

Dataset yang digunakan adalah dataset Abalone dari

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

Dataset ini digunakan untuk mengetahui usia sebuah Kerang (abalone) dimana ciri-ciri fisiknya menggambarkan usia kerang tersebut. Seperti jumlah ring-nya (garis-garis melingkar yang dapat ditemukan jika memotong cangkangnya).

a. Number of Instances : 4177

b. Number of Attributes: 9

c. Attribute Information:

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

d. Missing Attribute Values: None

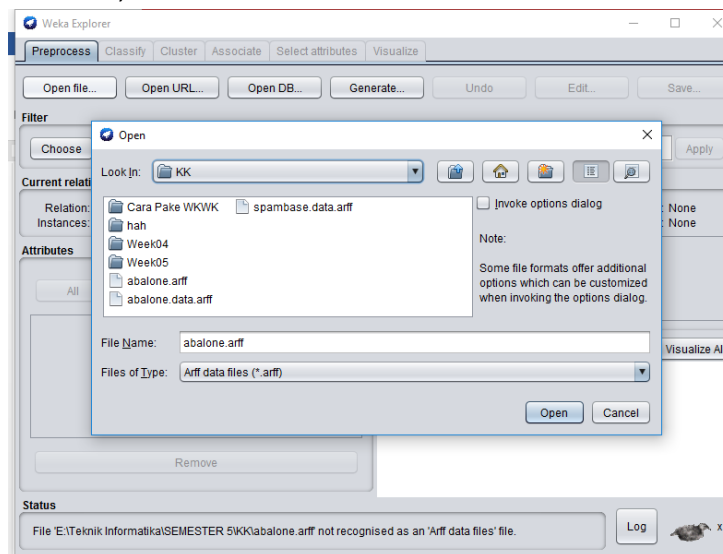
e. Class Distribution:

Class	Examples
----	-----
1	1
2	1
3	15
4	57
5	115
6	259
7	391
8	568
9	689
10	634
11	487
12	267
13	203

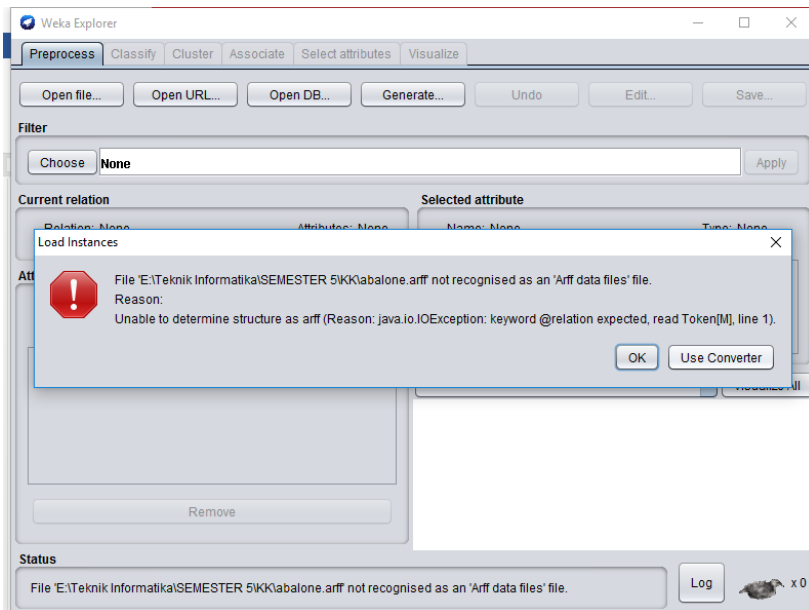
14	126
15	103
16	67
17	58
18	42
19	32
20	26
21	14
22	6
23	9
24	2
25	1
26	1
27	2
29	1
----	----
Total	4177

Soal Tugas DT

1. Describe your process and method.
 - a. Pertama, memilih dataset untuk di tes di WEKA

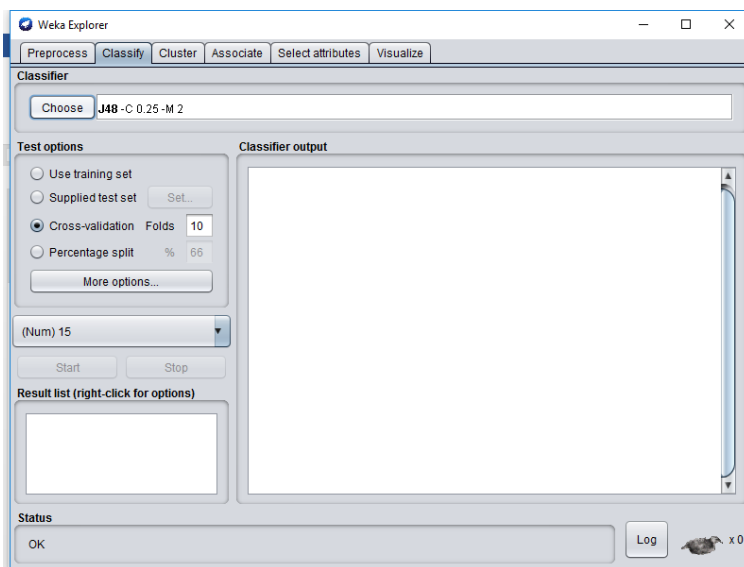


Sesuai yang ditentukan dari tugas, dataset yang digunakan adalah Dataset Abalone



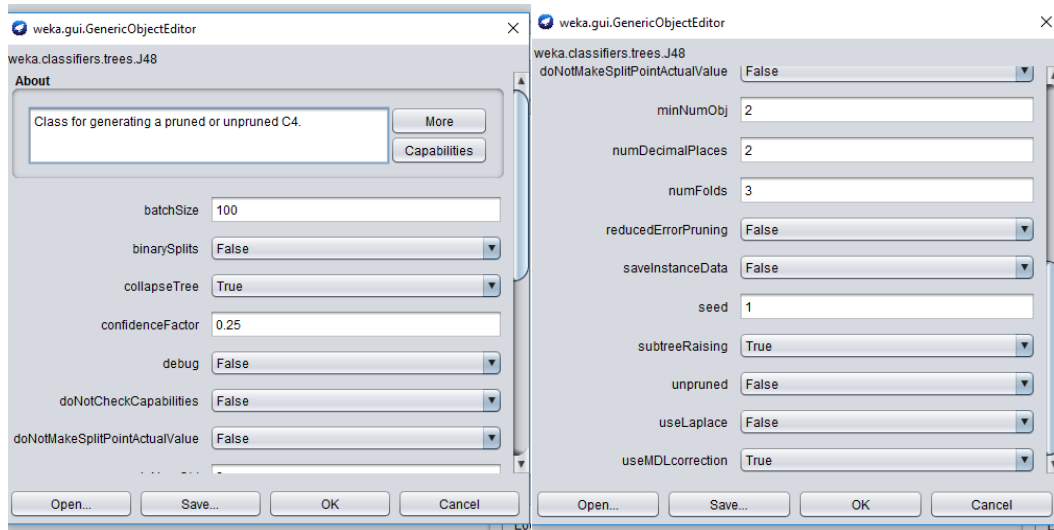
Jika muncul pop-up dialog seperti ini, pilih saja Use Converter, lalu klik OK

b. Kedua, pindah ke tab menu “Classify”



Pada tombol “Choose”, pilih Decision Tree J48

- c. Tentukan parameter dengan meng-klik tab **J58-C 0.25-M 2**

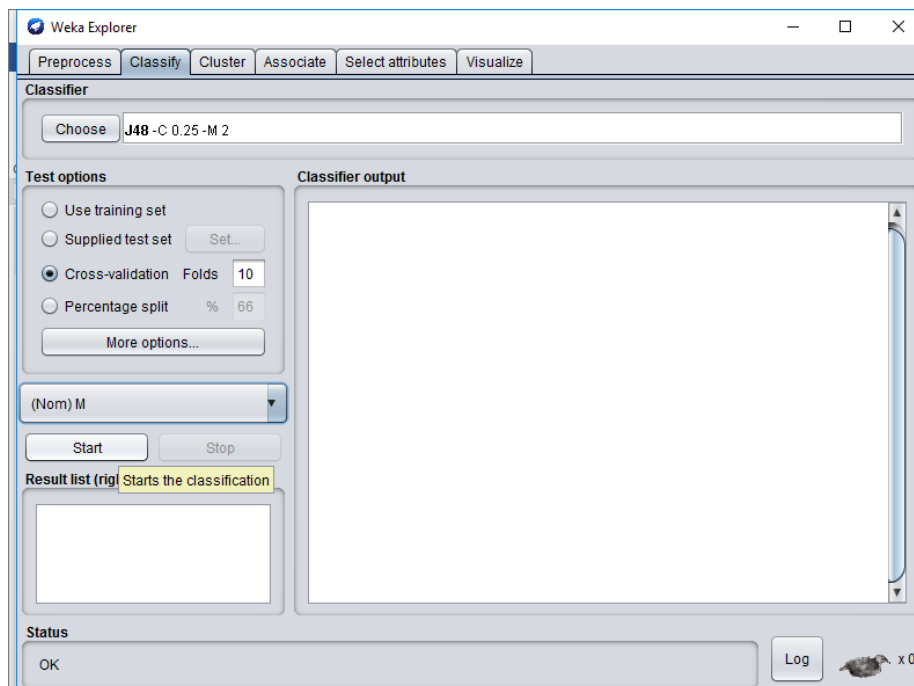


Parameter yang saya pilih default. Versi default ini menggunakan pendekatan Subtree Raising, dan tidak menghasilkan error pruning.

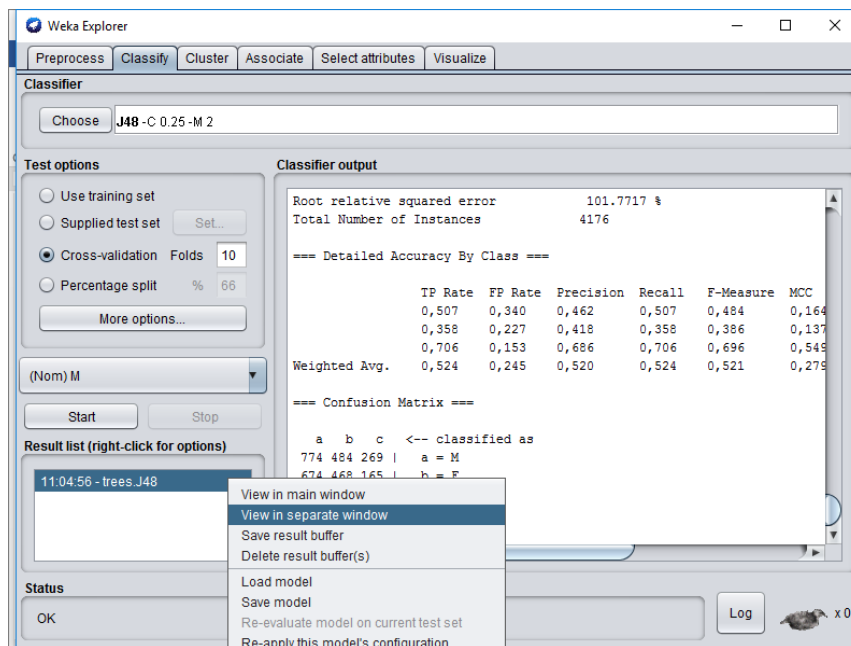
2. What were your result?

Untuk mengetahui perbandingan hasilnya, saya menggunakan 3 method:

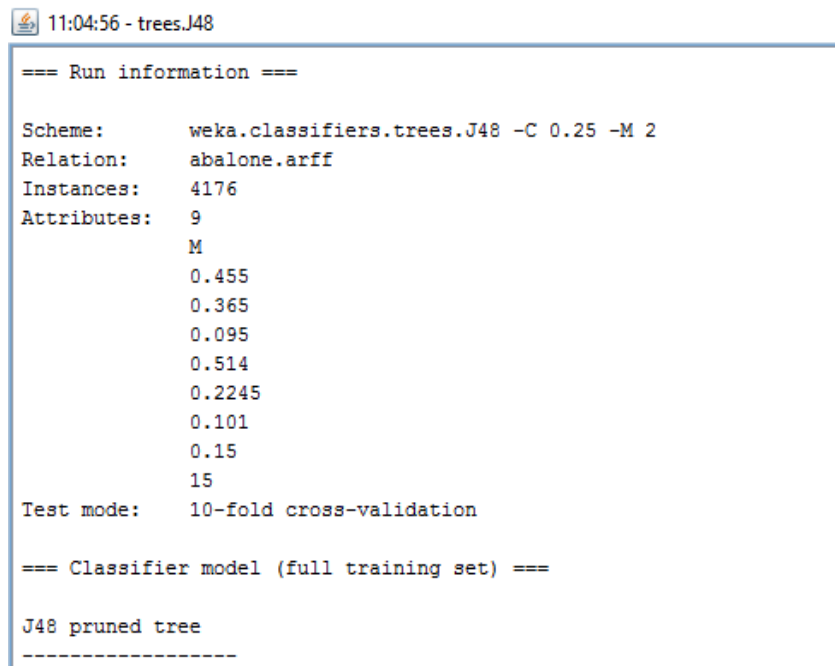
- a. Cross Validation

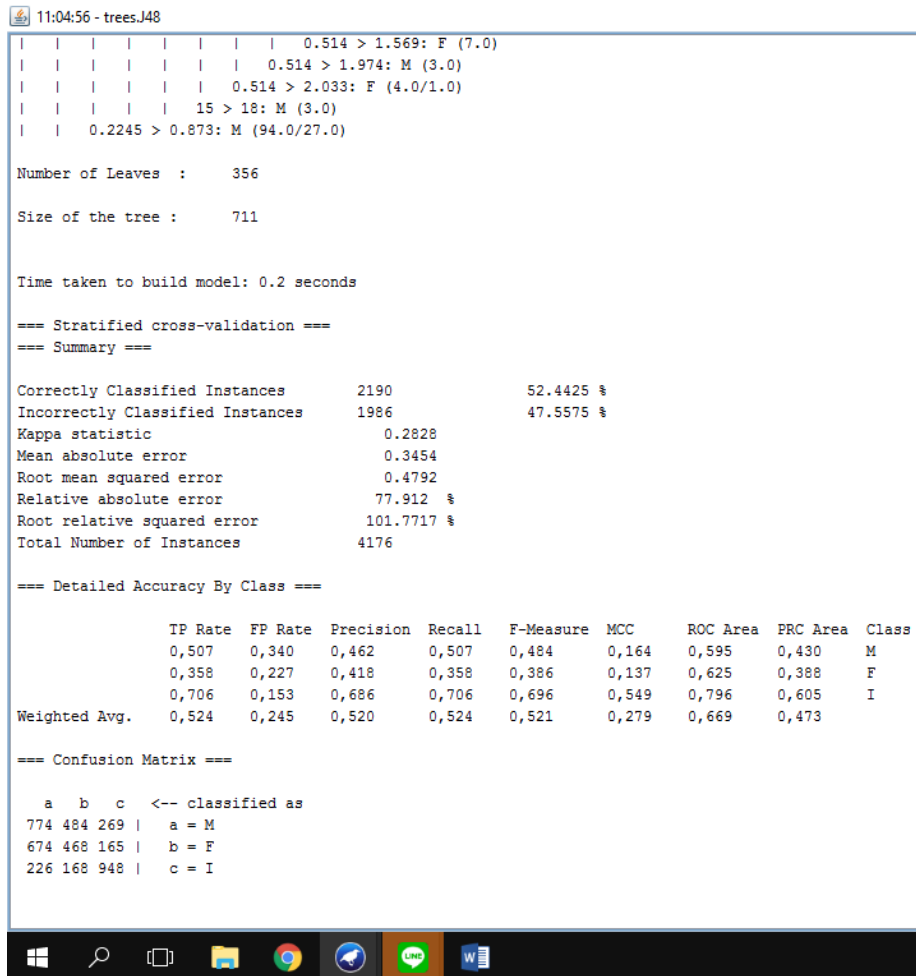


Di menu “Test Option” pada main panel, pilih Cross-Validation dengan Folds=10. Lalu klik “Start”.

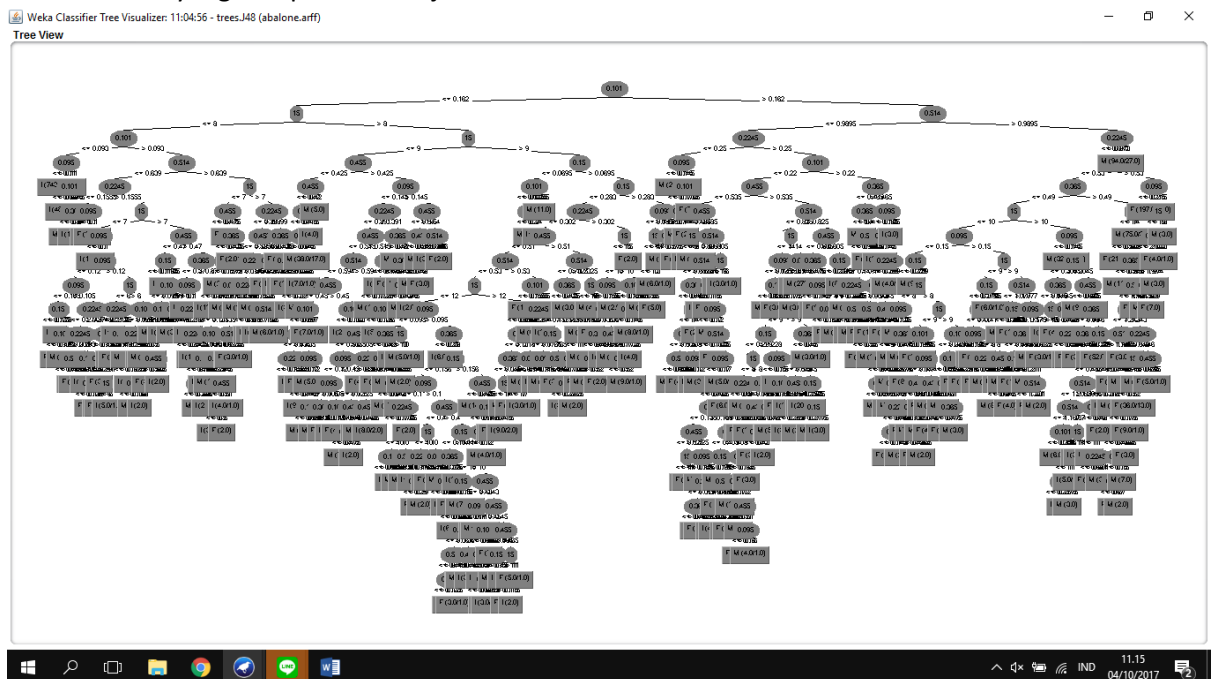


Hasilnya akan terlihat di panel “Classifier Output” Selanjutnya pilih “View in Separate Window” untuk melihat hasilnya agar lebih jelas.

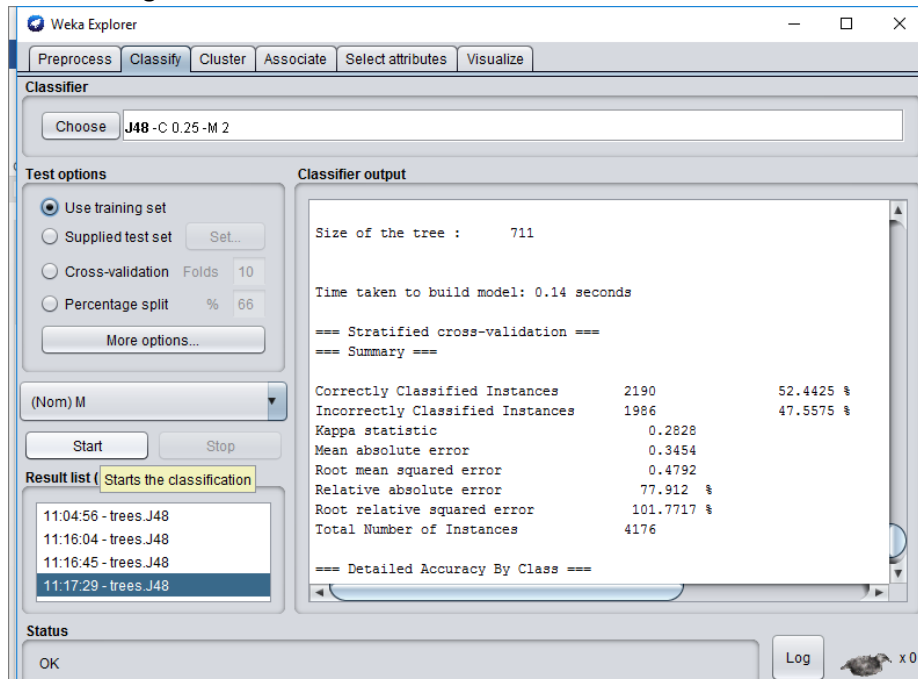




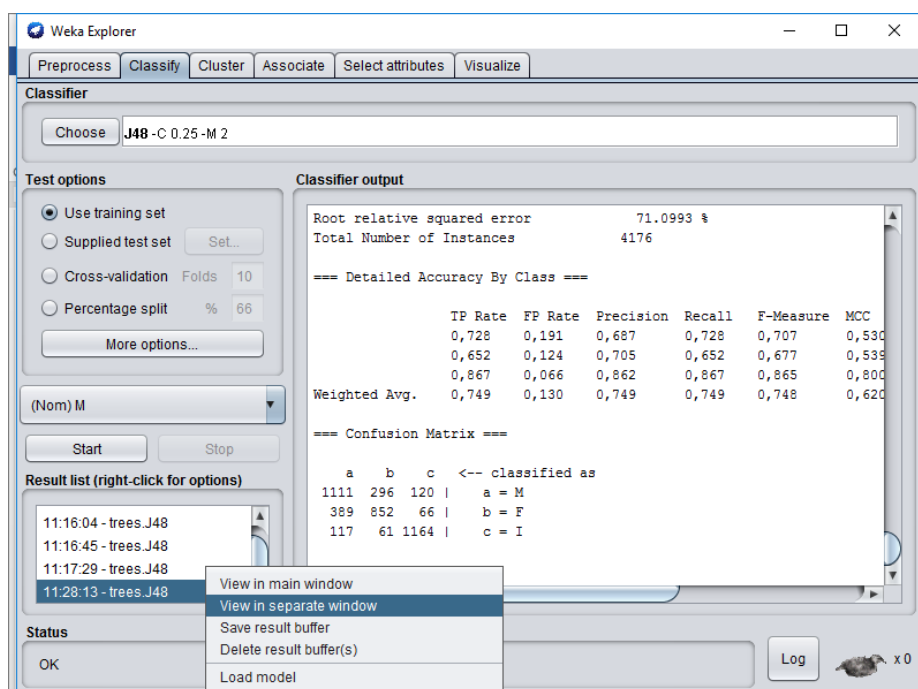
Decision Tree yang didapatkan menjadi



b. Use Training Set



Di menu “Test Option” pada main panel, pilih Use training set. Lalu klik “Start”.



Hasilnya akan terlihat di panel “Classifier Output” Selanjutnya pilih “View in Separate Window” untuk melihat hasilnya agar lebih jelas.

11:28:13 - trees.J48

```
| | | | | 15 > 18: M (3.0)
| | 0.2245 > 0.873: M (94.0/27.0)
```

Number of Leaves : 356

Size of the tree : 711

Time taken to build model: 0.13 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	3127	74.8803 %
Incorrectly Classified Instances	1049	25.1197 %
Kappa statistic	0.6216	
Mean absolute error	0.2241	
Root mean squared error	0.3348	
Relative absolute error	50.551 %	
Root relative squared error	71.0993 %	
Total Number of Instances	4176	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,728	0,191	0,687	0,728	0,707	0,530	0,867	0,793	M
	0,652	0,124	0,705	0,652	0,677	0,539	0,883	0,770	F
	0,867	0,066	0,862	0,867	0,865	0,800	0,959	0,889	I
Weighted Avg.	0,749	0,130	0,749	0,749	0,748	0,620	0,901	0,817	

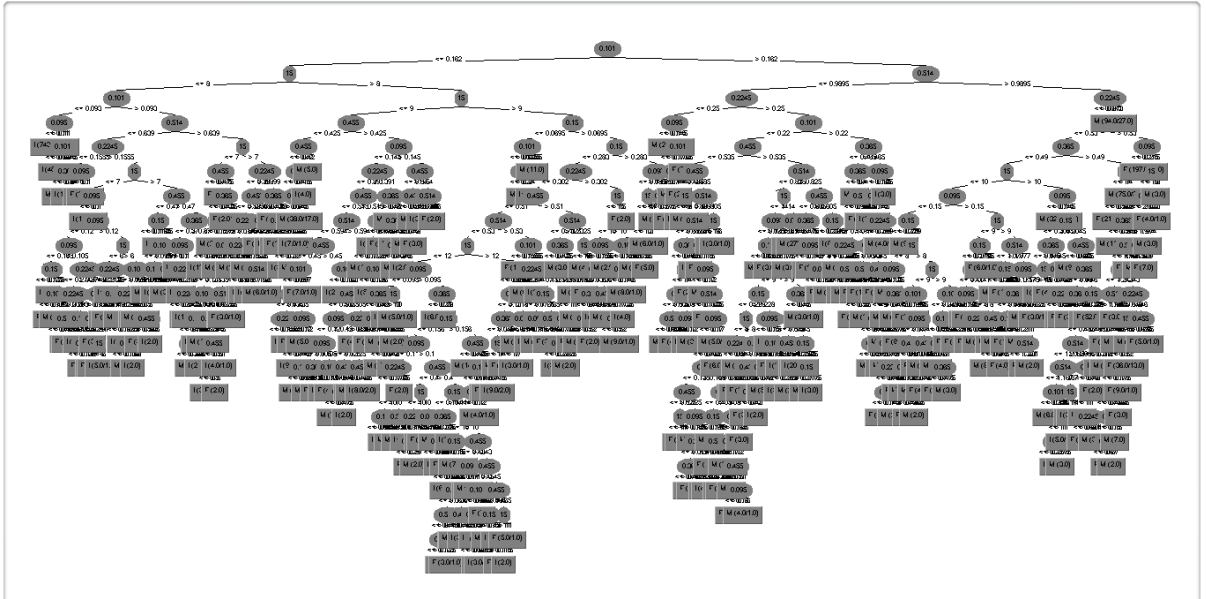
=== Confusion Matrix ===

```
a b c <-- classified as
1111 296 120 | a = M
389 852 66 | b = F
117 61 1164 | c = I
```

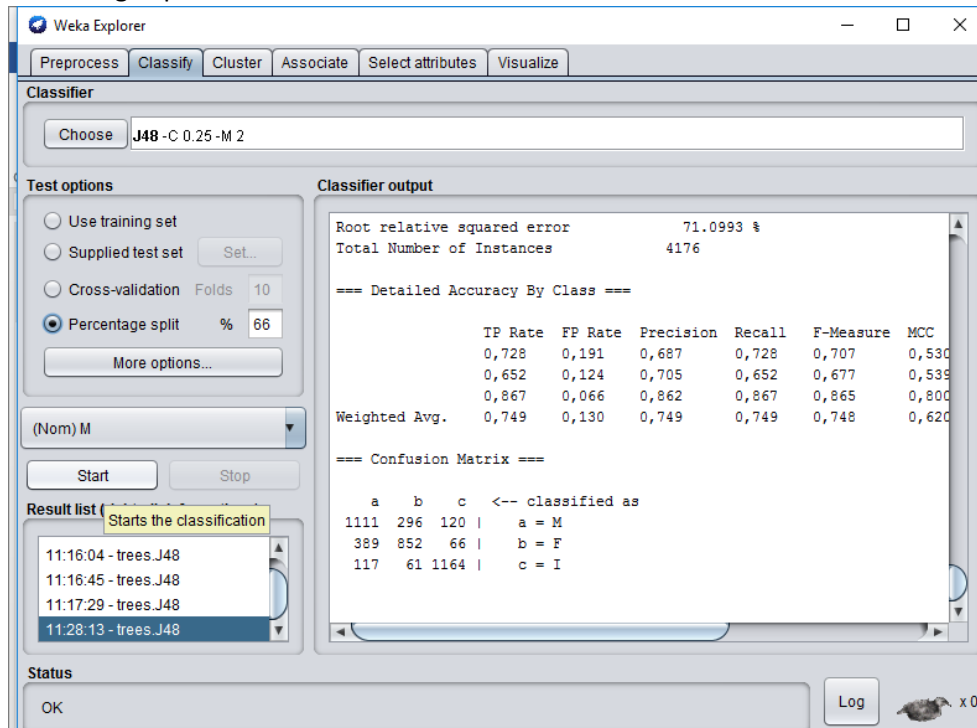
Decision Tree yang didapatkan menjadi

Weka Classifier Tree Visualizer: 11:28:13 - trees.J48 (abalone.aff)

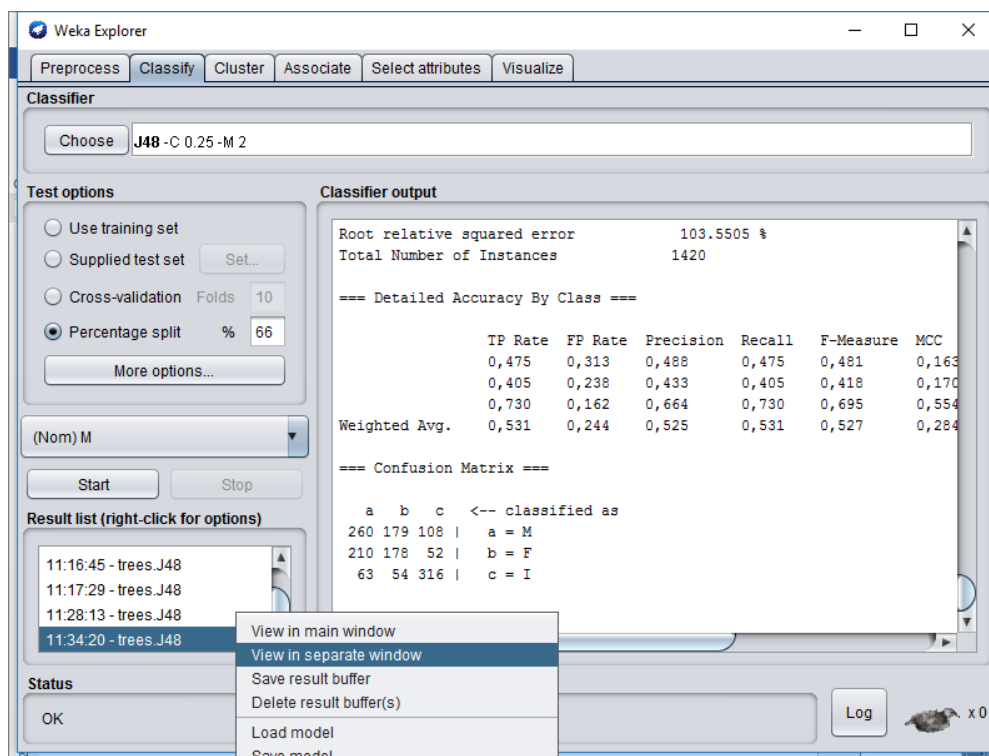
Tree View



c. Percentage Split



Di menu “Test Option” pada main panel, pilih Percentage split = 66%. Lalu klik “Start”.



Hasilnya akan terlihat di panel “Classifier Output” Selanjutnya pilih “View in Separate Window” untuk melihat hasilnya agar lebih jelas.

```
11:34:20 - trees.J48
| | | | | 15 > 18: M (3.0)
| | 0.2245 > 0.873: M (94.0/27.0)

Number of Leaves : 356

Size of the tree : 711

Time taken to build model: 0.2 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances 754 53.0986 %
Incorrectly Classified Instances 666 46.9014 %
Kappa statistic 0.2931
Mean absolute error 0.3366
Root mean squared error 0.4872
Relative absolute error 75.9178 %
Root relative squared error 103.5505 %
Total Number of Instances 1420

=== Detailed Accuracy By Class ===

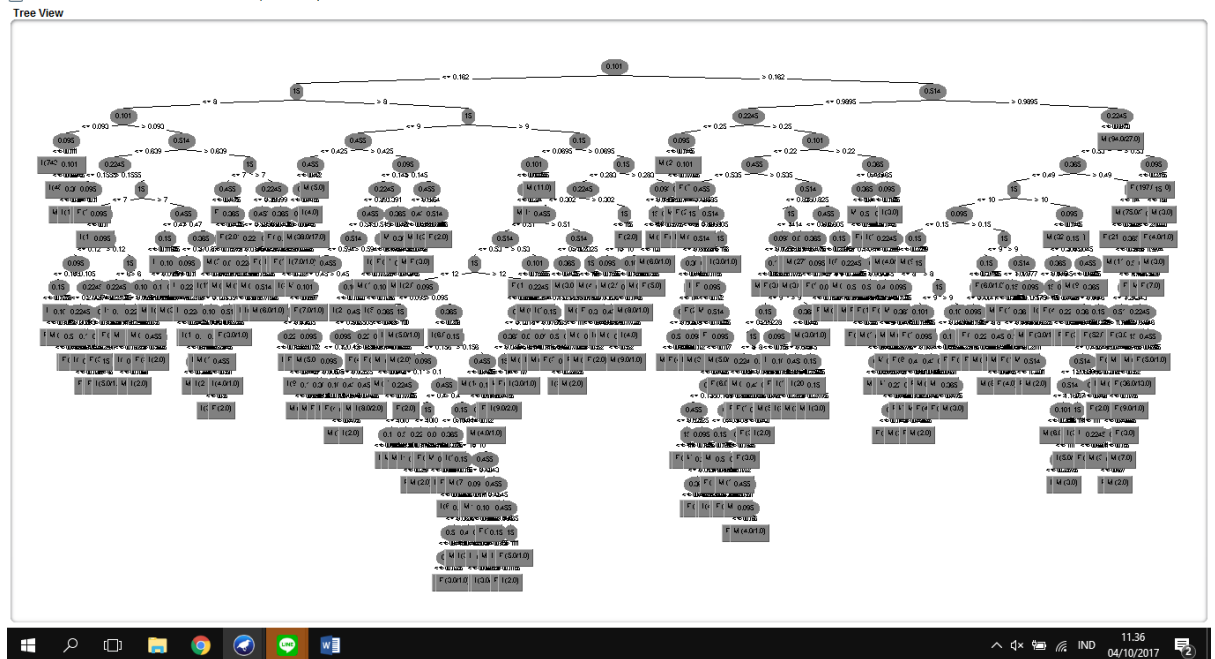
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0,475 0,313 0,488 0,475 0,481 0,163 0,589 0,452 M
0,405 0,238 0,433 0,405 0,418 0,170 0,631 0,397 F
0,730 0,162 0,664 0,730 0,695 0,554 0,794 0,604 I
Weighted Avg. 0,531 0,244 0,525 0,531 0,527 0,284 0,665 0,482

=== Confusion Matrix ===

a b c <-- classified as
260 179 108 | a = M
210 178 52 | b = F
63 54 316 | c = I
```

Decision Tree yang didapatkan menjadi

Weka Classifier Tree Visualizer: 11:34:20 - trees.J48 (abalone.arff)



3. What do the results tell us?

Dari 3 metode, terdapat perbedaan akurasi dari masing-masing metode.

Hasil akurasi adalah:

- a. Cross Validation = 52.445%
- b. Use Training Set = 74.8803 %
- c. Percentage Split = 53.0986 %

Pada Cross Validation, tingkat akurasi paling kecil. Hal ini mungkin pada preprocessing atau pada saat memilih parameter harus lebih benar lagi untuk di klasifikasikan.

Class yang menjadi samples Missclassified:

a. Cross Validation:

```
=== Confusion Matrix ===  
  
   a    b    c  <-- classified as  
774 484 269 |    a = M  
674 468 165 |    b = F  
226 168 948 |    c = I
```

b. Use Training Set:

```
=== Confusion Matrix ===  
  
   a    b    c  <-- classified as  
1111 296 120 |    a = M  
389  852  66 |    b = F  
117   61 1164 |    c = I
```

c. Pada Percentage Split:

```
=== Confusion Matrix ===  
  
   a    b    c  <-- classified as  
260 179 108 |    a = M  
210 178  52 |    b = F  
 63  54 316 |    c = I
```