**TUGAS**

**KECERDASAN KOMPUTASIONAL**

**“Visualising Tree Using Weka Application”**

**ABALON & BREAST CANCER**



**Dosen :**

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

**Oleh :**

**Aulia Teaku NR 5115100012**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA**

**2017**

1. **Get the “Abalone” dataset from UCI ML Repo**

Name of Dataset : Abalone.data

Number of Instance : 4177

Number of Attributes : 8

Attribute Information :

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Data Type | Measurement | Description |
| Sex | Nominal |  | M, F, I (infant) |
| Length | Continuous | mm | Pengukuran kerang terpanjang |
| Diameter | Continuous | mm | Tegak lurus dengan panjang |
| Height | Continuous | mm | Dengan daging di dalamnya |
| Whole weight | Continuous | grams | Berat keseluruhan |
| Shucked weight | Continuous | grams | Berat daging |
| Viscera weight | Continuous | grams | Bobot usus/anat |
| Shell weight | Continuous | grams | Bobot kerang |
| Rings | Integer |  | +1.5 tahun, umur, diperdiksi, apakah sebagai continuous atau classification problem |

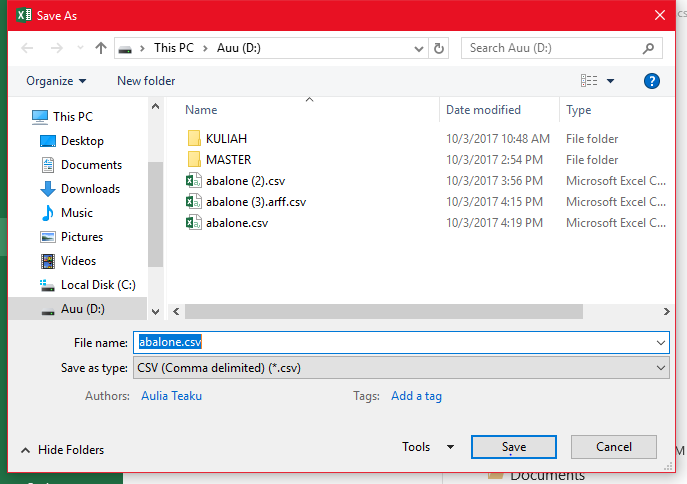
Missing attribute Values : none

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

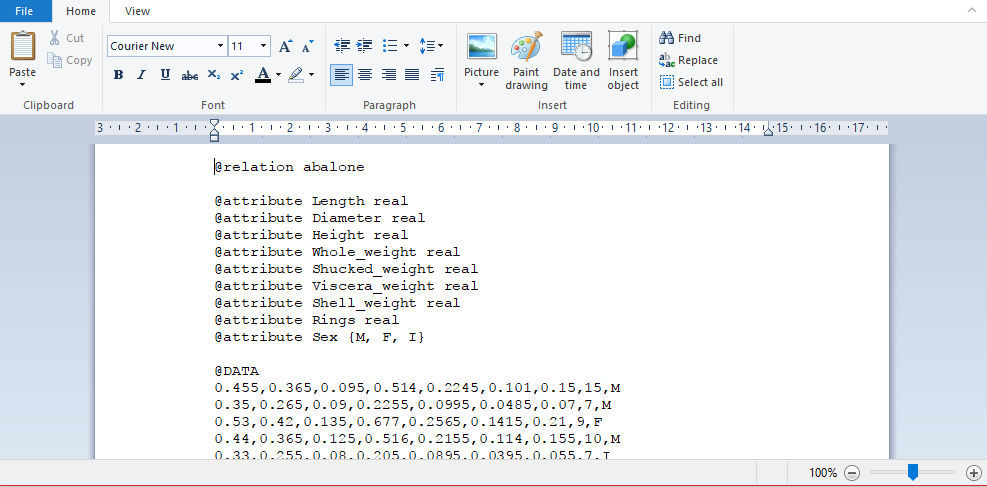
Langkah-langkah :

1. Merubah extensi dataset dari \*.data menjadi \*.csv

Bisa dengan cara rename, atau diopen di excel, save as Abalone.csv



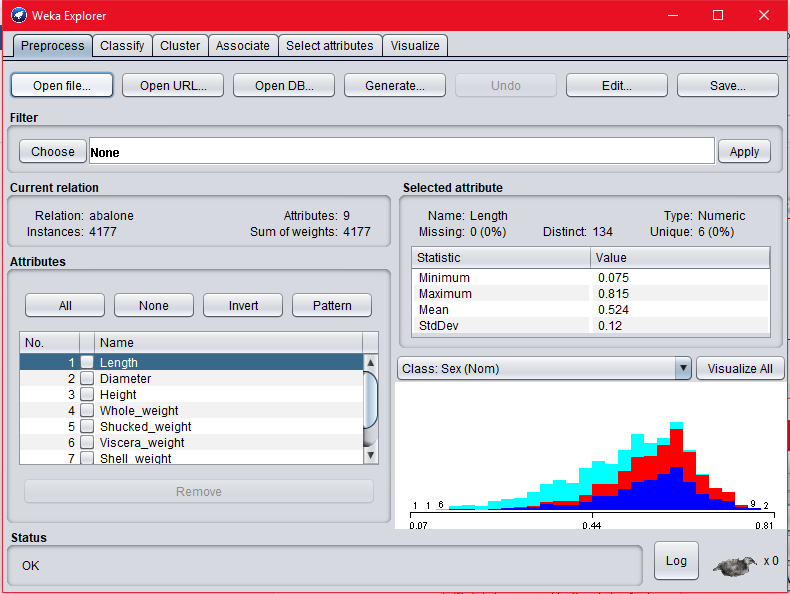
Abalone.csv di buka di note wordpad, dan tambahkan line seperti gambar dibawah, lalu save menjadi **Abalone.arff**



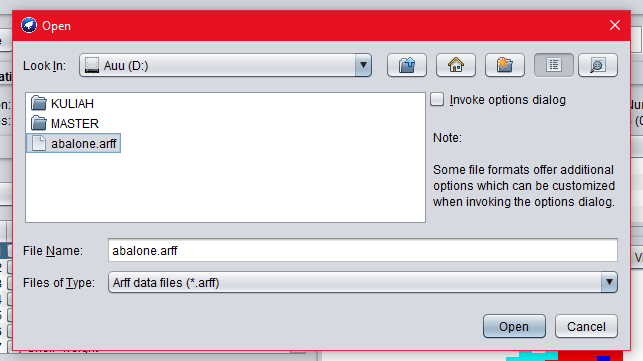
1. **Describe your process and methods. Parameter yang digunakan? Apa preproses yang dilakukan? Mengapa memilih parameter tersebut?**
2. Buka program weka, pilih menu EXPLORE



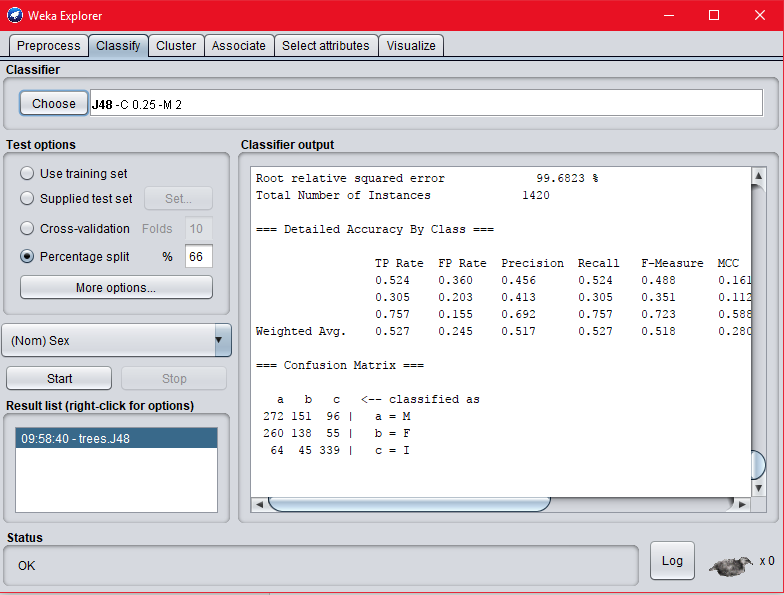
Akan muncul environment seperti gambar berikut : [PREPROCESS]



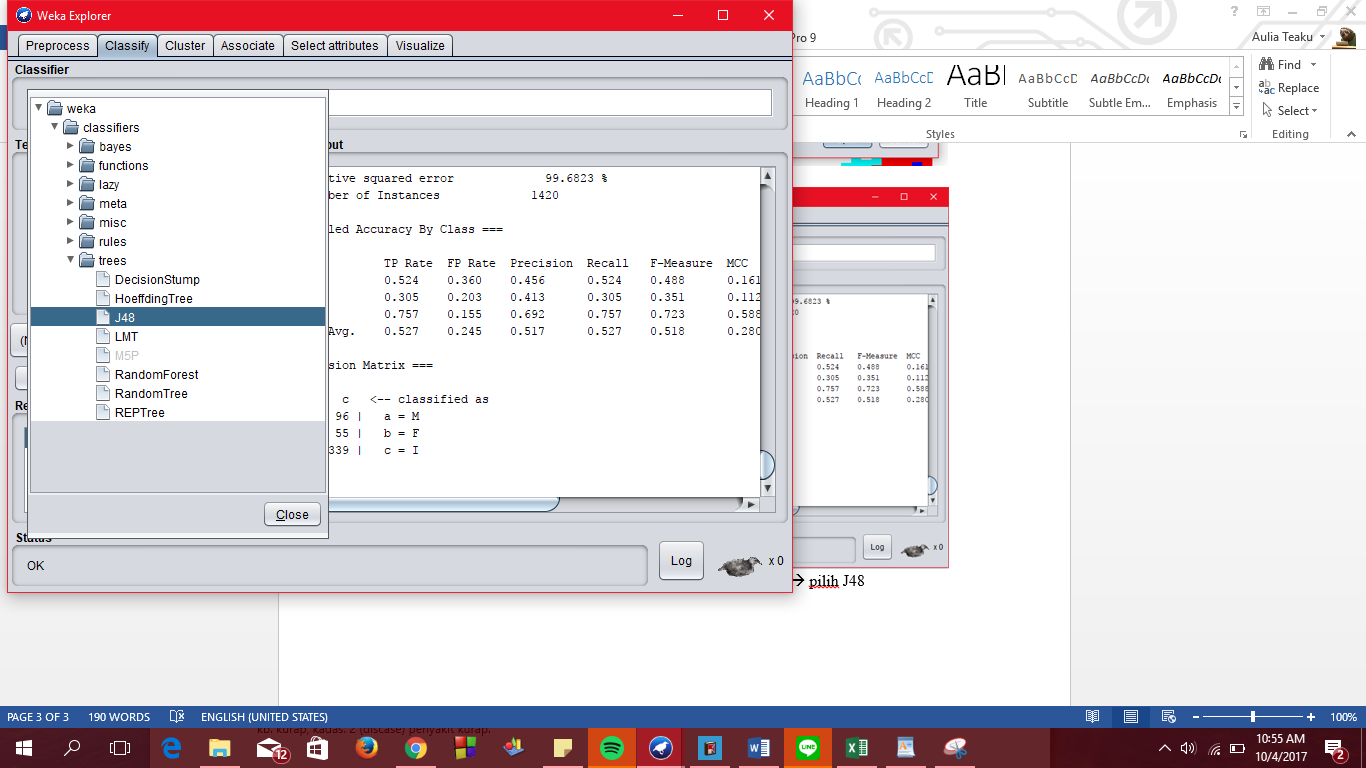
Lalu Open File 🡪 pilih Abalone.arff 🡪 OK



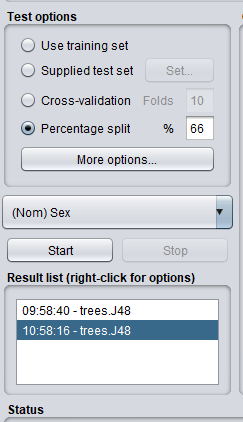
Pilih tab Classify, saya menggunakan parameter **TRAINING/TEST SIZE** dengat split size : 66% untuk training size nya.



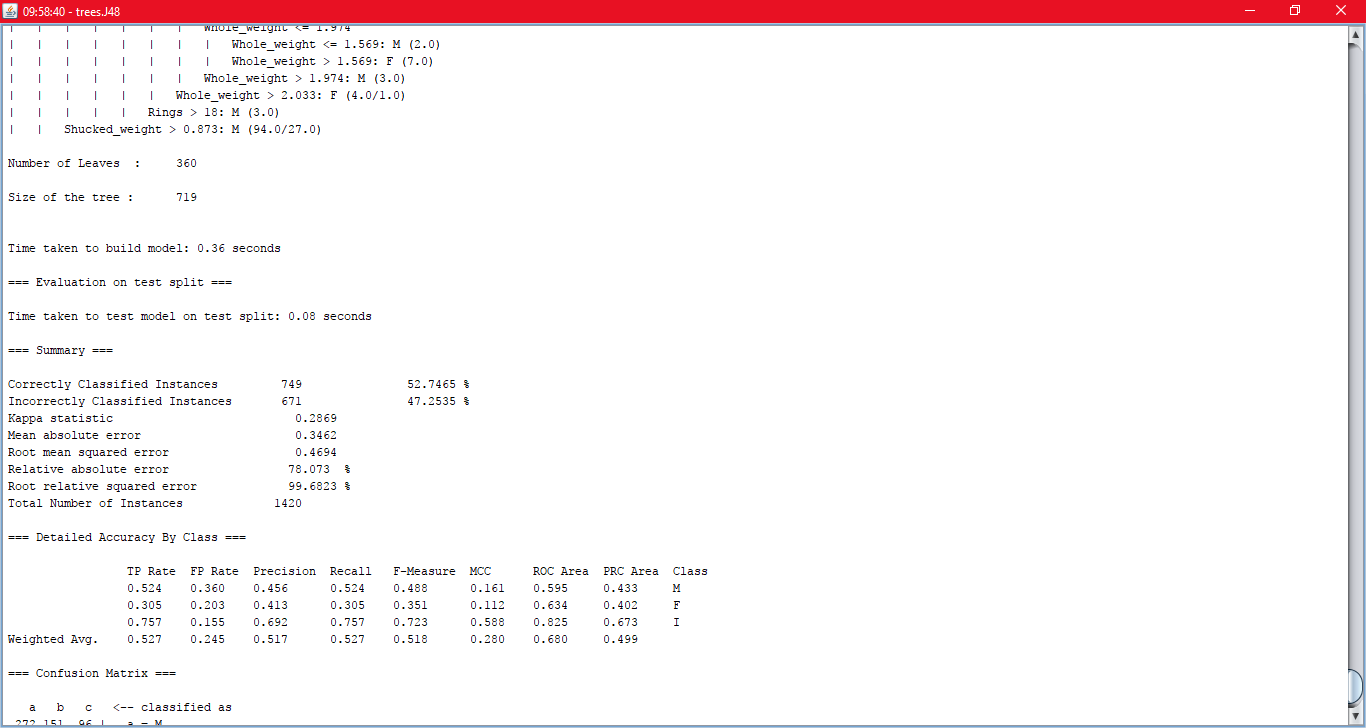
Pada tab classify, pilih Choose 🡪 klik Decision Tree 🡪 pilih J48



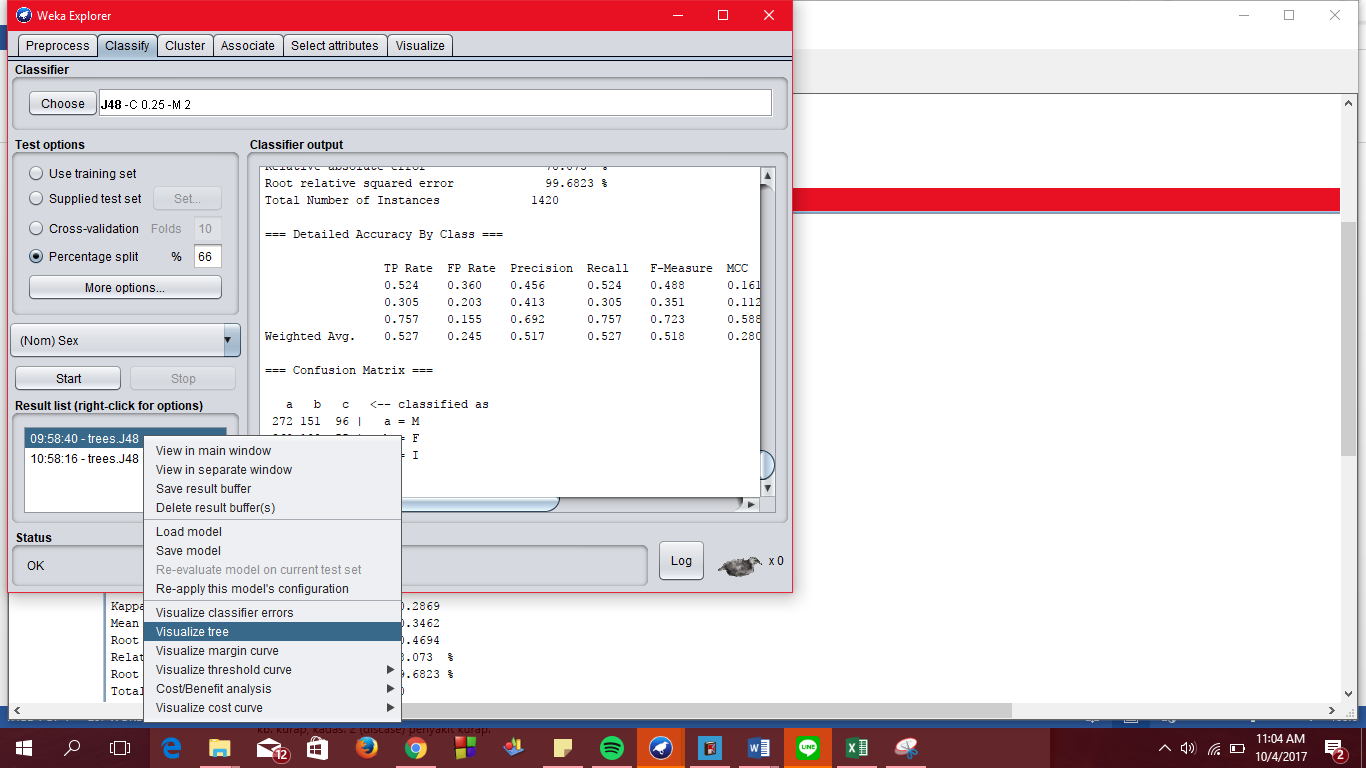
Pada table Test Option, pilih percentage split, lalu START



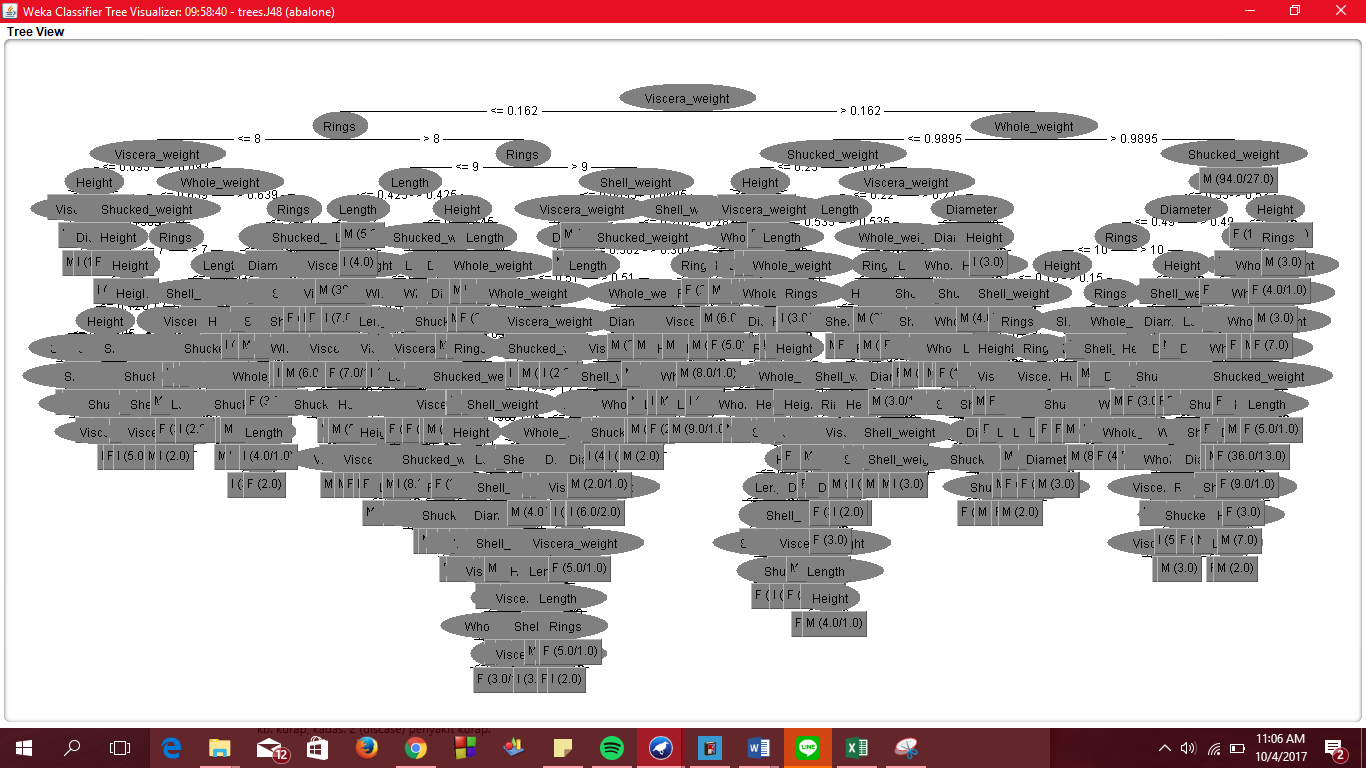
Maka akan muncul hasil seperti ini :



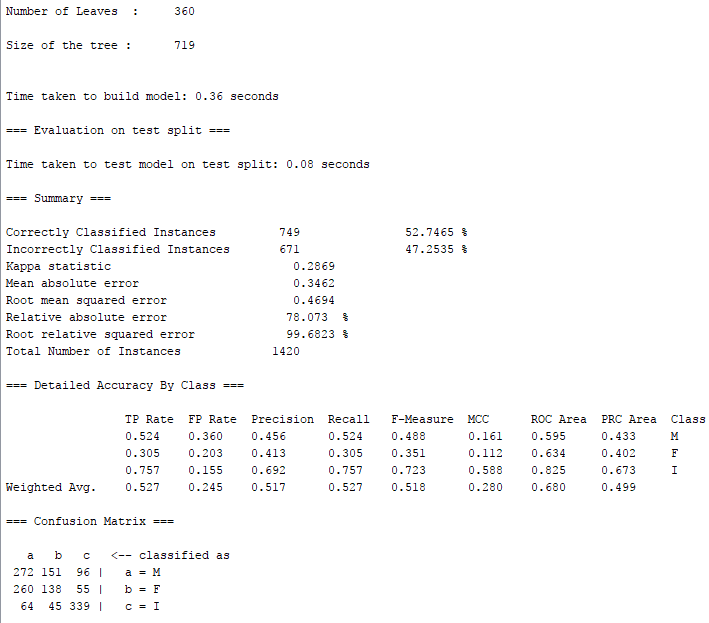
1. Untuk menampilkan tree, klik kanan pada resultnya, lalu pilih VISUALIZE TREE



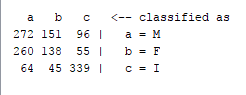
1. **What were your result? Show the decision tre you found.**



1. **What do the result tell us about?why are the result (in)accurate? Why did changing parameters improve/degrade accuracy?**



Kesimpulan : dengan menggunakan parameter training/test size, akurasi dari pengklasifikasian Abalone dataset adalah sebesar 52,75%. Number of rings menjadi classification problems, bukan continuous value, seperti pada Classifier Output pada WEKA dan bisa dilihat pada hasil tree. Keakurasian hanya sekitar 52% dan sekitar 47% data yang salah klasifikasi. Keakurasian ini disebabkan karena kita menggunakan J48 prunning tree. Pada dasarnya pruning tree mengurangi ukuran dari decision tree secara general maka akan mengurangi keakurasian secara general. Pruning mampu mengurangi overfitting dimana overfitting ini bias menyebabkan memperoleh akurasi yang sempurna. Selain itu, mengubah parameter juga mempengaruhi akurasi, karena tergantung berdasarkan apa kita mengklasifikasikan dataset. Sehingga mempengaruhi keakurasian. Berdasarkan hasil, beberapa sample diklasifikasian sebagai class berikut :



1. Using Breast Cancer Dataset

Name of Dataset : breastcancer.data

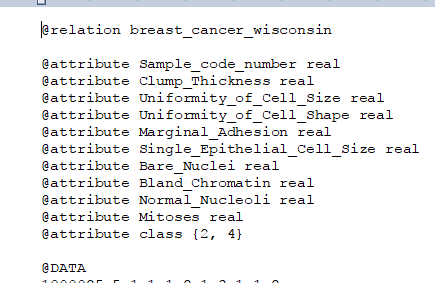
Number of instances : 699

Number of attribute : 10 plus class attribute

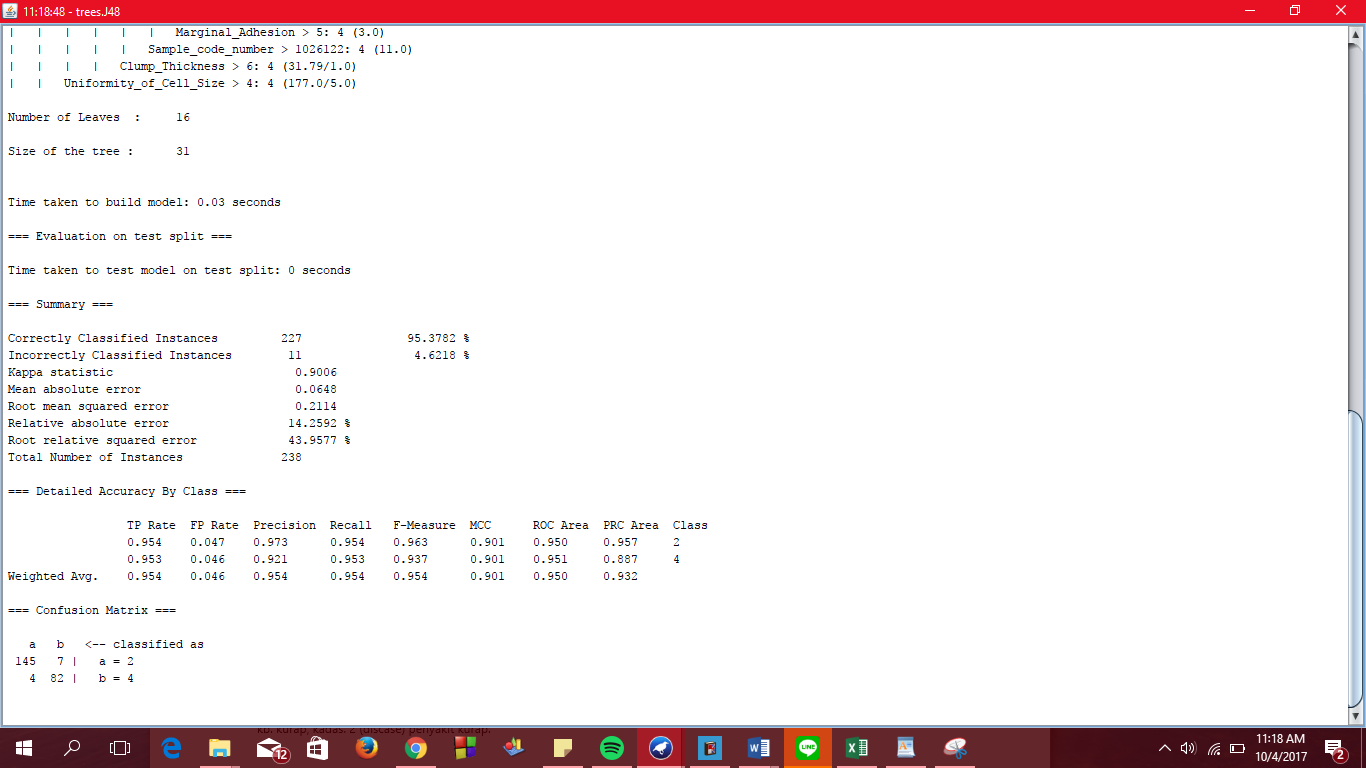
Attribute information :

|  |  |
| --- | --- |
| Attribute | Domain |
| Sample Code Number | Id number |
| Clump Thickness | 1-10 |
| Uniformity of cell size | 1-10 |
| Uniformity of cell shape | 1-10 |
| Marginal adhesion | 1-10 |
| Single epithelial cell size | 1-10 |
| Bare nuclei | 1-10 |
| Bland chromatin | 1-10 |
| Normal nucleoli | 1-10 |
| Mitoses | 1-10 |
| Class : | 2 for benign, 4 for malignant |

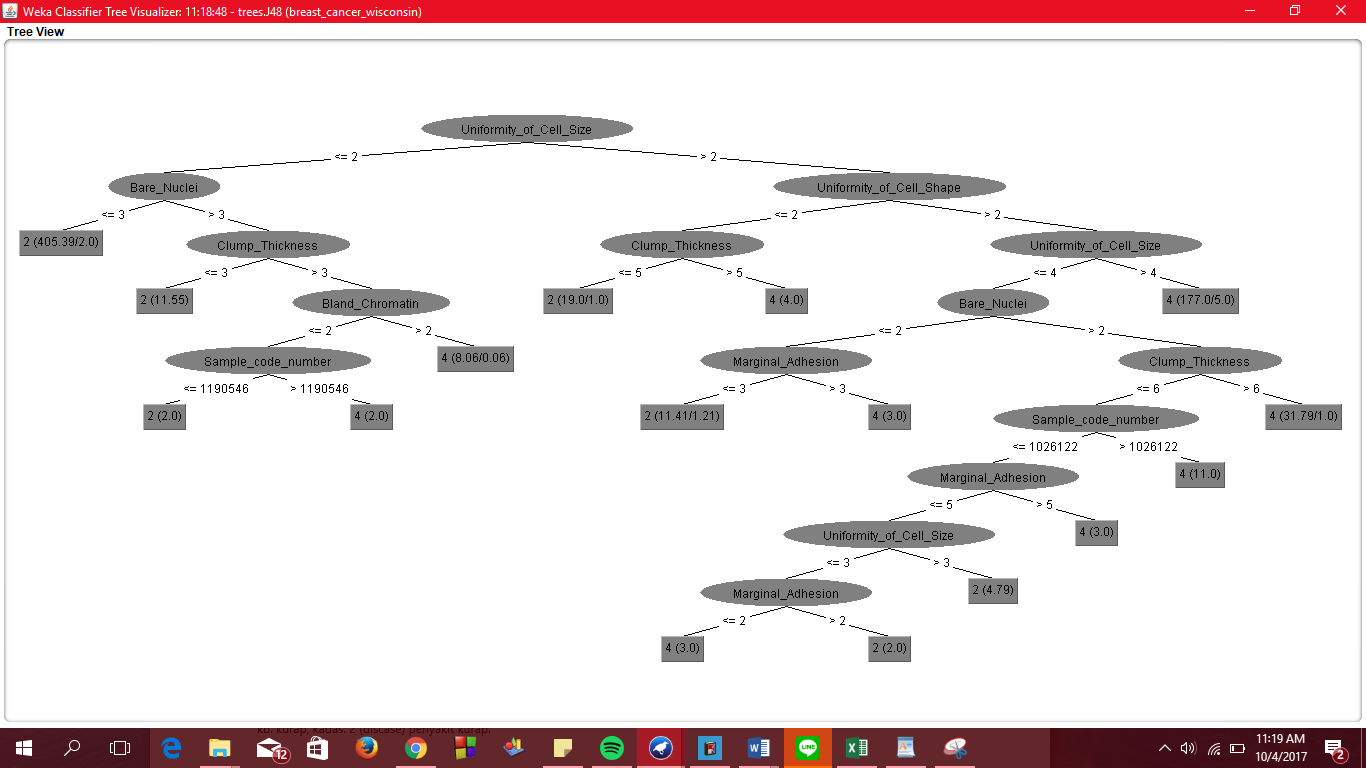
Langkah langkah, sama seperti diatas, perbedaannya pada file breastcancer.a*rff* ditambah line berikut :



Hasil dari classifier output adalah sebagai berikut :



Hasil tree nya :



Kesimpulannya : akurasi pengklasifikasian pada dataset breast cancer adalah sekitar 95%. Root dari tree nya adalah Uniformity of Cell Size.