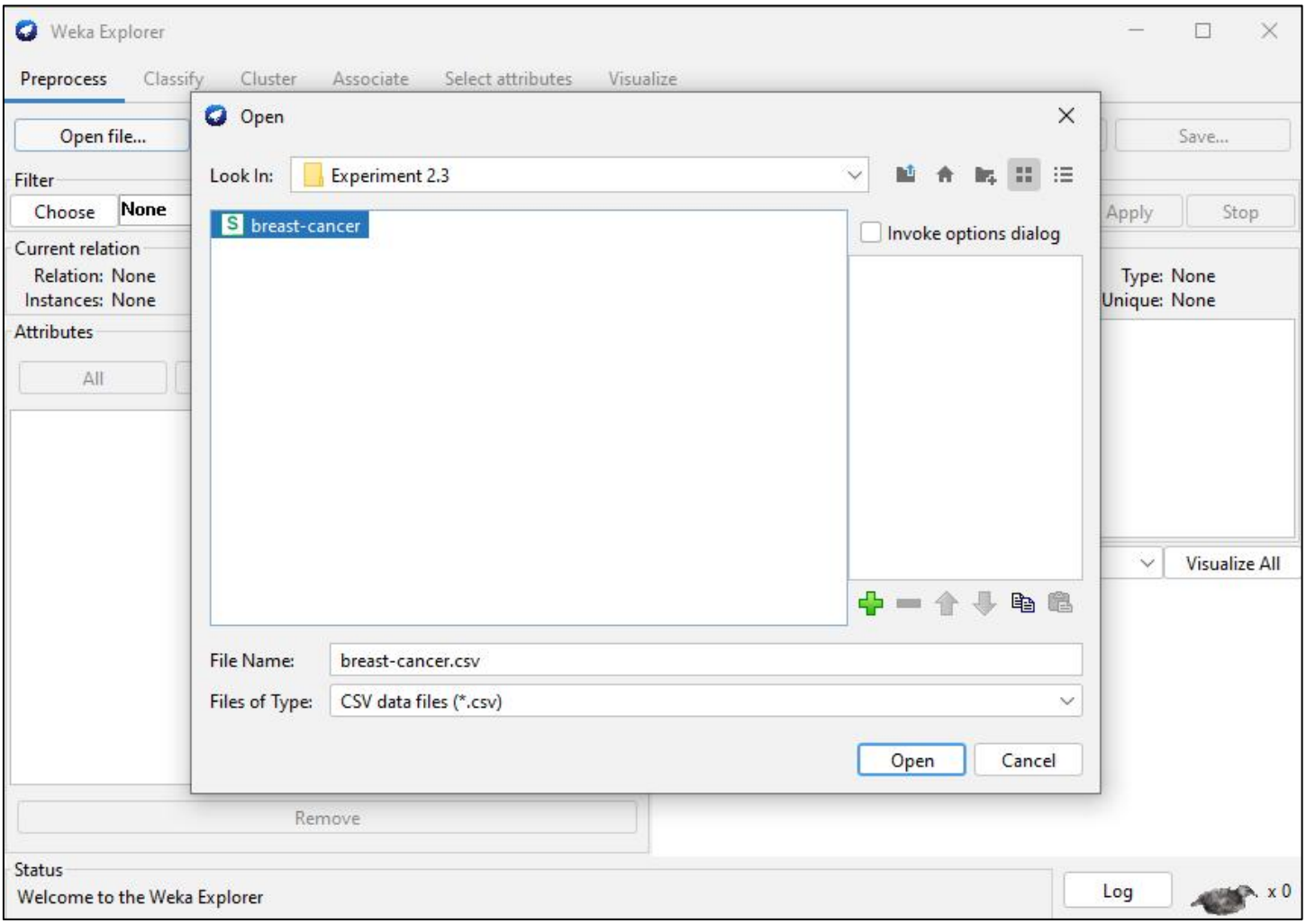
Menu | Home | Insert | Page Layout | Formulas | Data | Review | View | Tools | Smar

Format Painter | Paste | Calibri | 11 | Wrap Text | Orientation | Merge and Center

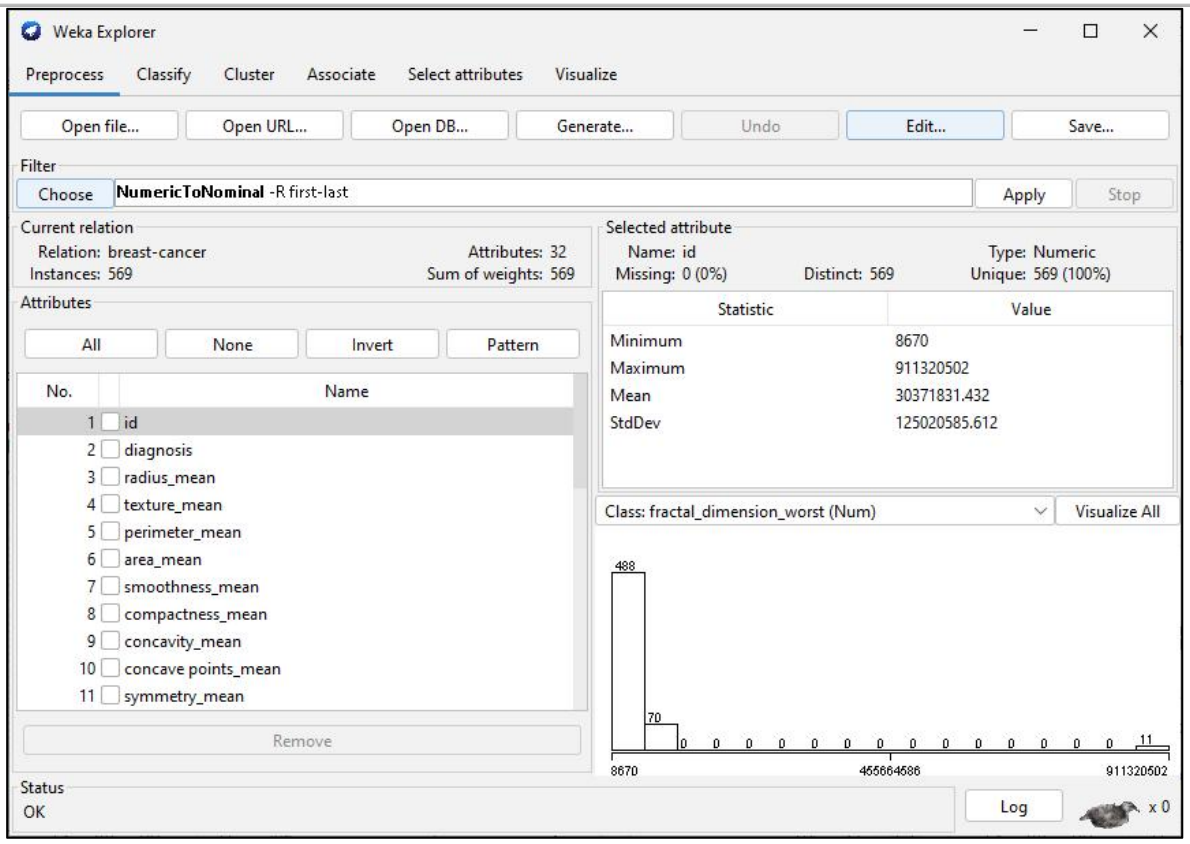
A1 | id

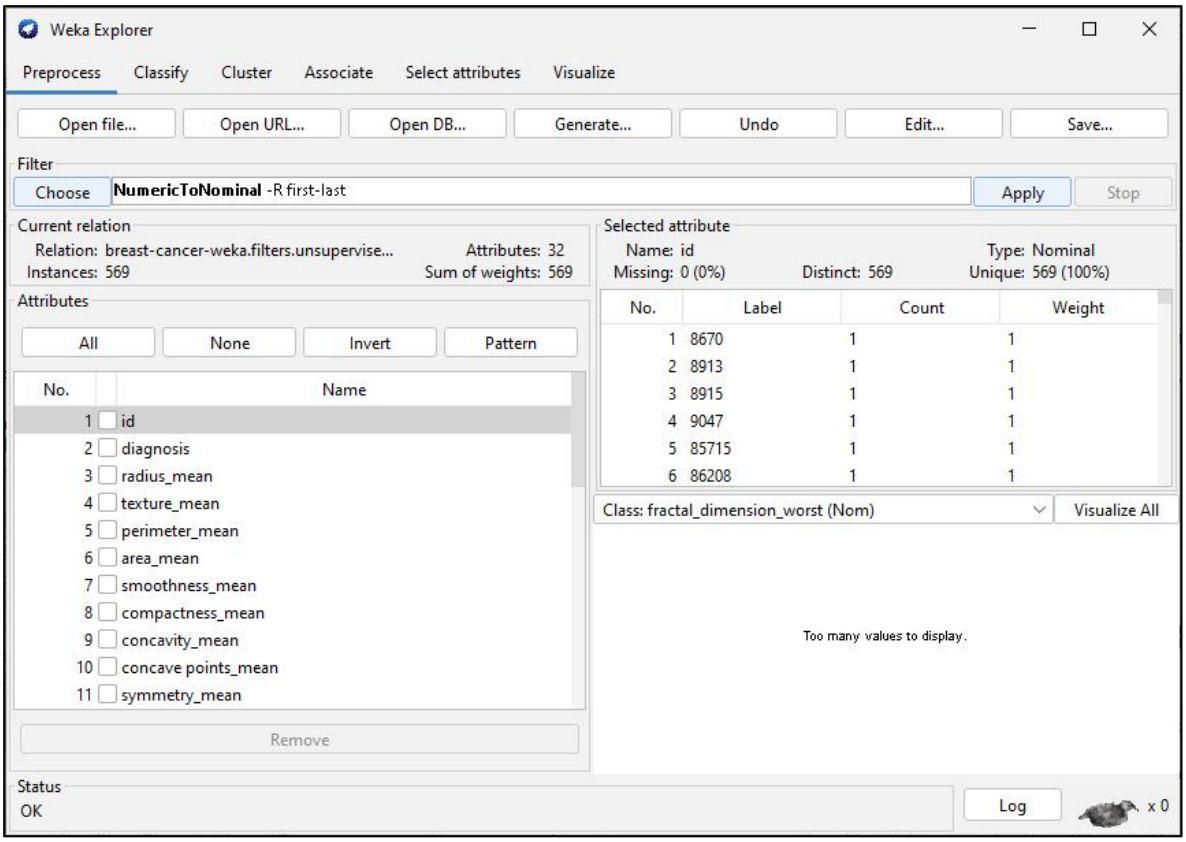
	A	B	C	D	E	F	G	H	I	J	K	L
1	id	diagnosis	radius_mea	texture_me	perimeter_	area_mean	smoothnes	compactne	concavity_r	concave po	symmetry_	fractal_dim
2	842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871
3	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667
4	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999
5	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744
6	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883
7	843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613
8	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742
9	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451
10	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389
11	84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243
12	845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697
13	84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082
14	846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078
15	846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338

Step 3: Click on the **'Open file'** Option >> Select Cancer dataset >> Click on **Open**.

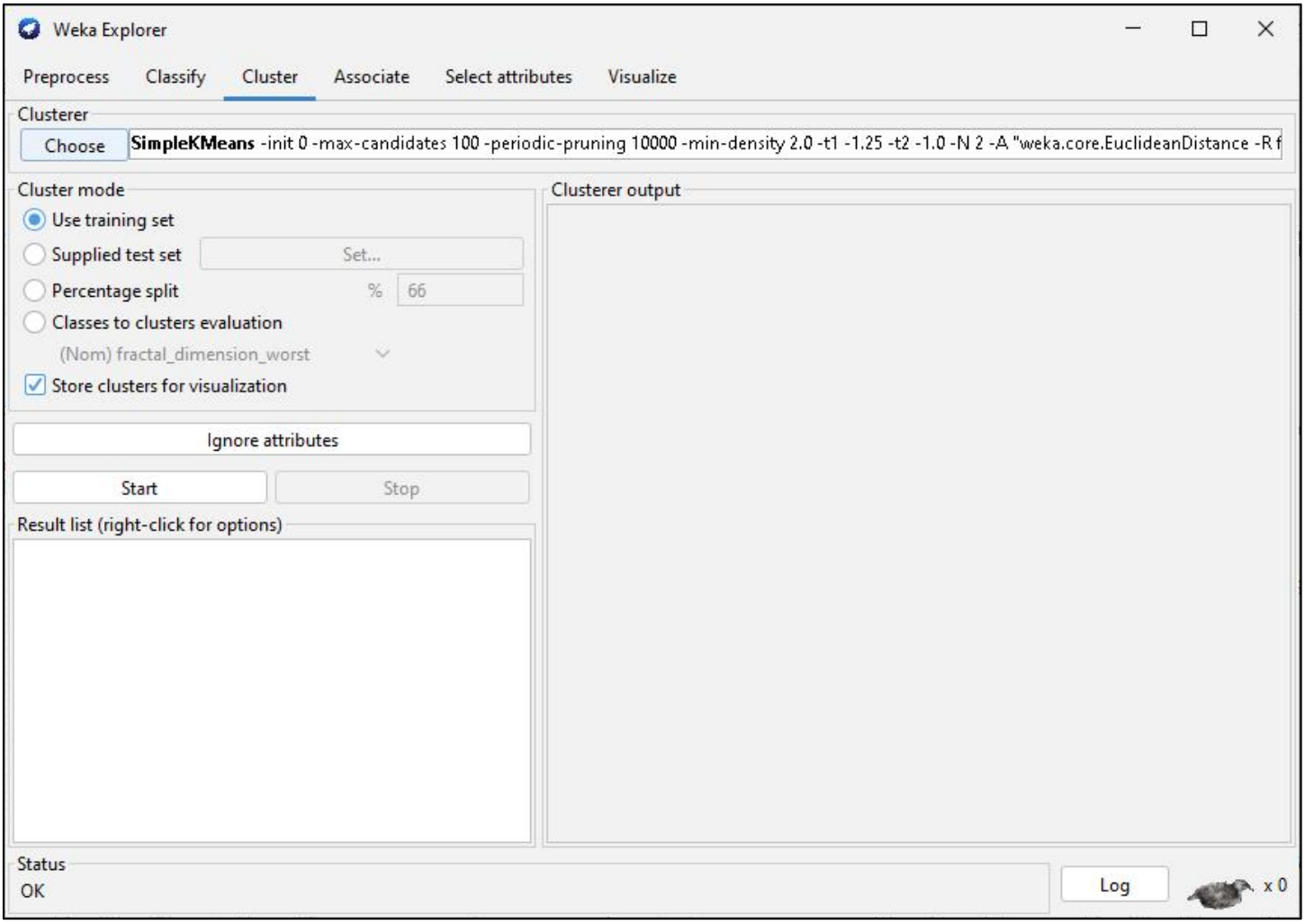


Step 4: Change the dataset from Numeric to Nominal. In **'Filter'** Section, click on **'Choose'** >> Unsupervised >> Attribute >> **Numeric to Nominal**. Click on **'Apply'** Option.

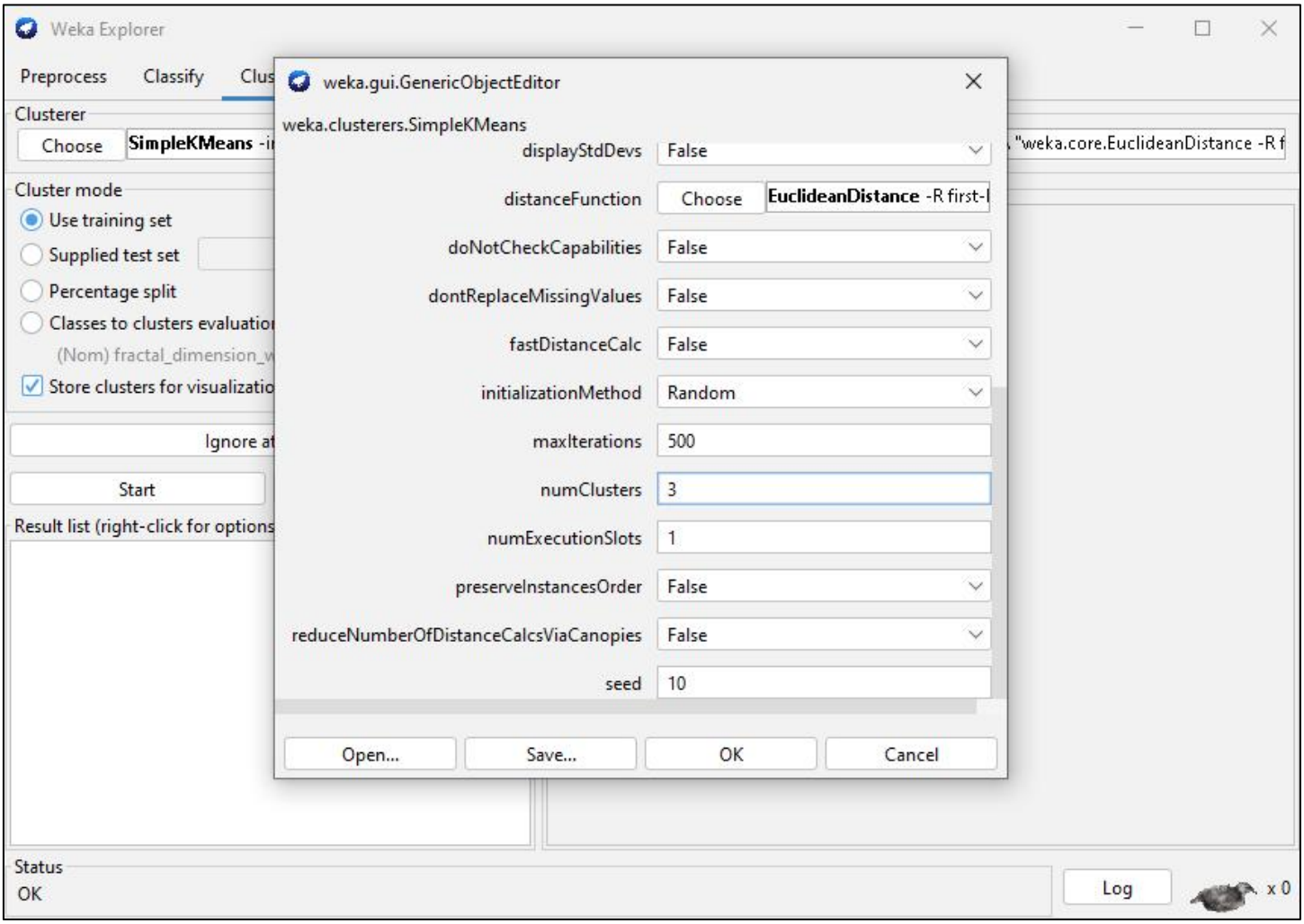

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Open file... Open URL... Open DB... Generate... Undo Edit... Save...
Filter
Choose NumericToNominal -R first-last Apply Stop
Current relation
Relation: breast-cancer Attributes: 32
Instances: 569 Sum of weights: 569
Attributes
All None Invert Pattern
No. Name
1 id
2 diagnosis
3 radius_mean
4 texture_mean
5 perimeter_mean
6 area_mean
7 smoothness_mean
8 compactness_mean
9 concavity_mean
10 concave points_mean
11 symmetry_mean
Remove
Selected attribute
Name: id Type: Numeric
Missing: 0 (0%) Distinct: 569 Unique: 569 (100%)
Statistic Value
Minimum 8670
Maximum 911320502
Mean 30371831.432
StdDev 125020585.612
Class: fractal_dimension_worst (Num) Visualize All
488
70
0 0 0 0 0 0 0 0 0 0 0 0 0 0 11
8670 455664586 911320502
Status
OK
Log x 0


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Open file... Open URL... Open DB... Generate... Undo Edit... Save...
Filter
Choose NumericToNominal -R first-last Apply Stop
Current relation
Relation: breast-cancer-weka.filters.unsupervise... Attributes: 32
Instances: 569 Sum of weights: 569
Attributes
All None Invert Pattern
No. Name
1 id
2 diagnosis
3 radius_mean
4 texture_mean
5 perimeter_mean
6 area_mean
7 smoothness_mean
8 compactness_mean
9 concavity_mean
10 concave points_mean
11 symmetry_mean
Remove
Selected attribute
Name: id Type: Nominal
Missing: 0 (0%) Distinct: 569 Unique: 569 (100%)
No. Label Count Weight
1 8670 1 1
2 8913 1 1
3 8915 1 1
4 9047 1 1
5 85715 1 1
6 86208 1 1
Class: fractal_dimension_worst (Nom) Visualize All
Too many values to display.
Status
OK
Log x 0

Step 5: To apply the K Means Clustering, Click on the **'Cluster'** Tab >> Choose '**K Means**' from the **'Clusterer'** Section.

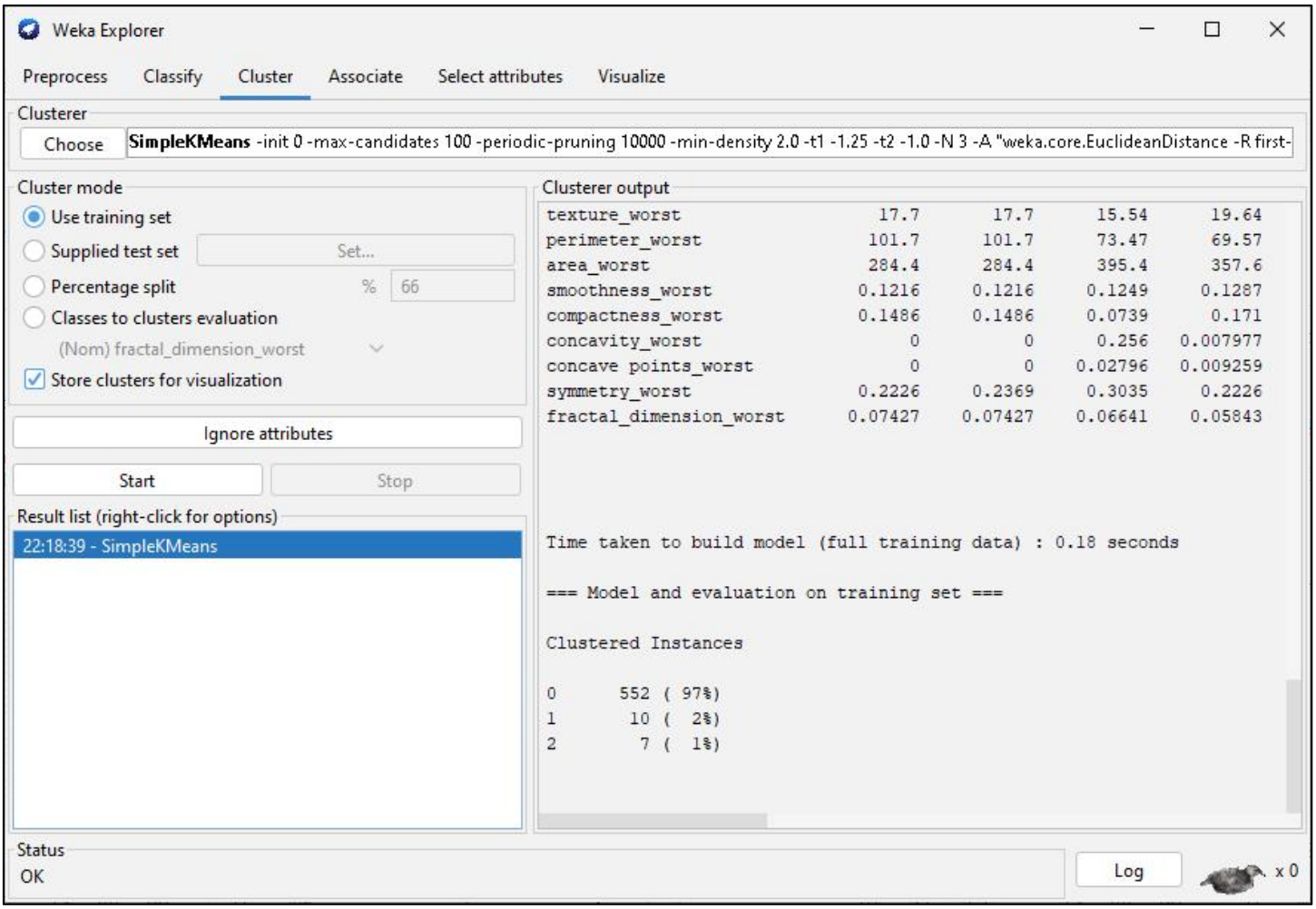


Step 6: Click on K Means, a dialog box will appear in that >> Set **numClusters** to 3.



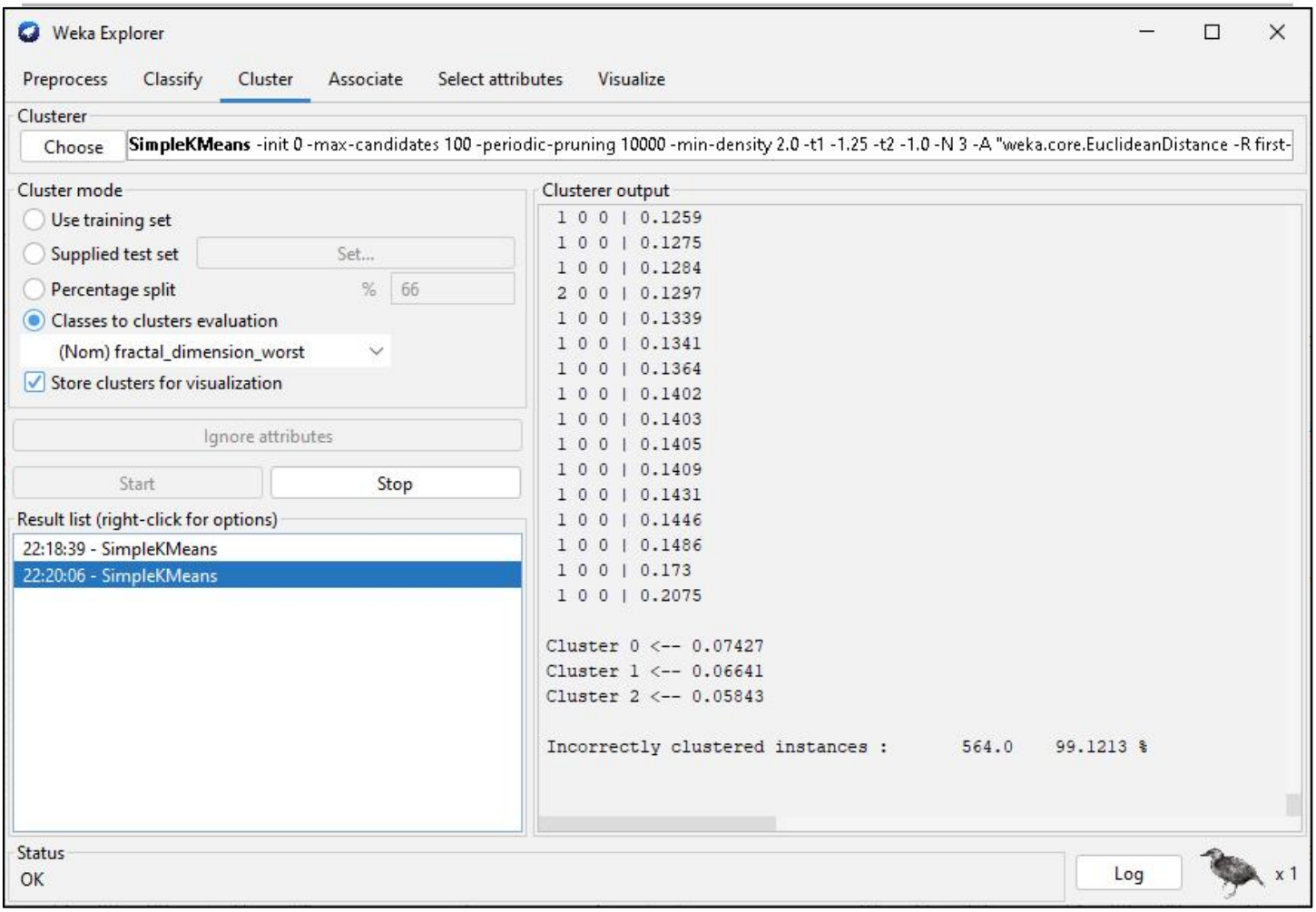
Step 7: In **'Cluster mode'** Section, Click on **'Use training set'** and Click on Start.

Step 8: Note the values for the applied method.



Step 9: In 'Cluster mode' Section, Click on 'Classes to clusters evaluation' and click on Start.

Step 10: Note the values for the applied method.



Learning outcomes (What I have learnt):

1. I learnt about the WEKA Tool and its applications.
2. I learnt about how to use Explorer Tab in WEKA Tool.
3. I learnt about how to change attributes from Numeric to Nominal.
4. I learnt about how to use the K Means Clustering in WEKA Tool.
5. I learnt about how to compare the accuracy of each clusters.