Predication the performance by using the ID3 and C4.5 Algorithms

# SURESH KUMAR

**MS(Software Engineering)**

**ROLL NO : 60247**

# RAPIDMINER :

Rapid Miner is an open source data mining tool that

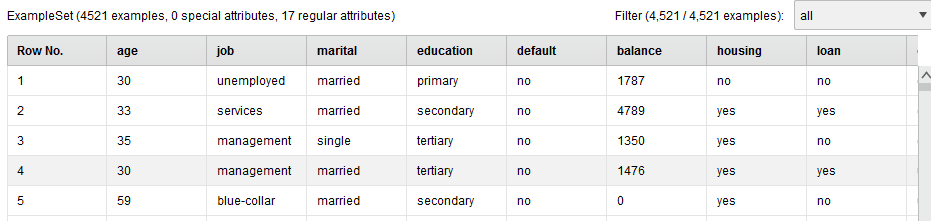
Provides data mining and machine learning procedures including data loading and transformation data preprocessing and visualization, modeling, evaluation, I have used Rapid Miner to generate decision trees of ID3 and C4.5 algorithms.

Training and Testing Data WITH (C4.5)

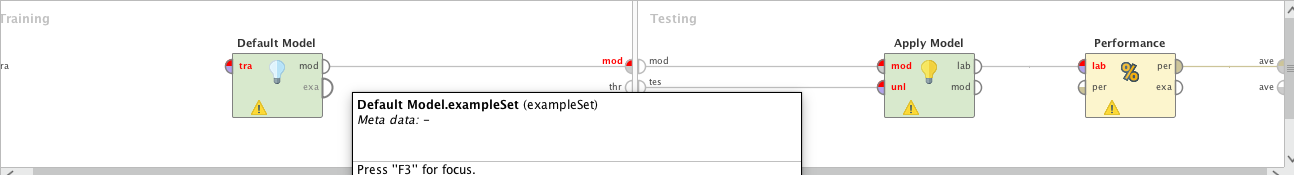
The operator I used for training and testing data in default modeling and itis the nested process . when we click on blue icon it further divide into process.

DATA SAMPLE:

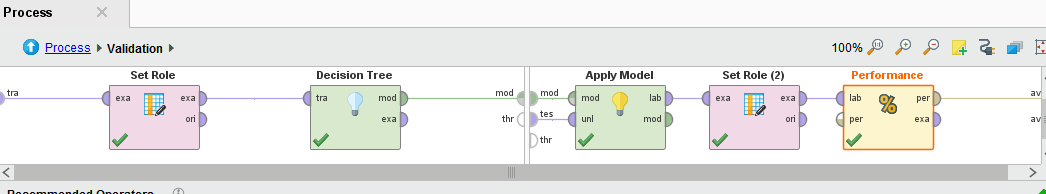
Data set contain banking information



While click on blue icon we further get another window

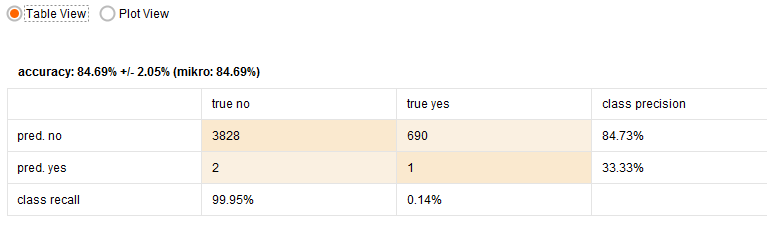


We can use any model, Use default model on training side and apply model on testing side. and other thing performance operator that evaluate the performance .



After Processing the performance model I got following results .

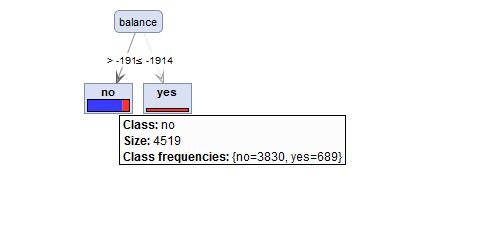
**ACCURACY**

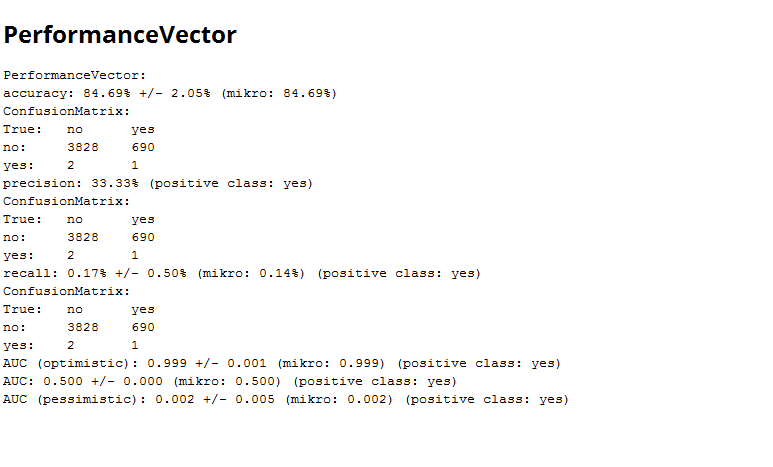


**Recall Progression**



**DECISSION TREE**



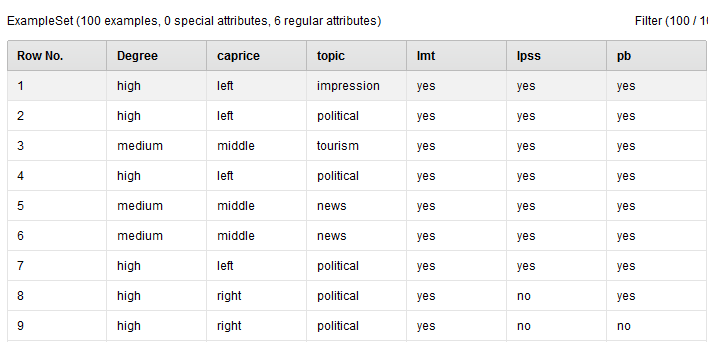
****

Testing and Training with ID3

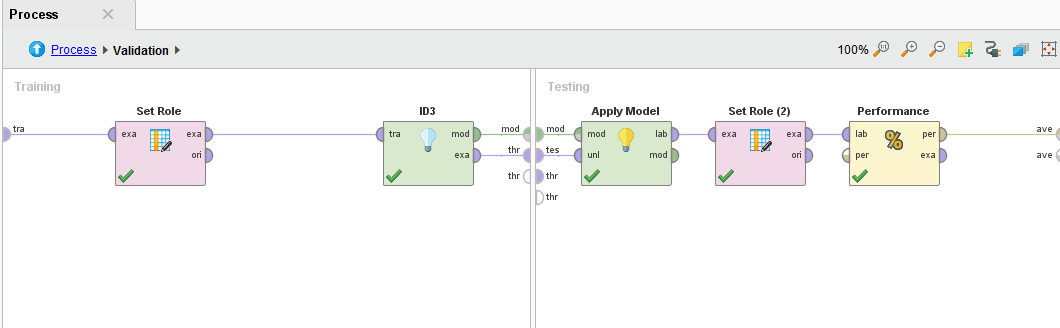
The operator I used for training and testing data in default modeling  **and its** is nested process . when we click on blue icon it further divide into process.

DATA SAMPLE:

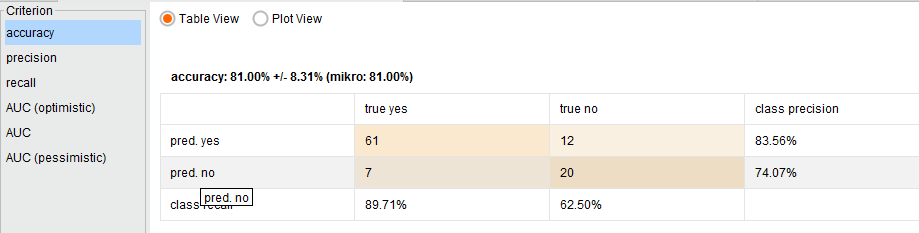
Data set contain following data



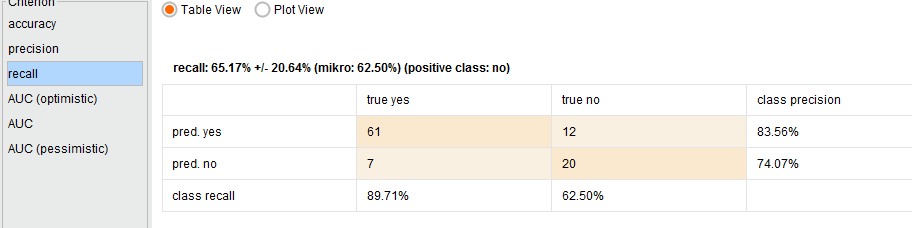
**PROCESS**

****

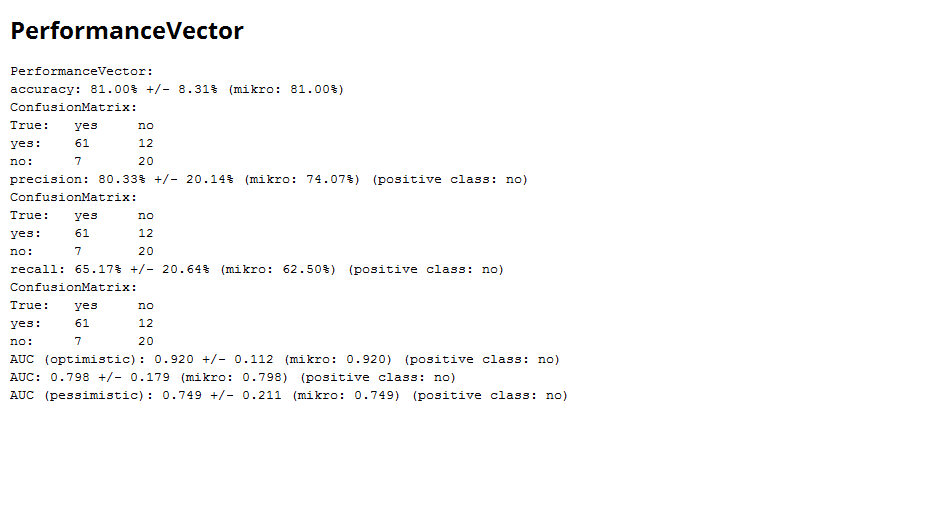
**RESULT**

****

**RECALL**

****

**RESULT**

****

# WEKA

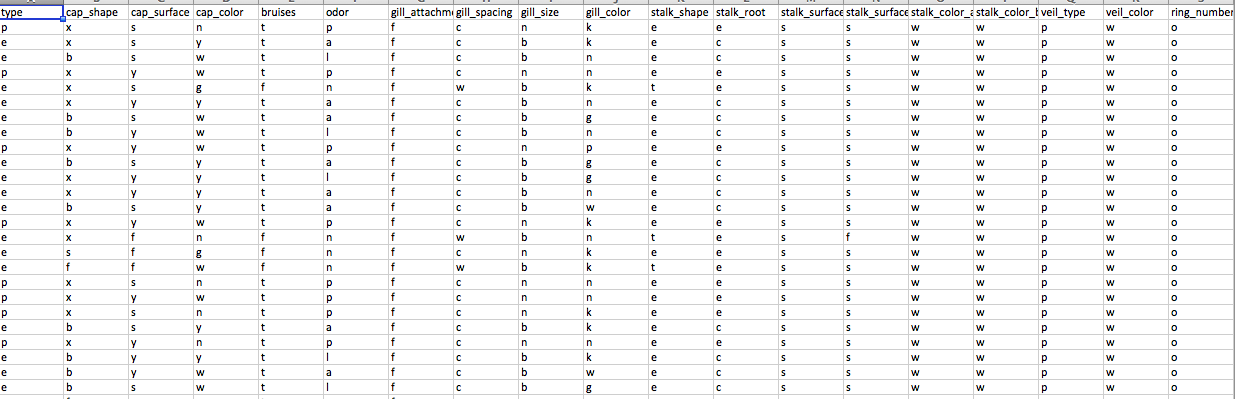
**Waikato Environment for Knowledge Analysis** (**Weka**) is a popular suite of machine learning software written in Java.

DATA SET NO1 :

**DATASET**

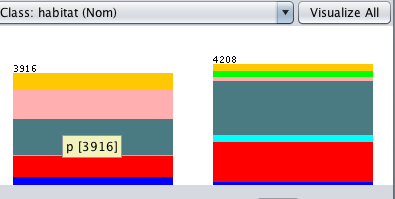
My dataset contain ingredients of different type of mushroom.

ROWS ==8000



## WITH ALGORITHM J48

**VISULIZE DATA**

****

**RESULT**

Time taken to build model: 0.07 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.15 seconds

=== Summary ===

Correctly Classified Instances 8124 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0

Root mean squared error 0

Relative absolute error 0 %

Root relative squared error 0 %

Total Number of Instances 8124

=== Detailed Accuracy By Class ===

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

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 p

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 e

Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000

**DESCISSION TREE**

J48 pruned tree

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

Number of Leaves : 24

Size of the tree : 29

Time taken to build model: 0.07 seconds

## ALGORITHM RANDOM TREE:

I use same data set with random tree results are given below.

**RESULT**

=== Summary ===

Correctly Classified Instances 8124 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0

Root mean squared error 0

Relative absolute error 0 %

Root relative squared error 0 %

Total Number of Instances 8124

=== Detailed Accuracy By Class ===

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

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 p

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 e

Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000

**DESCISION TREE**

RandomTree

==========

Size of the tree : 40

## Algorithm LMT:

I used the same data with LMT

**RESULT**

=== Summary ===

Correctly Classified Instances 8124 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0002

Root mean squared error 0.0015

Relative absolute error 0.0331 %

Root relative squared error 0.3048 %

Total Number of Instances 8124

=== Detailed Accuracy By Class ===

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

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 p

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 e

Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000

=== Confusion Matrix ===

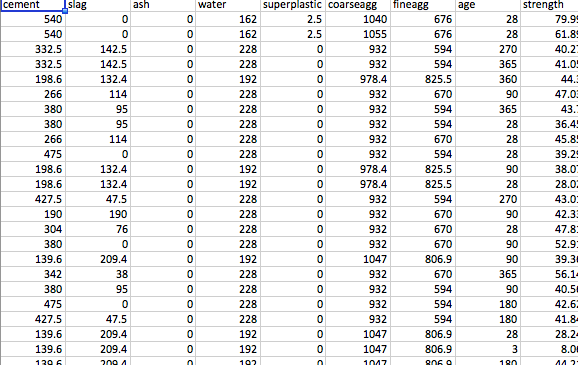
a b <-- classified as

3916 0 | a = p

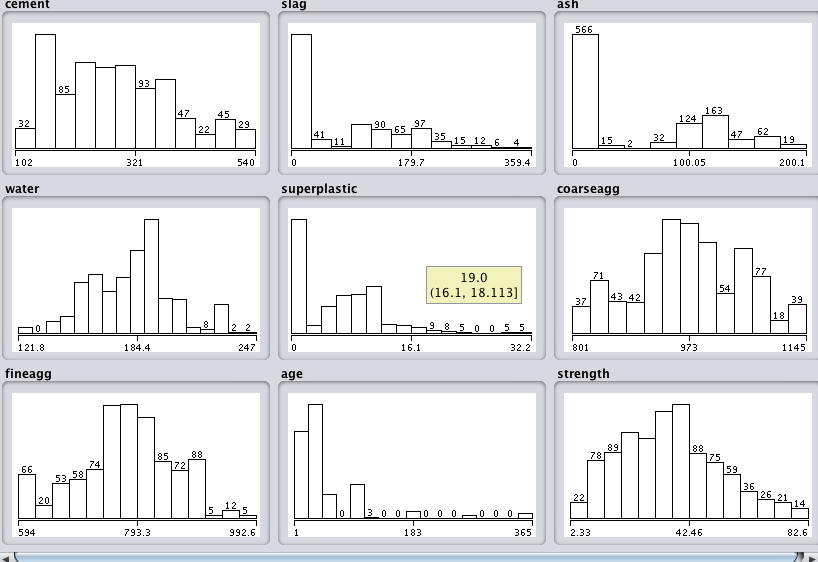
0 4208 | b = e

DATAS SET NO: 2

My second dataset is used to check the strength of concrete. Number of Records 1032



**DATA visualize:**



## ALGOIRTHM DESCISISON STUMP

**RESULT:**

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correlation coefficient 0.4981

Mean absolute error 11.5493

Root mean squared error 14.479

Relative absolute error 85.8007 %

Root relative squared error 86.7132 %

Total Number of Instances 1030

Decision Stump

Classifications

age <= 21.0 : 23.541234567901242

age > 21.0 : 41.452039660056606

age is missing : 35.817961165048494

## Random Tree Algorithm

**RESLUT:**

Time taken to build model: 0.12 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correlation coefficient 0.998

Mean absolute error 0.1992

Root mean squared error 1.0661

Relative absolute error 1.4799 %

Root relative squared error 6.3849 %

Total Number of Instances 1030

## REP TREE Algorithm

RESULT SET

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correlation coefficient 0.9506

Mean absolute error 3.6519

Root mean squared error 5.1816

Relative absolute error 27.1298 %

Root relative squared error 31.0321 %

Total Number of Instances 1030