

“Visualising Tree Using Weka Application”

ABALONE DATA

Diajukan untuk memenuhi Tugas Kecerdasan Komputasional kelas F



Dosen :

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

Oleh :

Mutia Rahmi Dewi

5115100048

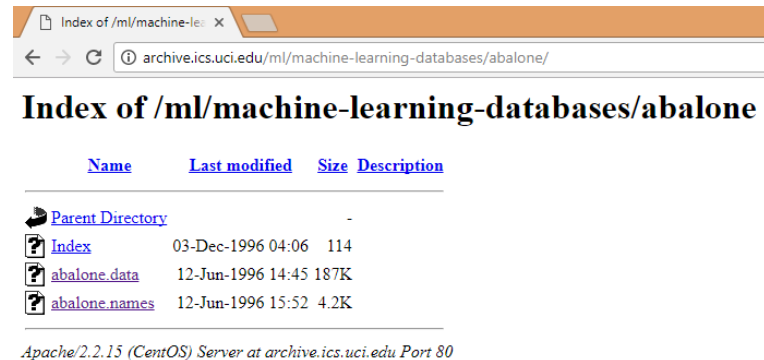
INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA

2017

- 1) Describe your process and methods. What parameters (testing/training size, classification target, subtree raising, etc.) did you try using with the decision tree? What, if anything, did you do to preprocess the dataset? Why did you select those parameters/preprocessing?

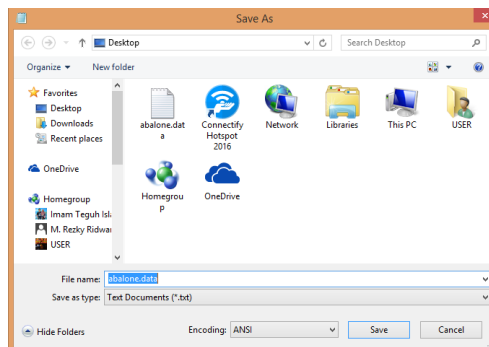
Langkah-langkah:

1. Pastikan Anda telah memiliki aplikasi Weka. Aplikasi Weka dapat didownload di: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
2. Dapatkan dataset "Abalone" dari UCI ML Repo. Link: <http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/>

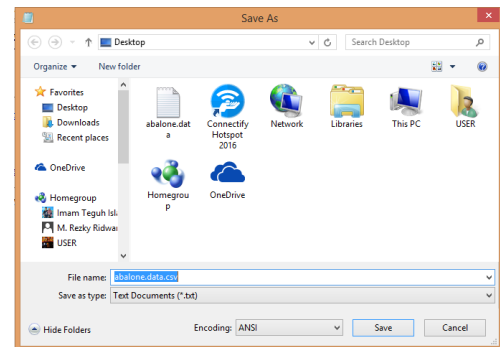


Pilih abalone.data. Kemudian simpan dengan format file .txt.

3. Ubah file abalone.data ke dalam format csv dengan menambahkan .csv.



(a) Sebelum diubah



(b) Setelah diubah

Kemudian klik save. Data abalone dengan format .csv pun diperoleh.

4. Buka abalone.data.csv. Kemudian tambahkan. Atribut di atasnya.

Urutan Kolom:

Kolom 1: Sex

Kolom 2: Length

Kolom 3: Diameter

Kolom 4: Height

Kolom 5: Whole weight

Kolom 6: Shucked weight

Kolom 7: Viscera weight

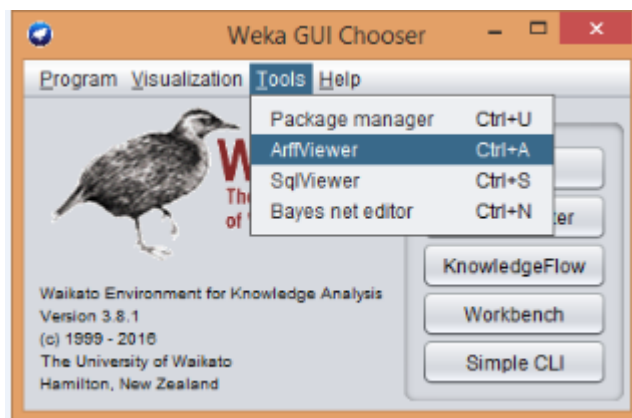
Kolom 8: Shell weight

Kolom 9: Rings

Setelah itu, ubah urutan atribut menjadi: Whole weight, Shell weight, Shucked weight, Viscera weight, Rings, Height, Diameter, Length. Hasil akhir:

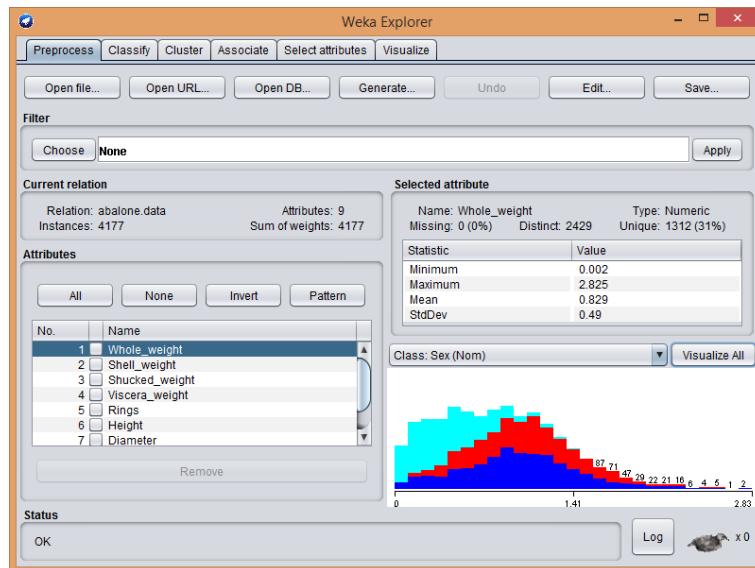
Whole_weight									
	A	B	C	D	E	F	G	H	I
1	Whole_weight	Shell_weight	Shucked_weight	Viscera_weight	Rings	Height	Diameter	Length	Sex
2	0.514	0.15	0.2245	0.101	15	0.095	0.365	0.455	M
3	0.2255	0.07	0.0995	0.0485	7	0.09	0.265	0.35	M
4	0.677	0.21	0.2565	0.1415	9	0.135	0.42	0.53	F
5	0.516	0.155	0.2155	0.114	10	0.125	0.365	0.44	M
6	0.205	0.055	0.0895	0.0395	7	0.08	0.255	0.33	I
7	0.3515	0.12	0.141	0.0775	8	0.095	0.3	0.425	I
8	0.7775	0.33	0.237	0.1415	20	0.15	0.415	0.53	F
9	0.768	0.26	0.294	0.1495	16	0.125	0.425	0.545	F
10	0.5095	0.165	0.2165	0.1125	9	0.125	0.37	0.475	M
11	0.8945	0.32	0.3145	0.151	19	0.15	0.44	0.55	F
12	0.6065	0.21	0.194	0.1475	14	0.14	0.38	0.525	F
13	0.406	0.135	0.1675	0.081	10	0.11	0.35	0.43	M
14	0.5415	0.19	0.2175	0.095	11	0.135	0.38	0.49	M
15	0.6845	0.205	0.2725	0.171	10	0.145	0.405	0.535	F
16	0.4755	0.185	0.1675	0.0805	10	0.1	0.355	0.47	F
17	0.6645	0.24	0.258	0.133	12	0.13	0.4	0.5	M
18	0.2905	0.115	0.095	0.0395	7	0.085	0.28	0.355	I
19	0.451	0.13	0.188	0.087	10	0.1	0.34	0.44	F
20	0.2555	0.1	0.097	0.043	7	0.08	0.295	0.365	M
21	0.381	0.115	0.1705	0.075	9	0.1	0.32	0.45	M
22	0.2455	0.075	0.0955	0.062	11	0.095	0.28	0.355	M
23	0.2255	0.085	0.08	0.049	10	0.1	0.275	0.38	I
24	0.9395	0.27	0.4275	0.214	12	0.155	0.44	0.565	F
25	0.7635	0.2	0.318	0.21	9	0.135	0.415	0.55	F

- Ubah file data.abalone.csv ke dalam format file .arff dengan menggunakan aplikasi Weka.

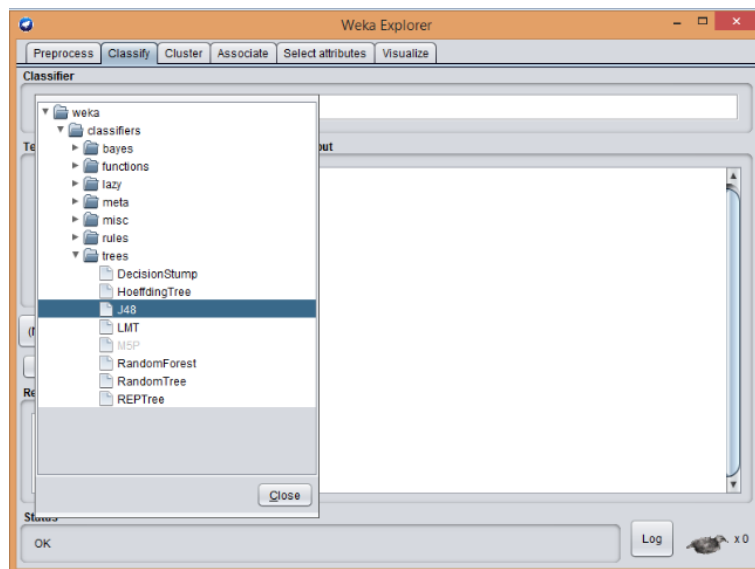


Setelah klik ArffViewer, buka file abalone.data.csv kemudian simpan dengan nama file: abalone.data.arff.

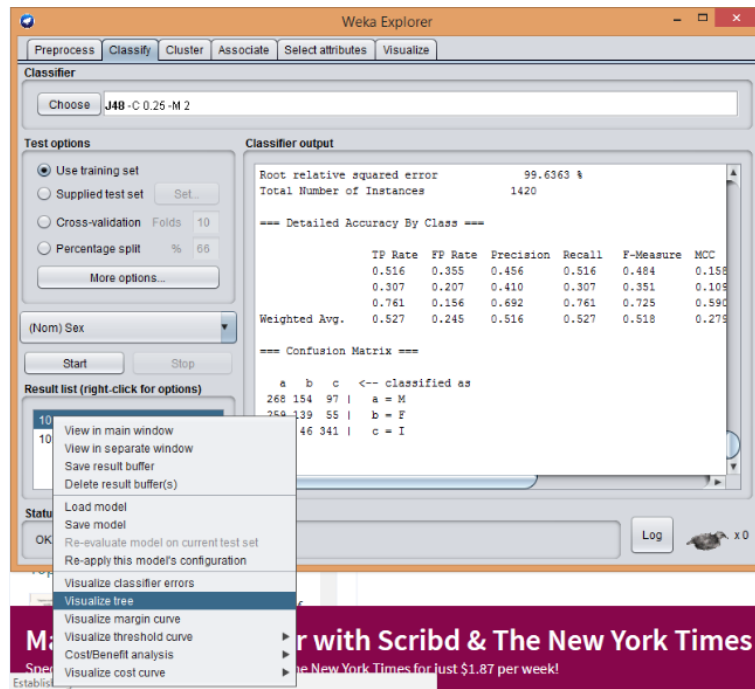
- Pilih Explore pada Weka. Kemudian klik Open file dan pilih abalone.data.arff. Maka diperoleh hasil:



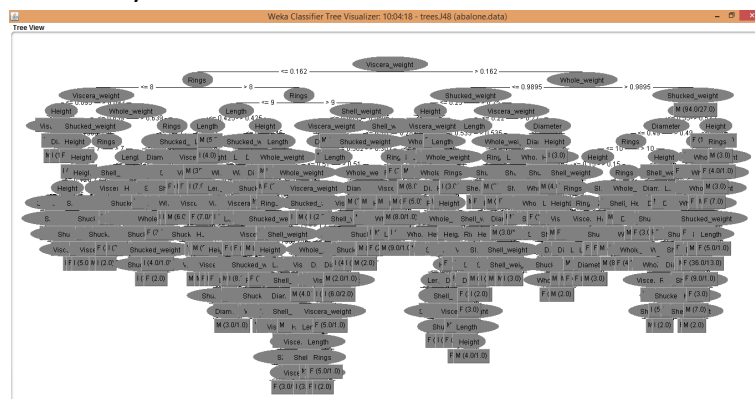
7. Klik Classify kemudian pilih algoritma (menggunakan algoritma J48) dengan mengklik Choose. Dan klik Test Option yang akan dipilih.



Setelah itu klik Start. Maka diperoleh hasil:



Untuk melihat gambar tree dengan klik kanan pada result list kemudian pilih Visualize tree. Hasilnya:



2) What were your result? Show what decision trees you found.

Kesimpulan: jika menggunakan **training set** maka diperoleh akurasi sebesar **74,982%** sedangkan jika menggunakan **percentage split 66%** maka diperoleh akurasi sebesar **52,6761%**. Jika menggunakan **cross-validation** maka diperoleh akurasi sebesar **52,7891%**.

```

=== Summary ===

Correctly Classified Instances      3132           74.982 %
Incorrectly Classified Instances    1045           25.018 %
Kappa statistic                    0.6232
Mean absolute error                0.2231
Root mean squared error            0.334
Relative absolute error             50.3148 %
Root relative squared error         70.933 %
Total Number of Instances          4177

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.727   0.189   0.689     0.727   0.708     0.532   0.868    0.795    M
                0.654   0.125   0.705     0.654   0.679     0.541   0.883    0.771    F
                0.869   0.066   0.862     0.869   0.866     0.802   0.959    0.890    I
Weighted Avg.   0.750   0.129   0.750     0.750   0.749     0.621   0.902    0.818

=== Confusion Matrix ===

  a    b    c  <-- classified as
1111  295  122 |    a = M
 388  855   64 |    b = F
 113   63 1166 |    c = I

```

(a) Jika menggunakan training set

```

=== Summary ===

Correctly Classified Instances      748           52.6761 %
Incorrectly Classified Instances    672           47.3239 %
Kappa statistic                    0.2862
Mean absolute error                0.3463
Root mean squared error            0.4692
Relative absolute error             78.0941 %
Root relative squared error         99.6363 %
Total Number of Instances          1420

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.516   0.355   0.456     0.516   0.484     0.158   0.594    0.434    M
                0.307   0.207   0.410     0.307   0.351     0.109   0.631    0.400    F
                0.761   0.156   0.692     0.761   0.725     0.590   0.827    0.674    I
Weighted Avg.   0.527   0.245   0.516     0.527   0.518     0.279   0.679    0.499

=== Confusion Matrix ===

  a    b    c  <-- classified as
268  154   97 |    a = M
259  139   55 |    b = F
 61   46 341 |    c = I

```

(b) Jika menggunakan Percentage split 66%

Pada gambar jawaban no. 2 terdapat Correctly Classifies Instances dan Incorrectly Classifies Instances atau (in)accurate. Terdapat hasil (in)accurate dikarenakan adanya misclassified.