**LAPORAN**

**KECERDASAN KOMPUTASIONAL**

**“Visualising Tree Using Weka Application”**

**ABALON & BREAST CANCER**



**Dosen :**

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

**Oleh :**

**Adi Darmawan 5115100079**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA**

**2017**

**Question**

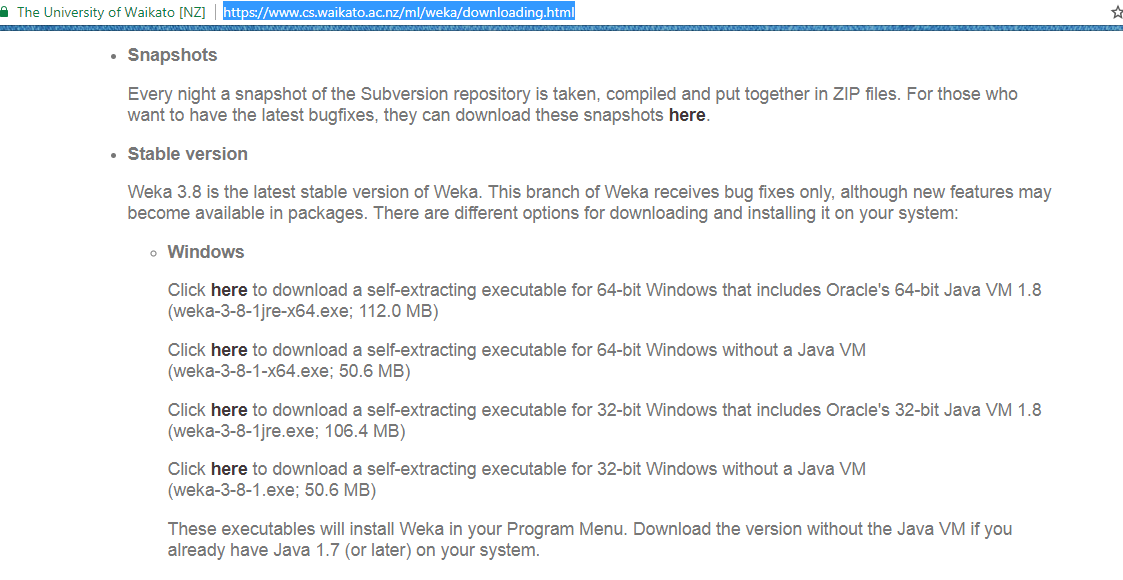
1. Get the "Abalone" dataset from UCI ML Repo
2. Use a decision tree (C4.5/J48) learner in WEKA to predict the number of rings Abalone will have
3. Write the description of your process and methods. What parameters (testing/training size, classification target, subtree raising etc.) did you use? Did you preprocess the dataset? Why did you select those parameters/preprocessing?
4. What were your results? Show what decision trees you found.

**Answer**

1. **Abalone Data**

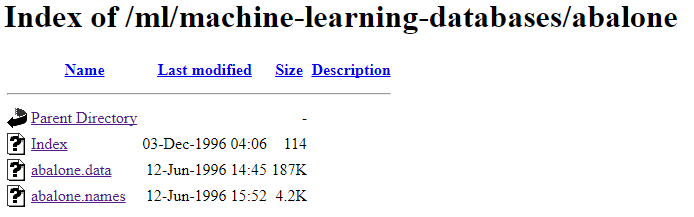
**Langkah-Langkah :**

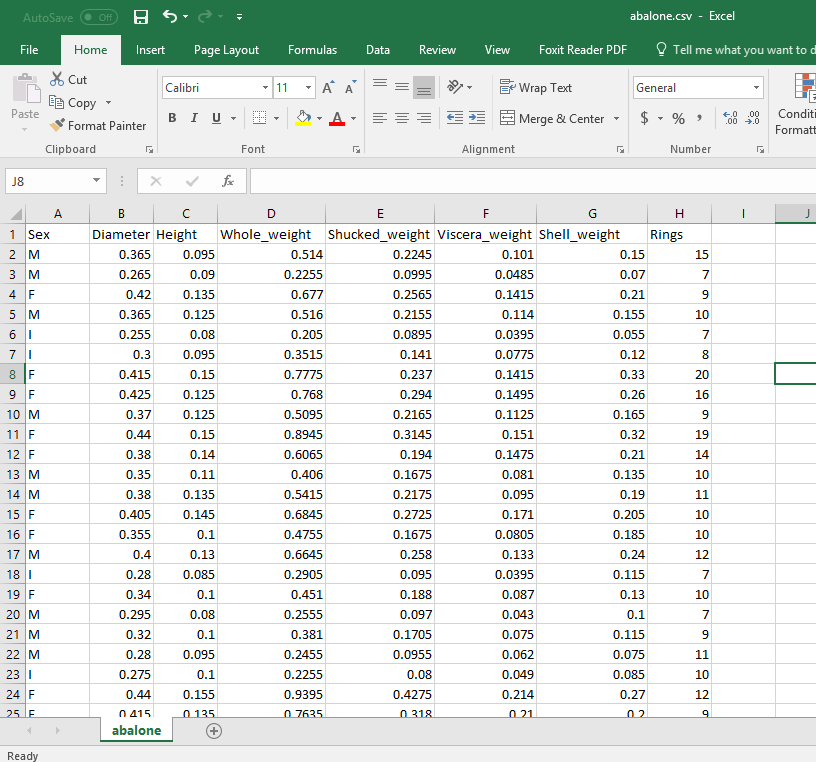
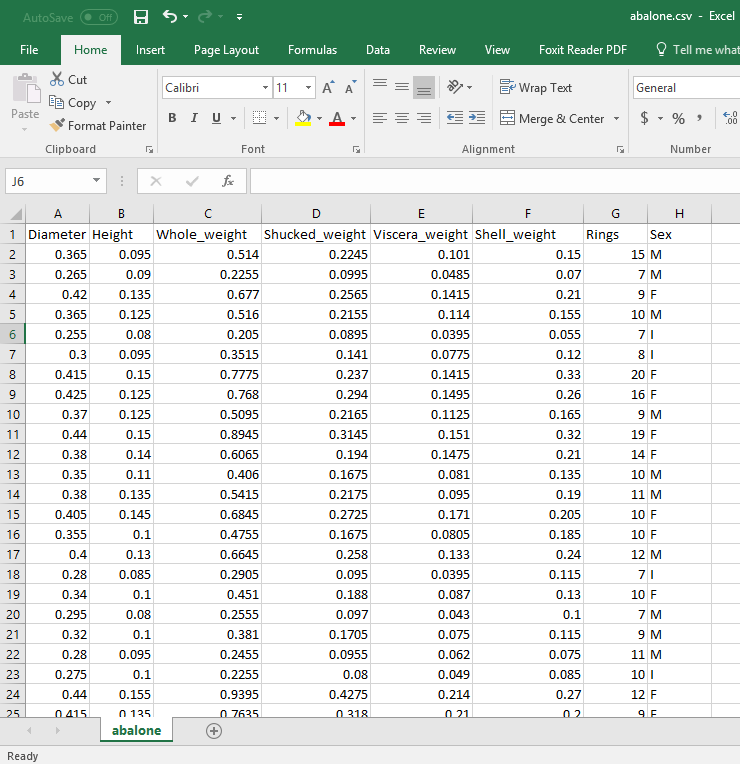
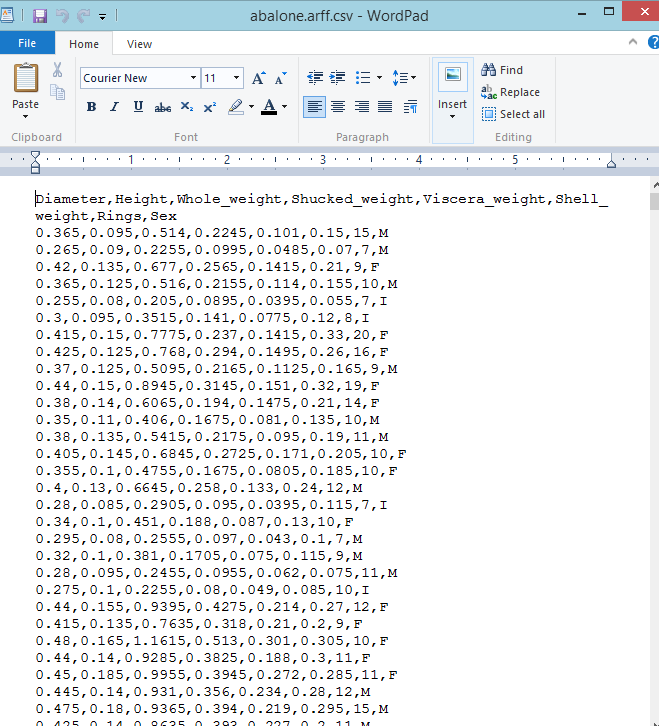
1. Download WEKA dari alamat dibawah ini, dan sesuaikan dengan system operasi pada laptop anda.

<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

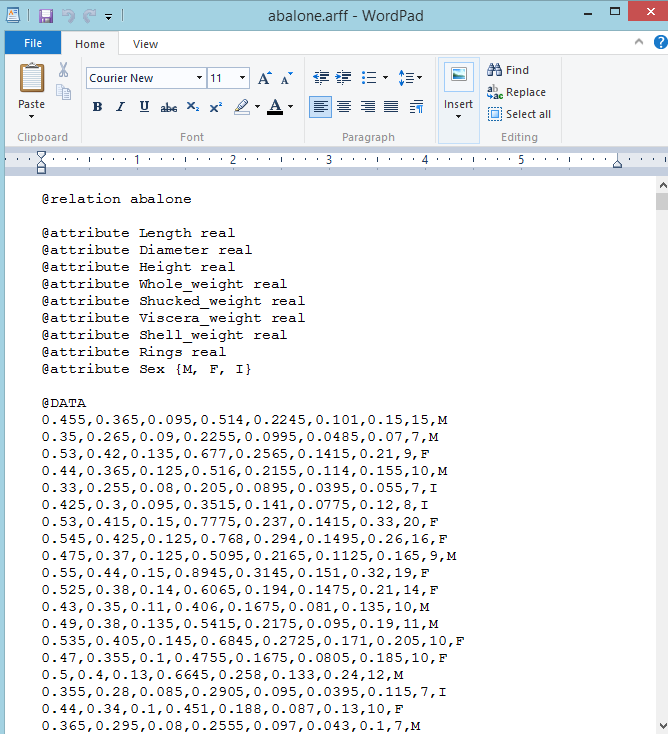
1. Download data Abalon UCI pada

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



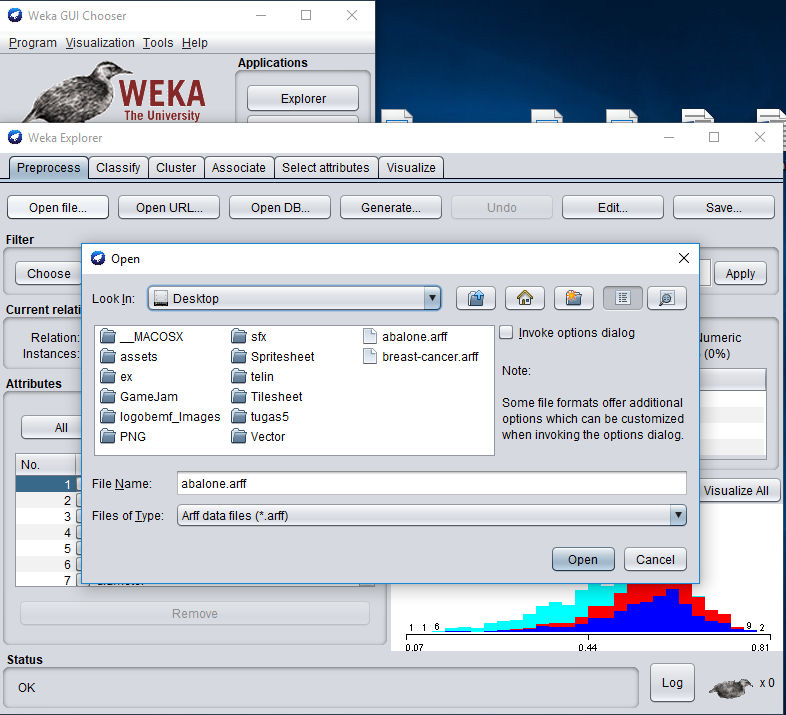
1. Setelah terdownload, rename “abalone.data” menjadi “abalone.csv” tanpa tanda kutip.
2. Langkah selanjutnya adalah buka file abalone.scv dengan Microsoft excel
3. Berilah nama tiap kolom attribute dan class nya sesuai dengan yang ada pada abalone.name pada database UCI.
4. Pindahlah kolom Sex dari paling kiri ke kolom paling kanan(setelah Rings)
5. Langkah selanjutnya adalah save as data dengan nama abalone.arrf dan selanjutnya buka file dengan Wordpad dan tambahkan @attribute dan real sebagai relation sebagai berikut:

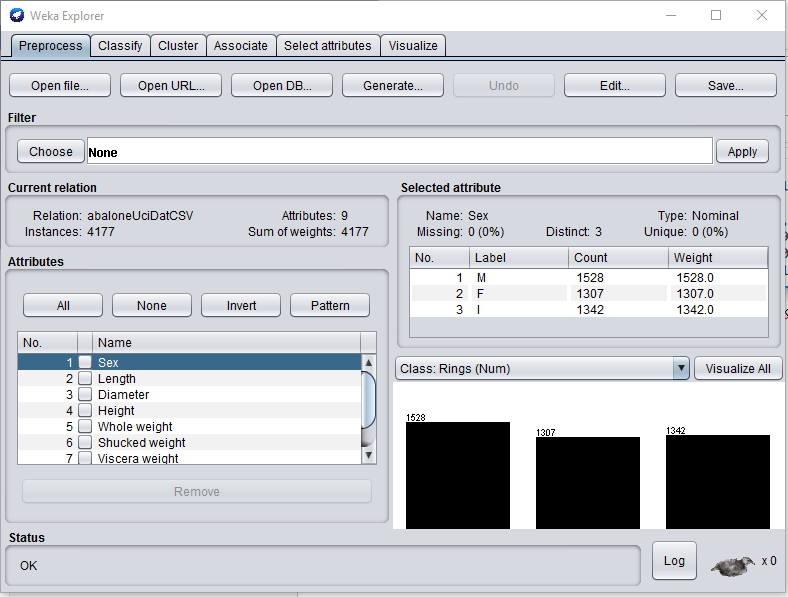
MENJADI



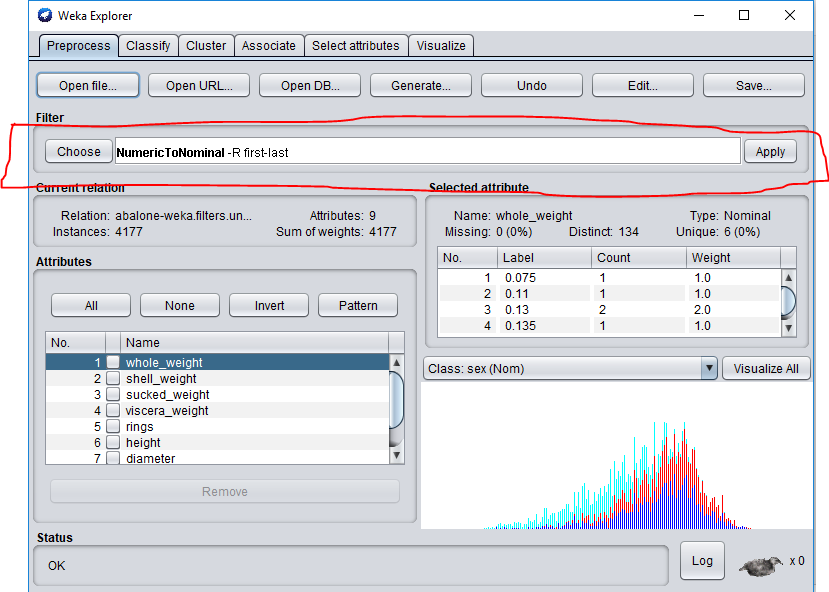
Dan simpan file abalone.arff

1. Setelah itu buka menu weka terdapat pilihan-pilihan, dan pilih menu Explorer
2. Setelah terbuka, load file abalone.arff dan ikuti langkah sebagai berikut :

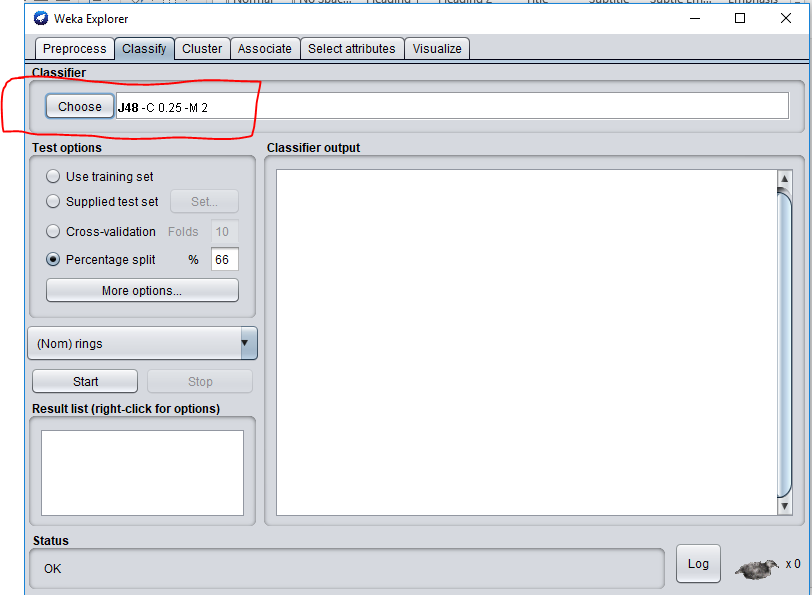




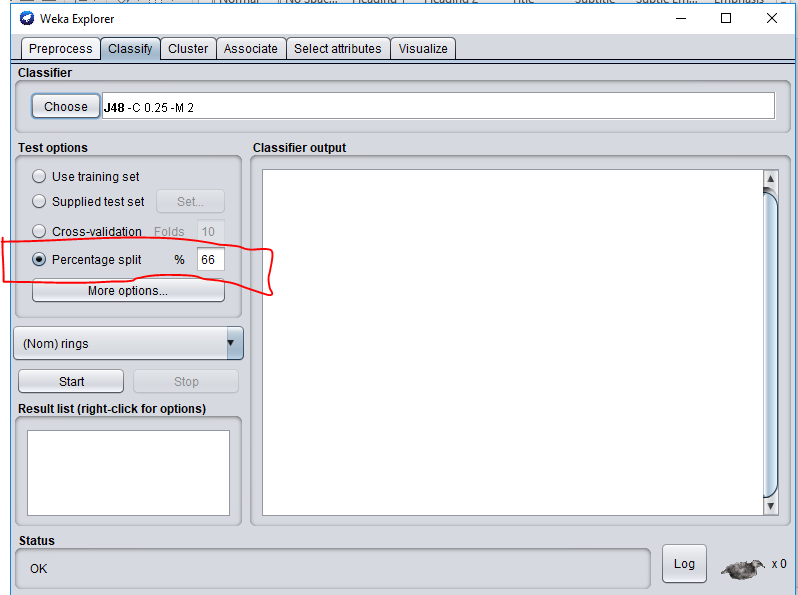
1. Tambahkan filter NumericalToNominal hal ini dikarenakan kita menggunakan file .csv



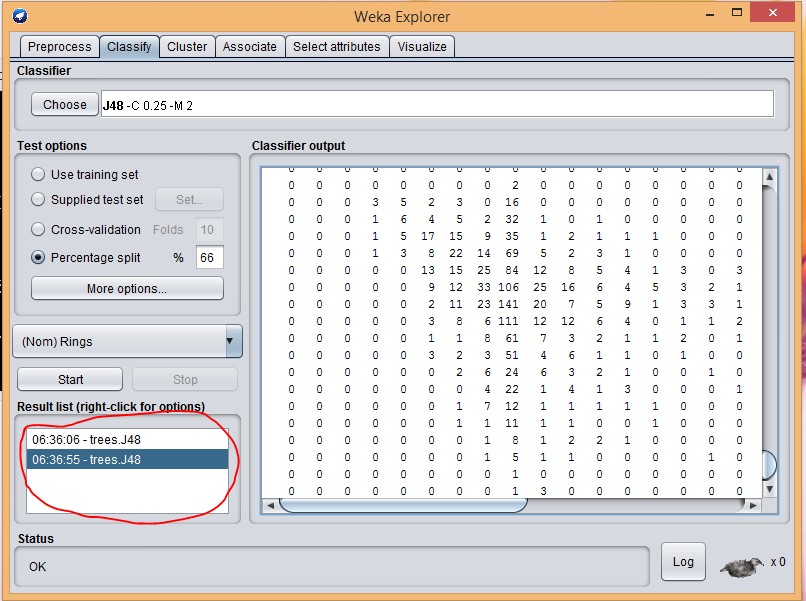
1. Masuk ke tab classify, Tambahkan Calssifier Tree lalu -> J48 -C 0.25 -M 2

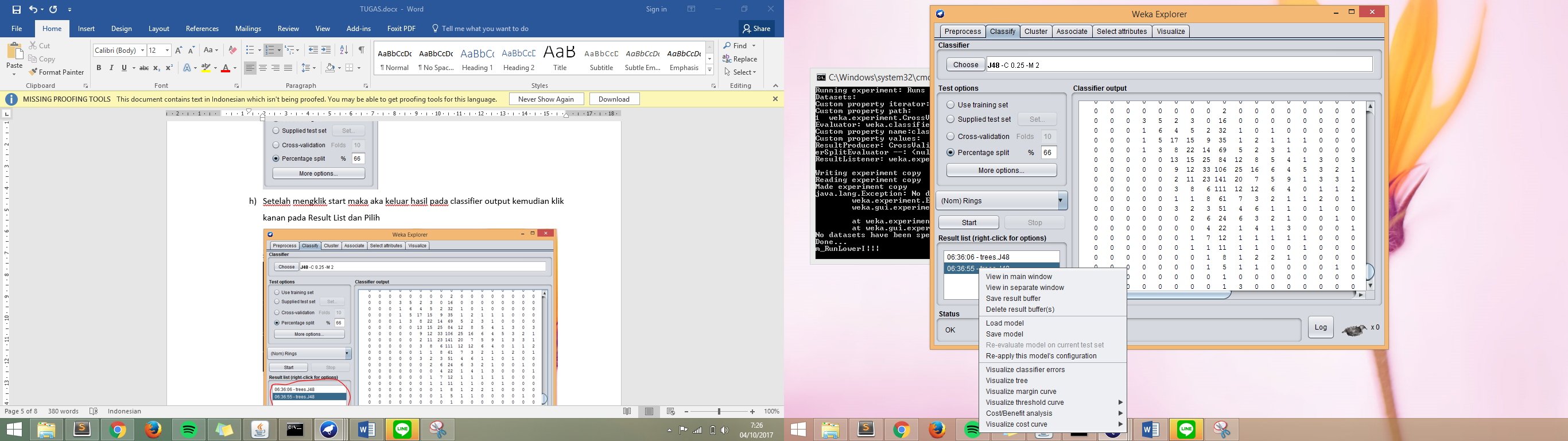


1. Dalam test options kita pilih percentage split sebesar 66% karena menginginkan data training sebanyak 2/3 dari datanya dan 1/3 sebagai data test. Kemudian Klik start

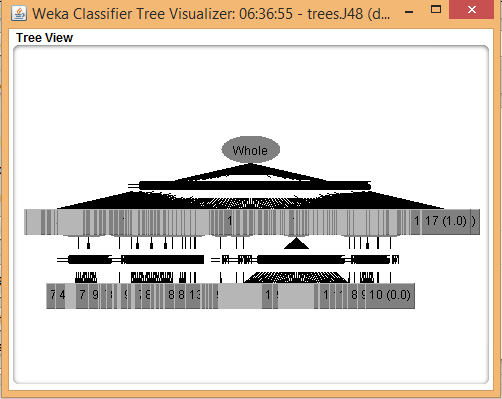


1. Setelah mengklik start maka aka keluar hasil pada classifier output kemudian klik kanan pada Result List dan Pilih Visualisze tree





1. Maka akan keluar hasil sebagai berikut



1. 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

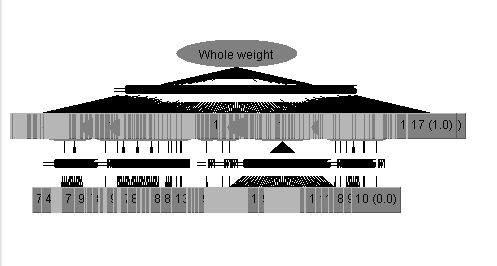
Mean absolute error 0.0622

Root mean squared error 0.2108

Relative absolute error 97.0872 %

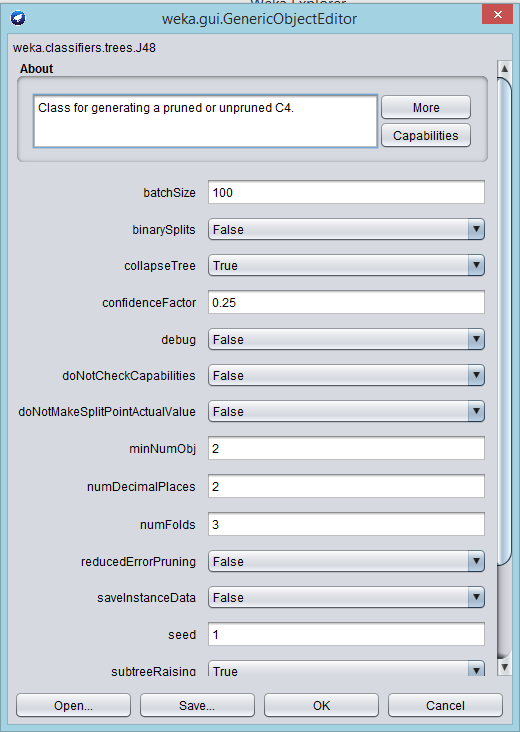
Root relative squared error 117.7184 %

Total Number of Instances 1420



1. Dari hasilnya terlihat clasifikasi yang tepat sangat kecil persentasinya 15.0704 % . Menurut pengamatan saya ini dikarenakan data ini tidak cocok menggunakan algoritma pencarian tree ini. Kemungkinan yang lain karena datanya pada awalnya berformat .data kemudian di convert ke csv sehingga menimbulkan beberapa error.

**PARAMETER**



1. **Breast Cancer Data**
2. Stepnya sama dengan abalone data. Tetapi data yang ada di repo sudah berformat “arff” yang didukung oleh weka

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

1. Number of Leaves : 4

Size of the tree : 6

Correctly Classified Instances 66 68.0412 %

Incorrectly Classified Instances 31 31.9588 %

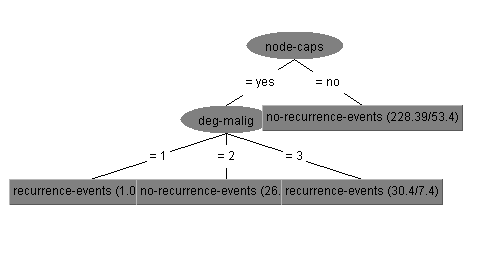
Kappa statistic 0.2001

Mean absolute error 0.3966

Root mean squared error 0.4879

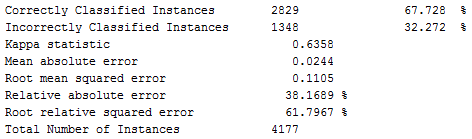
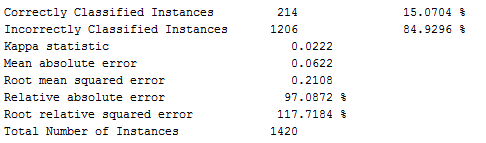
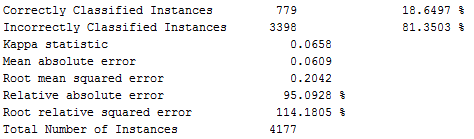
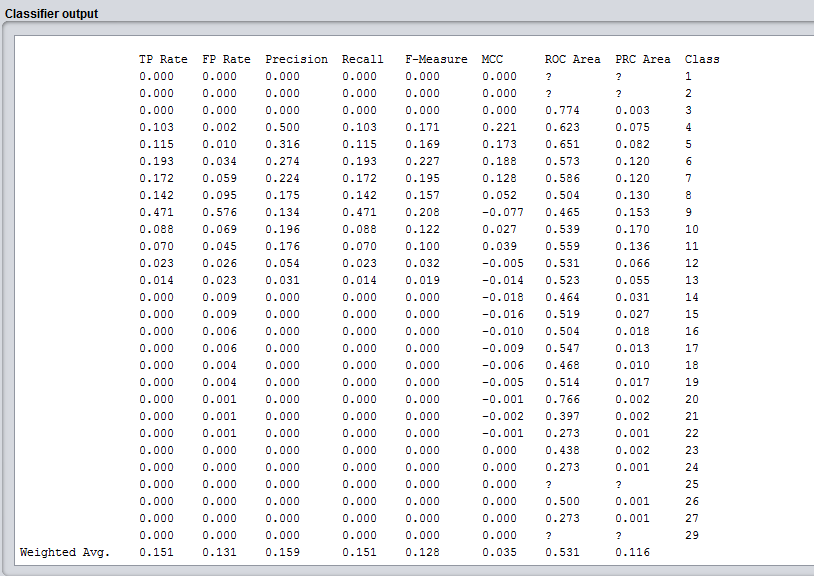
Relative absolute error 92.4804 %

Root relative squared error 102.0849 %

Total Number of Instances 97 

1. 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.
2. What do the results tell us? Why are the results (in)accurate? Why did changing parameter(s) improve/degrade accuracy?

**Answer**

* Use Training Set
* Cross Validation Folds: 10
* Percantage Splits 66%

Berdasarkan data dari hasil akurasi yang telah diujikan diatas terdapat perbedaan di masing-masing parameter. Hal tersebut bisa terjadi dikarenakan adanya *misclassified.*