**TUGAS**

**KECERDASAN KOMPUTASIONAL**

**“Visualising Tree Using Weka Application”**

**ABALONE & BREAST CANCER**



**Dosen :**

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

**Oleh :**

**Fajar Maulana Firdaus 5115100171**

With abalone dataset

Description

Number of Instances: 4177

Number of Attributes: 8

Attribute information:

Given is the attribute name, attribute type, the measurement unit and a

brief description. The number of rings is the value to predict: either

as a continuous value or as a classification problem.

Name Data Type Meas. Description

---- --------- ----- -----------

Sex nominal M, F, and I (infant)

Length continuous mm Longest shell measurement

Diameter continuous mm perpendicular to length

Height continuous mm with meat in shell

Whole weight continuous grams whole abalone

Shucked weight continuous grams weight of meat

Viscera weight continuous grams gut weight (after bleeding)

Shell weight continuous grams after being dried

Rings integer +1.5 gives the age in years

Statistics for numeric domains:

Length Diam Height Whole Shucked Viscera Shell Rings

Min 0.075 0.055 0.000 0.002 0.001 0.001 0.002 1

Max 0.815 0.650 1.130 2.826 1.488 0.760 1.005 29

Mean 0.524 0.408 0.140 0.829 0.359 0.181 0.239 9.934

SD 0.120 0.099 0.042 0.490 0.222 0.110 0.139 3.224

Correl 0.557 0.575 0.557 0.540 0.421 0.504 0.628 1.0

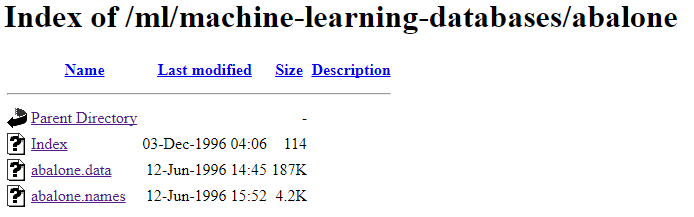
Missing Attribute Values: None

1. **Abalone Data**

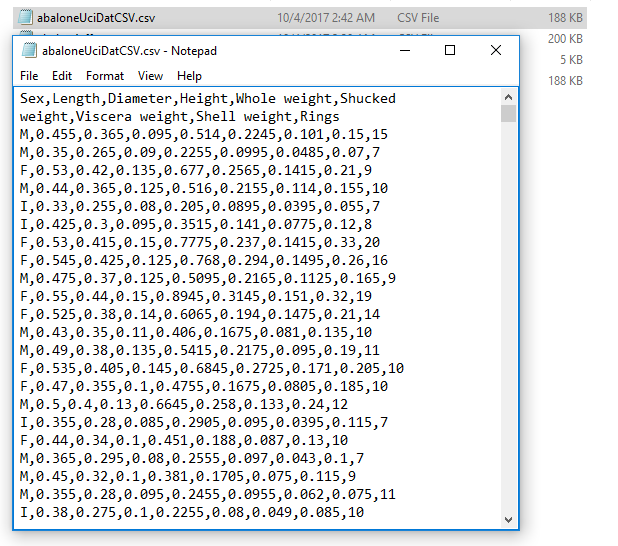
**Langkah-Langkah :**

1. Download data Abalone UCI pada

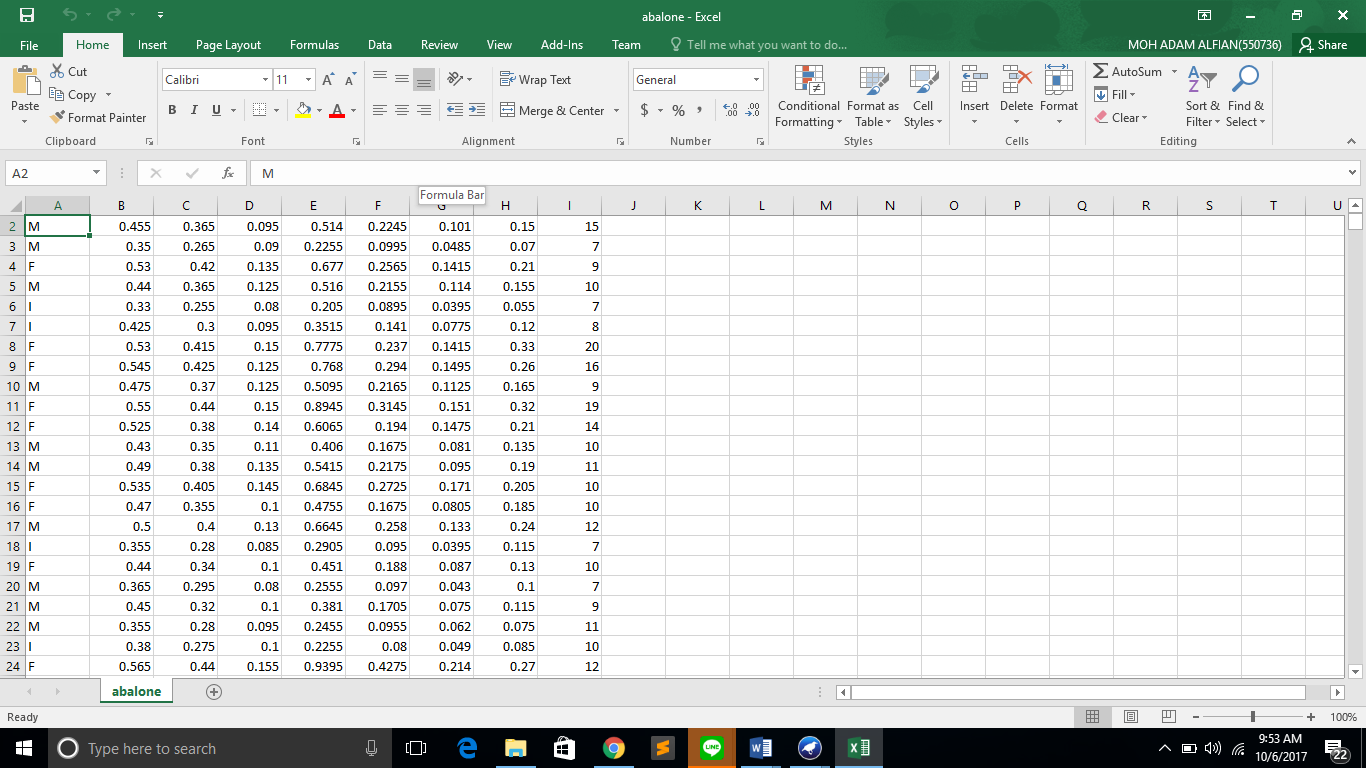
<http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/>



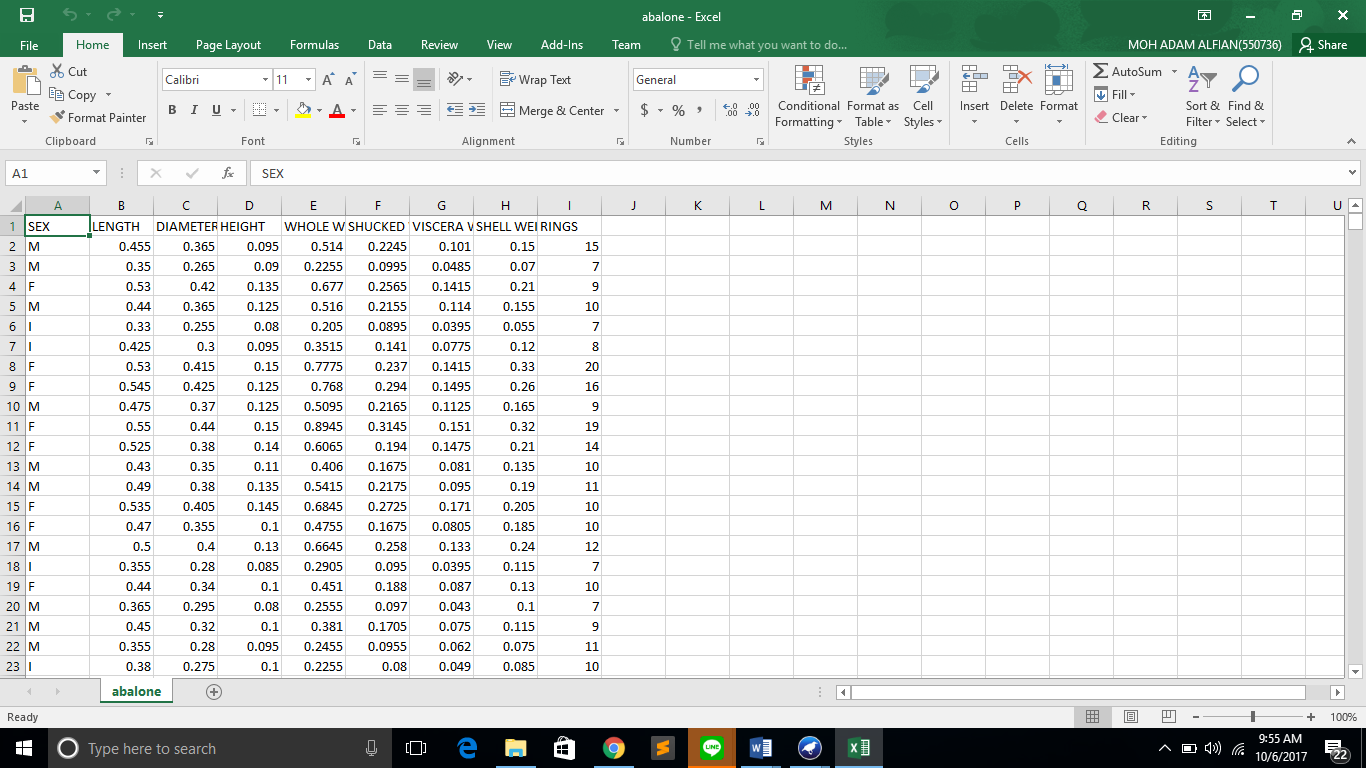
1. Simpan data dalam bentuk file ekstensi .csv agar weka dapat membaca file tersebut



Hasil dalam csv akan tampak seperti dibawah ini



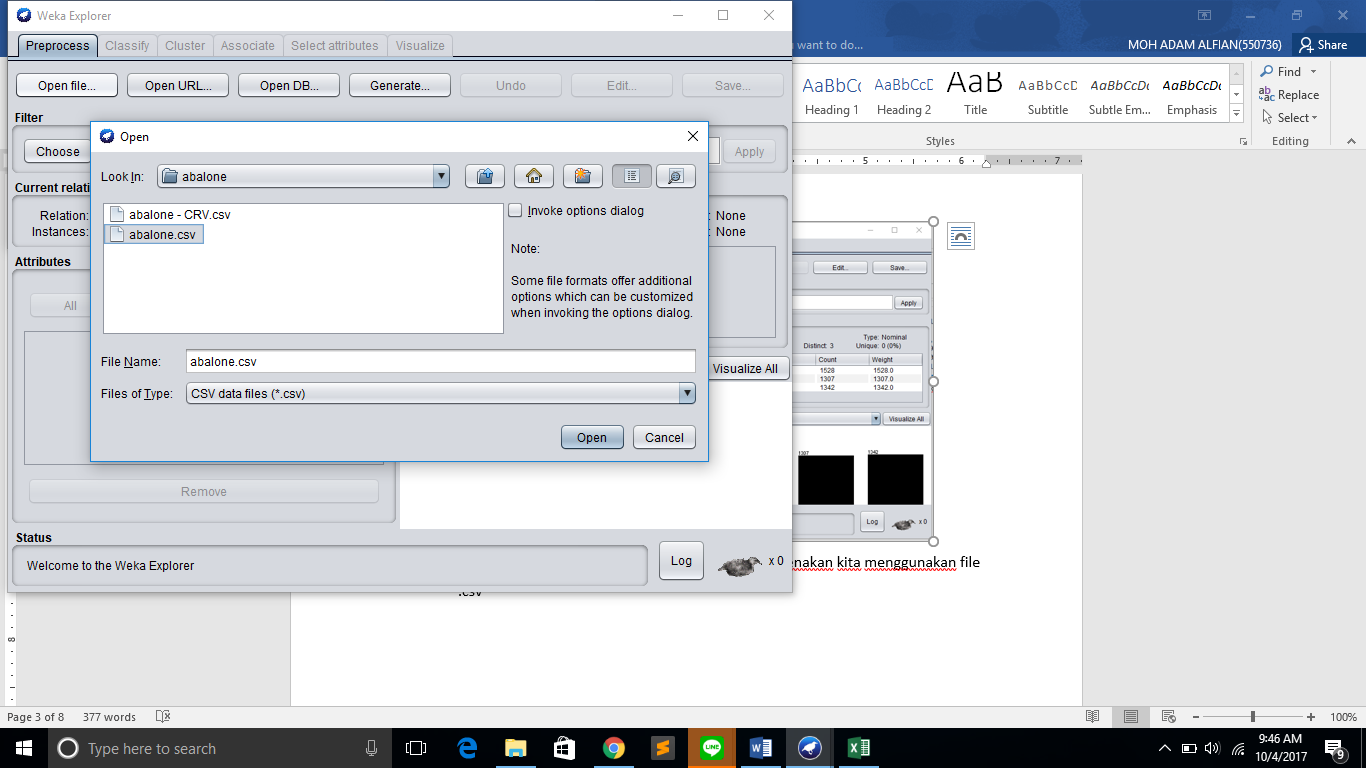
Kemudian dikasih atribut setiap kolom :

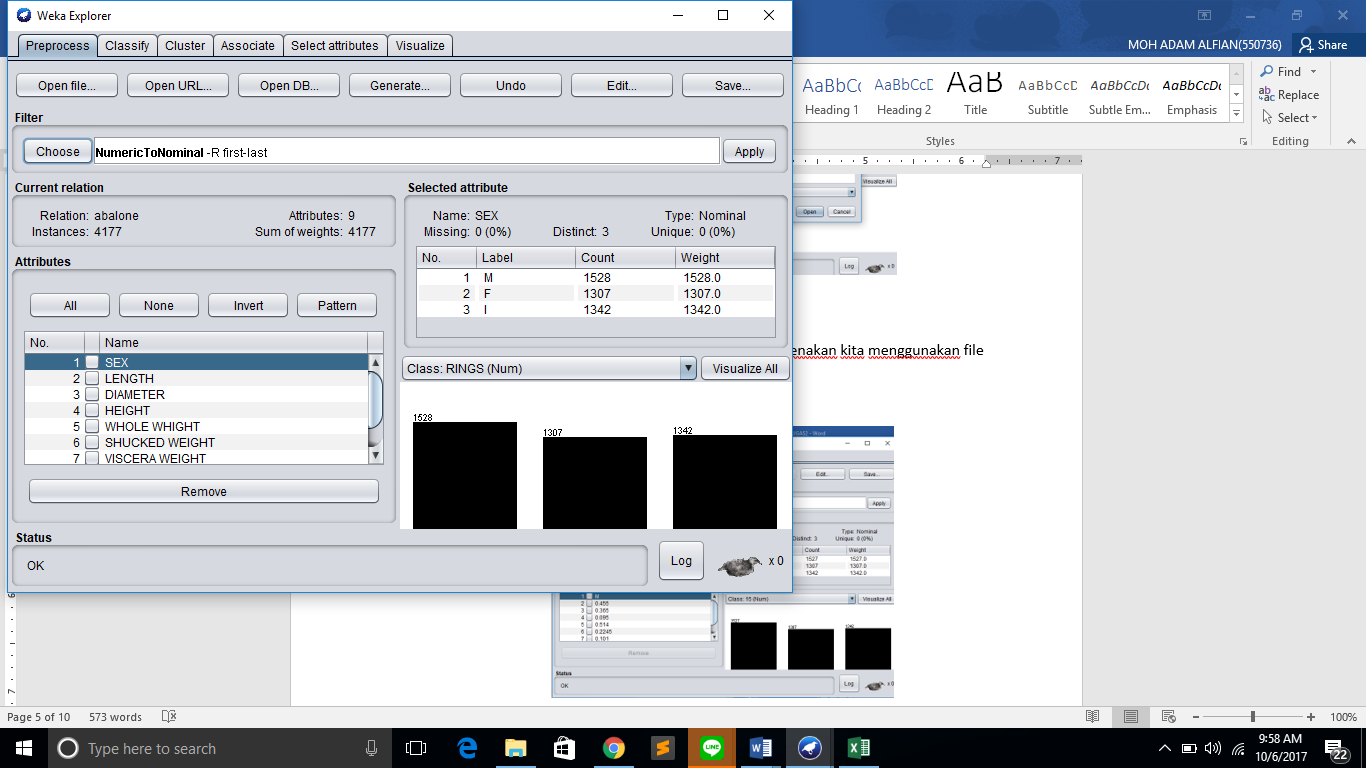


1. Setelah itu buka menu weka terdapat pilihan-pilihan

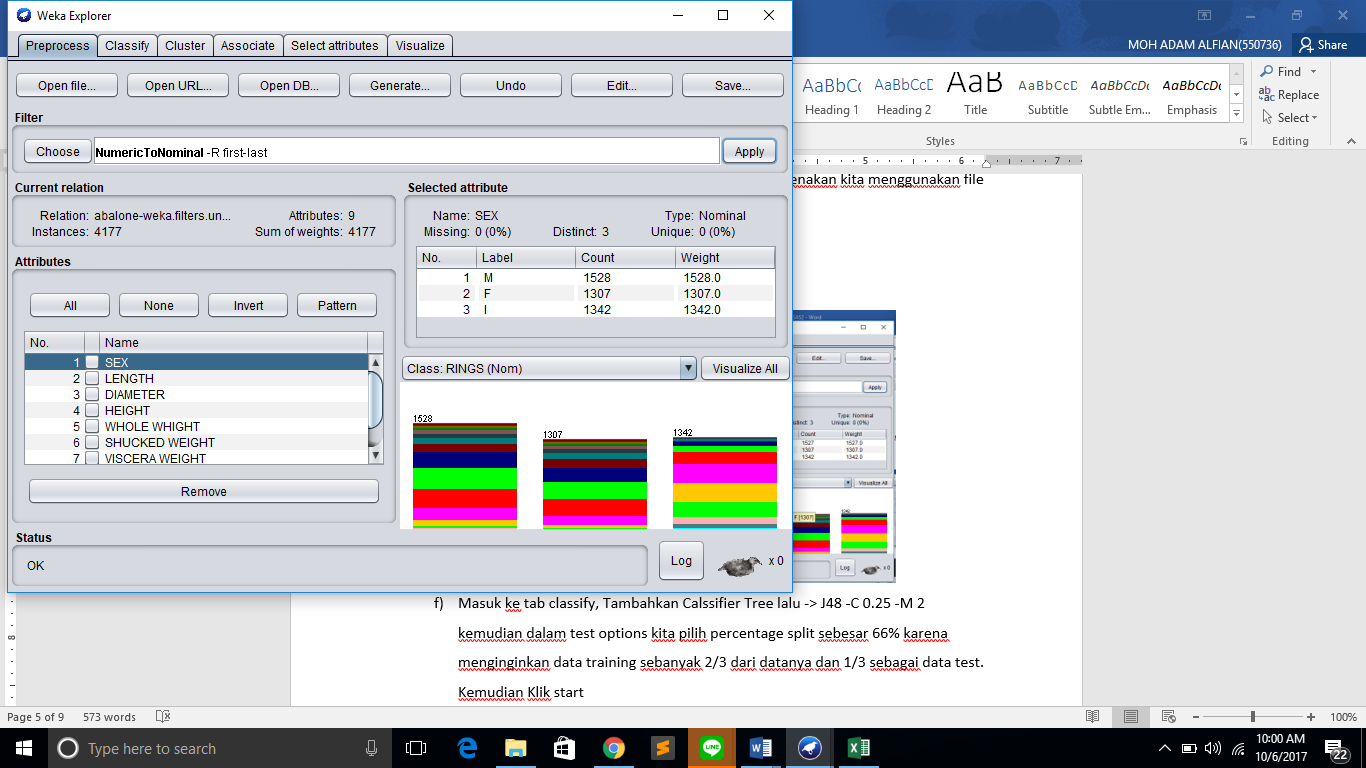


1. Pilih Menu explorer kemudian open file pilih format file csv dan masukan data csv yang tadi telah kita buat.

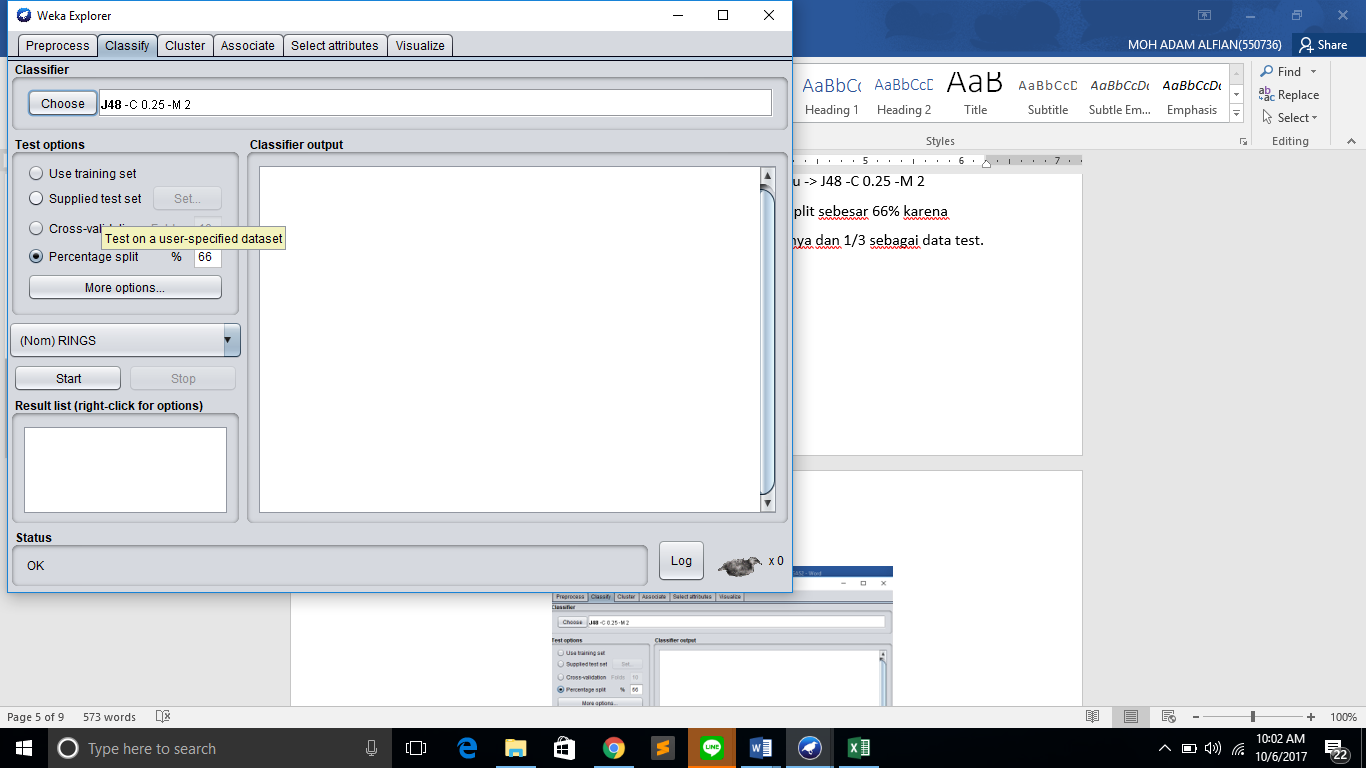




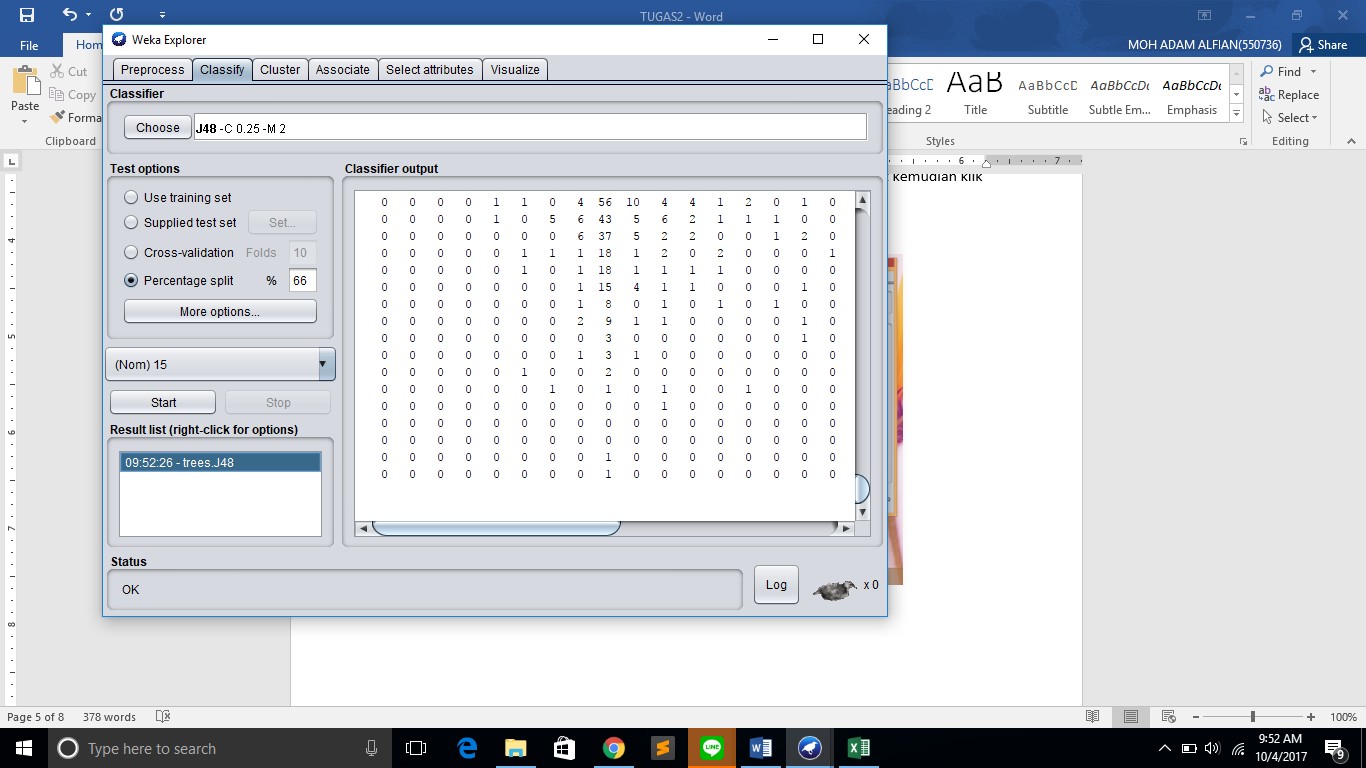
1. Tambahkan filter NumericalToNominal hal ini dikarenakan kita menggunakan file .csv

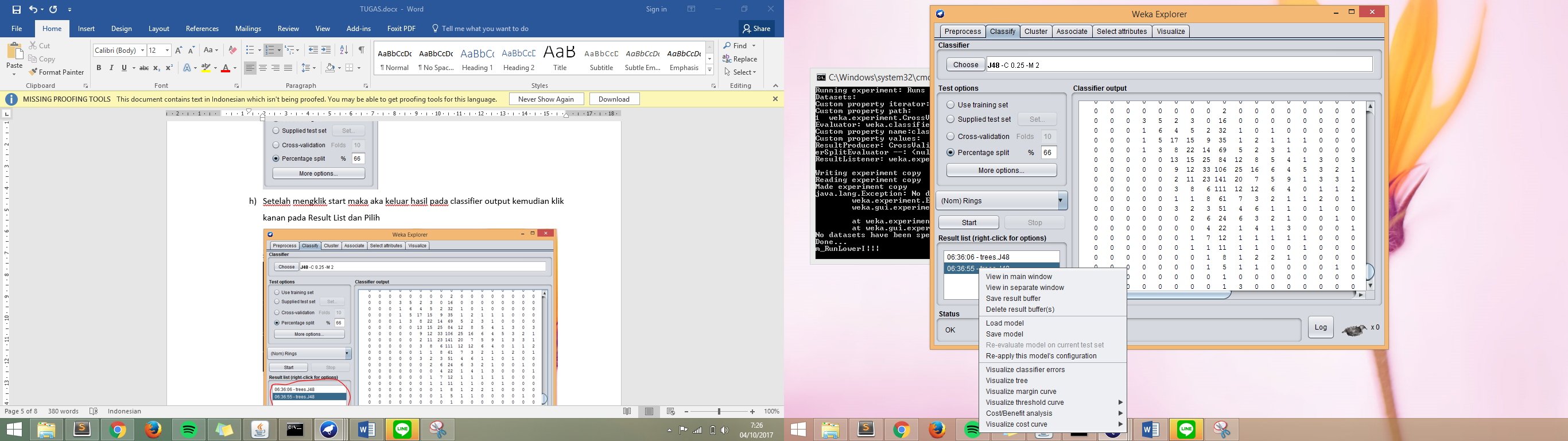


1. Masuk ke tab Classify, Tambahkan Classifier Tree lalu -> J48 -C 0.25 -M 2 kemudian dalam test options kita pilih percentage split sebesar 66% karena menginginkan data training sebanyak 2/3 dari datanya dan 1/3 sebagai data test. Kemudian Klik start

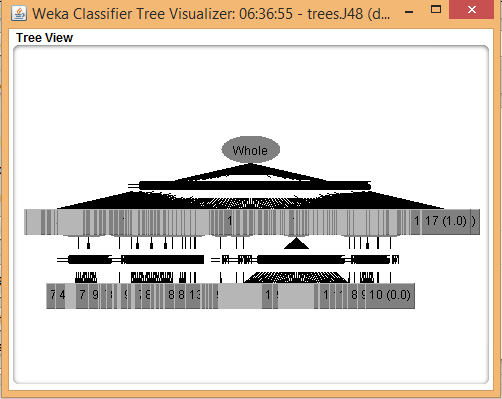


1. Setelah mengklik start maka aka keluar hasil pada classifier output kemudian klik kanan pada Result List dan Pilih Visualize tree





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

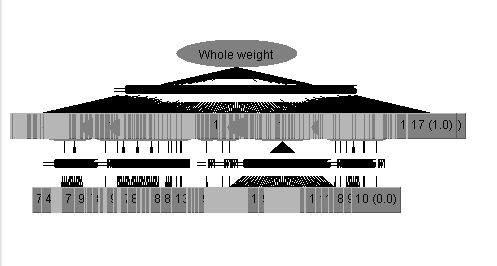
Mean absolute error 0.0622

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 pengconvertan data secara langsung akan mengganggu algoritma pencarian tree ini yang awalnya berformat .data kemudian di convert ke csv sehingga menimbulkan beberapa error.
2. **Breast Cancer Data**

Description

Number of Instances: 286

Number of Attributes: 9 + the class attribute

|  |  |  |
| --- | --- | --- |
| Attribute Information: | | |
|  | 1. Class: no-recurrence-events, recurrence-events | |
|  | 2. age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99. | |
|  | 3. menopause: lt40, ge40, premeno. | |
|  | 4. tumor-size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59. | |
|  |  | |
|  | 5. inv-nodes: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, | |
|  | 27-29, 30-32, 33-35, 36-39. | |
|  | 6. node-caps: yes, no. | |
|  | 7. deg-malig: 1, 2, 3. | |
|  | 8. breast: left, right. | |
|  | 9. breast-quad: left-up, left-low, right-up, right-low, central. | |
|  | 10. irradiat: yes, no. | |
| Class Distribution: | | |
|  | | 1. no-recurrence-events: 201 instances |
|  | | 2. recurrence-events: 85 instances |

|  |  |
| --- | --- |
| Num Instances: 286 | |
|  | Num Attributes: 10 |
|  | Num Continuous: 0 (Int 0 / Real 0) |
|  | Num Discrete: 10 |
|  | Missing values: 9 / 0.3% |

1. Langkah langkahnya sama seperti abalone data. Tetapi data sudah berformat “arff” yang didukung oleh Weka sehingga tidak perlu melakukan convert ke “csv”

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

1. Dari hasil pengamatan kami menendapatkan hasil

Number of Leaves : 4

Size of the tree : 6

Correctly Classified Instances 66 68.0412 %

Incorrectly Classified Instances 31 31.9588 %

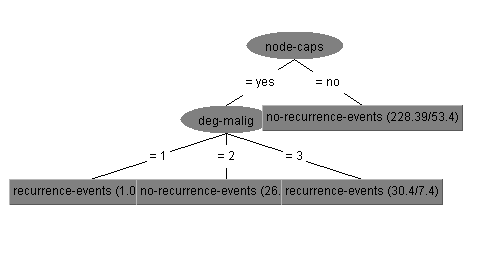
Kappa statistic 0.2001

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. setelah pemrosesan ditemukan presentase ketepatan yang tinggi yaitu 68.0412 %. Ini dikarenakan data ini cocok menggunakan algoritma ini ataupun juga karena file inputnya langsung dalam format arff tanpa mengconvert secara langsung.