ANDREAN JANUAR PRIATMOJO

5115100029

KK F

Objective :

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

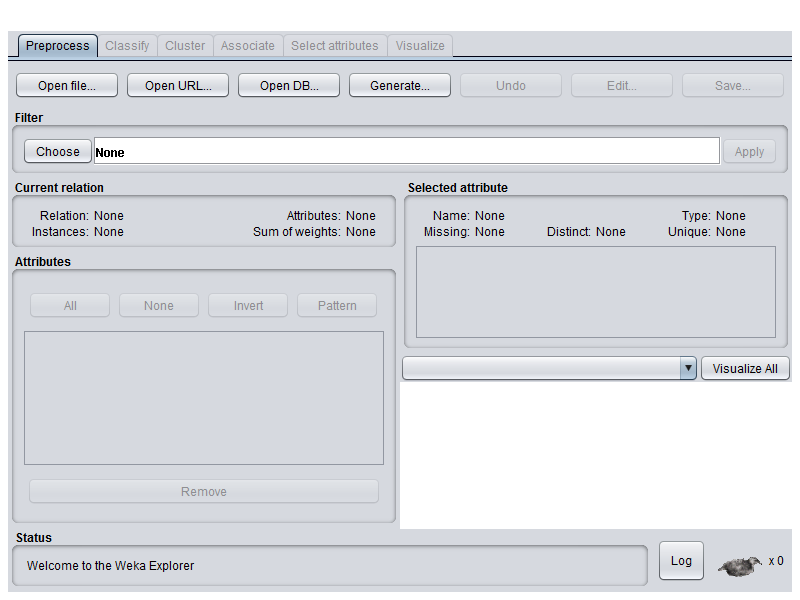
Missing Attribute Values: None

Langkah :

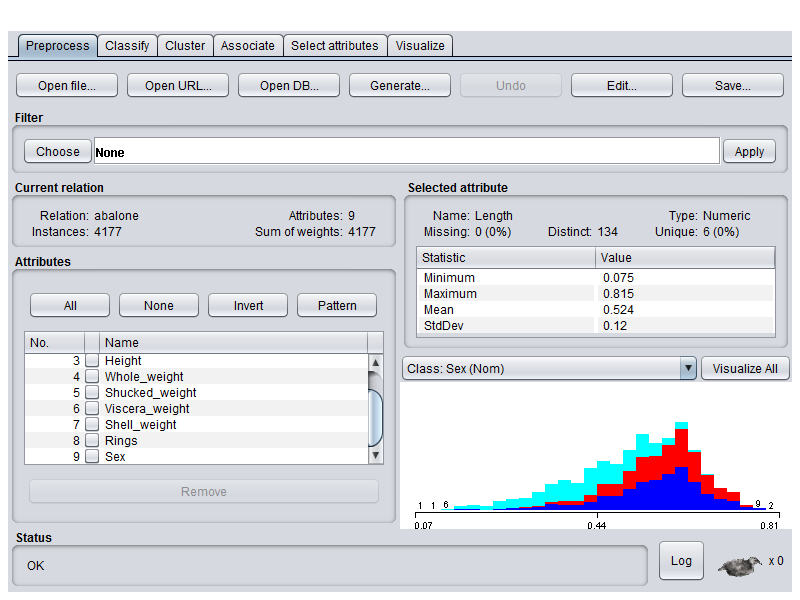
1. Buka weka



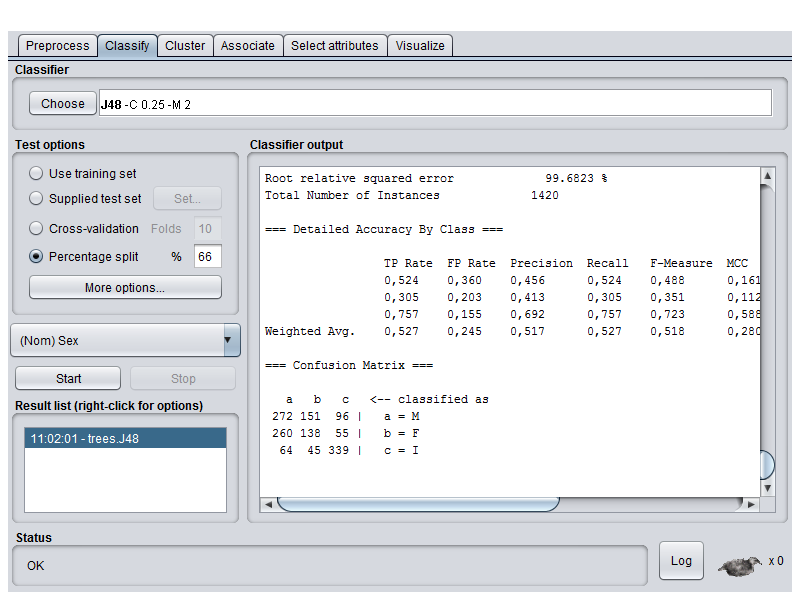
1. Buka explore



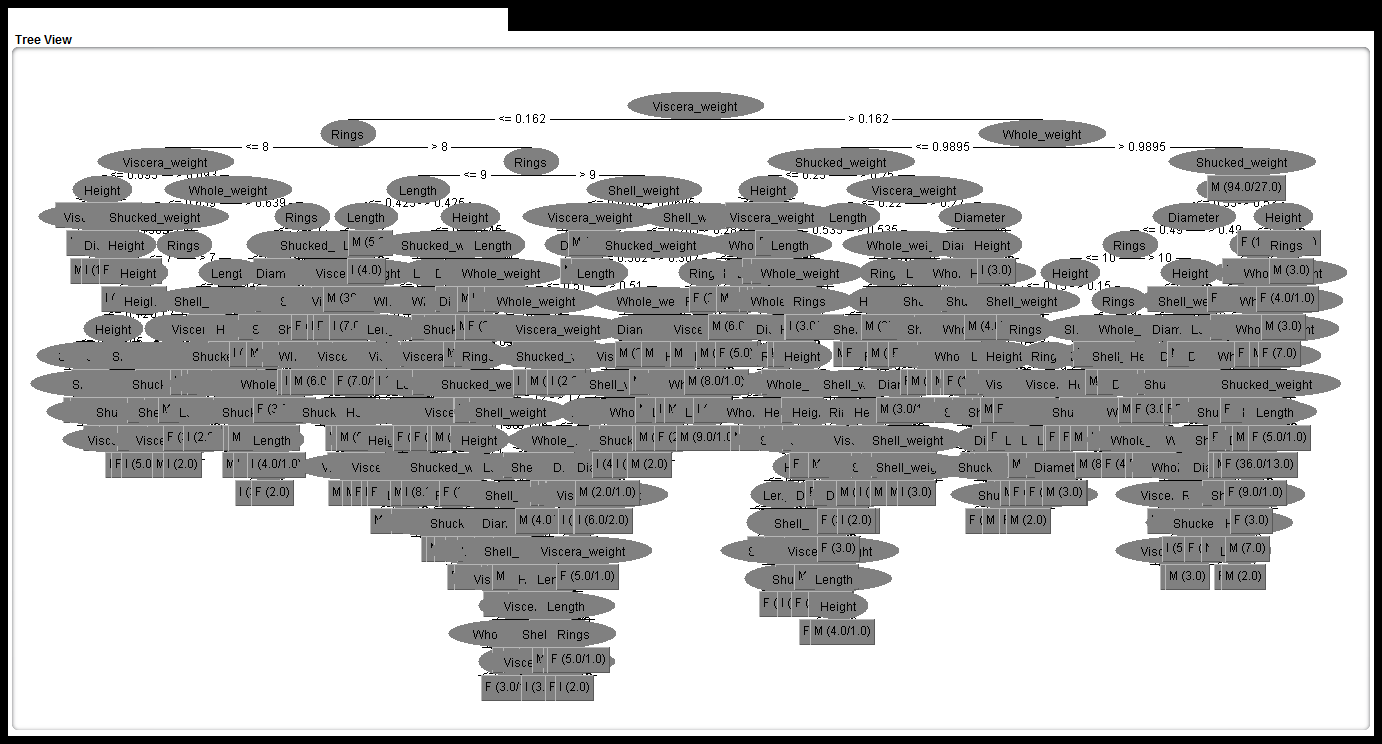
1. Open arff file



1. Pilih Classify >> J48 >> test option >> precentage split 66% >> start



1. Generate tree



Number of Leaves : 360

Size of the tree : 719

=== 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 |

Write up the following to turn in :

Using breast cancer dataset

Description

Number of Instances: 286

Number of Attributes: 9 + the class attribute

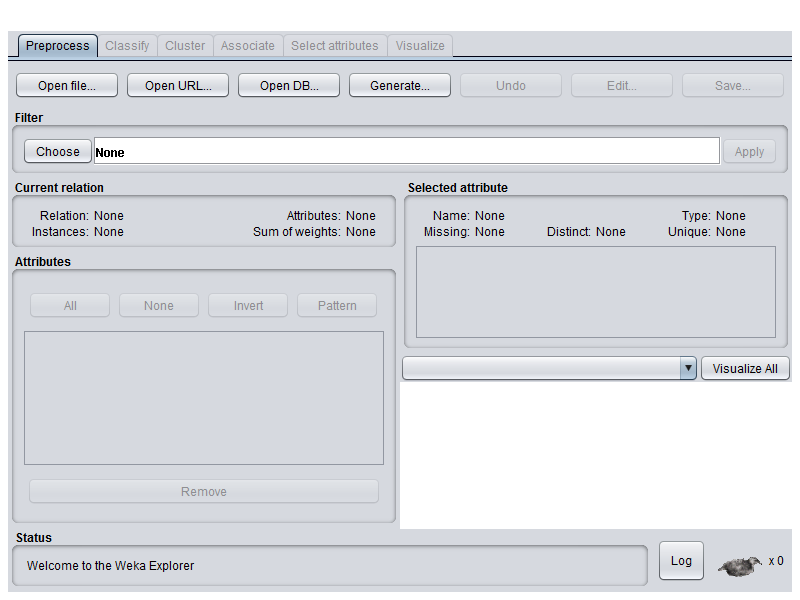
|  |
| --- |
| Attribute Information: |
|  | 1. Class: no-recurrence-events, recurrence-events | |
|  | 2. age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99. | |
|  | 3. menopause: lt40, ge40, premeno. | |
|  | 4. tumor-size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, | |
|  | 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% |

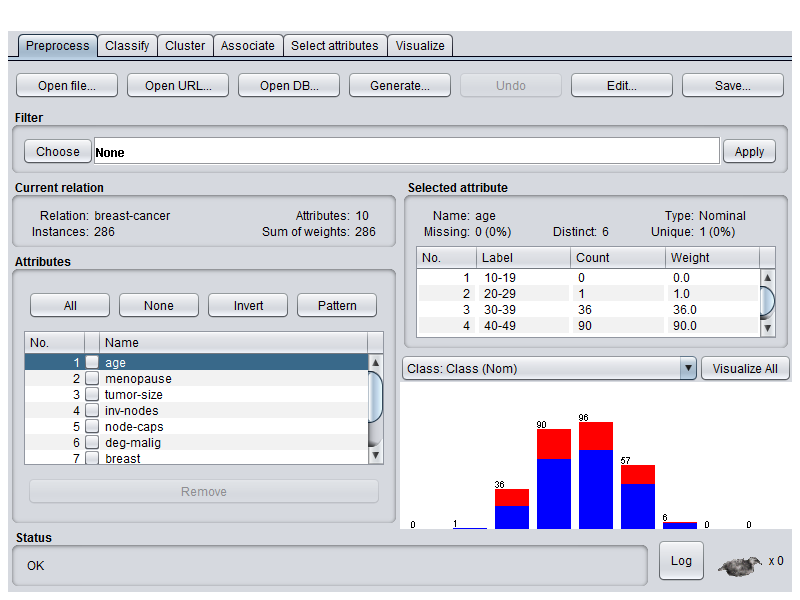
1. Buka weka



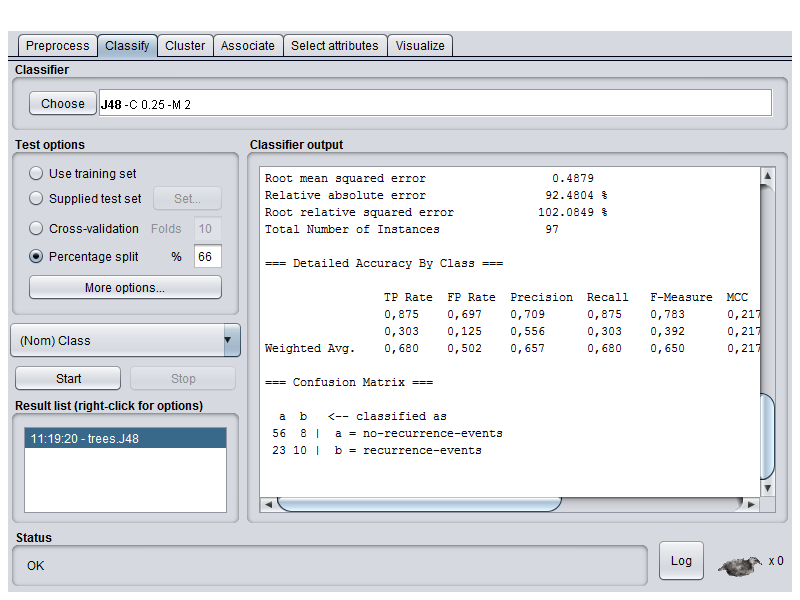
1. Buka explore



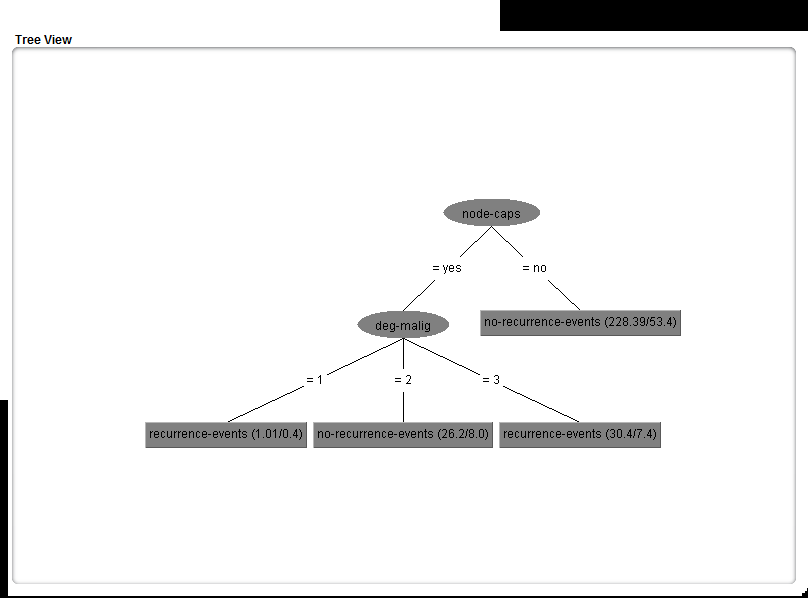
1. Open arff file



1. Classify >> choose >> test option precentage split >> start



1. Generate tree



=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: breast-cancer

Instances: 286

Attributes: 10

age

menopause

tumor-size

inv-nodes

node-caps

deg-malig

breast

breast-quad

irradiat

Class

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

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

node-caps = yes

| deg-malig = 1: recurrence-events (1.01/0.4)

| deg-malig = 2: no-recurrence-events (26.2/8.0)

| deg-malig = 3: recurrence-events (30.4/7.4)

node-caps = no: no-recurrence-events (228.39/53.4)

Number of Leaves : 4

Size of the tree : 6

=== Summary ===

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

=== Detailed Accuracy By Class ===

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|  | 0,875 | 0,697 | 0,709 | 0,875 | 0,783 | 0,217 | 0,603 | 0,717 | no-recurrence-events |
|  | 0,303 | 0,125 | 0,556 | 0,303 | 0,392 | 0,217 | 0,603 | 0,420 | recurrence-events |
| Weighted Avg. | 0,680 | 0,502 | 0,657 | 0,680 | 0,650 | 0,217 | 0,603 | 0,616 |  |

=== Confusion Matrix ===

|  |  |  |
| --- | --- | --- |
| a | B | <-- classified as |
| 56 | 8 | a = no-recurrence-events |
| 23 | 10 | b = recurrence-events |

3. Hasilnya menunjuk kan bahwa presentase ketepatan nya 68.0412 % . Karena dataset breast cancer memiliki missing value sebesar sebanyak 9 / 0.3%.