

TUGAS KECERDASAN KOMPUTASIONAL

VISUALISING TREE USING WEKA APPLICATION



Dosen :

Anny Yuniarti, S.Kom.,M.Comp.Sc

Oleh :

Findryan Kurnia P 5115100035

INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA

Surabaya

2017

1. ABALONE DATA

1.1 Penjelasan Data Set

- Number of Instances: 4177
- Number of Attributes: 8
- Missing Attribute Values: None
- Attribute information:

| NO | Name | Data type |
|----|---------|------------|
| 1 | Sex | nominal |
| 2 | Length | continuous |
| 3 | Diam | Continuous |
| 4 | Height | Continuous |
| 5 | Whole | Continuous |
| 6 | Shucked | Continuous |
| 7 | Viscera | Continuous |
| 8 | Shell | Continuous |
| 9 | Rings | integer |

- Class distribution:

| NO | Class | Jumlah |
|----|-------|--------|
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 3 | 15 |
| 4 | 4 | 57 |
| 5 | 5 | 115 |
| 6 | 6 | 259 |
| 7 | 7 | 391 |
| 8 | 8 | 568 |
| 9 | 9 | 689 |
| 10 | 10 | 634 |
| 11 | 11 | 487 |
| 12 | 12 | 267 |
| 13 | 13 | 203 |
| 14 | 14 | 126 |
| 15 | 15 | 103 |
| 16 | 16 | 67 |
| 17 | 17 | 58 |
| 18 | 18 | 42 |
| 19 | 19 | 32 |
| 20 | 20 | 26 |
| 21 | 21 | 14 |
| 22 | 22 | 6 |
| 23 | 23 | 9 |
| 24 | 24 | 2 |
| 25 | 25 | 1 |
| 26 | 26 | 1 |
| 27 | 27 | 2 |
| 28 | 28 | 0 |
| 29 | 29 | 1 |

1.2 Langkah-Langkah :

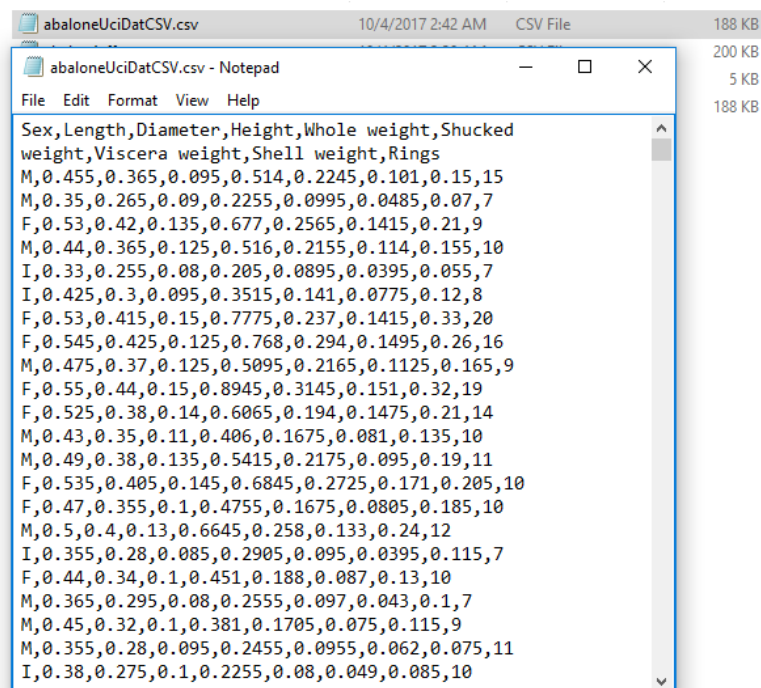
- a) Download data Abalon UCI pada

<http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/>

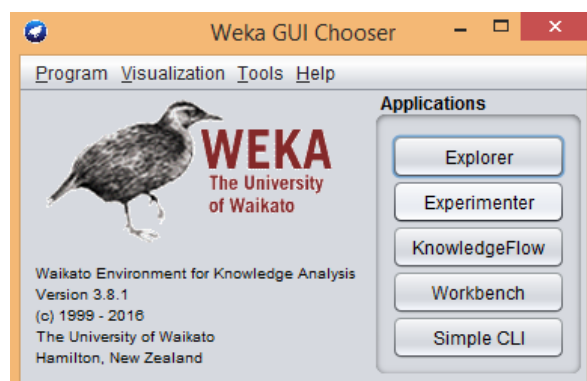
Index of /ml/machine-learning-databases/abalone

| Name | Last modified | Size | Description |
|----------------------------------|-------------------------------|----------------------|-----------------------------|
| Parent Directory | - | - | - |
| Index | 03-Dec-1996 04:06 | 114 | |
| abalone.data | 12-Jun-1996 14:45 | 187K | |
| abalone.names | 12-Jun-1996 15:52 | 4.2K | |

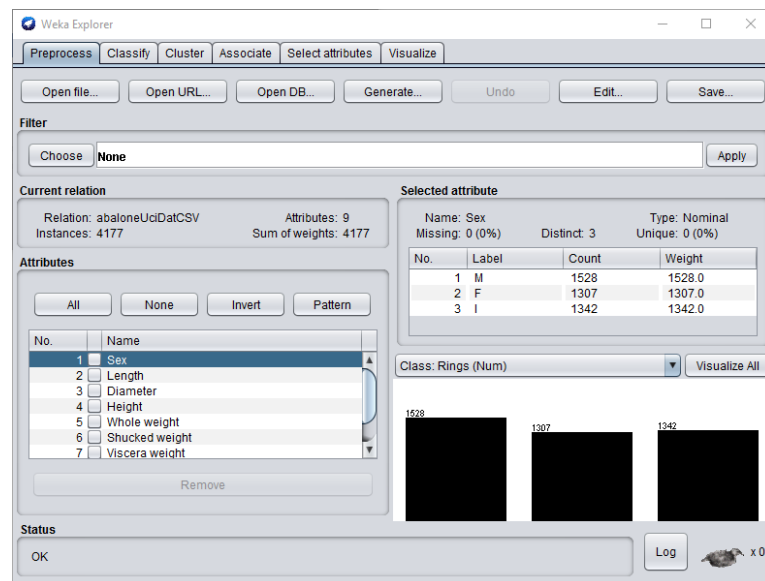
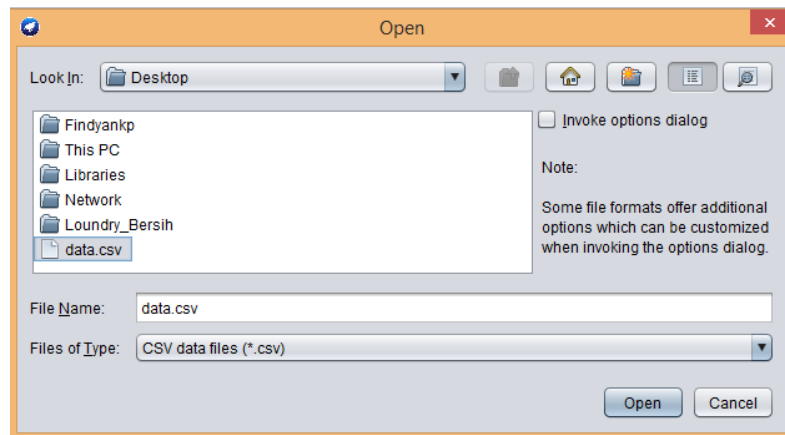
- b) Simpan data dalam bentuk file ekstensi .csv agar weka dapat membaca file tersebut



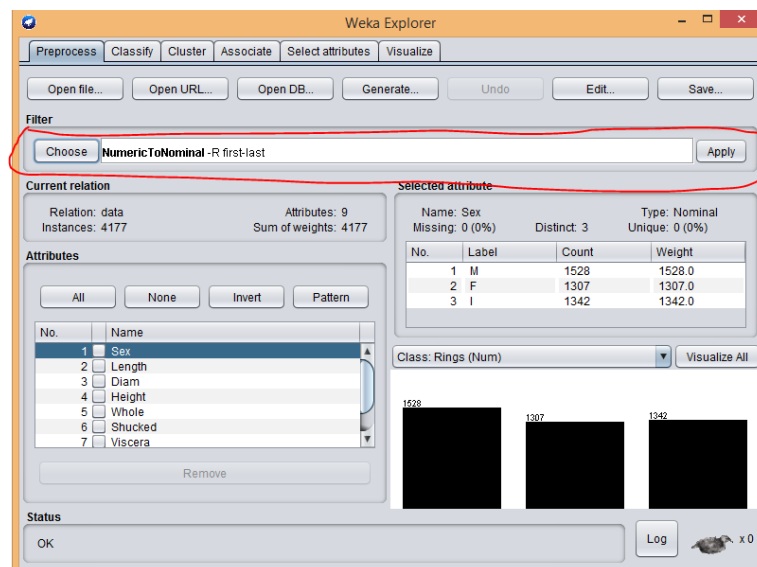
- c) Setelah itu buka menu weka terdapat pilihan-pilihan



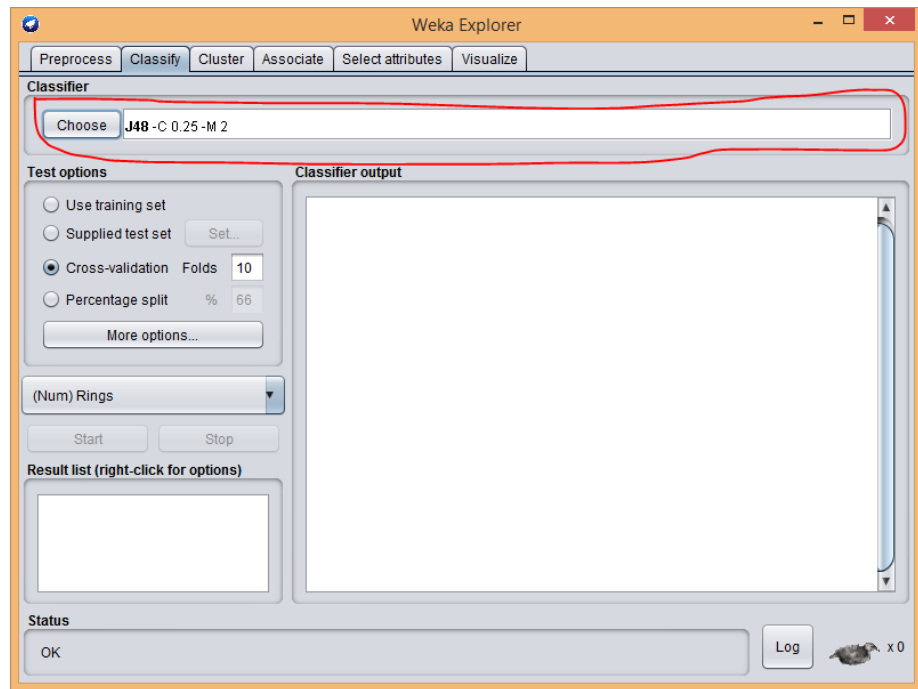
- d) Pilih Menu explorer kemudian open file pilih format file csv -dan masukan data csv yang tadi telah kita buat.



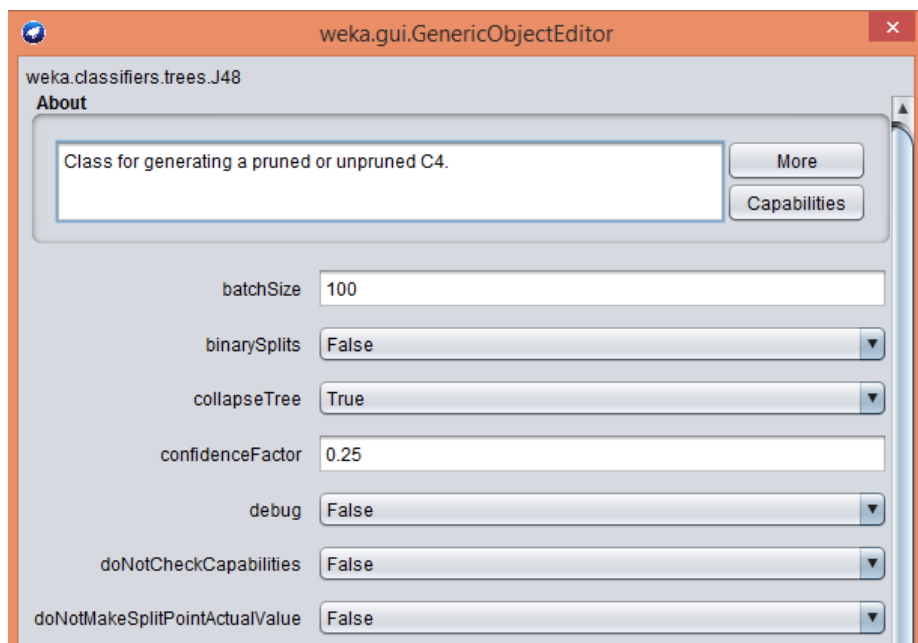
e) Tambahkan filter NumericalToNominal ini dikarenakan kita menggunakan file .CSV

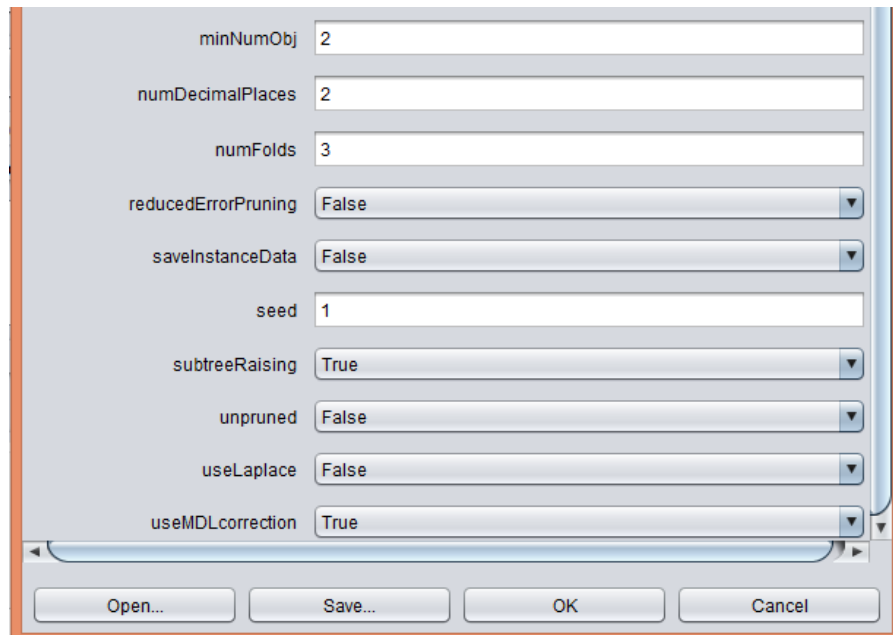


f) Masuk ke tab classify, Tambahkan Calssifier Tree lalu -> J48 -C 0.25 -M 2

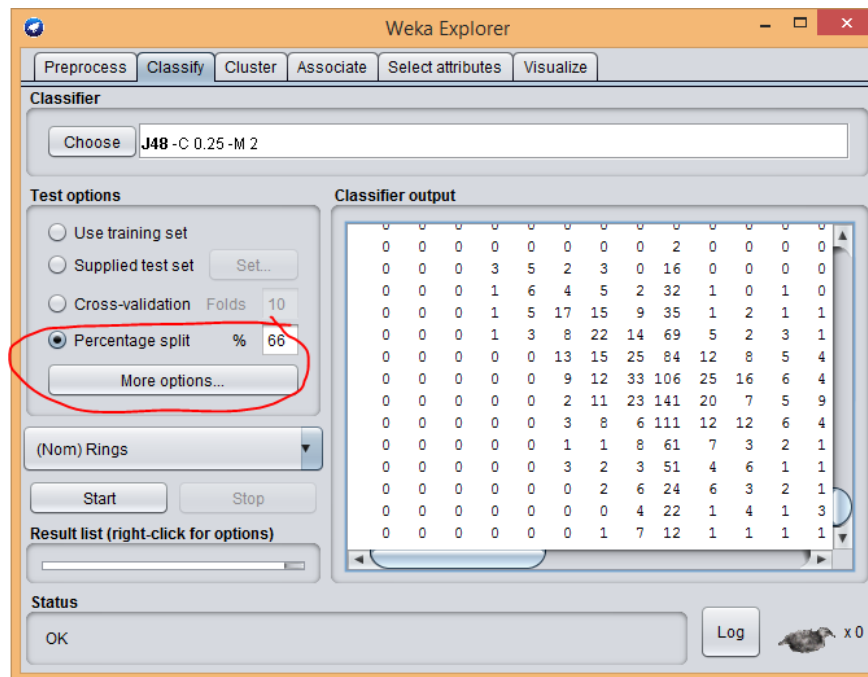


g) Mengatur setiap parameter yang hendak digunakan

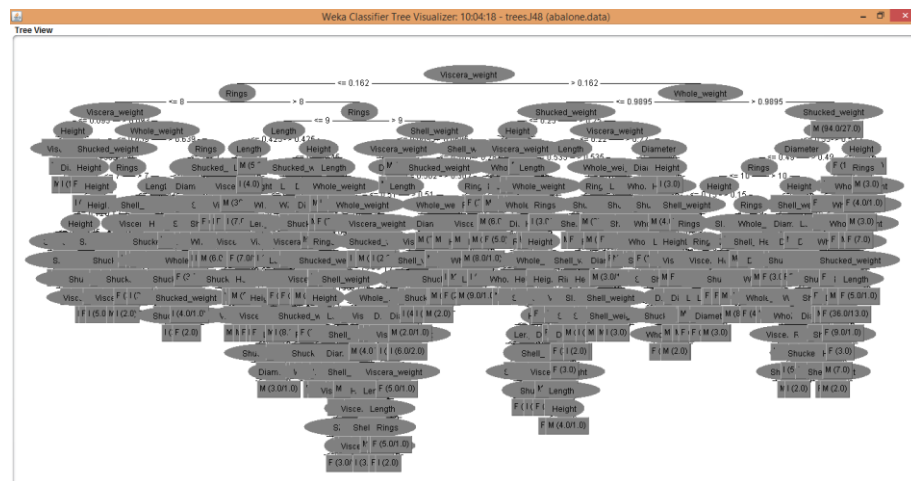




- h) Dalam test options kita memilih percentage split sebesar 66% karena menginginkan data training sebanyak 2/3 dari datanya dan 1/3 sebagai data test. Kemudian Klik start

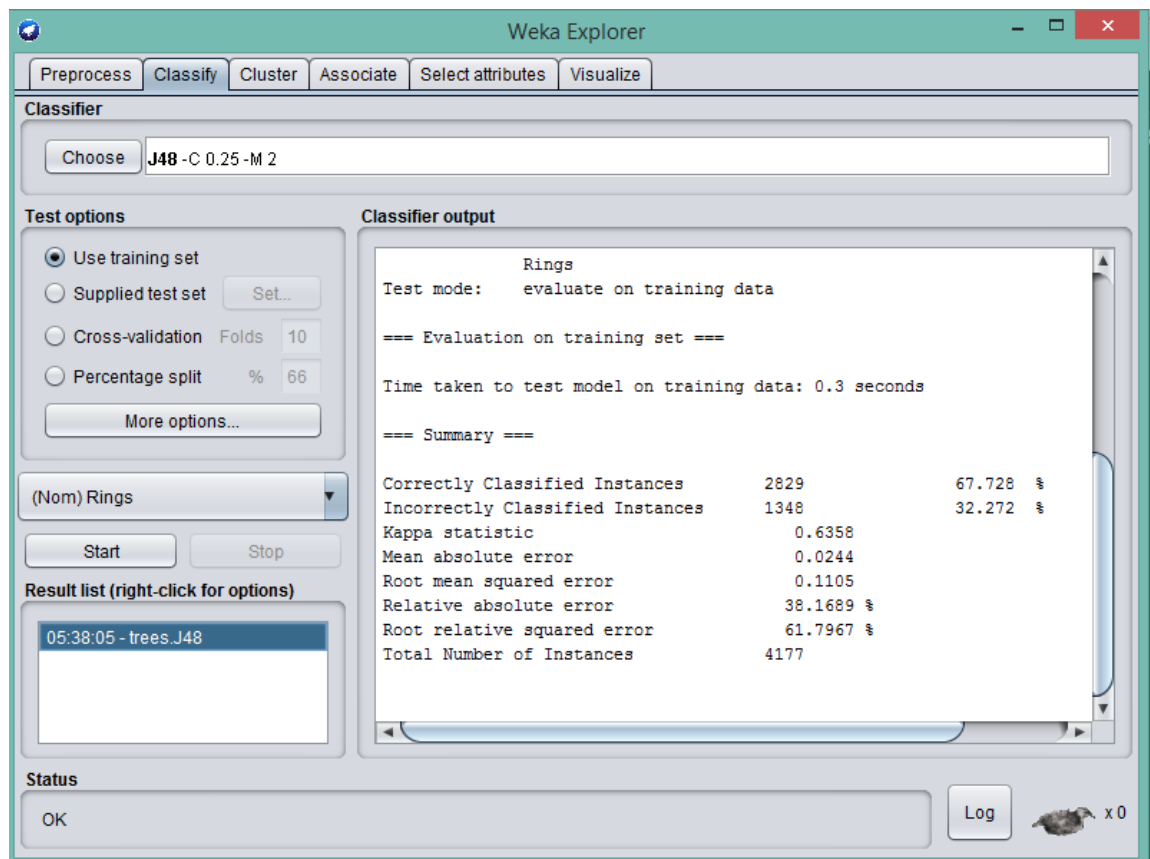


- i) Setelah mengklik start maka akan keluar hasil pada classifier output kemudian klik kanan pada Result List dan Pilih Visualize tree



- Number of Leaves : 4084
 Size of the tree : 4110

| | | |
|----------------------------------|------------|-----------|
| Correctly Classified Instances | 214 | 15.0704 % |
| Incorrectly Classified Instances | 1206 | 84.9296 % |
| Kappa statistic | 0.0222 | |
| Mean absolute error | 0.0622 | |
| Root mean squared error | 0.2108 | |
| Relative absolute error | 97.0872 % | |
| Root relative squared error | 117.7184 % | |
| Total Number of Instances | 1420 | |
- Dari hasilnya terlihat klasifikasi yang tepat sangat kecil persentasinya 15.0704 % .Ini bisa disimpulkan tidak cocok disimpulkan sebagai tree.
- Namun apabila menggunakan parameter lain seperti akan menghasilkan hasil yang lebih besar dikarenakan data set yang digunakan sebagai training juga digunakan sebagai data test sehingga membuat hasil lebih baik



2. BREAST CANCER DATA

2.1 Penjelasan Data Set

- Number of Instances: 286
- Number of Attributes: 9 + the class attribute
- Missing Attribute Values: 9 / 0.3%
- Attribute information:

| NO | Name | Data type |
|----|-------------|-----------|
| 1 | Age | Enum |
| 2 | Menopause | Enum |
| 3 | Tumor-size | Enum |
| 4 | Inv- nodes | Enum |
| 5 | Node-caps | Enum |
| 6 | Deg-malig | Enum |
| 7 | Breast | Enum |
| 8 | Breast-quad | Enum |
| 9 | Irradiat | Enum |
| 10 | class | Enum |

- Class distribution:

| NO | Class | Jumlah |
|----|----------------------|--------|
| 1 | recurrence-events | 85 |
| 2 | no-recurrence-events | 201 |

2.2 Langkah-Langkah

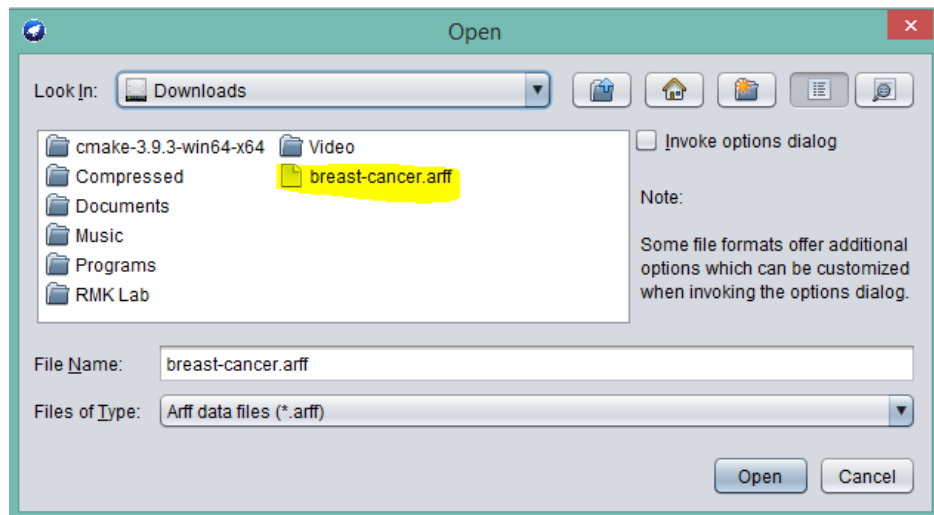
- a) Ambil data dari breast-cancer di

<http://repository.seasr.org/Datasets/UCI/arff/breast-cancer.arff>

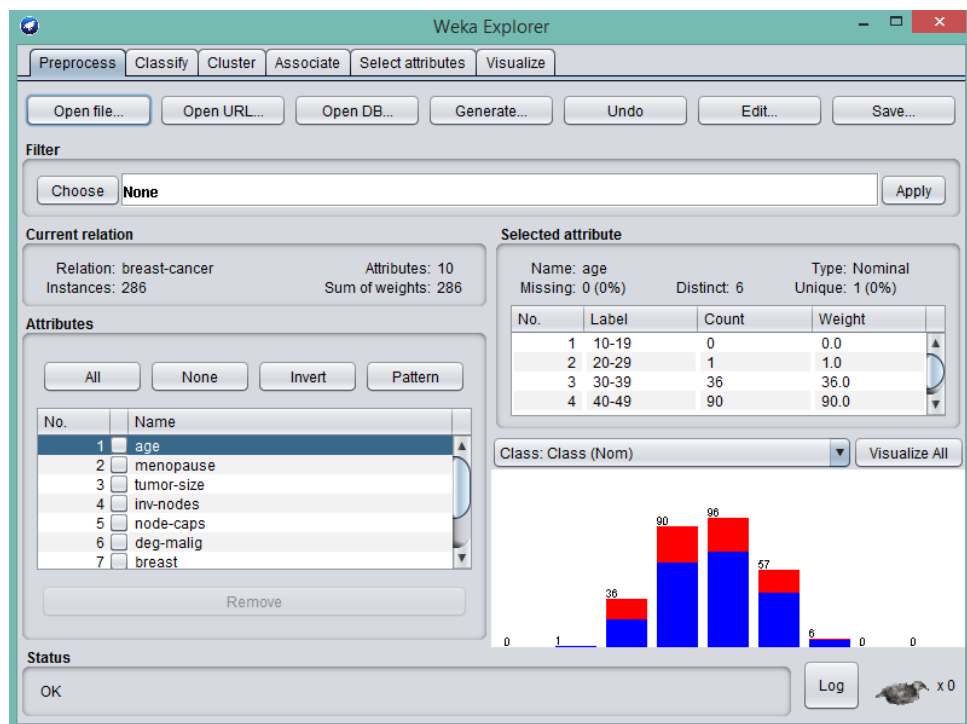
Index of /Datasets/UCI/arff/

| | | |
|---------------------------------------|-------------------|--------|
| ../ | | |
| anneal.ORIG.arff | 15-Jul-2008 15:27 | 84017 |
| anneal.arff | 15-Jul-2008 15:27 | 143336 |
| arrhythmia.arff | 15-Jul-2008 15:27 | 418221 |
| audiology.arff | 15-Jul-2008 15:27 | 45903 |
| autos.arff | 15-Jul-2008 15:27 | 30676 |
| balance-scale.arff | 15-Jul-2008 15:27 | 8714 |
| breast-cancer.arff | 15-Jul-2008 15:27 | 29418 |
| breast-w.arff | 15-Jul-2008 15:27 | 19167 |
| bridges_version1.arff | 15-Jul-2008 15:27 | 11911 |
| bridges_version2.arff | 15-Jul-2008 15:27 | 12313 |
| car.arff | 15-Jul-2008 15:27 | 55474 |

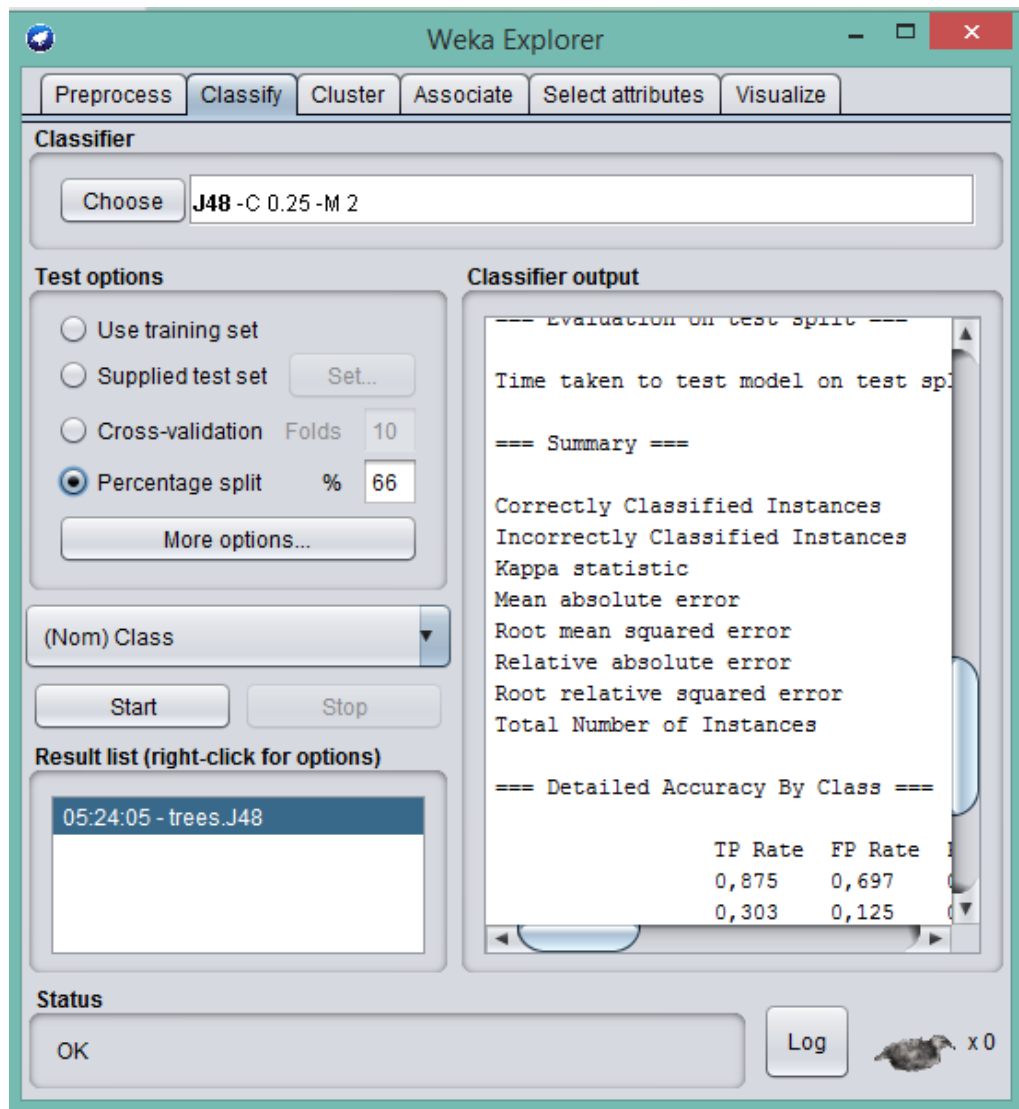
- b) Kemudian ikuti langkah-langkah seperti dengan abalone data diatas. Tetapi data yang ada di repo sudah berformat “arff” yang didukung oleh Weka sehingga tidak perlu melakukan convert ke “csv”



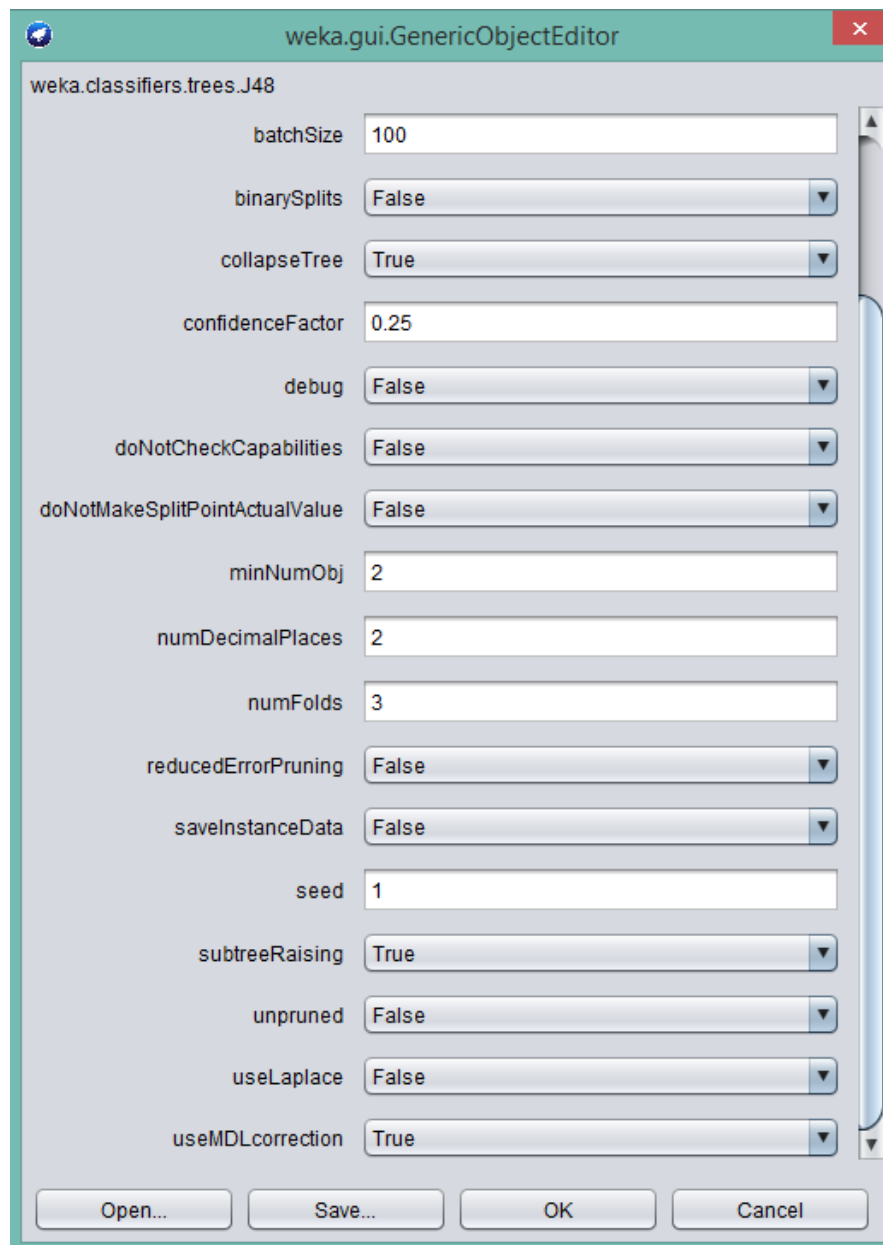
c) Maka akan menampilkan hasil seperti berikut



d) Kemudian masuk ke Classify dan mrngatur sebagai berikut

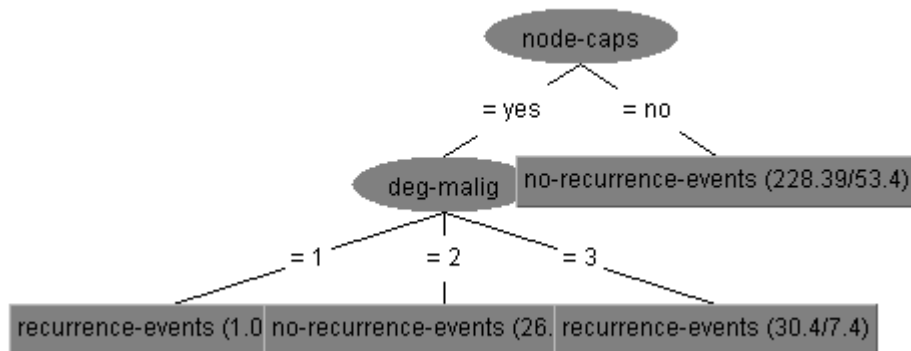
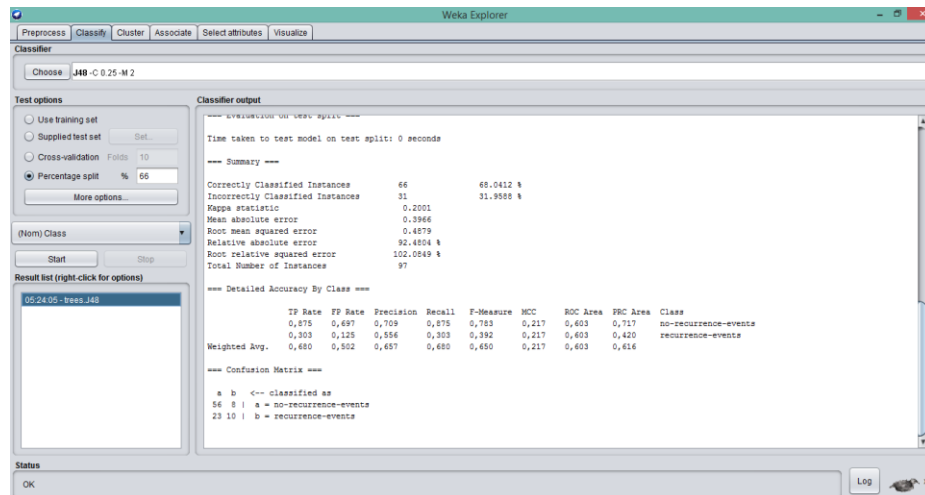


e) Dengan menggunakan parameter



2.3 Hasil dan Kesimpulan

Number of Leaves : 4
 Size of the tree : 6
 Correctly Classified Instances 66 68.0412 %
 Incorrectly Classified Instances 31 31.9588 %
 Kappa statistic 0.2001
 Mean absolute error 0.3966
 Root mean squared error 0.4879
 Relative absolute error 92.4804 %
 Root relative squared error 102.0849 %
 Total Number of Instances 97



- Dalam penggunaan data pada breast cancer, setelah pemrosesan ditemukan persentase ketepatan yang lumayan tinggi yaitu 68.0412 %. Ini dikarenakan data ini cocok menggunakan algoritma ini ataupun juga karena file inputnya langsung dalam format arff.
- Namun apabila menggunakan parameter lain seperti akan menghasilkan hasil yang lebih besar dikarenakan data set yang digunakan sebagai training juga digunakan sebagai data test sehingga membuat hasil lebih baik

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Classifier

ChooseZeroR

Test options

☒ Use training set

☐ Supplied test set

☐ Cross-validation

☐ Percentage split

Set...

Folds10

%66

More options...

(Nom) Class

StartStop

Result list (right-click for options)

05:33:23 - rules.ZeroR

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.03 seconds

=== Summary ===

| | | |
|----------------------------------|--------|-----------|
| Correctly Classified Instances | 201 | 70.2797 % |
| Incorrectly Classified Instances | 85 | 29.7203 % |
| Kappa statistic | 0 | |
| Mean absolute error | 0.4183 | |
| Root mean squared error | 0.457 | |
| Relative absolute error | 100 | % |
| Root relative squared error | 100 | % |
| Total Number of Instances | 286 | |

=== Detailed Accuracy By Class ===

Status

OK

Log

x 0