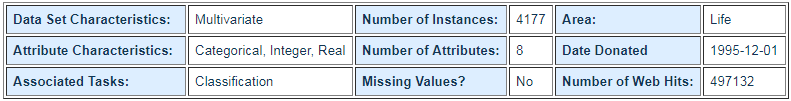
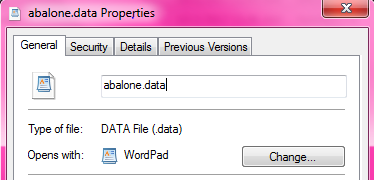
Download dataset abalone dari <https://archive.ics.uci.edu/ml/datasets/abalone>.



Dataset abalone berisi 8 atribut, yaitu: sex, length, diameter, heigh, whole weight, shucked weight, viscera weight, shell weight, dan rings.

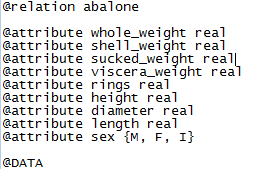
Download abalone.data, lalu ubah extensinya menjadi abalone.arff

1. Ubah extensi abalone.data menjadi abalone.arff



Ubah menjadi abalone.arff

1. Tambahkan syntax:

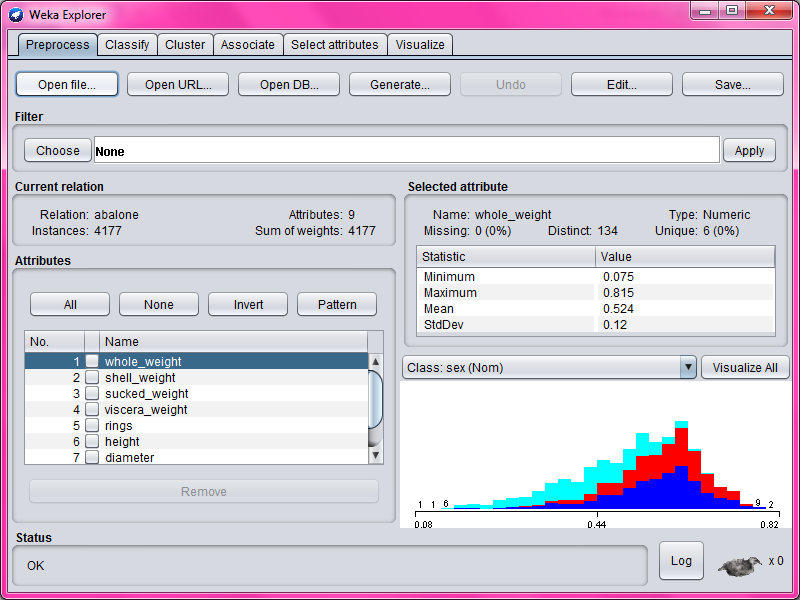


Kemudian save.

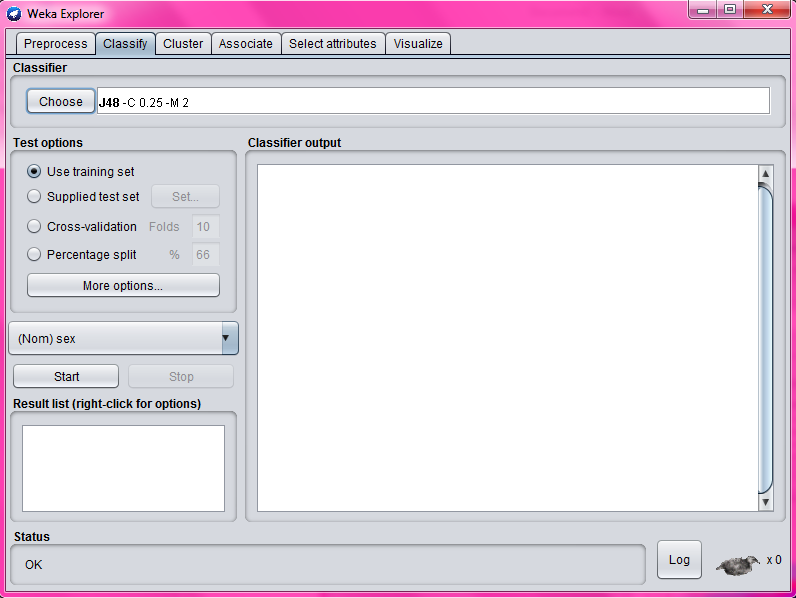
1. Buka aplikasi Weka, pilih explorer



1. Pada Preprocessor klik Open file, lalu pilih abalone.arff



1. Pada classify:



Pada classify:

1. Pilih algoritma yang akan digunakan pada kolom classifier
2. Klik tombol choose
3. Klik tree untuk decision tree, lalu pilih J48(C4.5)
4. Selanjutnya pilihlah scenario pengukuran akurasi pada test option.

Terdapat 4 pilihan pada rest option:

• Use training set : data uji dengan data latih sama  
• Supplied test set : data uji dengan data latih benar-benar berbeda  
• Cross-Validation : membagi data menjadi k-subset. Misalnya Folds yang digunakan 10, 9 akan  digunakan sebagai data training dan 1 sebagai data uji sampai seluruh data.  
• Percentage Split : Membagi data sesuai dengan parameter yang akan menjadi data training.

1. Menggunakan use training set

Number of Leaves : 360

Size of the tree : 719

Time taken to build model: 5.41 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.04 seconds

=== Summary ===

Correctly Classified Instances 3131 74.9581 %

Incorrectly Classified Instances 1046 25.0419 %

Kappa statistic 0.6228

Mean absolute error 0.2232

Root mean squared error 0.3341

Relative absolute error 50.3422 %

Root relative squared error 70.9523 %

Total Number of Instances 4177

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.726 0.190 0.689 0.726 0.707 0.531 0.868 0.795 M

0.653 0.124 0.706 0.653 0.679 0.541 0.884 0.772 F

0.870 0.066 0.861 0.870 0.865 0.801 0.959 0.890 I

Weighted Avg. 0.750 0.129 0.749 0.750 0.749 0.621 0.902 0.818

=== Confusion Matrix ===

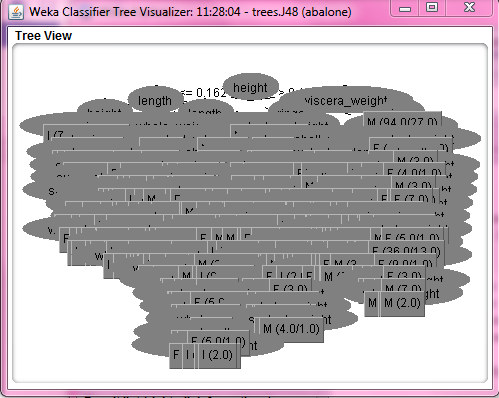
a b c <-- classified as

1110 295 123 | a = M

388 854 65 | b = F

114 61 1167 | c = I

Hasil visualize tree:



1. Menggunakan cross-validation

Number of Leaves : 360

Size of the tree : 719

Time taken to build model: 7.62 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 2206 52.813 %

Incorrectly Classified Instances 1971 47.187 %

Kappa statistic 0.2886

Mean absolute error 0.3405

Root mean squared error 0.4806

Relative absolute error 76.7932 %

Root relative squared error 102.0728 %

Total Number of Instances 4177

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.506 0.336 0.465 0.506 0.485 0.168 0.602 0.437 M

0.376 0.223 0.434 0.376 0.403 0.159 0.626 0.396 F

0.702 0.156 0.681 0.702 0.691 0.542 0.797 0.606 I

Weighted Avg. 0.528 0.243 0.525 0.528 0.525 0.285 0.672 0.479

=== Confusion Matrix ===

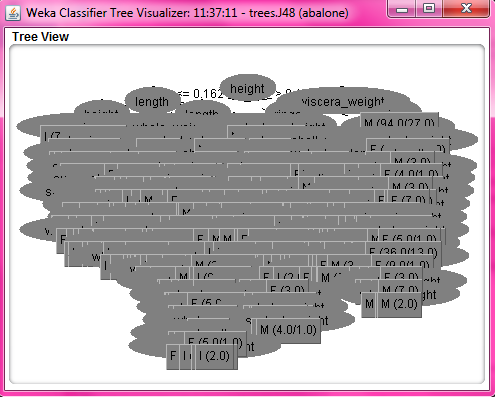
a b c <-- classified as

773 489 266 | a = M

641 491 175 | b = F

248 152 942 | c = I

Hasil visualize tree:



1. Menggunakan percentage split 66%

Number of Leaves : 360

Size of the tree : 719

Time taken to build model: 0.55 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances 749 52.7465 %

Incorrectly Classified Instances 671 47.2535 %

Kappa statistic 0.2869

Mean absolute error 0.3462

Root mean squared error 0.4694

Relative absolute error 78.073 %

Root relative squared error 99.6823 %

Total Number of Instances 1420

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.524 0.360 0.456 0.524 0.488 0.161 0.595 0.433 M

0.305 0.203 0.413 0.305 0.351 0.112 0.634 0.402 F

0.757 0.155 0.692 0.757 0.723 0.588 0.825 0.673 I

Weighted Avg. 0.527 0.245 0.517 0.527 0.518 0.280 0.680 0.499

=== Confusion Matrix ===

a b c <-- classified as

272 151 96 | a = M

260 138 55 | b = F

64 45 339 | c = I

Hasil visualize tree:

