**COMPUTATIONAL INTELLIGENCE**

**“Predict the Number of Rings Abalone Will Have Using Weka Application”**



**By :**

**Agatha Putri Adwitya 5115100137**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA**

**2017**

1. **Informasi Data Set**

* **Deskripsi**

Memprediksi usia dari Abalone berdasarkan pengukuran fisik (*length, diameter, height, whole weight, shucked weight, viscera weight, shell weight,* dan *rings*) dimana data yang *missing value* sudah dihapus/ tidak ada *missing value*.

* **Atribut**

*Name / Data Type / Measurement Unit / 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*

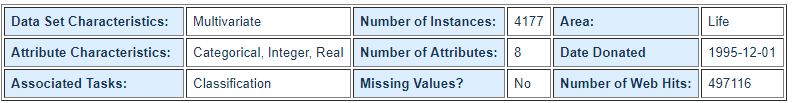
*Shucked weight / continuous / grams / weight of meat*

*Viscera weight / continuous / grams / gut weight (after bleeding)*

*Shell weight / continuous / grams / after being dried*

*Rings / integer / -- / +1.5 gives the age in years*

* **Keterangan Lain**

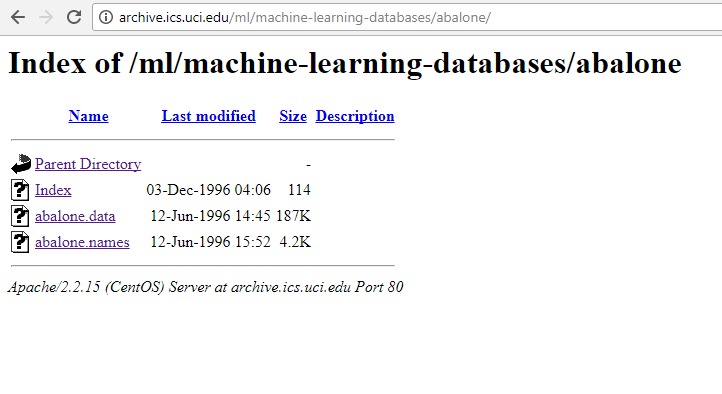


* **Sumber**

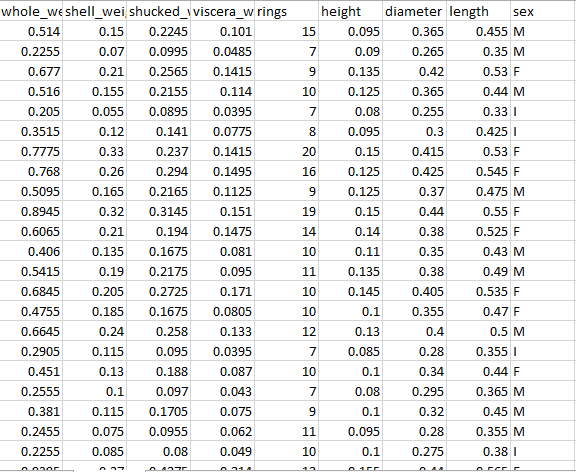
<http://archive.ics.uci.edu/ml/datasets/Abalone>

1. **Pertanyaan**
2. *Get the "Abalone" dataset from UCI ML Repo*
3. Langkah-langkah:
4. Download abalone.data pada

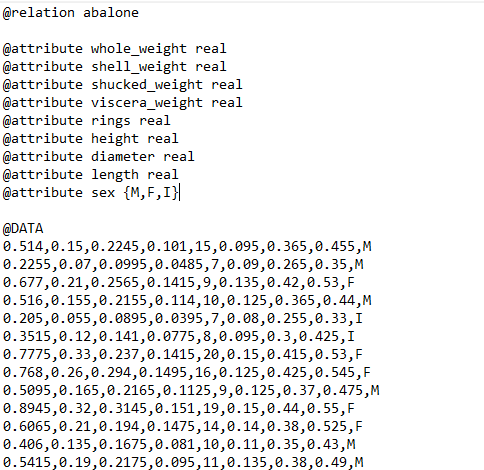
<http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/>



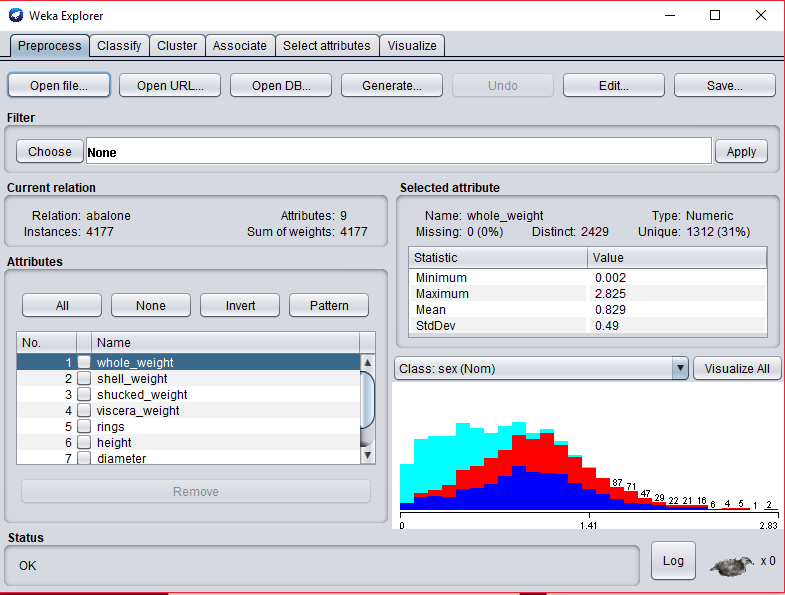
1. Ubah ekstensi *abalone.data* menjadi *abalone.csv* kemudian ubah urutan atribut menjadi seperti berikut:



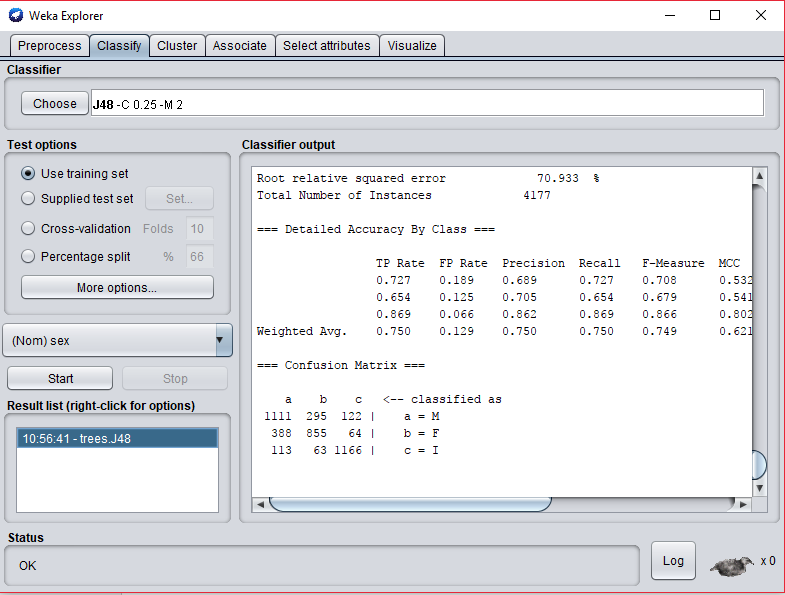
1. Lalu save data, kemudian buka data kembali menggunakan notepad. Edit header sesuai gambar berikut:



1. Ubah ekstensi menjadi *abalone.arff.* Lalu buka aplikasi Weka, kemudian buka file abalone.arff



1. *Use a decision tree (C4.5/J48) learner in WEKA to predict the number of rings Abalone will have*

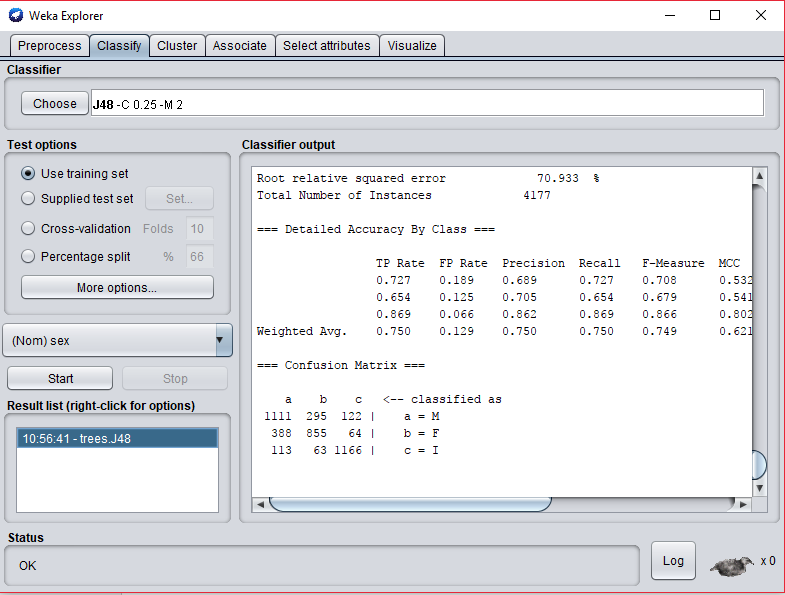
Buka tab Classify, lalu pilih J48 sebagai Classifier.

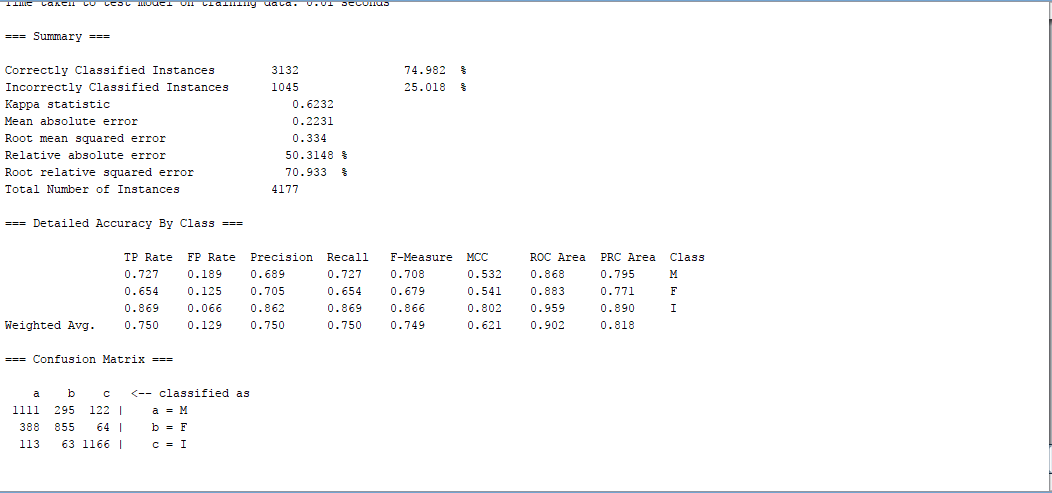
Dan hasilnya adalah sebagai berikut:

1. *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*

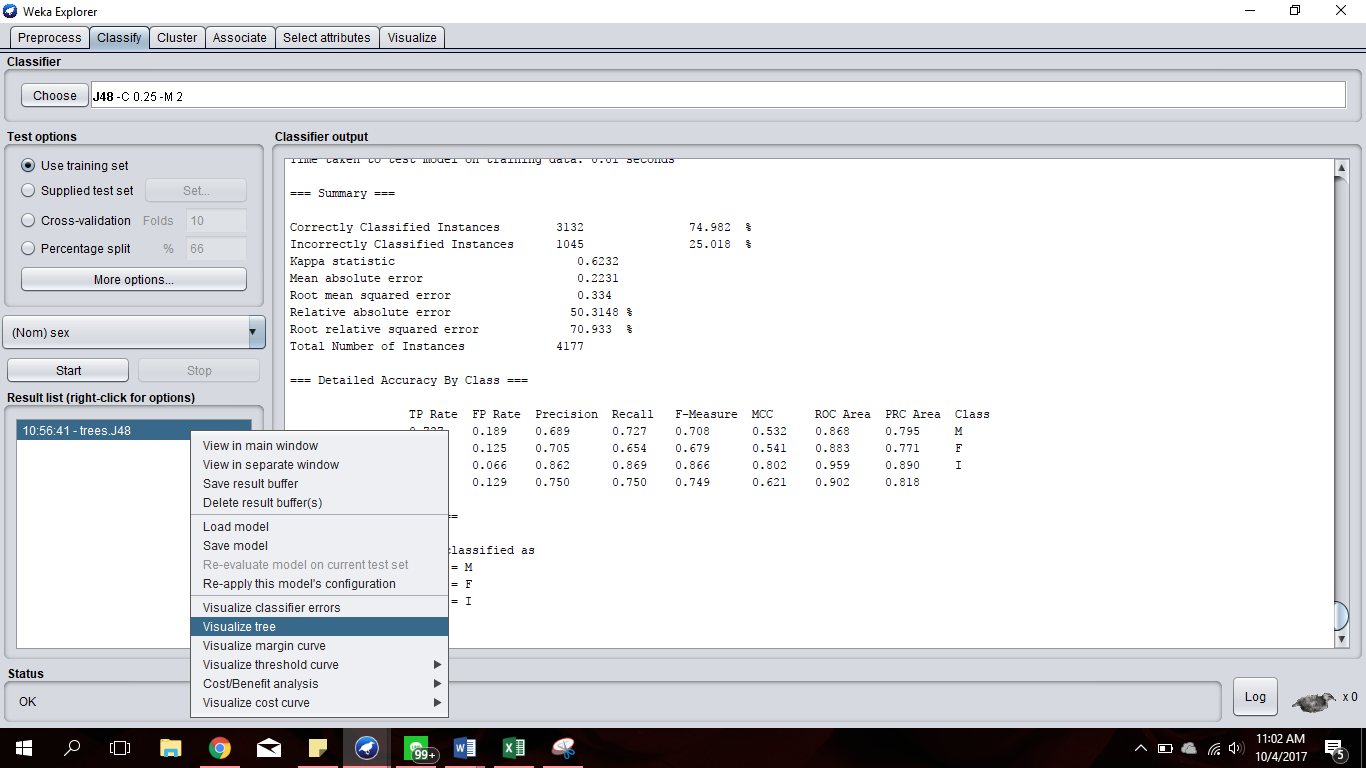
Setelah memilih J48 sebagai *classifier*, Lalu pilih metode yang akan digunakan untuk pengujian pada *test options*:

1. Menggunakan *Use Training Set*





Lalu klik kanan dan pilih *visualize tree*:

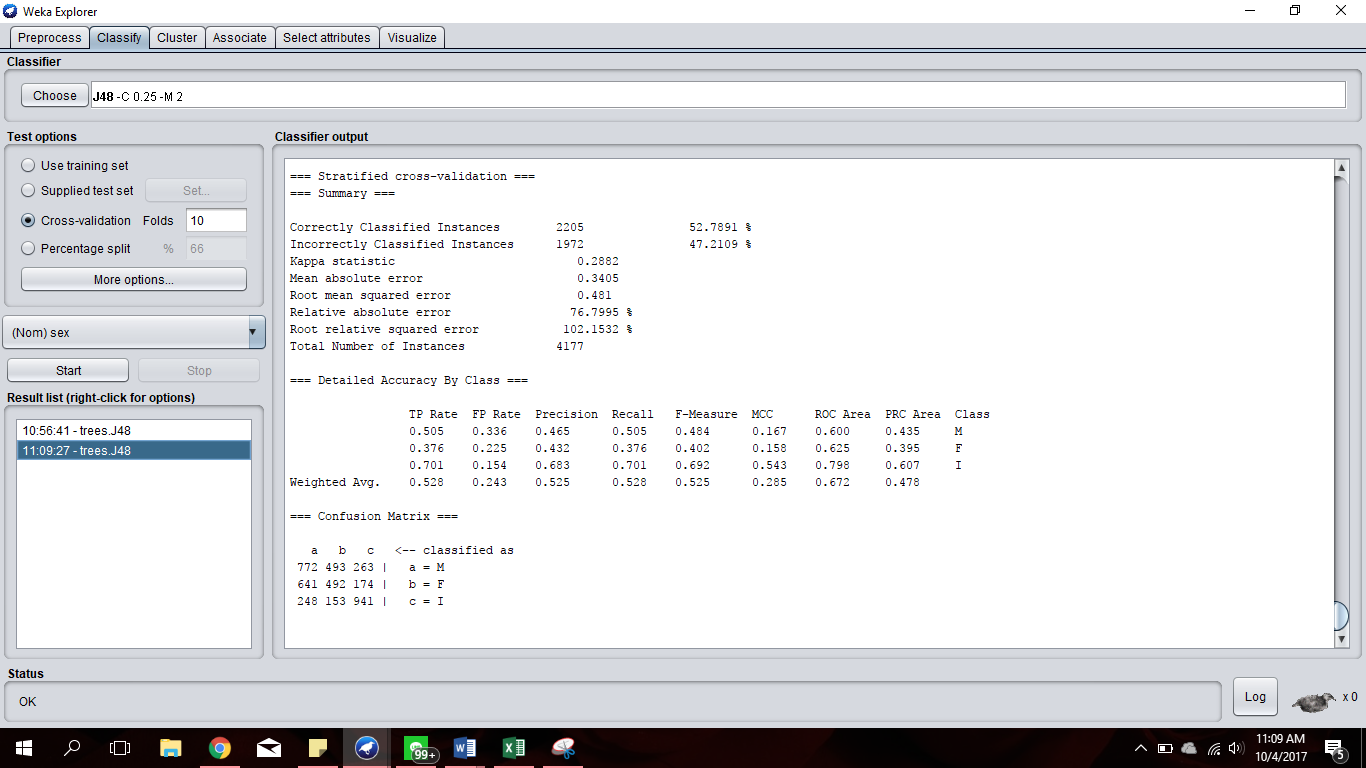


Hasil *visualize tree* adalah sebagai berikut:

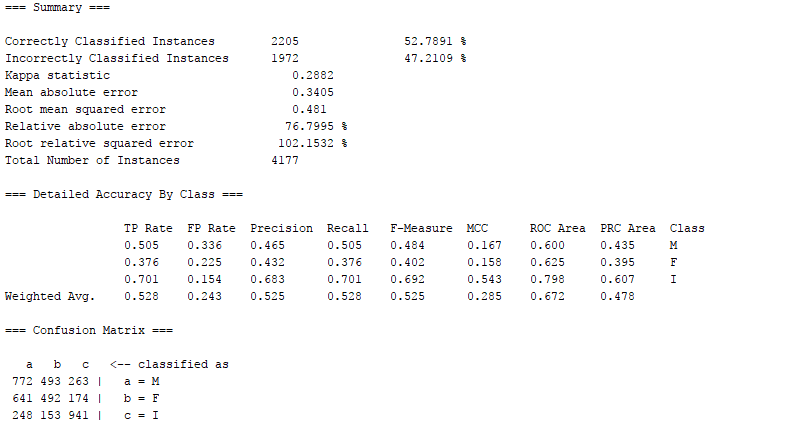


Jadi kesimpulan dengan menggunakan *Use Training Set* yaitu akurasi yang dihasilkan sebesar **74,982%**

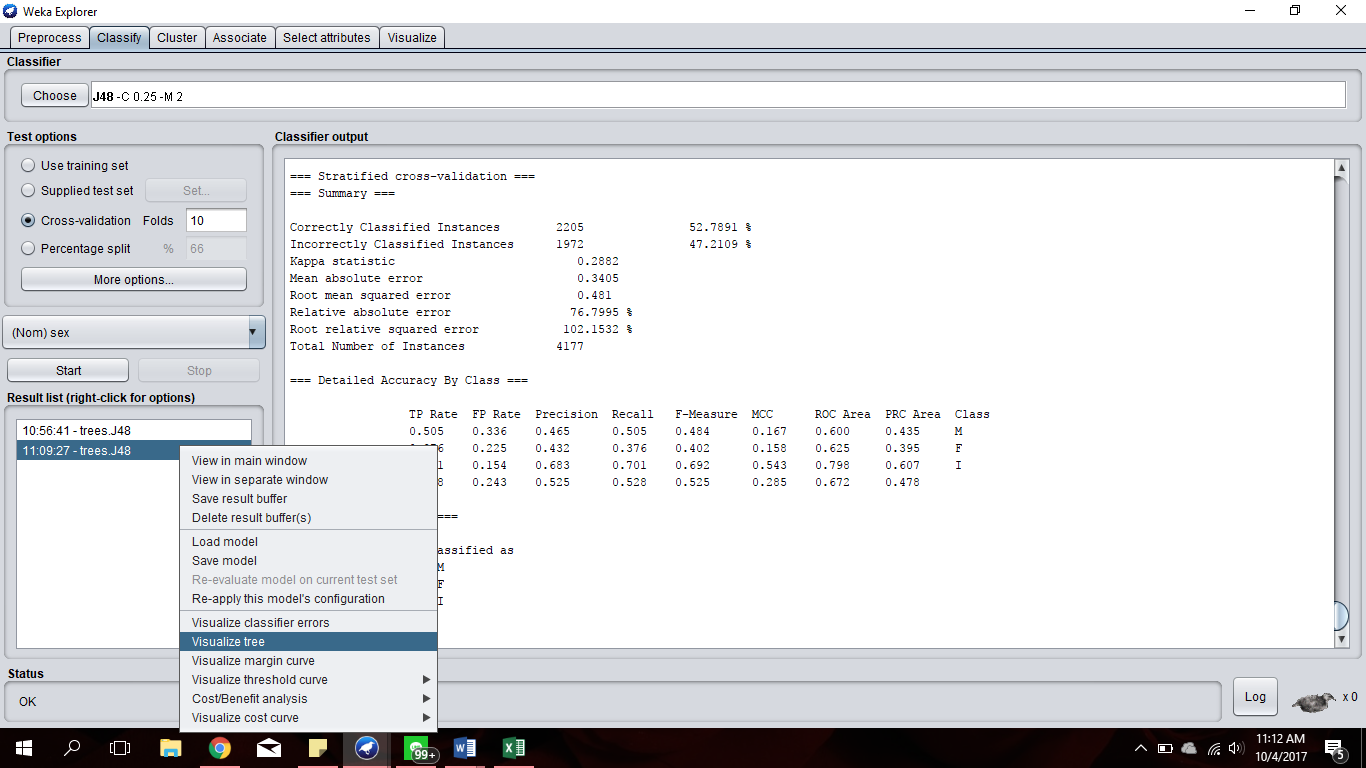
1. Menggunakan *Cross Validation Folds 10*



Dan hasilnya adalah sebagai berikut:



Lalu klik kanan dan pilih *visualize tree*:

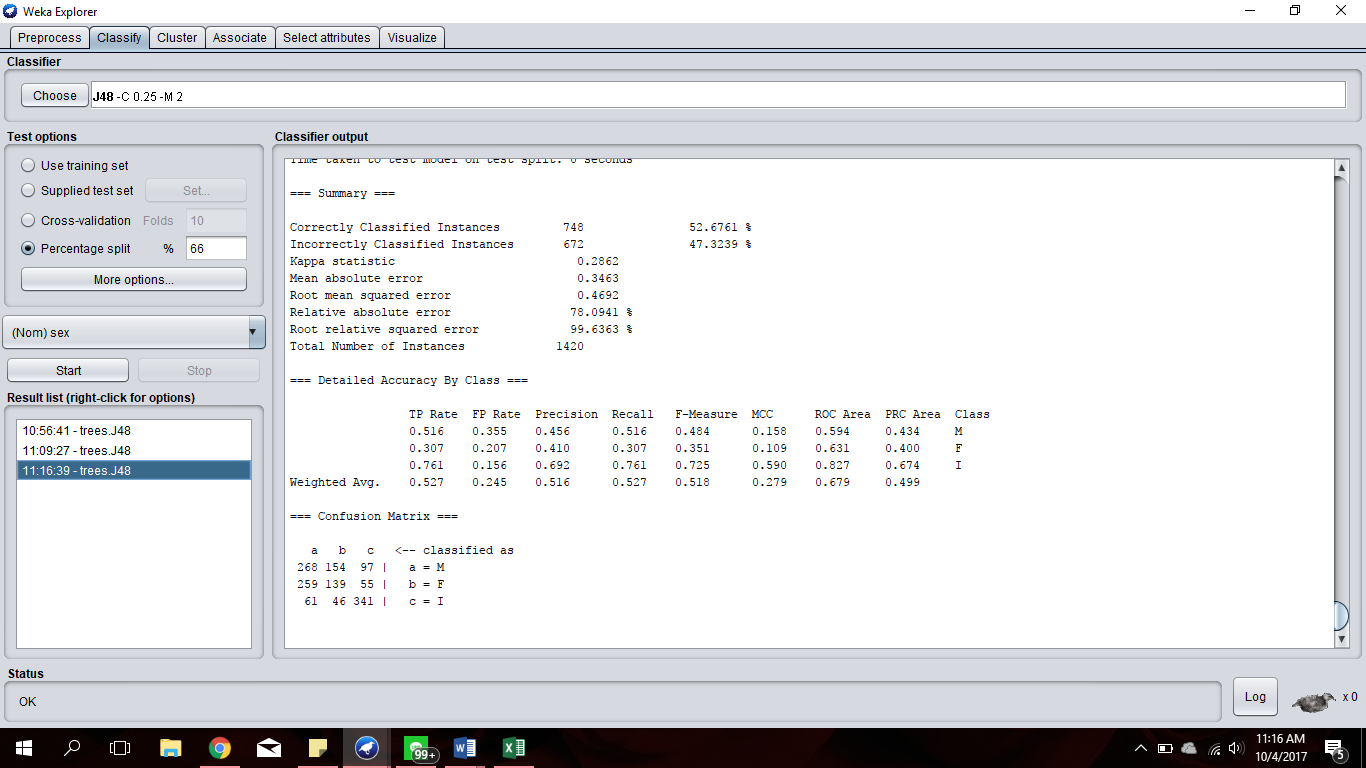


Hasil *visualize tree* nya adalah sebagai berikut:

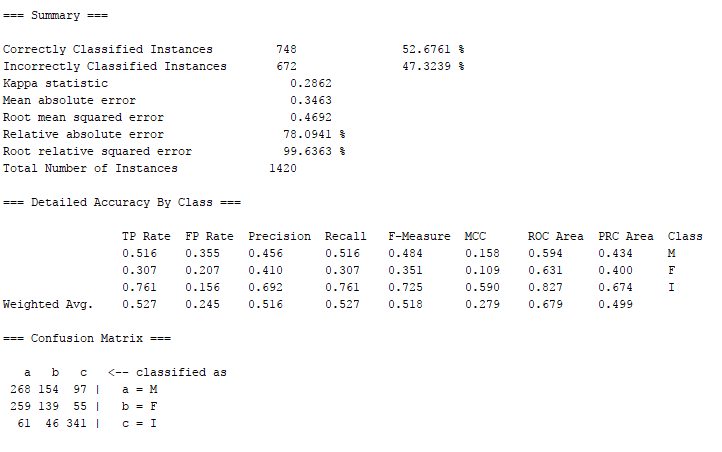


Jadi kesimpulan dengan menggunakan *Cross Validation Folds 10* yaitu akurasi yang dihasilkan sebesar **52,7891%**

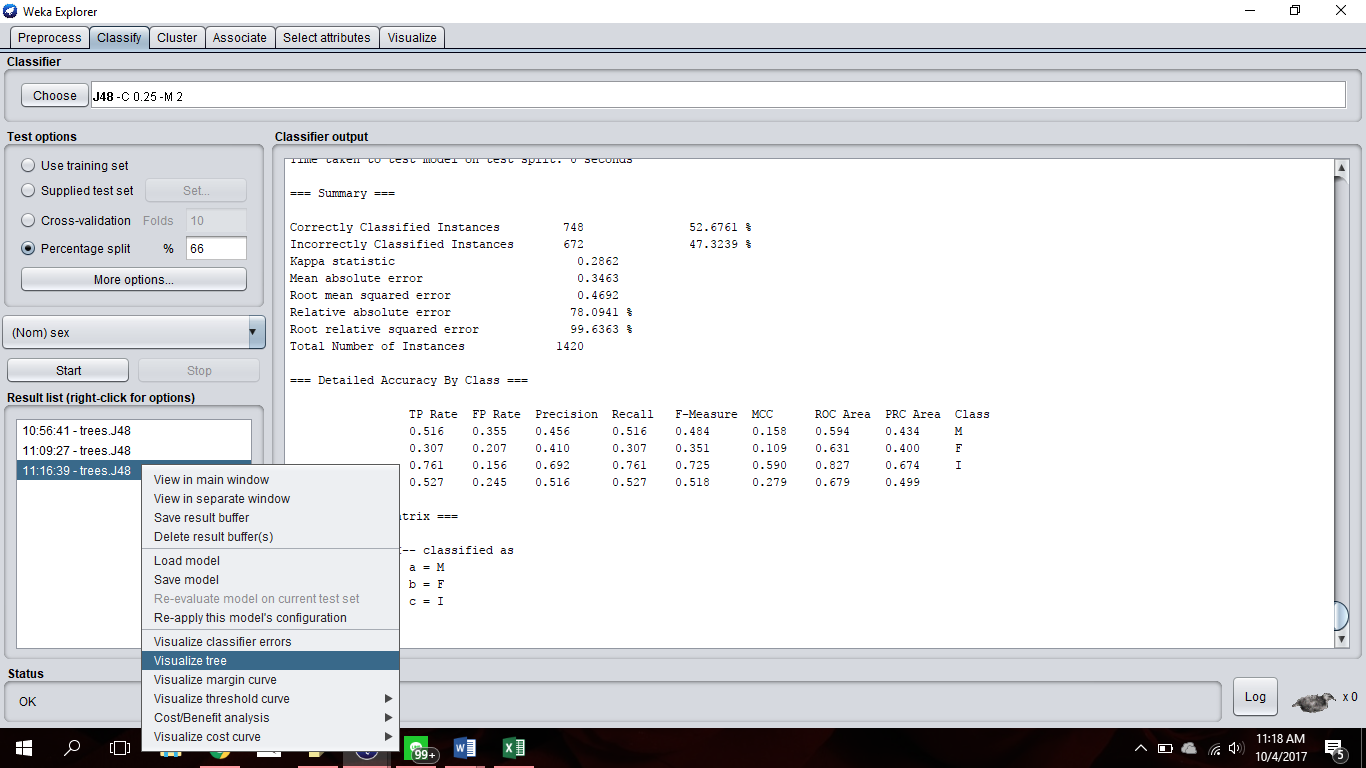
1. Menggunakan *Percentage Split*



Hasilnya adalah sebagai berikut:



Lalu klik kanan dan pilih *visualize tree*:



Hasil *visualize tree* nya adalah sebagai berikut:



Jadi kesimpulan dengan menggunakan *Percentage Split* yaitu akurasi yang dihasilkan sebesar **52,6761%**

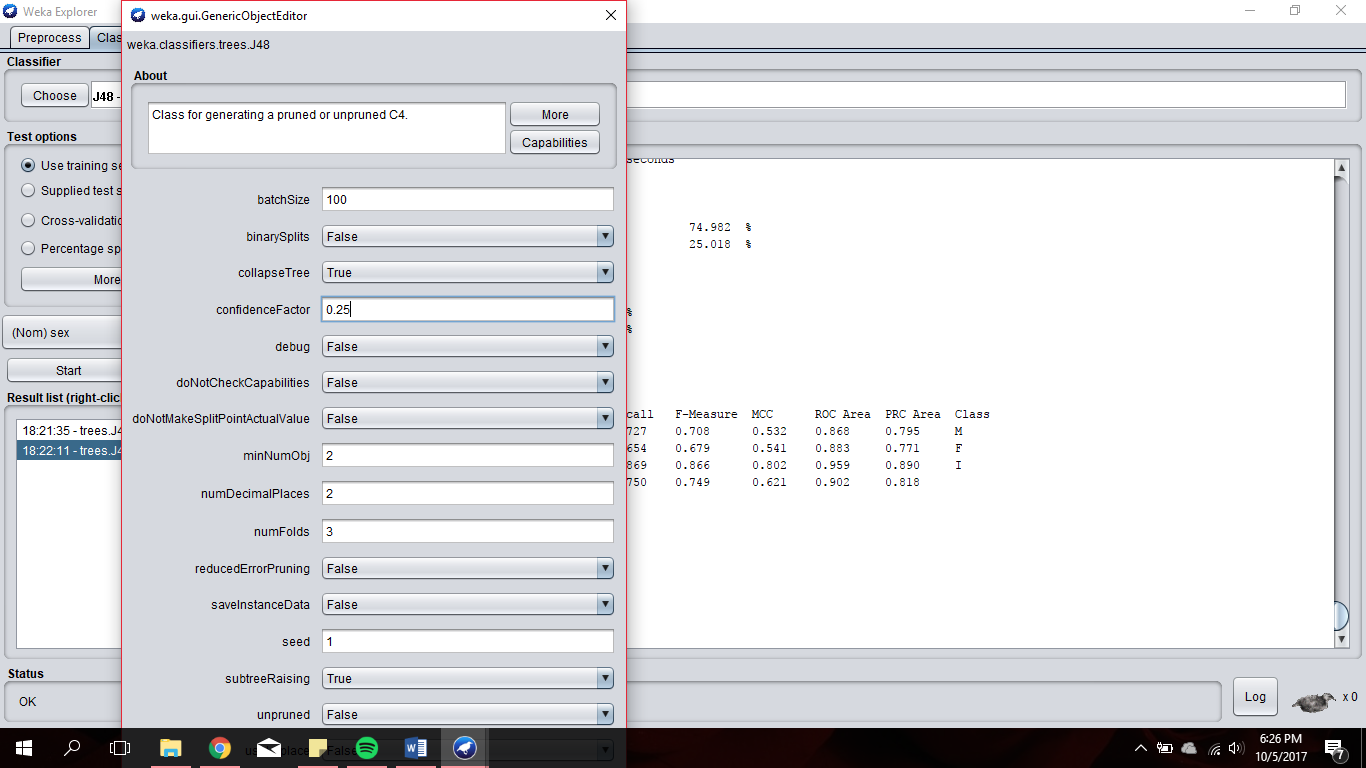
1. *What were your results? Show what decision trees you found.*

Berdasarkan 3 metode tes dan hasil decision tree yang telah dilakukan pada nomor 3, yaitu *Use Training Set,* *Cross Validation Folds 10* dan *Percentage Split,* akurasi terbesar dihasilkan oleh metode *Use Training Set* yaitu sebesar 74,982%

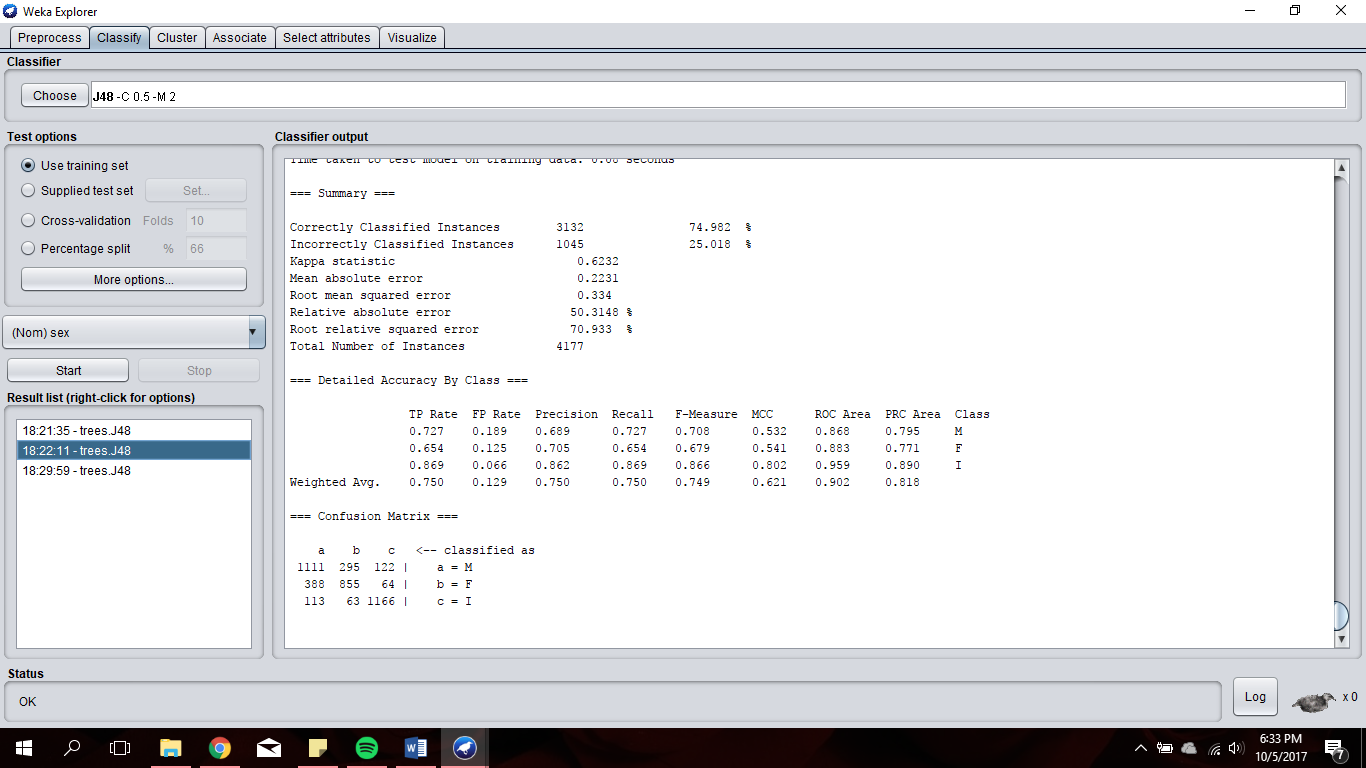
1. *What do the results tell us? Why are the results (in)accurate? Why did changing parameter(s) improve/degrade accuracy?*

Berdasarkan kesimpulan pada nomor 4, metode *Use Training Set* menghasilkan akurasi sebesar 74,982% dan data yang tidak akurat sebesar 25,018%. Data yang tidak akurat disebabkan oleh misclassified. Karena data yang diolah sangat banyak, maka terdapat kemungkinan terdapat data yang terselip/ tidak diolah sehingga menyebabkan data tidak akurat. Hal ini juga dialami oleh metode yang lain.

Parameter default pada Weka adalah sebagai berikut:



Semisal dengan menggunakan metode *Use Training Set,* nilai *confidence factor* pada parameter yaitu 0.25, dan akurasi yang dihasilkan yaitu sebesar **74,982%.**



Namun jika nilai *confidence factor* pada parameter dirubah menjadi 0.5 maka akurasi yang dihasilkan meningkat menjadi **76.7537%** .Sesuai gambar berikut:

