AMERICAN INTERNATIONAL UNIVERSITY–BANGLADESH (AIUB)

FACULTY OF SCIENCE & TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

DATA WAREHOUSING AND DATA MINING

Summer 2021-2022

Section: C

Supervised By

Akinul Islam Jony

Submitted By

Name	ID
Either Rahman	19-39750-1
Hasan Sanjary Islam	19-39589-1
Farahn Sadik Ferdous	20-42072-1
Tapu Biswas	20-42073-1

Date of Submission: August 7, 2022

Project Overview

Data mining is a process where data are extracted from huge datasets according to one's usefulness. Data mining is basically done to gather knowledge from data. Extracted data can be used for more effective marketing strategies, increasing sales, decreasing costs, fraud detection, product design, weather forecasting, etc. Data mining generally breaks down patterns and connections in data based on what information users request. KNN, Naïve Bayes, and Decision Tree are some of the algorithms used in data mining. Naïve Bayes and KNN are classification method that only deals with categorical attributes or classified dataset. Naïve Bayes is a supervised learning algorithm that is used for solving classification problems. Naïve Bayes uses probability theory to find the most likely of the possible classifications. KNN is a classifier algorithm where the classification is based on the nearest Neighbor's data (distance between instances). A decision tree is a type of probability tree that enables you to make a decision on a process followed by some decision rule. Decision rules can be combined together to form a tree structure. Prediction accuracy defines the predicted values match the actual values of the target field. Prediction accuracy is used to estimate the performance of a classifier. Prediction accuracy determines the proportion of a set of unseen instances that it correctly classifies. The confusion matrix gives an overview of prediction accuracy thoroughly like instances of the correct class being correctly classified as the correct class or misclassified as some other class and its breakdown. In this project, we have determined the accuracy of a dataset. We have collected a dataset that is Bank marketing response prediction. We have applied three algorithms KNN, Naïve Bayes, and Decision Tree to WEKA software. To predict the accuracy and its breakdown we have used prediction accuracy and confusion matrix.

Dataset Overview

i. Data source with valid URL

Bank Marketing Response Predict

https://www.kaggle.com/datasets/kukuroo3/bank-marketing-response-predict

ii. Description about dataset

The dataset is collected from Kaggle. We have taken a classified dataset from Kaggle. The data is from a marketing campaign (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit. This dataset has 12870 instances and 17 attributes. The class attribute here is the attribute y. There is no missing data here. We have applied three algorithms KNN, Naïve Bayes, and Decision Tree on WEKA software in this dataset. To predict the accuracy and its breakdown we have used a prediction accuracy and confusion matrix.

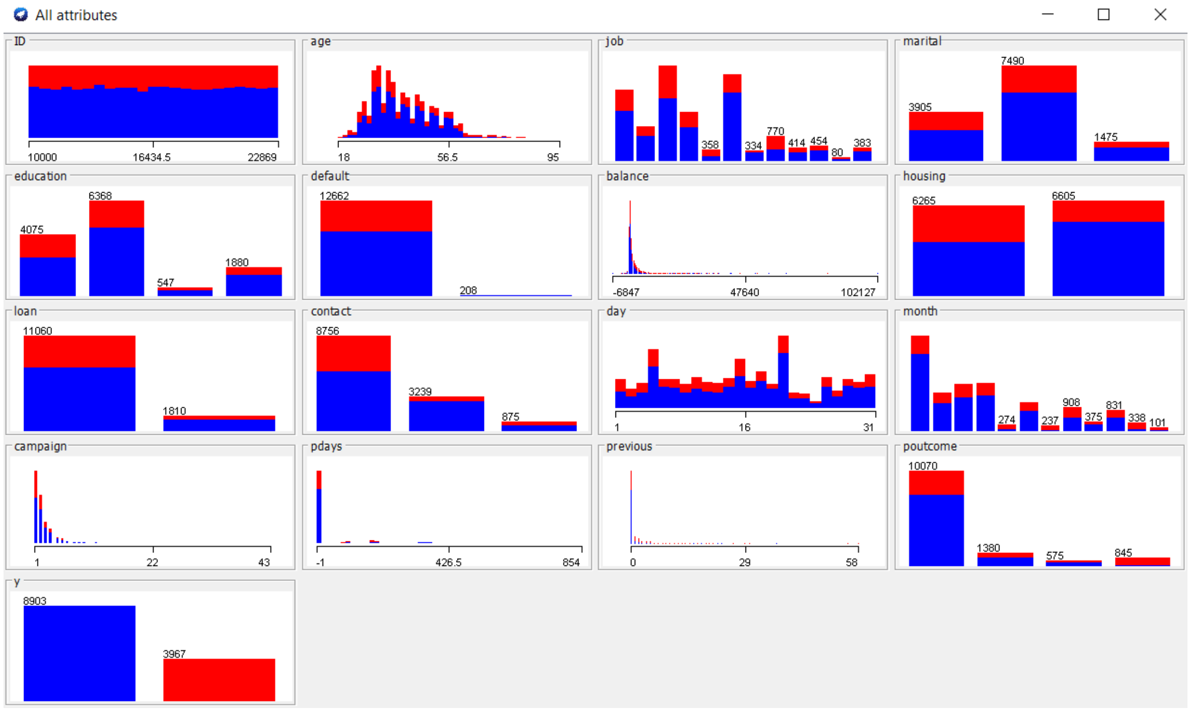


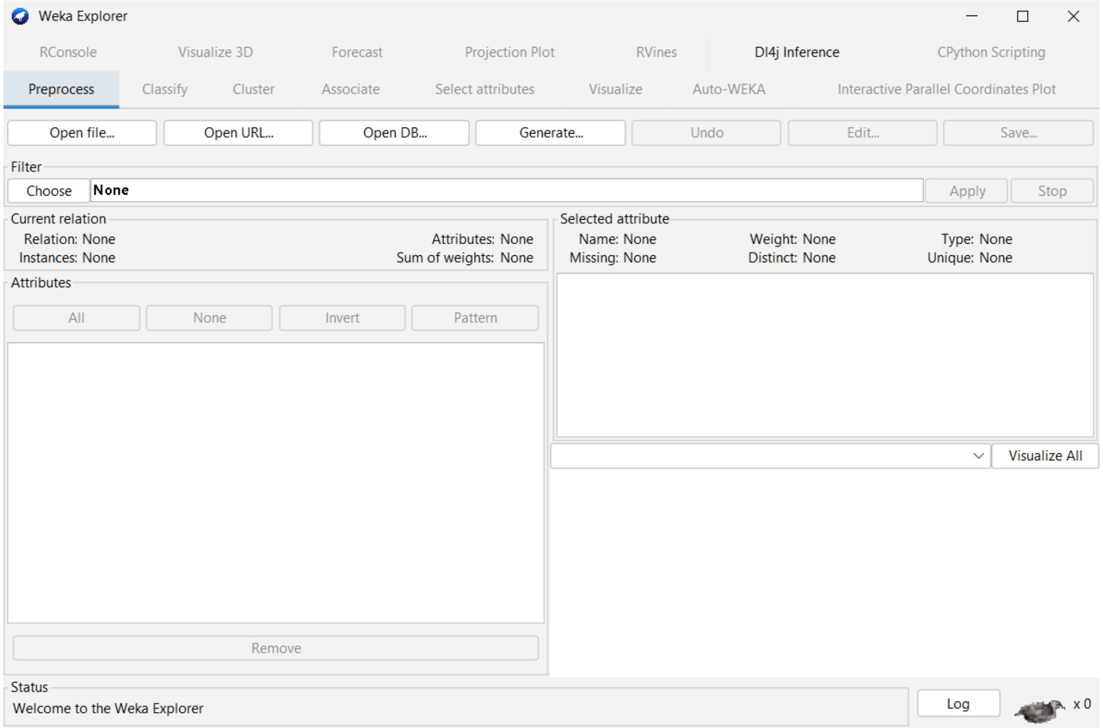
Figure: Bank Marketing Response Predict Dataset

Model Development

Classification in data mining is a common technique that separates data points into different classes. It allows you to organize data sets of all sorts, including complex and large datasets as well as small and simple ones. It primarily involves using algorithms that you can easily modify to improve the data quality. Here, we have applied three algorithms Naïve Bayes, KNN and Decision Tree on WEKA software in this dataset. To predict the accuracy and its breakdown we have used a prediction accuracy and confusion matrix.

Before inserting the dataset into the WEKA software, the .csv file was converted into .arff file. Procedure of inserting dataset into the WEKA software is given below:

1. First, Weka 3.8.6 was opened and ‘Explorer’ option was chosen and new window named Weka Explorer was opened.



Filter: None

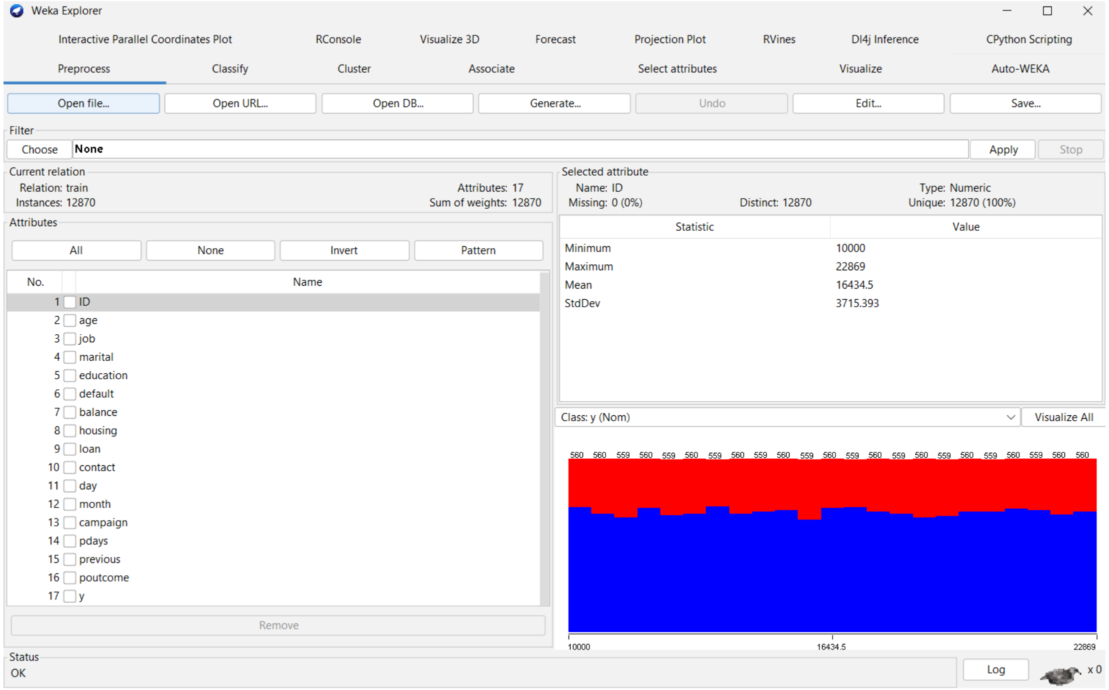
Relation: None
Instances: None
Attributes: None
Sum of weights: None

Name: None
Missing: None
Weight: None
Distinct: None
Type: None
Unique: None

Welcome to the Weka Explorer

Figure: Weka Explorer

2. Then, train dataset was selected from the device.



Filter: None

Relation: train
Instances: 12870
Attributes: 17
Sum of weights: 12870

Name: ID
Missing: 0 (0%)
Distinct: 12870
Type: Numeric
Unique: 12870 (100%)

Statistic	Value
Minimum	10000
Maximum	22869
Mean	16434.5
StdDev	3715.393

No.		Name
1	□	ID
2	□	age
3	□	job
4	□	marital
5	□	education
6	□	default
7	□	balance
8	□	housing
9	□	loan
10	□	contact
11	□	day
12	□	month
13	□	campaign
14	□	pdays
15	□	previous
16	□	poutcome
17	□	y

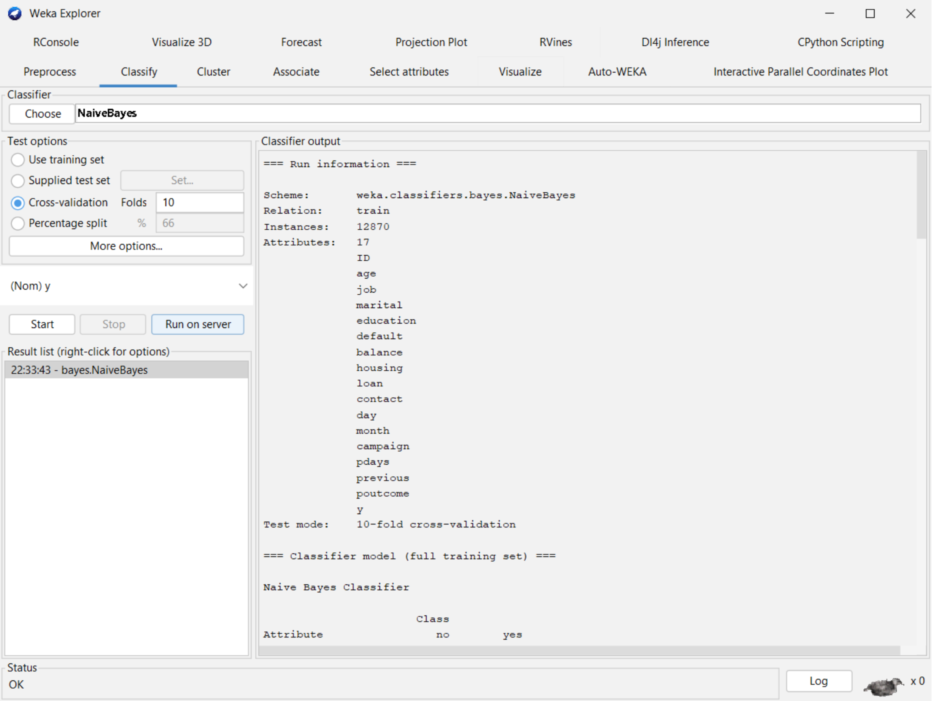
Class: y (Nom)

OK

Figure: Train dataset

Naïve Bayes classifier:

1. Firstly, classify was selected, ‘Choose’ option was pressed then NaiveBayes was chosen, and at last start button was pressed.



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10

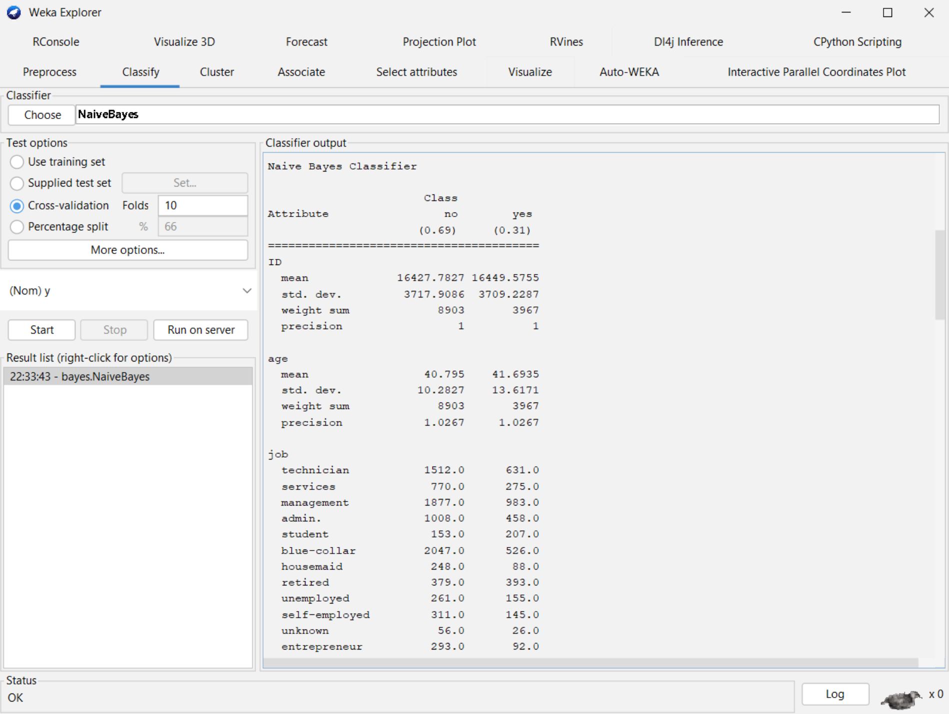
(Nom) y

Result list (right-click for options): 22:33:43 - bayes.NaiveBayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     train
Instances:    12870
Attributes:   17
              ID
              age
              job
              marital
              education
              default
              balance
              housing
              loan
              contact
              day
              month
              campaign
              pdays
              previous
              poutcome
              y
Test mode:    10-fold cross-validation

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

Naive Bayes Classifier

                       Class
Attribute                 no        yes
```

Status: OK

Figure: Naïve Bayes Algorithm



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10

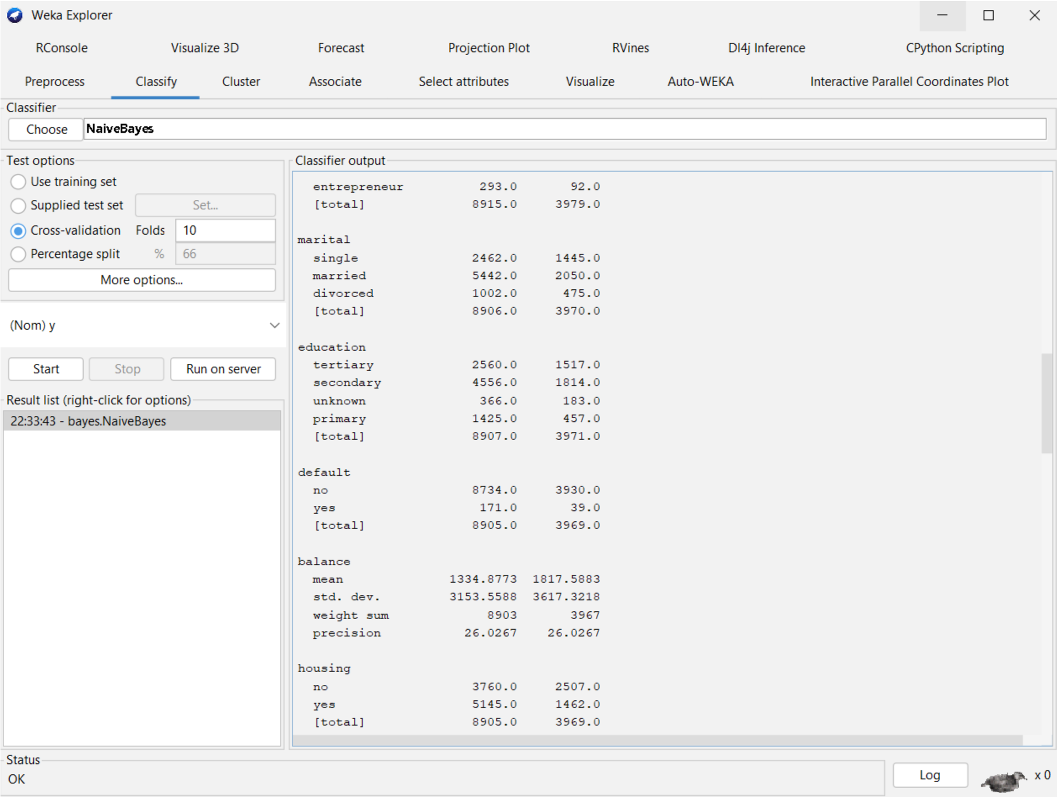
(Nom) y

Result list (right-click for options): 22:33:43 - bayes.NaiveBayes

```
Naive Bayes Classifier

                       Class
Attribute                 no        yes
                      (0.69)     (0.31)
=========================================
ID
  mean            16427.7827 16449.5755
  std. dev.        3717.9086  3709.2287
  weight sum            8903       3967
  precision                1          1

age
  mean                40.795    41.6935
  std. dev.          10.2827    13.6171
  weight sum            8903       3967
  precision           1.0267     1.0267

job
  technician          1512.0      631.0
  services             770.0      275.0
  management          1877.0      983.0
  admin.              1008.0      458.0
  student              153.0      207.0
  blue-collar         2047.0      526.0
  housemaid            248.0       88.0
  retired              379.0      393.0
  unemployed           261.0      155.0
  self-employed        311.0      145.0
  unknown               56.0       26.0
  entrepreneur         293.0       92.0
```

Status: OK

Figure: Naïve Bayes Algorithm



Classifier: NaiveBayes

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

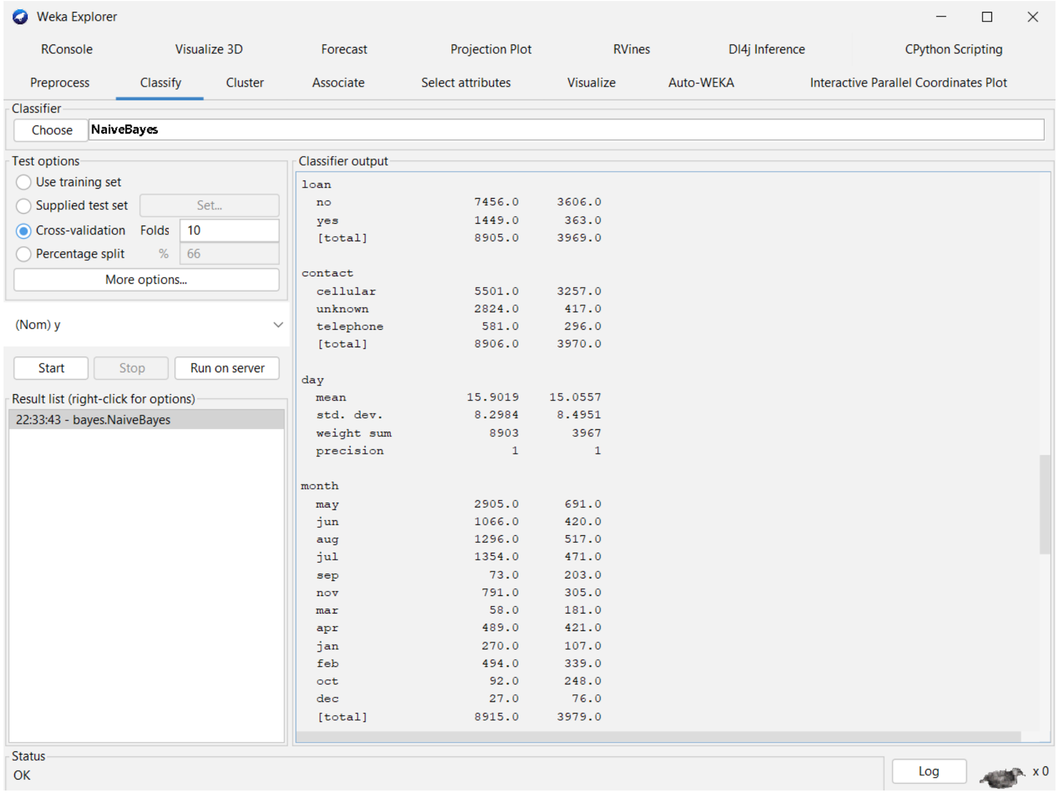
Result list (right-click for options): 22:33:43 - bayes.NaiveBayes

Status: OK

Classifier output

```
  entrepreneur        293.0      92.0
  [total]            8915.0    3979.0

marital
  single             2462.0    1445.0
  married            5442.0    2050.0
  divorced           1002.0     475.0
  [total]            8906.0    3970.0

education
  tertiary           2560.0    1517.0
  secondary          4556.0    1814.0
  unknown             366.0     183.0
  primary            1425.0     457.0
  [total]            8907.0    3971.0

default
  no                 8734.0    3930.0
  yes                 171.0      39.0
  [total]            8905.0    3969.0

balance
  mean            1334.8773 1817.5883
  std. dev.       3153.5588 3617.3218
  weight sum           8903      3967
  precision         26.0267   26.0267

housing
  no                 3760.0    2507.0
  yes                5145.0    1462.0
  [total]            8905.0    3969.0
```

Figure: Naïve Bayes Algorithm



Classifier: NaiveBayes

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

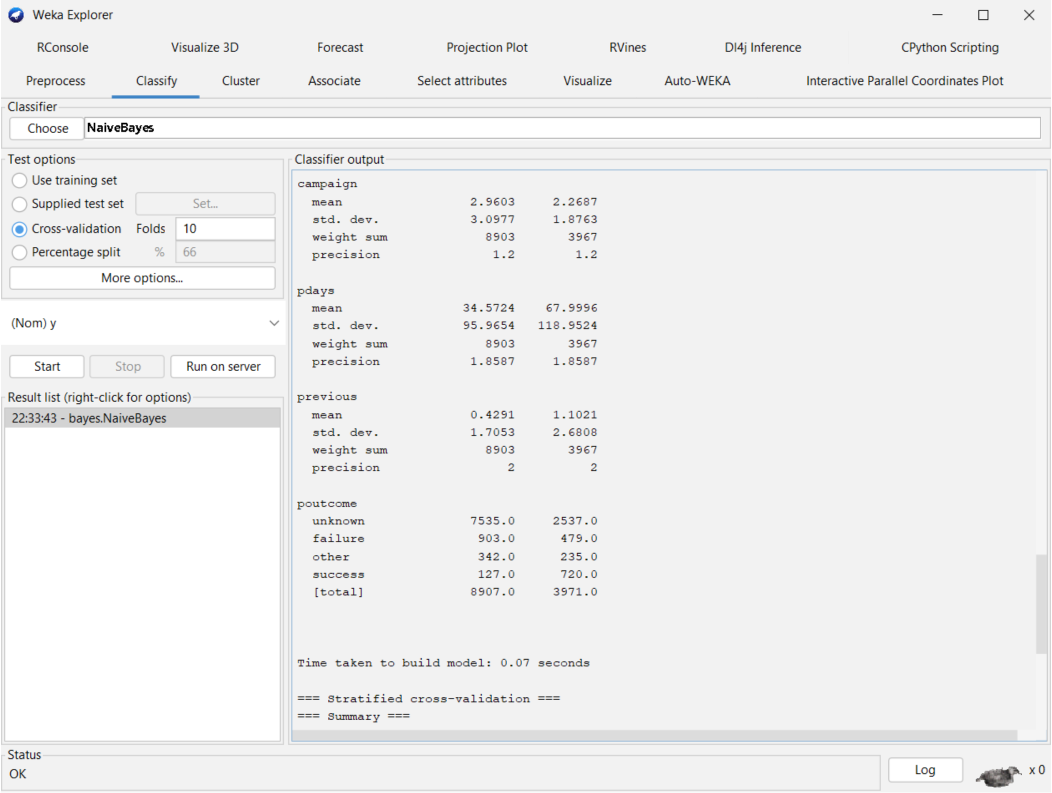
Result list (right-click for options): 22:33:43 - bayes.NaiveBayes

Status: OK

Classifier output

```
loan
  no                 7456.0    3606.0
  yes                1449.0     363.0
  [total]            8905.0    3969.0

contact
  cellular           5501.0    3257.0
  unknown            2824.0     417.0
  telephone           581.0     296.0
  [total]            8906.0    3970.0

day
  mean              15.9019   15.0557
  std. dev.          8.2984    8.4951
  weight sum           8903      3967
  precision               1         1

month
  may                2905.0     691.0
  jun                1066.0     420.0
  aug                1296.0     517.0
  jul                1354.0     471.0
  sep                  73.0     203.0
  nov                 791.0     305.0
  mar                  58.0     181.0
  apr                 489.0     421.0
  jan                 270.0     107.0
  feb                 494.0     339.0
  oct                  92.0     248.0
  dec                  27.0      76.0
  [total]            8915.0    3979.0
```

Figure: Naïve Bayes Algorithm



Classifier: NaiveBayes

● Cross-validation Folds 10

(Nom) y

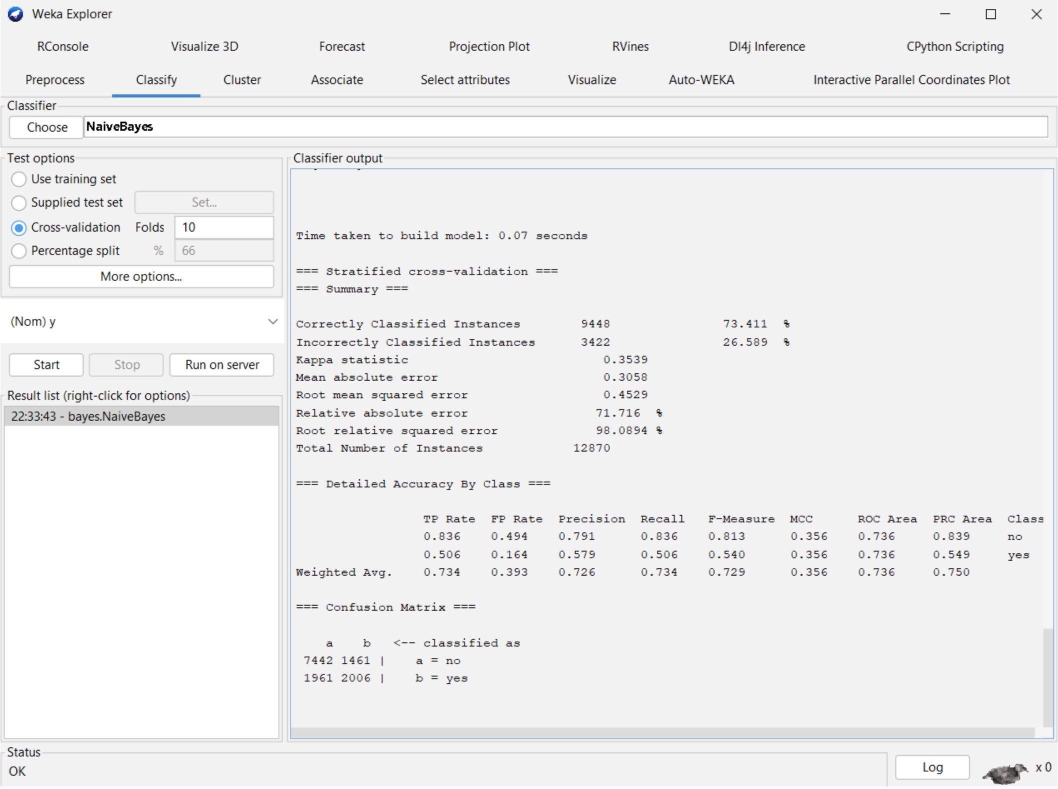
Result list (right-click for options)
22:33:43 - bayes.NaiveBayes

Classifier output

```
campaign
  mean                 2.9603     2.2687
  std. dev.            3.0977     1.8763
  weight sum             8903       3967
  precision               1.2        1.2

pdays
  mean                34.5724    67.9996
  std. dev.           95.9654   118.9524
  weight sum             8903       3967
  precision            1.8587     1.8587

previous
  mean                 0.4291     1.1021
  std. dev.            1.7053     2.6808
  weight sum             8903       3967
  precision                 2          2

poutcome
  unknown              7535.0     2537.0
  failure               903.0      479.0
  other                 342.0      235.0
  success               127.0      720.0
  [total]              8907.0     3971.0


Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===
```

Status
OK

Figure: Naïve Bayes Algorithm



Classifier: NaiveBayes

● Cross-validation Folds 10

(Nom) y

Result list (right-click for options)
22:33:43 - bayes.NaiveBayes

Classifier output

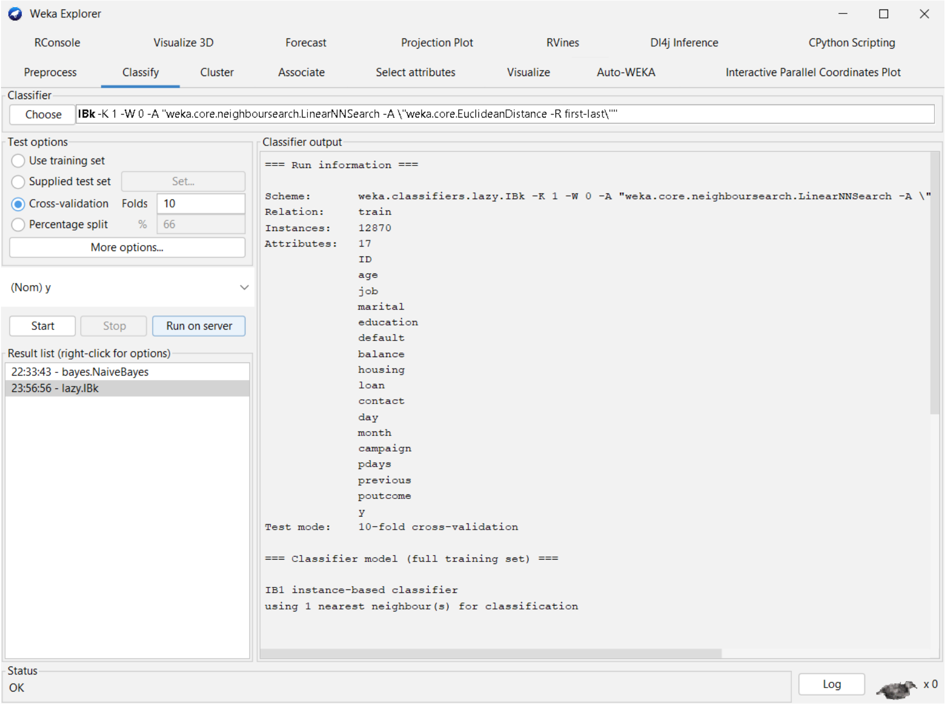
```
Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        9448               73.411  %
Incorrectly Classified Instances      3422               26.589  %
Kappa statistic                          0.3539
Mean absolute error                      0.3058
Root mean squared error                  0.4529
Relative absolute error                 71.716  %
Root relative squared error             98.0894 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.836    0.494    0.791      0.836    0.813      0.356    0.736     0.839     no
                 0.506    0.164    0.579      0.506    0.540      0.356    0.736     0.549     yes
Weighted Avg.    0.734    0.393    0.726      0.734    0.729      0.356    0.736     0.750

=== Confusion Matrix ===

    a    b   <-- classified as
 7442 1461 |    a = no
 1961 2006 |    b = yes
```

Status
OK

Figure: Naïve Bayes Algorithm

K-Nearest Neighbour Classification:

1. Firstly, classify was selected, ‘Choose’ option was pressed then IBK was chosen, and at last start button was pressed.



Classifier: IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

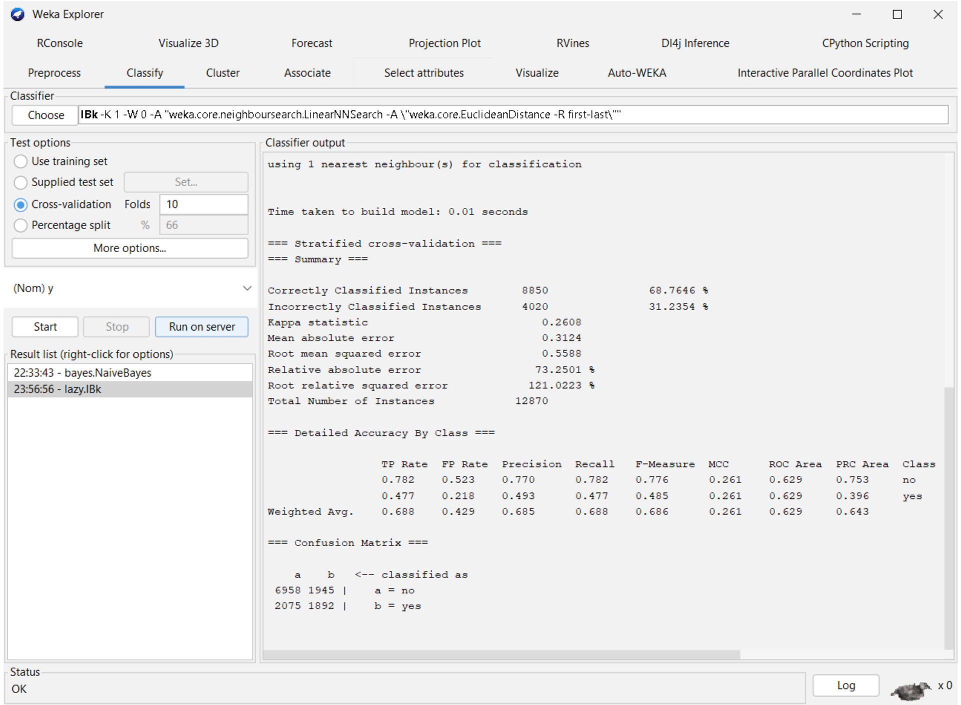
Result list (right-click for options)
22:33:43 - bayes.NaiveBayes
23:56:56 - lazy.IBk

Status: OK

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"
Relation:     train
Instances:    12870
Attributes:   17
              ID
              age
              job
              marital
              education
              default
              balance
              housing
              loan
              contact
              day
              month
              campaign
              pdays
              previous
              poutcome
              y
Test mode:    10-fold cross-validation

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

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification
```

Figure: KNN Algorithm



Classifier: IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

Result list (right-click for options)
22:33:43 - bayes.NaiveBayes
23:56:56 - lazy.IBk

Status: OK

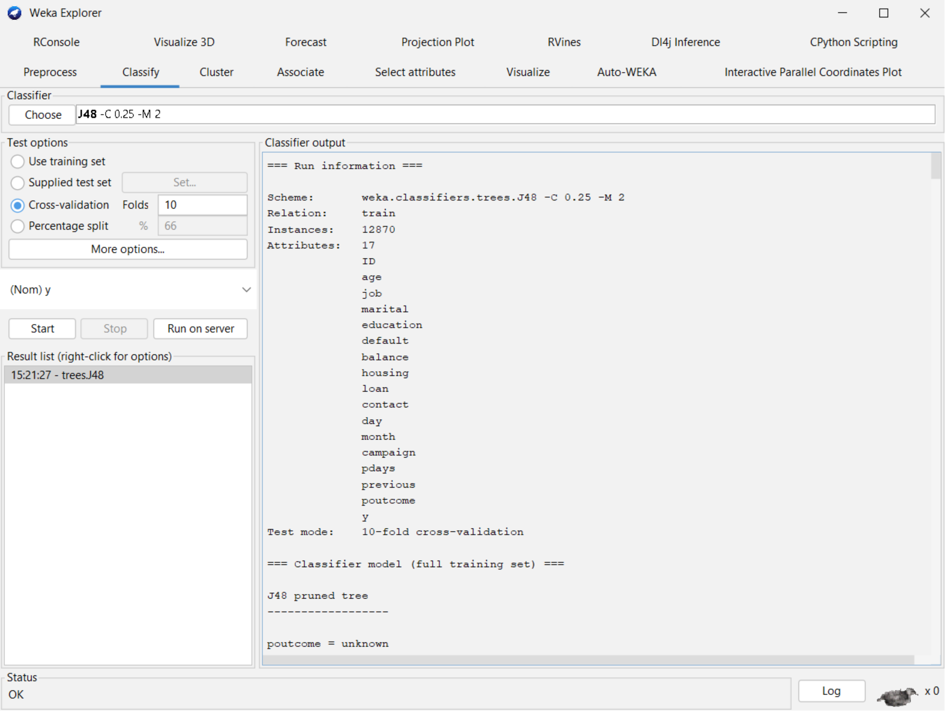
Classifier output

```
using 1 nearest neighbour(s) for classification


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        8850               68.7646 %
Incorrectly Classified Instances      4020               31.2354 %
Kappa statistic                          0.2608
Mean absolute error                      0.3124
Root mean squared error                  0.5588
Relative absolute error                 73.2501 %
Root relative squared error            121.0223 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.782    0.523    0.770      0.782    0.776      0.261    0.629     0.753     no
                 0.477    0.218    0.493      0.477    0.485      0.261    0.629     0.396     yes
Weighted Avg.    0.688    0.429    0.685      0.688    0.686      0.261    0.629     0.643

=== Confusion Matrix ===

    a    b   <-- classified as
 6958 1945 |    a = no
 2075 1892 |    b = yes
```

Figure: KNN Algorithm

Decision Trees:

1. Firstly, classify was selected, ‘Choose’ option was pressed then J48 was chosen, and at last start button was pressed.



Classifier
Choose **J48** -C 0.25 -M 2

Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

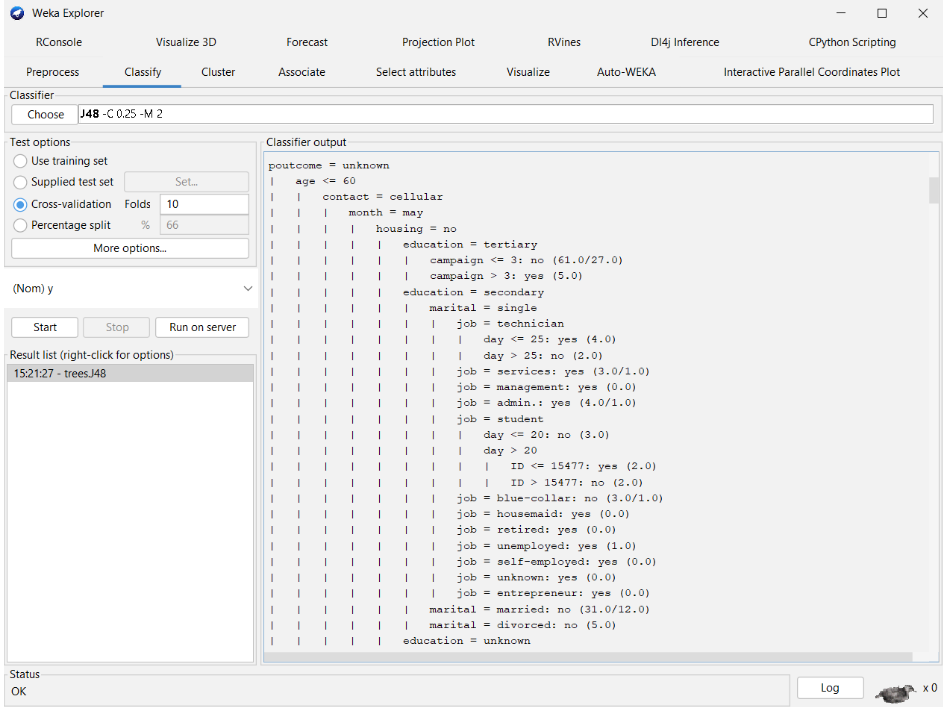
Result list (right-click for options)
15:21:27 - trees.J48

Status
OK

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     train
Instances:    12870
Attributes:   17
              ID
              age
              job
              marital
              education
              default
              balance
              housing
              loan
              contact
              day
              month
              campaign
              pdays
              previous
              poutcome
              y
Test mode:    10-fold cross-validation

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

J48 pruned tree
------------------

poutcome = unknown
```

Figure: Decision Tree Algorithm



Classifier
Choose **J48** -C 0.25 -M 2

Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

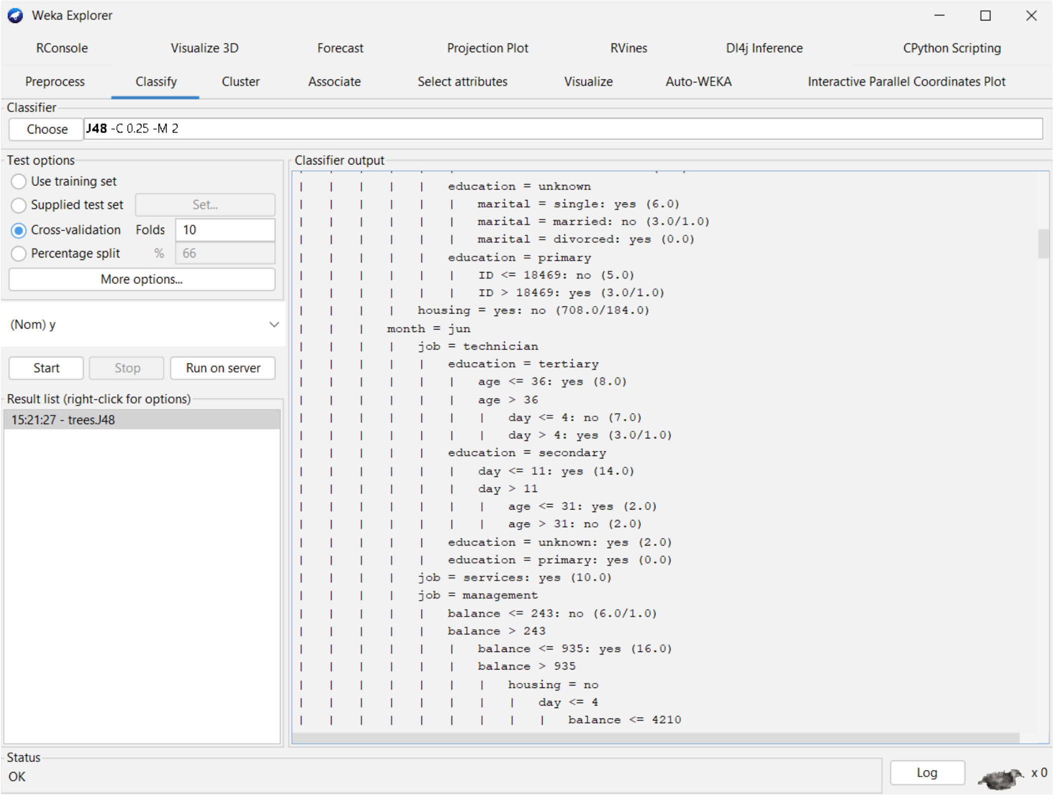
Result list (right-click for options)
15:21:27 - trees.J48

Status
OK

Classifier output

```
poutcome = unknown
|   age <= 60
|   |   contact = cellular
|   |   |   month = may
|   |   |   |   housing = no
|   |   |   |   |   education = tertiary
|   |   |   |   |   |   campaign <= 3: no (61.0/27.0)
|   |   |   |   |   |   campaign > 3: yes (5.0)
|   |   |   |   |   education = secondary
|   |   |   |   |   |   marital = single
|   |   |   |   |   |   |   job = technician
|   |   |   |   |   |   |   |   day <= 25: yes (4.0)
|   |   |   |   |   |   |   |   day > 25: no (2.0)
|   |   |   |   |   |   |   job = services: yes (3.0/1.0)
|   |   |   |   |   |   |   job = management: yes (0.0)
|   |   |   |   |   |   |   job = admin.: yes (4.0/1.0)
|   |   |   |   |   |   |   job = student
|   |   |   |   |   |   |   |   day <= 20: no (3.0)
|   |   |   |   |   |   |   |   day > 20
|   |   |   |   |   |   |   |   |   ID <= 15477: yes (2.0)
|   |   |   |   |   |   |   |   |   ID > 15477: no (2.0)
|   |   |   |   |   |   |   job = blue-collar: no (3.0/1.0)
|   |   |   |   |   |   |   job = housemaid: yes (0.0)
|   |   |   |   |   |   |   job = retired: yes (0.0)
|   |   |   |   |   |   |   job = unemployed: yes (1.0)
|   |   |   |   |   |   |   job = self-employed: yes (0.0)
|   |   |   |   |   |   |   job = unknown: yes (0.0)
|   |   |   |   |   |   |   job = entrepreneur: yes (0.0)
|   |   |   |   |   |   marital = married: no (31.0/12.0)
|   |   |   |   |   |   marital = divorced: no (5.0)
|   |   |   |   |   education = unknown
```

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options

- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) y

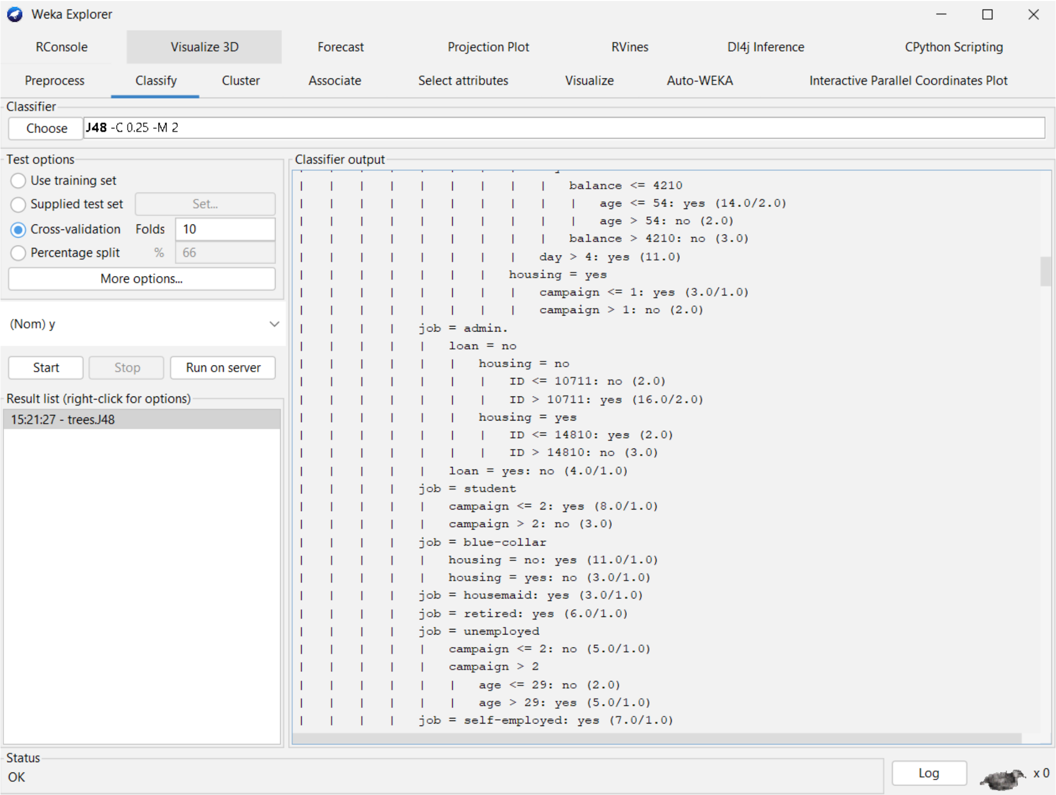
Result list (right-click for options): 15:21:27 - trees.J48

Classifier output

```
|   |   |   |   |   education = unknown
|   |   |   |   |   |   marital = single: yes (6.0)
|   |   |   |   |   |   marital = married: no (3.0/1.0)
|   |   |   |   |   |   marital = divorced: yes (0.0)
|   |   |   |   |   education = primary
|   |   |   |   |   |   ID <= 18469: no (5.0)
|   |   |   |   |   |   ID > 18469: yes (3.0/1.0)
|   |   |   |   housing = yes: no (708.0/184.0)
|   |   |   month = jun
|   |   |   |   job = technician
|   |   |   |   |   education = tertiary
|   |   |   |   |   |   age <= 36: yes (8.0)
|   |   |   |   |   |   age > 36
|   |   |   |   |   |   |   day <= 4: no (7.0)
|   |   |   |   |   |   |   day > 4: yes (3.0/1.0)
|   |   |   |   |   education = secondary
|   |   |   |   |   |   day <= 11: yes (14.0)
|   |   |   |   |   |   day > 11
|   |   |   |   |   |   |   age <= 31: yes (2.0)
|   |   |   |   |   |   |   age > 31: no (2.0)
|   |   |   |   |   education = unknown: yes (2.0)
|   |   |   |   |   education = primary: yes (0.0)
|   |   |   |   job = services: yes (10.0)
|   |   |   |   job = management
|   |   |   |   |   balance <= 243: no (6.0/1.0)
|   |   |   |   |   balance > 243
|   |   |   |   |   |   balance <= 935: yes (16.0)
|   |   |   |   |   |   balance > 935
|   |   |   |   |   |   |   housing = no
|   |   |   |   |   |   |   |   day <= 4
|   |   |   |   |   |   |   |   |   balance <= 4210
```

Status: OK

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options

- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) y

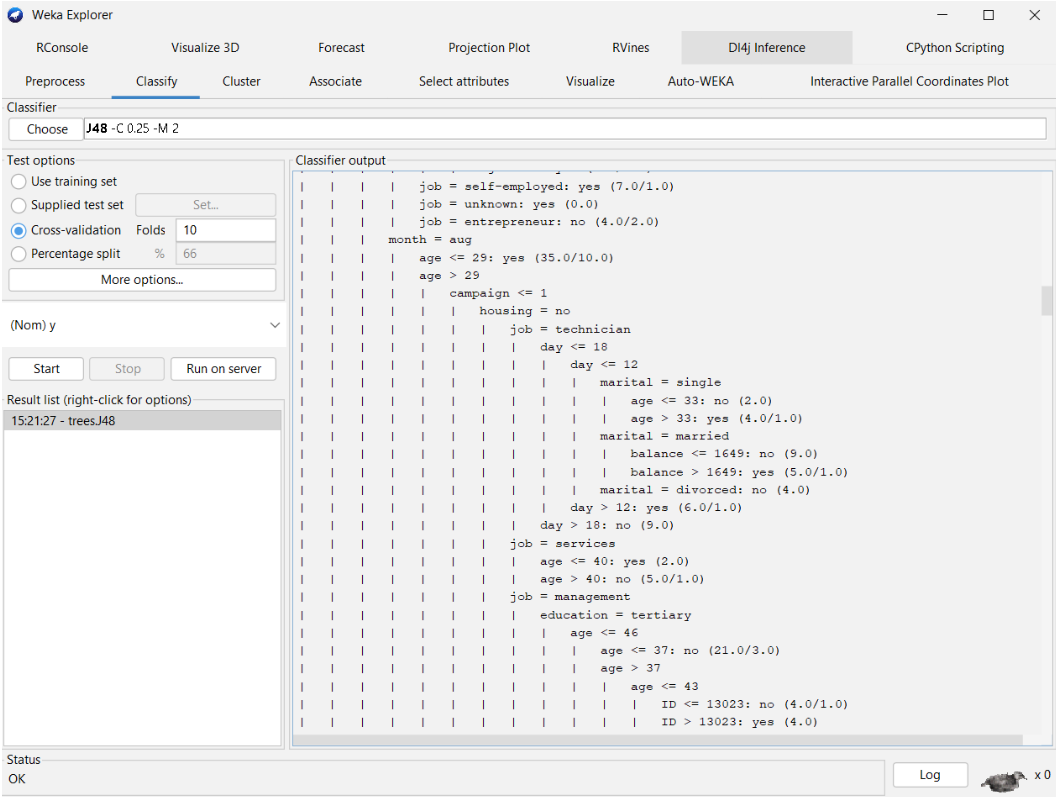
Result list (right-click for options): 15:21:27 - trees.J48

Classifier output

```
|   |   |   |   |   |   |   |   |   balance <= 4210
|   |   |   |   |   |   |   |   |   |   age <= 54: yes (14.0/2.0)
|   |   |   |   |   |   |   |   |   |   age > 54: no (2.0)
|   |   |   |   |   |   |   |   |   balance > 4210: no (3.0)
|   |   |   |   |   |   |   |   day > 4: yes (11.0)
|   |   |   |   |   |   |   housing = yes
|   |   |   |   |   |   |   |   campaign <= 1: yes (3.0/1.0)
|   |   |   |   |   |   |   |   campaign > 1: no (2.0)
|   |   |   |   job = admin.
|   |   |   |   |   loan = no
|   |   |   |   |   |   housing = no
|   |   |   |   |   |   |   ID <= 10711: no (2.0)
|   |   |   |   |   |   |   ID > 10711: yes (16.0/2.0)
|   |   |   |   |   |   housing = yes
|   |   |   |   |   |   |   ID <= 14810: yes (2.0)
|   |   |   |   |   |   |   ID > 14810: no (3.0)
|   |   |   |   |   loan = yes: no (4.0/1.0)
|   |   |   |   job = student
|   |   |   |   |   campaign <= 2: yes (8.0/1.0)
|   |   |   |   |   campaign > 2: no (3.0)
|   |   |   |   job = blue-collar
|   |   |   |   |   housing = no: yes (11.0/1.0)
|   |   |   |   |   housing = yes: no (3.0/1.0)
|   |   |   |   job = housemaid: yes (3.0/1.0)
|   |   |   |   job = retired: yes (6.0/1.0)
|   |   |   |   job = unemployed
|   |   |   |   |   campaign <= 2: no (5.0/1.0)
|   |   |   |   |   campaign > 2
|   |   |   |   |   |   age <= 29: no (2.0)
|   |   |   |   |   |   age > 29: yes (5.0/1.0)
|   |   |   |   job = self-employed: yes (7.0/1.0)
```

Status: OK

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options:
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

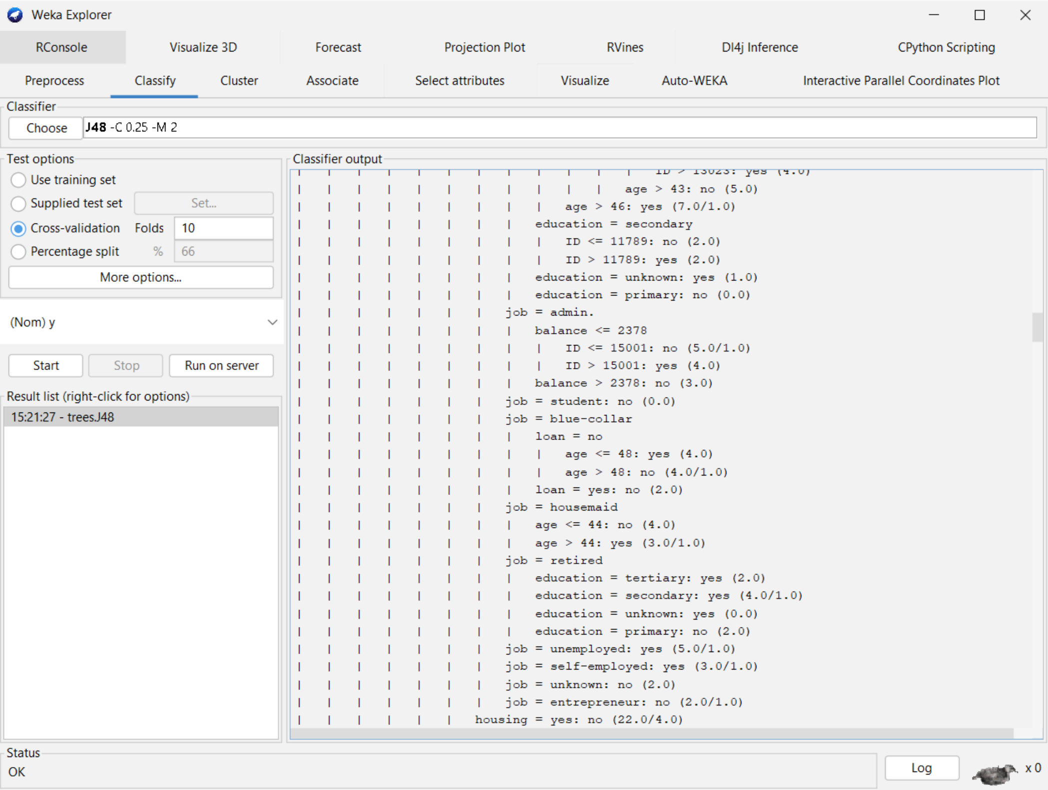
(Nom) y

Result list: 15:21:27 - trees.J48

Status: OK

```
|   |   |   |   job = self-employed: yes (7.0/1.0)
|   |   |   |   job = unknown: yes (0.0)
|   |   |   |   job = entrepreneur: no (4.0/2.0)
|   |   |   month = aug
|   |   |   |   age <= 29: yes (35.0/10.0)
|   |   |   |   age > 29
|   |   |   |   |   campaign <= 1
|   |   |   |   |   |   housing = no
|   |   |   |   |   |   |   job = technician
|   |   |   |   |   |   |   |   day <= 18
|   |   |   |   |   |   |   |   |   day <= 12
|   |   |   |   |   |   |   |   |   |   marital = single
|   |   |   |   |   |   |   |   |   |   |   age <= 33: no (2.0)
|   |   |   |   |   |   |   |   |   |   |   age > 33: yes (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   marital = married
|   |   |   |   |   |   |   |   |   |   |   balance <= 1649: no (9.0)
|   |   |   |   |   |   |   |   |   |   |   balance > 1649: yes (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   marital = divorced: no (4.0)
|   |   |   |   |   |   |   |   |   day > 12: yes (6.0/1.0)
|   |   |   |   |   |   |   |   day > 18: no (9.0)
|   |   |   |   |   |   |   job = services
|   |   |   |   |   |   |   |   age <= 40: yes (2.0)
|   |   |   |   |   |   |   |   age > 40: no (5.0/1.0)
|   |   |   |   |   |   |   job = management
|   |   |   |   |   |   |   |   education = tertiary
|   |   |   |   |   |   |   |   |   age <= 46
|   |   |   |   |   |   |   |   |   |   age <= 37: no (21.0/3.0)
|   |   |   |   |   |   |   |   |   |   age > 37
|   |   |   |   |   |   |   |   |   |   |   age <= 43
|   |   |   |   |   |   |   |   |   |   |   |   ID <= 13023: no (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   ID > 13023: yes (4.0)
```

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options:
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

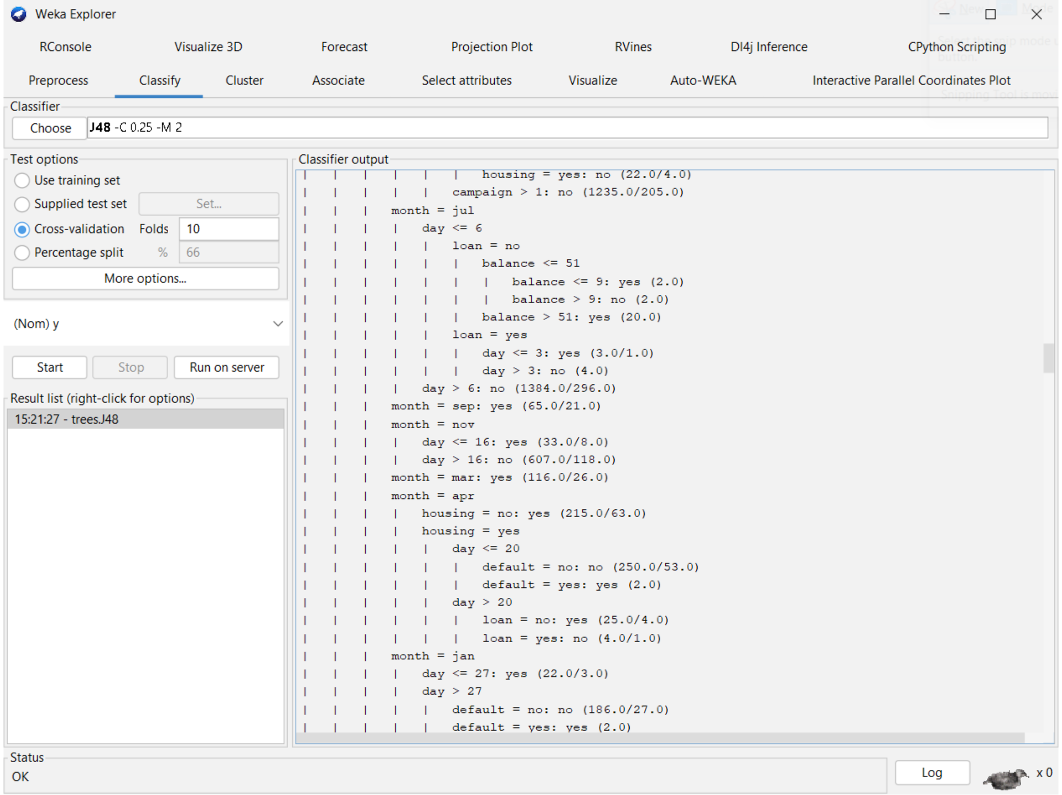
(Nom) y

Result list: 15:21:27 - trees.J48

Status: OK

```
|   |   |   |   |   |   |   |   |   |   |   |   ID > 13023: yes (4.0)
|   |   |   |   |   |   |   |   |   |   |   age > 43: no (5.0)
|   |   |   |   |   |   |   |   |   age > 46: yes (7.0/1.0)
|   |   |   |   |   |   |   |   education = secondary
|   |   |   |   |   |   |   |   |   ID <= 11789: no (2.0)
|   |   |   |   |   |   |   |   |   ID > 11789: yes (2.0)
|   |   |   |   |   |   |   |   education = unknown: yes (1.0)
|   |   |   |   |   |   |   |   education = primary: no (0.0)
|   |   |   |   |   |   |   job = admin.
|   |   |   |   |   |   |   |   balance <= 2378
|   |   |   |   |   |   |   |   |   ID <= 15001: no (5.0/1.0)
|   |   |   |   |   |   |   |   |   ID > 15001: yes (4.0)
|   |   |   |   |   |   |   |   balance > 2378: no (3.0)
|   |   |   |   |   |   |   job = student: no (0.0)
|   |   |   |   |   |   |   job = blue-collar
|   |   |   |   |   |   |   |   loan = no
|   |   |   |   |   |   |   |   |   age <= 48: yes (4.0)
|   |   |   |   |   |   |   |   |   age > 48: no (4.0/1.0)
|   |   |   |   |   |   |   |   loan = yes: no (2.0)
|   |   |   |   |   |   |   job = housemaid
|   |   |   |   |   |   |   |   age <= 44: no (4.0)
|   |   |   |   |   |   |   |   age > 44: yes (3.0/1.0)
|   |   |   |   |   |   |   job = retired
|   |   |   |   |   |   |   |   education = tertiary: yes (2.0)
|   |   |   |   |   |   |   |   education = secondary: yes (4.0/1.0)
|   |   |   |   |   |   |   |   education = unknown: yes (0.0)
|   |   |   |   |   |   |   |   education = primary: no (2.0)
|   |   |   |   |   |   |   job = unemployed: yes (5.0/1.0)
|   |   |   |   |   |   |   job = self-employed: yes (3.0/1.0)
|   |   |   |   |   |   |   job = unknown: no (2.0)
|   |   |   |   |   |   |   job = entrepreneur: no (2.0/1.0)
|   |   |   |   |   |   housing = yes: no (22.0/4.0)
```

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set

○ Supplied test set

● Cross-validation Folds 10

○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

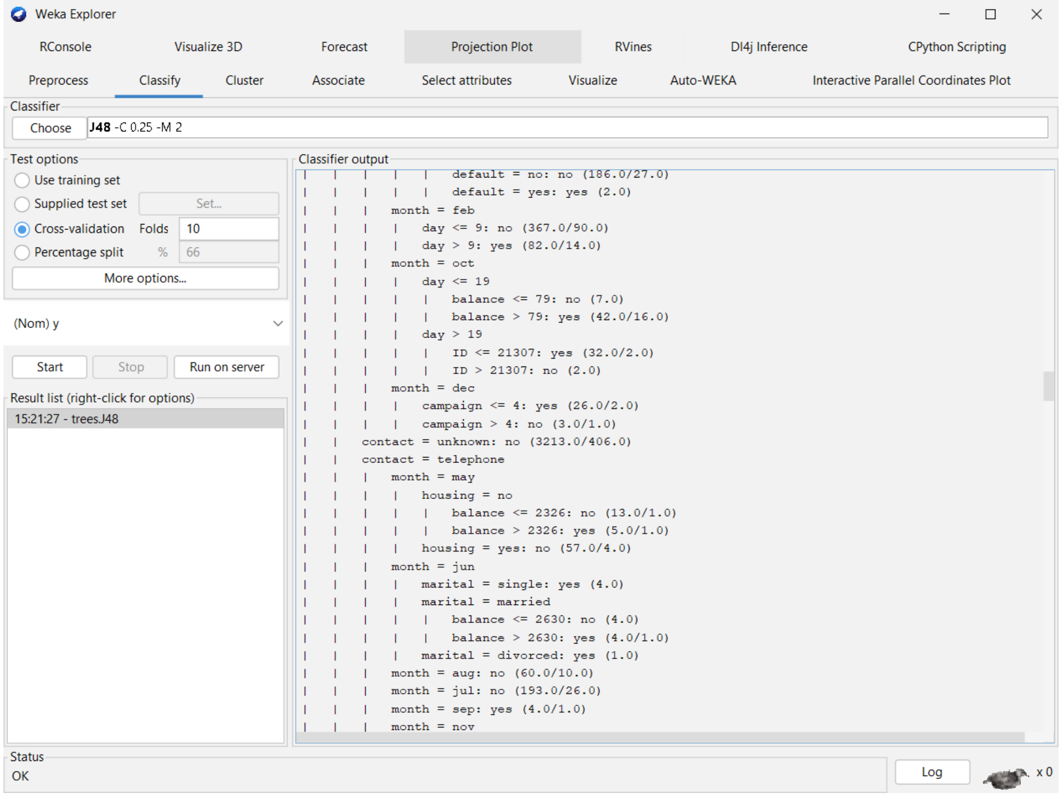
Classifier output

```
|   |   |   |   |   |   housing = yes: no (22.0/4.0)
|   |   |   |   |   campaign > 1: no (1235.0/205.0)
|   |   |   month = jul
|   |   |   |   day <= 6
|   |   |   |   |   loan = no
|   |   |   |   |   |   balance <= 51
|   |   |   |   |   |   |   balance <= 9: yes (2.0)
|   |   |   |   |   |   |   balance > 9: no (2.0)
|   |   |   |   |   |   balance > 51: yes (20.0)
|   |   |   |   |   loan = yes
|   |   |   |   |   |   day <= 3: yes (3.0/1.0)
|   |   |   |   |   |   day > 3: no (4.0)
|   |   |   |   day > 6: no (1384.0/296.0)
|   |   |   month = sep: yes (65.0/21.0)
|   |   |   month = nov
|   |   |   |   day <= 16: yes (33.0/8.0)
|   |   |   |   day > 16: no (607.0/118.0)
|   |   |   month = mar: yes (116.0/26.0)
|   |   |   month = apr
|   |   |   |   housing = no: yes (215.0/63.0)
|   |   |   |   housing = yes
|   |   |   |   |   day <= 20
|   |   |   |   |   |   default = no: no (250.0/53.0)
|   |   |   |   |   |   default = yes: yes (2.0)
|   |   |   |   |   day > 20
|   |   |   |   |   |   loan = no: yes (25.0/4.0)
|   |   |   |   |   |   loan = yes: no (4.0/1.0)
|   |   |   month = jan
|   |   |   |   day <= 27: yes (22.0/3.0)
|   |   |   |   day > 27
|   |   |   |   |   default = no: no (186.0/27.0)
|   |   |   |   |   default = yes: yes (2.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set

○ Supplied test set

● Cross-validation Folds 10

○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

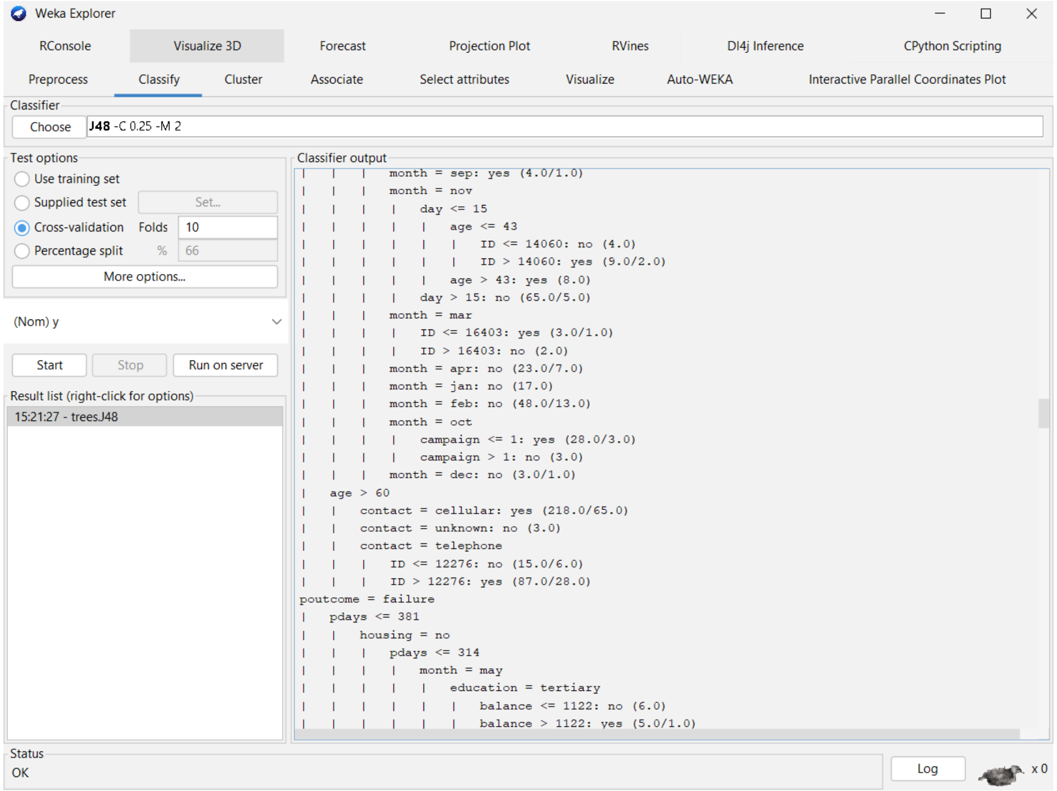
Classifier output

```
|   |   |   |   |   default = no: no (186.0/27.0)
|   |   |   |   |   default = yes: yes (2.0)
|   |   |   month = feb
|   |   |   |   day <= 9: no (367.0/90.0)
|   |   |   |   day > 9: yes (82.0/14.0)
|   |   |   month = oct
|   |   |   |   day <= 19
|   |   |   |   |   balance <= 79: no (7.0)
|   |   |   |   |   balance > 79: yes (42.0/16.0)
|   |   |   |   day > 19
|   |   |   |   |   ID <= 21307: yes (32.0/2.0)
|   |   |   |   |   ID > 21307: no (2.0)
|   |   |   month = dec
|   |   |   |   campaign <= 4: yes (26.0/2.0)
|   |   |   |   campaign > 4: no (3.0/1.0)
|   |   contact = unknown: no (3213.0/406.0)
|   |   contact = telephone
|   |   |   month = may
|   |   |   |   housing = no
|   |   |   |   |   balance <= 2326: no (13.0/1.0)
|   |   |   |   |   balance > 2326: yes (5.0/1.0)
|   |   |   |   housing = yes: no (57.0/4.0)
|   |   |   month = jun
|   |   |   |   marital = single: yes (4.0)
|   |   |   |   marital = married
|   |   |   |   |   balance <= 2630: no (4.0)
|   |   |   |   |   balance > 2630: yes (4.0/1.0)
|   |   |   |   marital = divorced: yes (1.0)
|   |   |   month = aug: no (60.0/10.0)
|   |   |   month = jul: no (193.0/26.0)
|   |   |   month = sep: yes (4.0/1.0)
|   |   |   month = nov
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

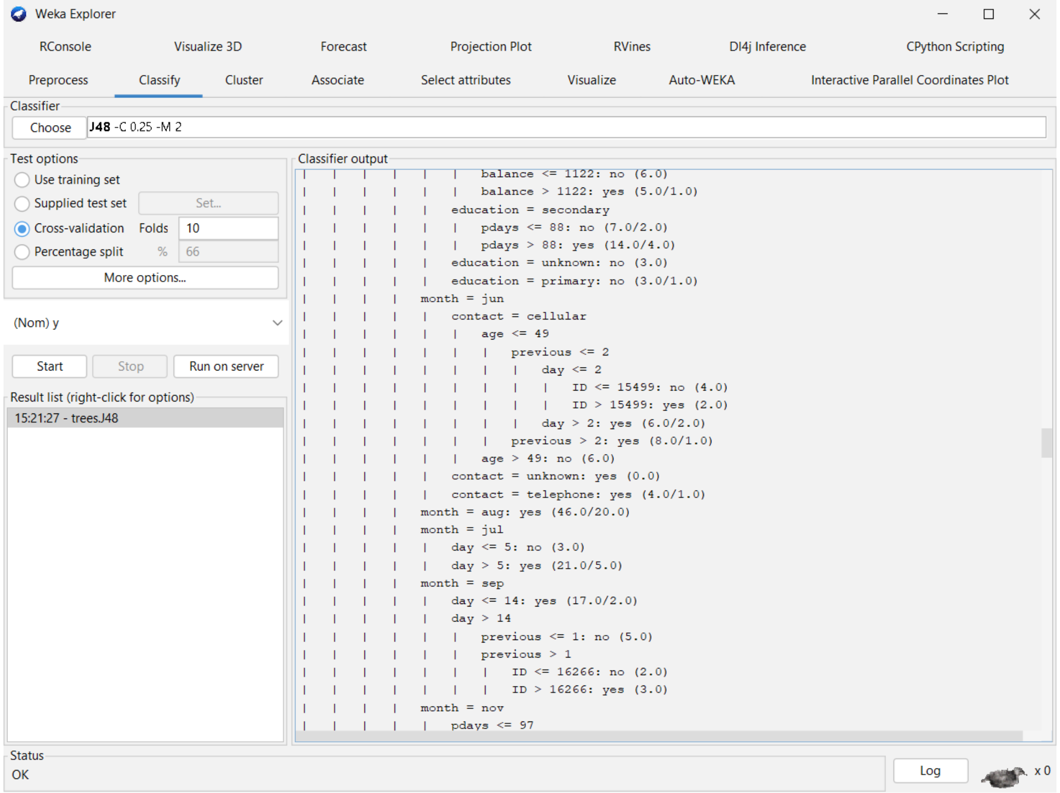
Classifier output

```
|   |   |   month = sep: yes (4.0/1.0)
|   |   |   month = nov
|   |   |   |   day <= 15
|   |   |   |   |   age <= 43
|   |   |   |   |   |   ID <= 14060: no (4.0)
|   |   |   |   |   |   ID > 14060: yes (9.0/2.0)
|   |   |   |   |   age > 43: yes (8.0)
|   |   |   |   day > 15: no (65.0/5.0)
|   |   |   month = mar
|   |   |   |   ID <= 16403: yes (3.0/1.0)
|   |   |   |   ID > 16403: no (2.0)
|   |   |   month = apr: no (23.0/7.0)
|   |   |   month = jan: no (17.0)
|   |   |   month = feb: no (48.0/13.0)
|   |   |   month = oct
|   |   |   |   campaign <= 1: yes (28.0/3.0)
|   |   |   |   campaign > 1: no (3.0)
|   |   |   month = dec: no (3.0/1.0)
|   age > 60
|   |   contact = cellular: yes (218.0/65.0)
|   |   contact = unknown: no (3.0)
|   |   contact = telephone
|   |   |   ID <= 12276: no (15.0/6.0)
|   |   |   ID > 12276: yes (87.0/28.0)
poutcome = failure
|   pdays <= 381
|   |   housing = no
|   |   |   pdays <= 314
|   |   |   |   month = may
|   |   |   |   |   education = tertiary
|   |   |   |   |   |   balance <= 1122: no (6.0)
|   |   |   |   |   |   balance > 1122: yes (5.0/1.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

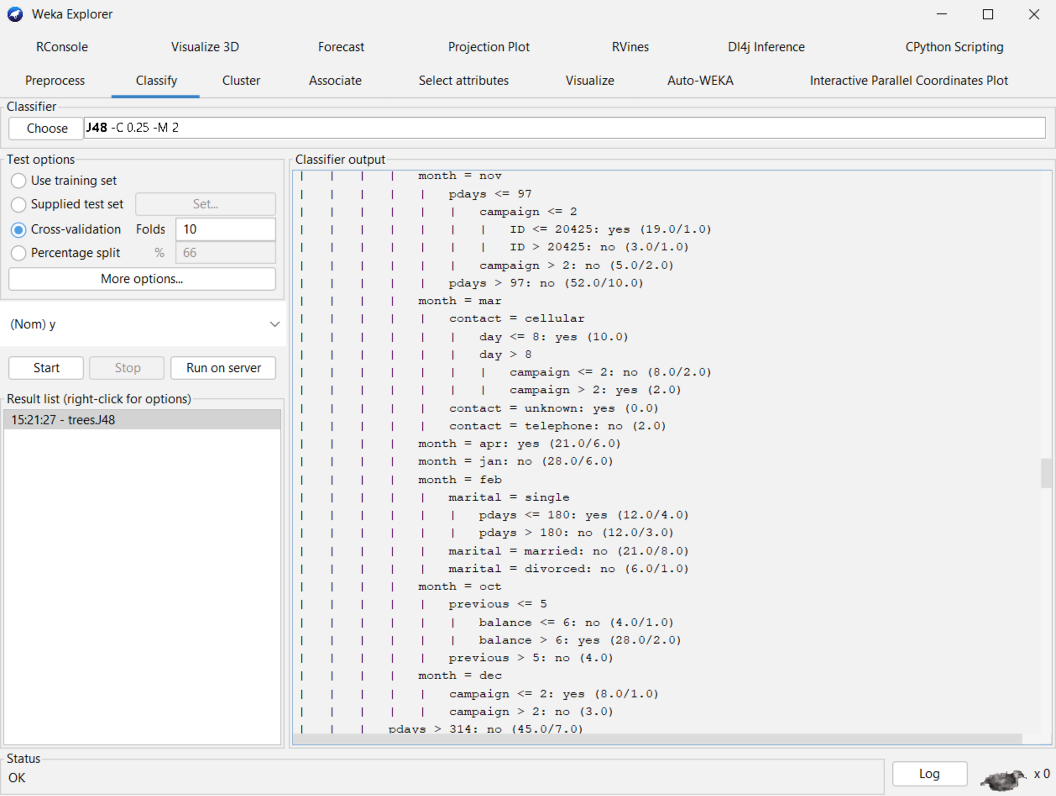
Classifier output

```
|   |   |   |   |   |   balance <= 1122: no (6.0)
|   |   |   |   |   |   balance > 1122: yes (5.0/1.0)
|   |   |   |   |   education = secondary
|   |   |   |   |   |   pdays <= 88: no (7.0/2.0)
|   |   |   |   |   |   pdays > 88: yes (14.0/4.0)
|   |   |   |   |   education = unknown: no (3.0)
|   |   |   |   |   education = primary: no (3.0/1.0)
|   |   |   |   month = jun
|   |   |   |   |   contact = cellular
|   |   |   |   |   |   age <= 49
|   |   |   |   |   |   |   previous <= 2
|   |   |   |   |   |   |   |   day <= 2
|   |   |   |   |   |   |   |   |   ID <= 15499: no (4.0)
|   |   |   |   |   |   |   |   |   ID > 15499: yes (2.0)
|   |   |   |   |   |   |   |   day > 2: yes (6.0/2.0)
|   |   |   |   |   |   |   previous > 2: yes (8.0/1.0)
|   |   |   |   |   |   age > 49: no (6.0)
|   |   |   |   |   contact = unknown: yes (0.0)
|   |   |   |   |   contact = telephone: yes (4.0/1.0)
|   |   |   |   month = aug: yes (46.0/20.0)
|   |   |   |   month = jul
|   |   |   |   |   day <= 5: no (3.0)
|   |   |   |   |   day > 5: yes (21.0/5.0)
|   |   |   |   month = sep
|   |   |   |   |   day <= 14: yes (17.0/2.0)
|   |   |   |   |   day > 14
|   |   |   |   |   |   previous <= 1: no (5.0)
|   |   |   |   |   |   previous > 1
|   |   |   |   |   |   |   ID <= 16266: no (2.0)
|   |   |   |   |   |   |   ID > 16266: yes (3.0)
|   |   |   |   month = nov
|   |   |   |   |   pdays <= 97
```

Status

OK

Figure: Decision Tree Algorithm



Classifier
J48 -C 0.25 -M 2

Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

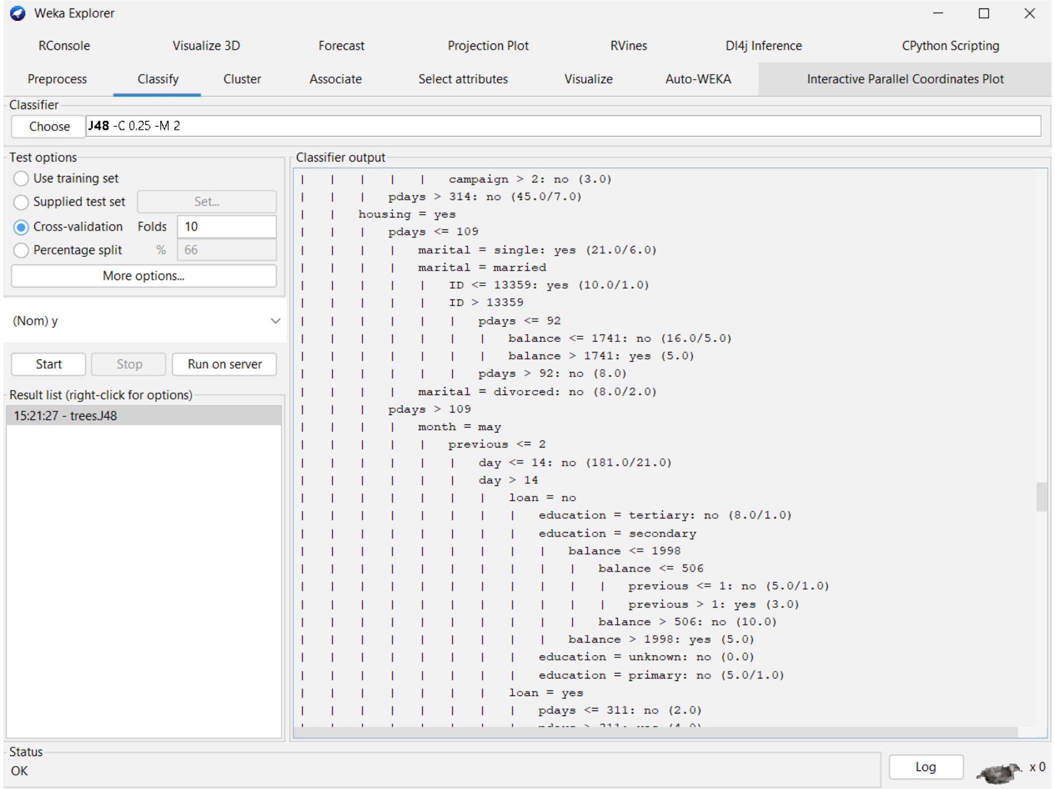
Result list (right-click for options)
15:21:27 - trees.J48

Classifier output

```
|   |   |   |   month = nov
|   |   |   |   |   pdays <= 97
|   |   |   |   |   |   campaign <= 2
|   |   |   |   |   |   |   ID <= 20425: yes (19.0/1.0)
|   |   |   |   |   |   |   ID > 20425: no (3.0/1.0)
|   |   |   |   |   |   campaign > 2: no (5.0/2.0)
|   |   |   |   |   pdays > 97: no (52.0/10.0)
|   |   |   |   month = mar
|   |   |   |   |   contact = cellular
|   |   |   |   |   |   day <= 8: yes (10.0)
|   |   |   |   |   |   day > 8
|   |   |   |   |   |   |   campaign <= 2: no (8.0/2.0)
|   |   |   |   |   |   |   campaign > 2: yes (2.0)
|   |   |   |   |   contact = unknown: yes (0.0)
|   |   |   |   |   contact = telephone: no (2.0)
|   |   |   |   month = apr: yes (21.0/6.0)
|   |   |   |   month = jan: no (28.0/6.0)
|   |   |   |   month = feb
|   |   |   |   |   marital = single
|   |   |   |   |   |   pdays <= 180: yes (12.0/4.0)
|   |   |   |   |   |   pdays > 180: no (12.0/3.0)
|   |   |   |   |   marital = married: no (21.0/8.0)
|   |   |   |   |   marital = divorced: no (6.0/1.0)
|   |   |   |   month = oct
|   |   |   |   |   previous <= 5
|   |   |   |   |   |   balance <= 6: no (4.0/1.0)
|   |   |   |   |   |   balance > 6: yes (28.0/2.0)
|   |   |   |   |   previous > 5: no (4.0)
|   |   |   |   month = dec
|   |   |   |   |   campaign <= 2: yes (8.0/1.0)
|   |   |   |   |   campaign > 2: no (3.0)
|   |   |   pdays > 314: no (45.0/7.0)
```

Status
OK

Figure: Decision Tree Algorithm



Classifier
J48 -C 0.25 -M 2

Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

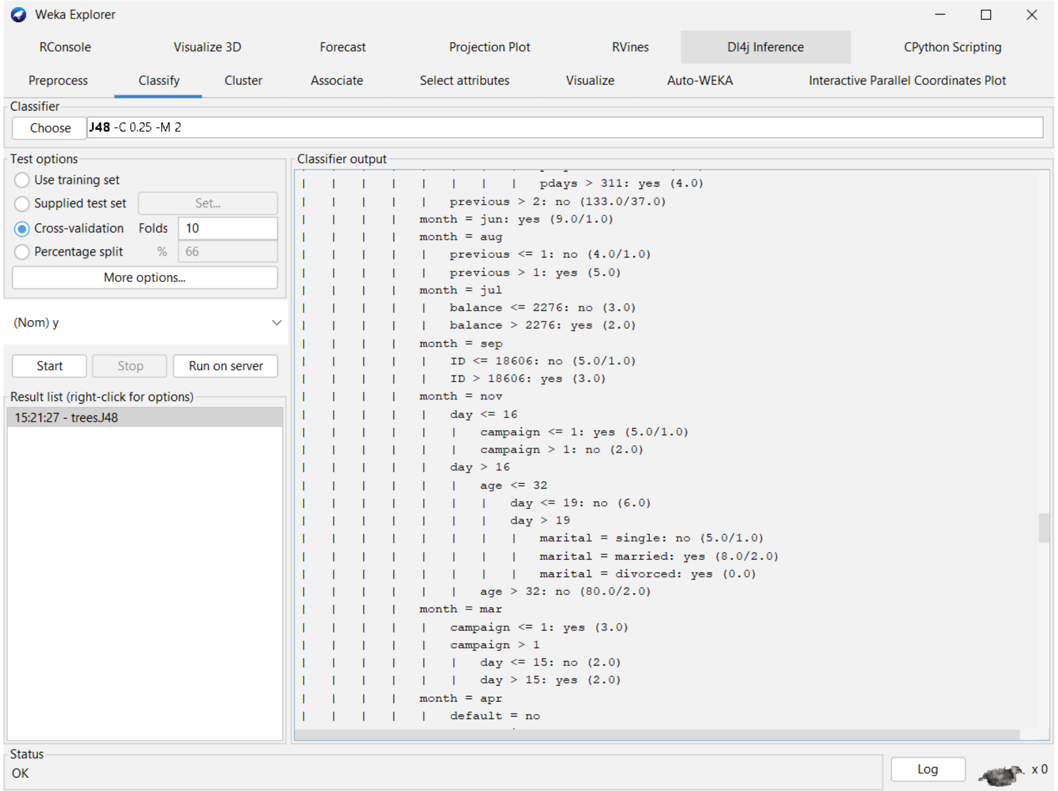
Result list (right-click for options)
15:21:27 - trees.J48

Classifier output

```
|   |   |   |   |   campaign > 2: no (3.0)
|   |   |   pdays > 314: no (45.0/7.0)
|   |   housing = yes
|   |   |   pdays <= 109
|   |   |   |   marital = single: yes (21.0/6.0)
|   |   |   |   marital = married
|   |   |   |   |   ID <= 13359: yes (10.0/1.0)
|   |   |   |   |   ID > 13359
|   |   |   |   |   |   pdays <= 92
|   |   |   |   |   |   |   balance <= 1741: no (16.0/5.0)
|   |   |   |   |   |   |   balance > 1741: yes (5.0)
|   |   |   |   |   |   pdays > 92: no (8.0)
|   |   |   |   marital = divorced: no (8.0/2.0)
|   |   |   pdays > 109
|   |   |   |   month = may
|   |   |   |   |   previous <= 2
|   |   |   |   |   |   day <= 14: no (181.0/21.0)
|   |   |   |   |   |   day > 14
|   |   |   |   |   |   |   loan = no
|   |   |   |   |   |   |   |   education = tertiary: no (8.0/1.0)
|   |   |   |   |   |   |   |   education = secondary
|   |   |   |   |   |   |   |   |   balance <= 1998
|   |   |   |   |   |   |   |   |   |   balance <= 506
|   |   |   |   |   |   |   |   |   |   |   previous <= 1: no (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   previous > 1: yes (3.0)
|   |   |   |   |   |   |   |   |   |   balance > 506: no (10.0)
|   |   |   |   |   |   |   |   |   balance > 1998: yes (5.0)
|   |   |   |   |   |   |   |   education = unknown: no (0.0)
|   |   |   |   |   |   |   |   education = primary: no (5.0/1.0)
|   |   |   |   |   |   |   loan = yes
|   |   |   |   |   |   |   |   pdays <= 311: no (2.0)
|   |   |   |   |   |   |   |   pdays > 311: yes (4.0)
```

Status
OK

Figure: Decision Tree Algorithm



J48 -C 0.25 -M 2

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

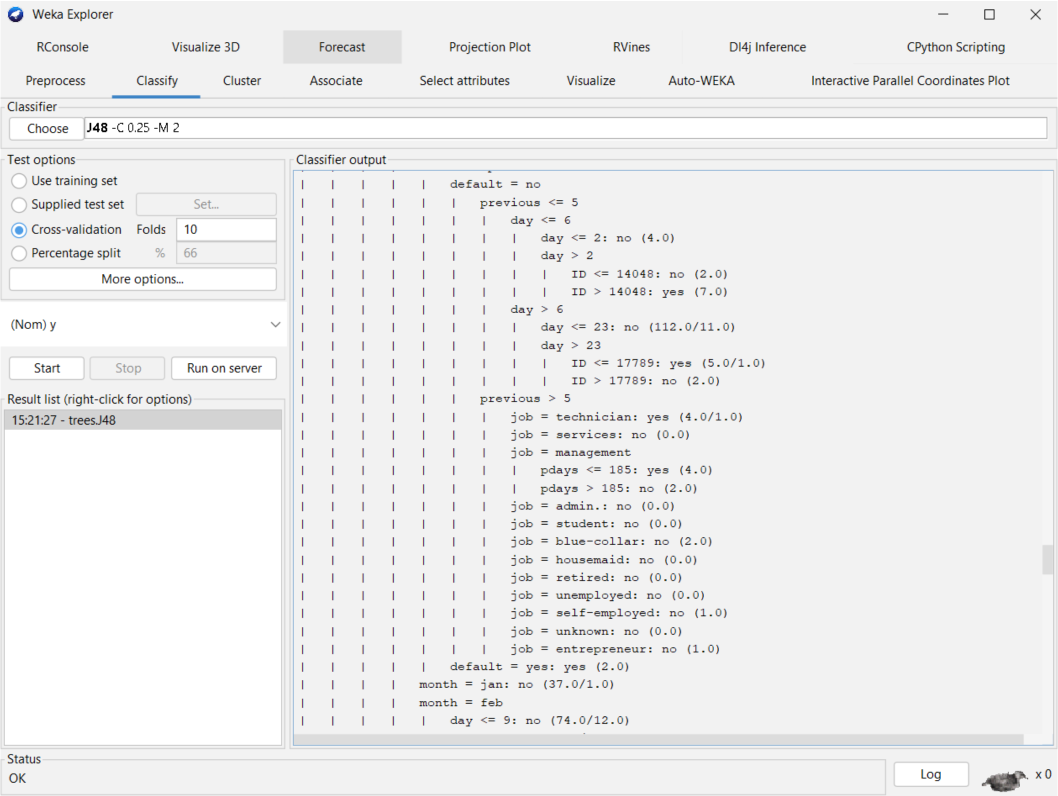
(Nom) y

15:21:27 - trees.J48

OK

```
|   |   |   |   |   |   |   |   pdays > 311: yes (4.0)
|   |   |   |   |   previous > 2: no (133.0/37.0)
|   |   |   |   month = jun: yes (9.0/1.0)
|   |   |   |   month = aug
|   |   |   |   |   previous <= 1: no (4.0/1.0)
|   |   |   |   |   previous > 1: yes (5.0)
|   |   |   |   month = jul
|   |   |   |   |   balance <= 2276: no (3.0)
|   |   |   |   |   balance > 2276: yes (2.0)
|   |   |   |   month = sep
|   |   |   |   |   ID <= 18606: no (5.0/1.0)
|   |   |   |   |   ID > 18606: yes (3.0)
|   |   |   |   month = nov
|   |   |   |   |   day <= 16
|   |   |   |   |   |   campaign <= 1: yes (5.0/1.0)
|   |   |   |   |   |   campaign > 1: no (2.0)
|   |   |   |   |   day > 16
|   |   |   |   |   |   age <= 32
|   |   |   |   |   |   |   day <= 19: no (6.0)
|   |   |   |   |   |   |   day > 19
|   |   |   |   |   |   |   |   marital = single: no (5.0/1.0)
|   |   |   |   |   |   |   |   marital = married: yes (8.0/2.0)
|   |   |   |   |   |   |   |   marital = divorced: yes (0.0)
|   |   |   |   |   |   age > 32: no (80.0/2.0)
|   |   |   |   month = mar
|   |   |   |   |   campaign <= 1: yes (3.0)
|   |   |   |   |   campaign > 1
|   |   |   |   |   |   day <= 15: no (2.0)
|   |   |   |   |   |   day > 15: yes (2.0)
|   |   |   |   month = apr
|   |   |   |   |   default = no
```

Figure: Decision Tree Algorithm



J48 -C 0.25 -M 2

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

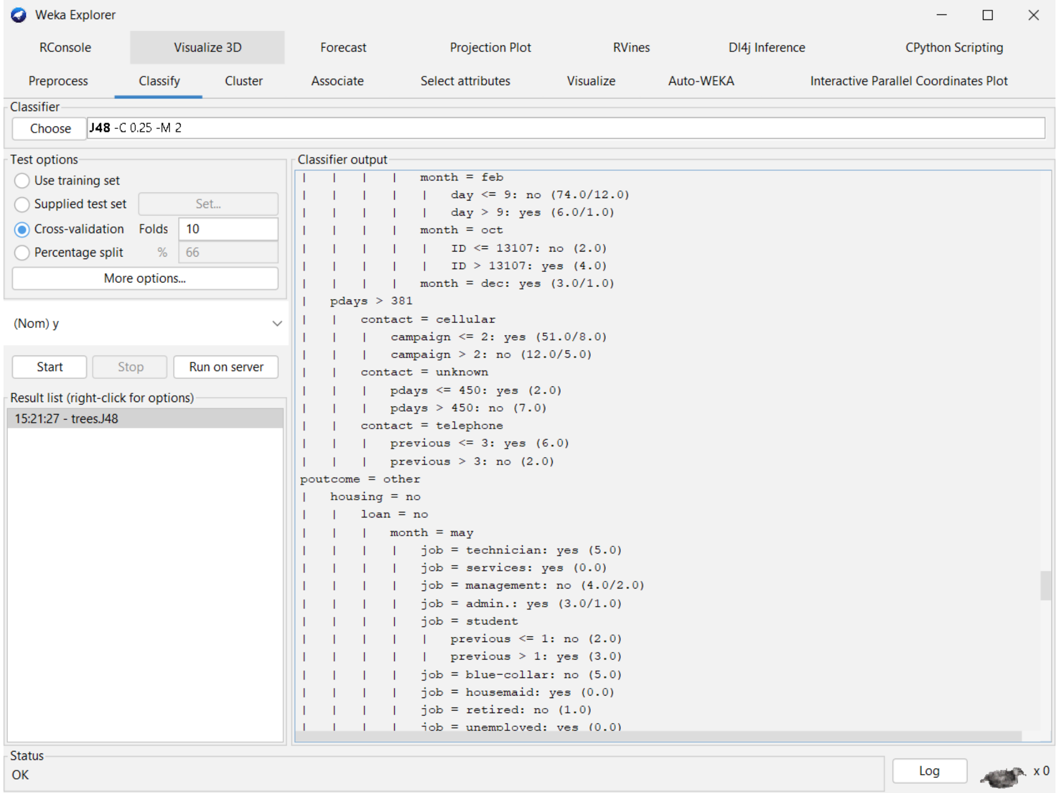
(Nom) y

15:21:27 - trees.J48

OK

```
|   |   |   |   |   default = no
|   |   |   |   |   |   previous <= 5
|   |   |   |   |   |   |   day <= 6
|   |   |   |   |   |   |   |   day <= 2: no (4.0)
|   |   |   |   |   |   |   |   day > 2
|   |   |   |   |   |   |   |   |   ID <= 14048: no (2.0)
|   |   |   |   |   |   |   |   |   ID > 14048: yes (7.0)
|   |   |   |   |   |   |   day > 6
|   |   |   |   |   |   |   |   day <= 23: no (112.0/11.0)
|   |   |   |   |   |   |   |   day > 23
|   |   |   |   |   |   |   |   |   ID <= 17789: yes (5.0/1.0)
|   |   |   |   |   |   |   |   |   ID > 17789: no (2.0)
|   |   |   |   |   |   previous > 5
|   |   |   |   |   |   |   job = technician: yes (4.0/1.0)
|   |   |   |   |   |   |   job = services: no (0.0)
|   |   |   |   |   |   |   job = management
|   |   |   |   |   |   |   |   pdays <= 185: yes (4.0)
|   |   |   |   |   |   |   |   pdays > 185: no (2.0)
|   |   |   |   |   |   |   job = admin.: no (0.0)
|   |   |   |   |   |   |   job = student: no (0.0)
|   |   |   |   |   |   |   job = blue-collar: no (2.0)
|   |   |   |   |   |   |   job = housemaid: no (0.0)
|   |   |   |   |   |   |   job = retired: no (0.0)
|   |   |   |   |   |   |   job = unemployed: no (0.0)
|   |   |   |   |   |   |   job = self-employed: no (1.0)
|   |   |   |   |   |   |   job = unknown: no (0.0)
|   |   |   |   |   |   |   job = entrepreneur: no (1.0)
|   |   |   |   |   default = yes: yes (2.0)
|   |   |   |   month = jan: no (37.0/1.0)
|   |   |   |   month = feb
|   |   |   |   |   day <= 9: no (74.0/12.0)
```

Figure: Decision Tree Algorithm



Classifier

Choose **J48** -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

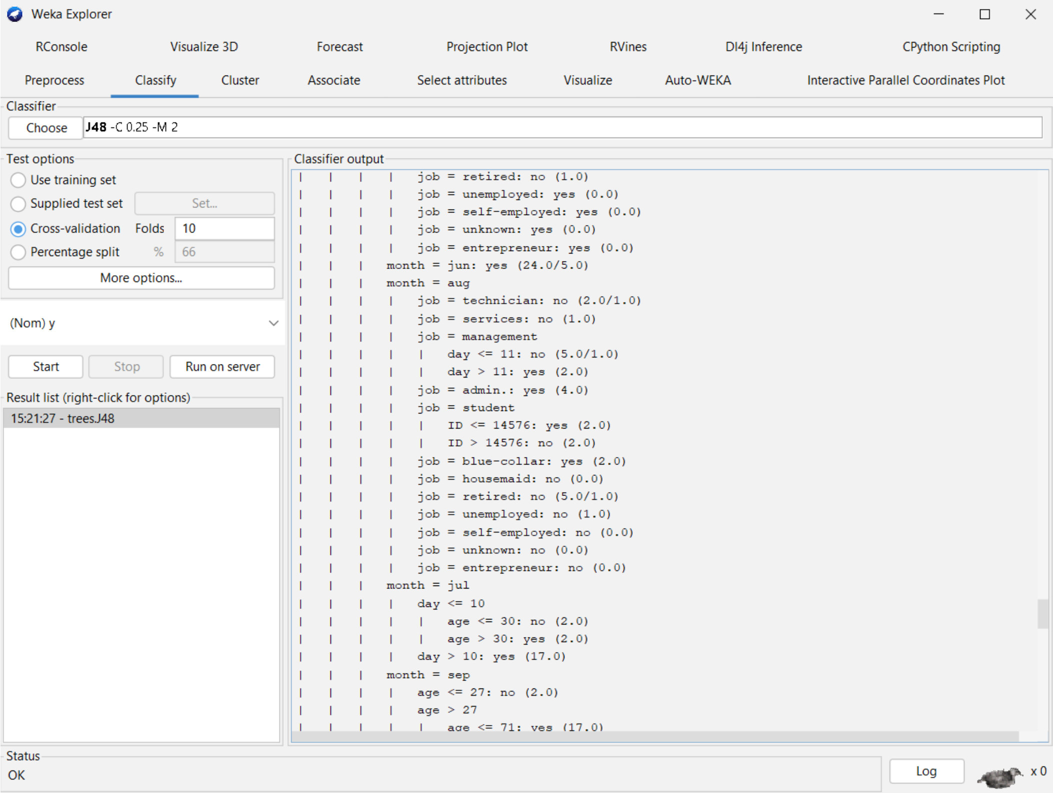
Classifier output

```
|   |   |   |   month = feb
|   |   |   |   |   day <= 9: no (74.0/12.0)
|   |   |   |   |   day > 9: yes (6.0/1.0)
|   |   |   |   month = oct
|   |   |   |   |   ID <= 13107: no (2.0)
|   |   |   |   |   ID > 13107: yes (4.0)
|   |   |   |   month = dec: yes (3.0/1.0)
|   pdays > 381
|   |   contact = cellular
|   |   |   campaign <= 2: yes (51.0/8.0)
|   |   |   campaign > 2: no (12.0/5.0)
|   |   contact = unknown
|   |   |   pdays <= 450: yes (2.0)
|   |   |   pdays > 450: no (7.0)
|   |   contact = telephone
|   |   |   previous <= 3: yes (6.0)
|   |   |   previous > 3: no (2.0)
poutcome = other
|   housing = no
|   |   loan = no
|   |   |   month = may
|   |   |   |   job = technician: yes (5.0)
|   |   |   |   job = services: yes (0.0)
|   |   |   |   job = management: no (4.0/2.0)
|   |   |   |   job = admin.: yes (3.0/1.0)
|   |   |   |   job = student
|   |   |   |   |   previous <= 1: no (2.0)
|   |   |   |   |   previous > 1: yes (3.0)
|   |   |   |   job = blue-collar: no (5.0)
|   |   |   |   job = housemaid: yes (0.0)
|   |   |   |   job = retired: no (1.0)
|   |   |   |   job = unemployed: yes (0.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

Choose **J48** -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

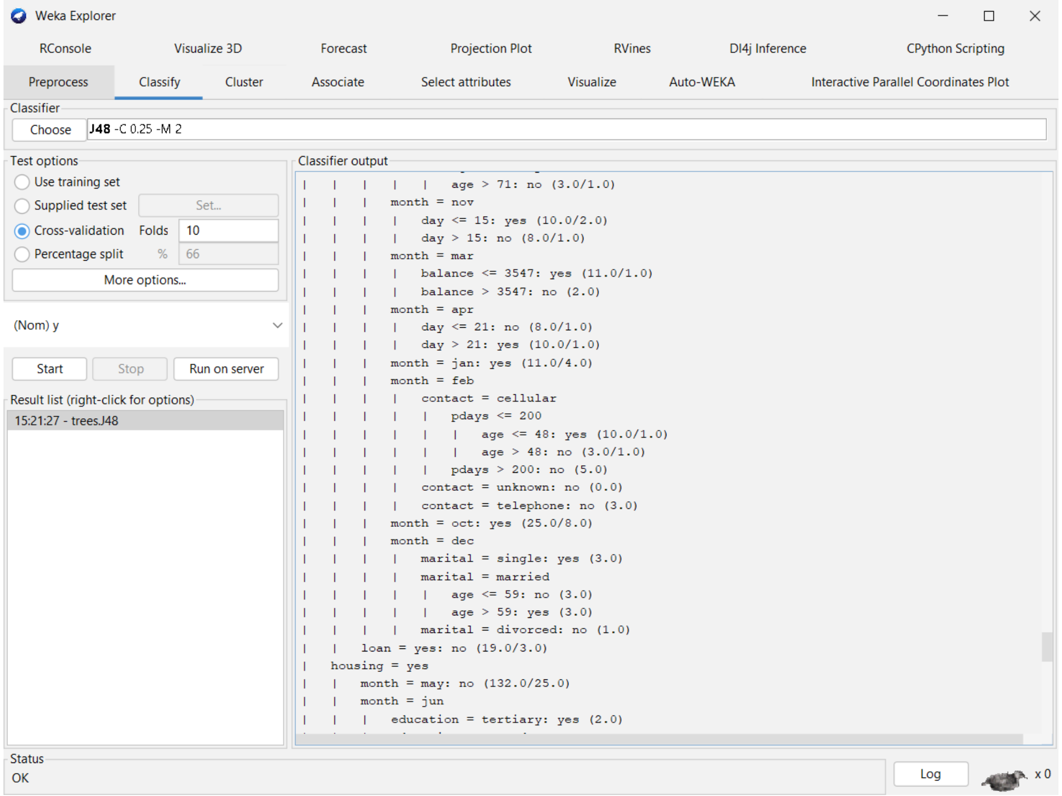
Classifier output

```
|   |   |   |   job = retired: no (1.0)
|   |   |   |   job = unemployed: yes (0.0)
|   |   |   |   job = self-employed: yes (0.0)
|   |   |   |   job = unknown: yes (0.0)
|   |   |   |   job = entrepreneur: yes (0.0)
|   |   |   month = jun: yes (24.0/5.0)
|   |   |   month = aug
|   |   |   |   job = technician: no (2.0/1.0)
|   |   |   |   job = services: no (1.0)
|   |   |   |   job = management
|   |   |   |   |   day <= 11: no (5.0/1.0)
|   |   |   |   |   day > 11: yes (2.0)
|   |   |   |   job = admin.: yes (4.0)
|   |   |   |   job = student
|   |   |   |   |   ID <= 14576: yes (2.0)
|   |   |   |   |   ID > 14576: no (2.0)
|   |   |   |   job = blue-collar: yes (2.0)
|   |   |   |   job = housemaid: no (0.0)
|   |   |   |   job = retired: no (5.0/1.0)
|   |   |   |   job = unemployed: no (1.0)
|   |   |   |   job = self-employed: no (0.0)
|   |   |   |   job = unknown: no (0.0)
|   |   |   |   job = entrepreneur: no (0.0)
|   |   |   month = jul
|   |   |   |   day <= 10
|   |   |   |   |   age <= 30: no (2.0)
|   |   |   |   |   age > 30: yes (2.0)
|   |   |   |   day > 10: yes (17.0)
|   |   |   month = sep
|   |   |   |   age <= 27: no (2.0)
|   |   |   |   age > 27
|   |   |   |   |   age <= 71: yes (17.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set

○ Supplied test set

● Cross-validation Folds 10

○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

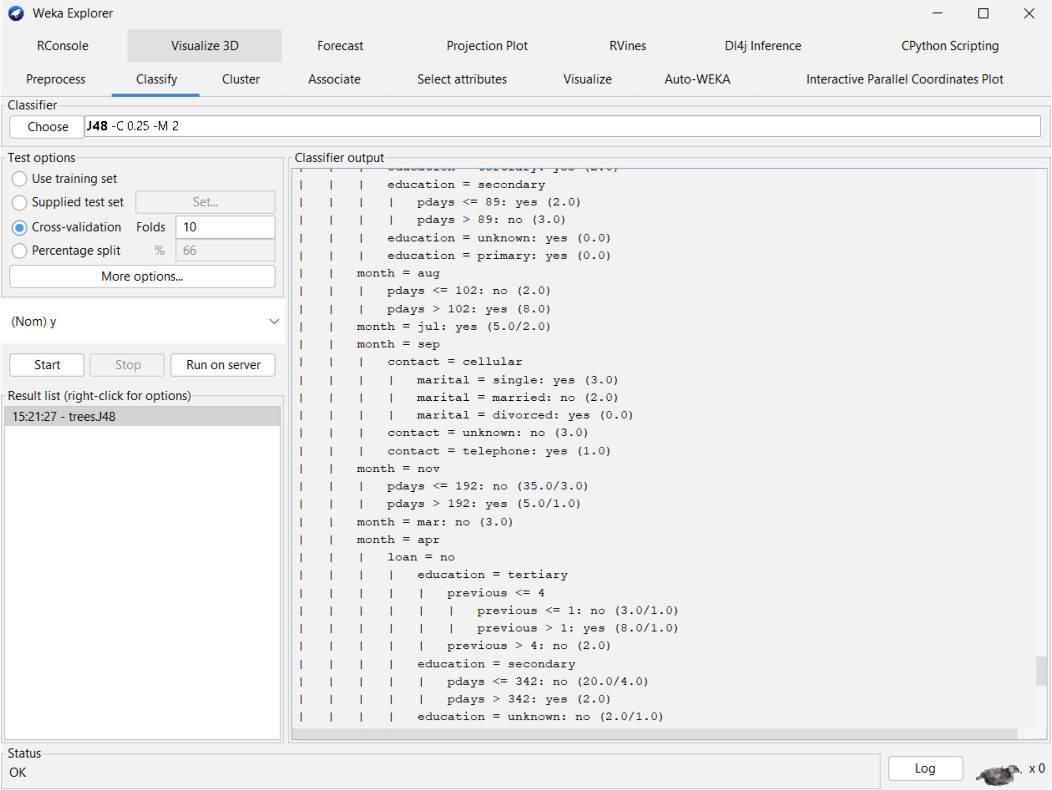
Classifier output

```
|   |   |   |   |   age > 71: no (3.0/1.0)
|   |   |   month = nov
|   |   |   |   day <= 15: yes (10.0/2.0)
|   |   |   |   day > 15: no (8.0/1.0)
|   |   |   month = mar
|   |   |   |   balance <= 3547: yes (11.0/1.0)
|   |   |   |   balance > 3547: no (2.0)
|   |   |   month = apr
|   |   |   |   day <= 21: no (8.0/1.0)
|   |   |   |   day > 21: yes (10.0/1.0)
|   |   |   month = jan: yes (11.0/4.0)
|   |   |   month = feb
|   |   |   |   contact = cellular
|   |   |   |   |   pdays <= 200
|   |   |   |   |   |   age <= 48: yes (10.0/1.0)
|   |   |   |   |   |   age > 48: no (3.0/1.0)
|   |   |   |   |   pdays > 200: no (5.0)
|   |   |   |   contact = unknown: no (0.0)
|   |   |   |   contact = telephone: no (3.0)
|   |   |   month = oct: yes (25.0/8.0)
|   |   |   month = dec
|   |   |   |   marital = single: yes (3.0)
|   |   |   |   marital = married
|   |   |   |   |   age <= 59: no (3.0)
|   |   |   |   |   age > 59: yes (3.0)
|   |   |   |   marital = divorced: no (1.0)
|   |   loan = yes: no (19.0/3.0)
|   housing = yes
|   |   month = may: no (132.0/25.0)
|   |   month = jun
|   |   |   education = tertiary: yes (2.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier

J48 -C 0.25 -M 2

Test options

○ Use training set

○ Supplied test set

● Cross-validation Folds 10

○ Percentage split % 66

(Nom) y

Result list (right-click for options)

15:21:27 - trees.J48

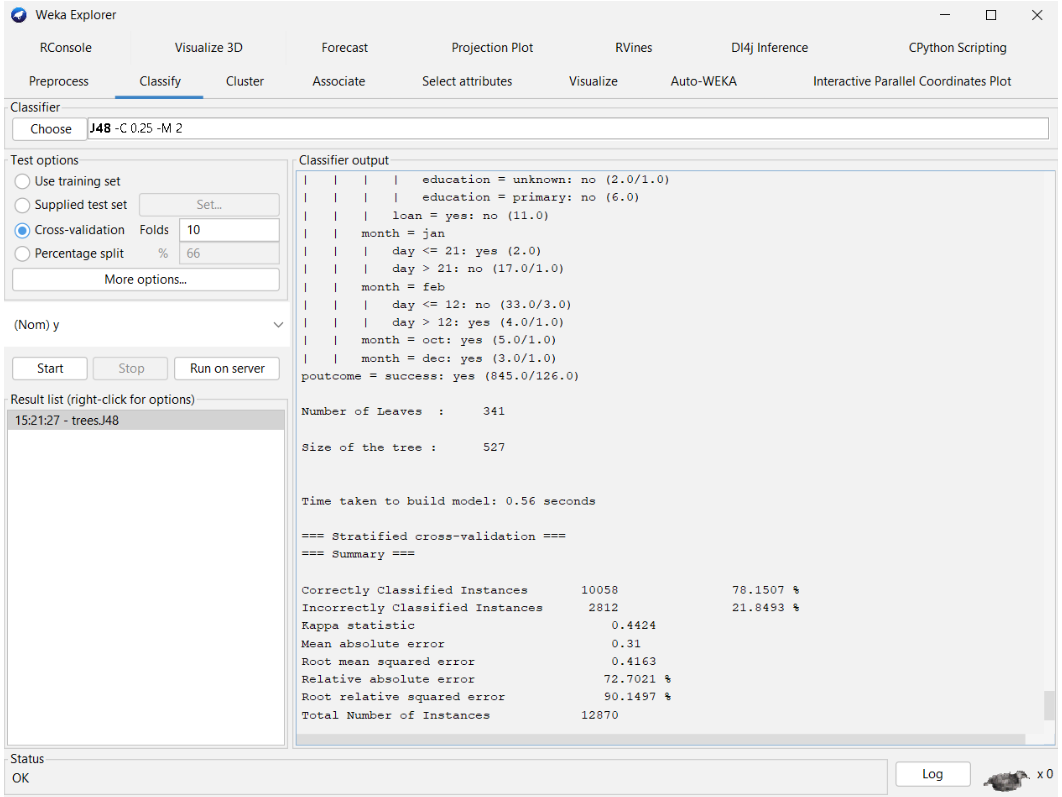
Classifier output

```
|   |   |   education = secondary
|   |   |   |   pdays <= 89: yes (2.0)
|   |   |   |   pdays > 89: no (3.0)
|   |   |   education = unknown: yes (0.0)
|   |   |   education = primary: yes (0.0)
|   |   month = aug
|   |   |   pdays <= 102: no (2.0)
|   |   |   pdays > 102: yes (8.0)
|   |   month = jul: yes (5.0/2.0)
|   |   month = sep
|   |   |   contact = cellular
|   |   |   |   marital = single: yes (3.0)
|   |   |   |   marital = married: no (2.0)
|   |   |   |   marital = divorced: yes (0.0)
|   |   |   contact = unknown: no (3.0)
|   |   |   contact = telephone: yes (1.0)
|   |   month = nov
|   |   |   pdays <= 192: no (35.0/3.0)
|   |   |   pdays > 192: yes (5.0/1.0)
|   |   month = mar: no (3.0)
|   |   month = apr
|   |   |   loan = no
|   |   |   |   education = tertiary
|   |   |   |   |   previous <= 4
|   |   |   |   |   |   previous <= 1: no (3.0/1.0)
|   |   |   |   |   |   previous > 1: yes (8.0/1.0)
|   |   |   |   |   previous > 4: no (2.0)
|   |   |   |   education = secondary
|   |   |   |   |   pdays <= 342: no (20.0/4.0)
|   |   |   |   |   pdays > 342: yes (2.0)
|   |   |   |   education = unknown: no (2.0/1.0)
```

Status

OK

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options
- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) y

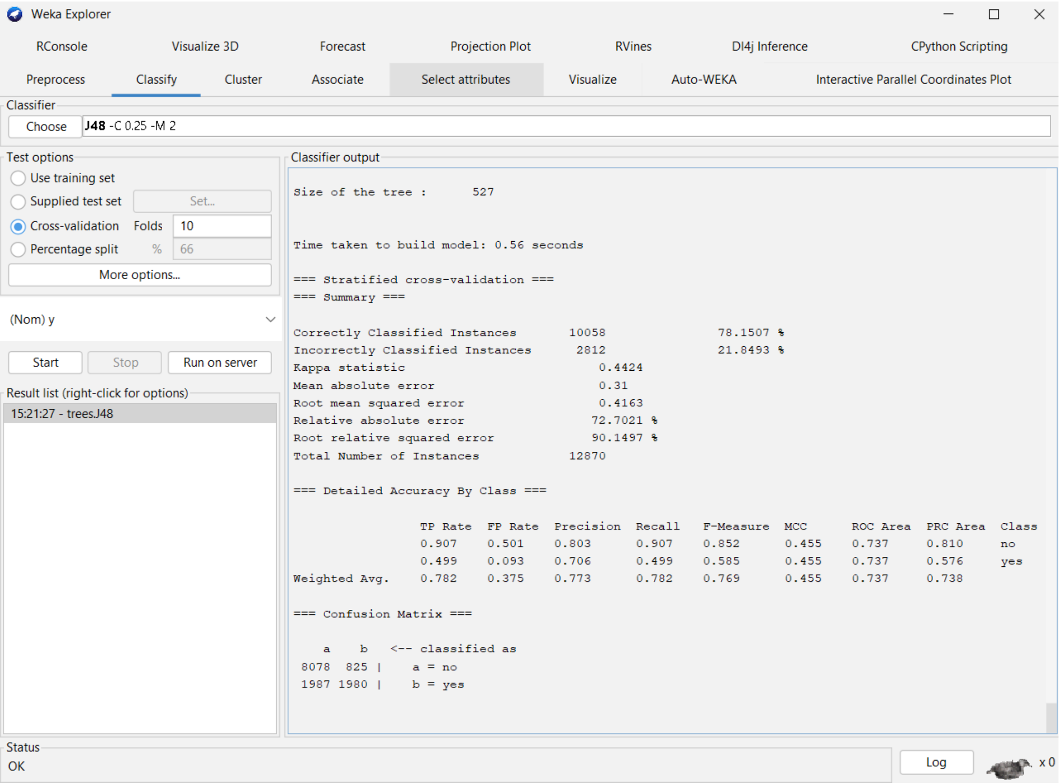
Result list (right-click for options): 15:21:27 - trees.J48

Status: OK

Classifier output

```
|   |   |   |   education = unknown: no (2.0/1.0)
|   |   |   |   education = primary: no (6.0)
|   |   |   loan = yes: no (11.0)
|   |   month = jan
|   |   |   day <= 21: yes (2.0)
|   |   |   day > 21: no (17.0/1.0)
|   |   month = feb
|   |   |   day <= 12: no (33.0/3.0)
|   |   |   day > 12: yes (4.0/1.0)
|   |   month = oct: yes (5.0/1.0)
|   |   month = dec: yes (3.0/1.0)
poutcome = success: yes (845.0/126.0)

Number of Leaves  :     341

Size of the tree :      527


Time taken to build model: 0.56 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       10058               78.1507 %
Incorrectly Classified Instances      2812               21.8493 %
Kappa statistic                          0.4424
Mean absolute error                      0.31
Root mean squared error                  0.4163
Relative absolute error                 72.7021 %
Root relative squared error             90.1497 %
Total Number of Instances            12870
```

Figure: Decision Tree Algorithm



Classifier: J48 -C 0.25 -M 2

Test options
- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) y

Result list (right-click for options): 15:21:27 - trees.J48

Status: OK

Classifier output

```
Size of the tree :      527


Time taken to build model: 0.56 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       10058               78.1507 %
Incorrectly Classified Instances      2812               21.8493 %
Kappa statistic                          0.4424
Mean absolute error                      0.31
Root mean squared error                  0.4163
Relative absolute error                 72.7021 %
Root relative squared error             90.1497 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.907    0.501    0.803      0.907    0.852      0.455    0.737     0.810     no
                 0.499    0.093    0.706      0.499    0.585      0.455    0.737     0.576     yes
Weighted Avg.    0.782    0.375    0.773      0.782    0.769      0.455    0.737     0.738

=== Confusion Matrix ===

    a    b   <-- classified as
 8078  825 |    a = no
 1987 1980 |    b = yes
```

Figure: Decision Tree Algorithm

2. Then trees.J48 was left click and Visualize tree was selected.

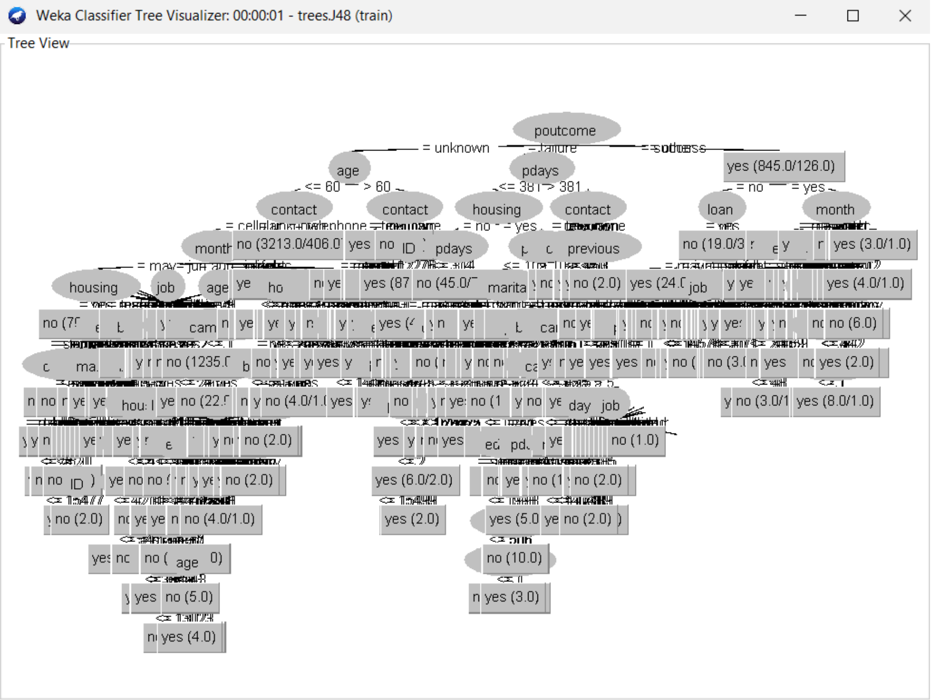
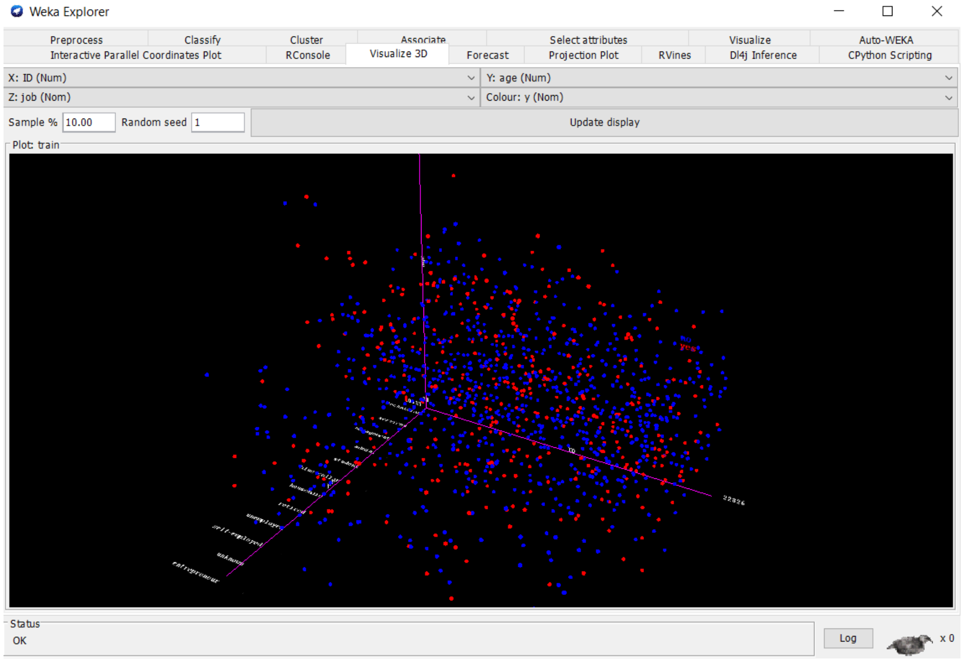


Figure: Decision Tree

Visualization Plotting:

1. Firstly, visualization 3D was selected, 'Sample %' was set to10.00% then update display was button was pressed.



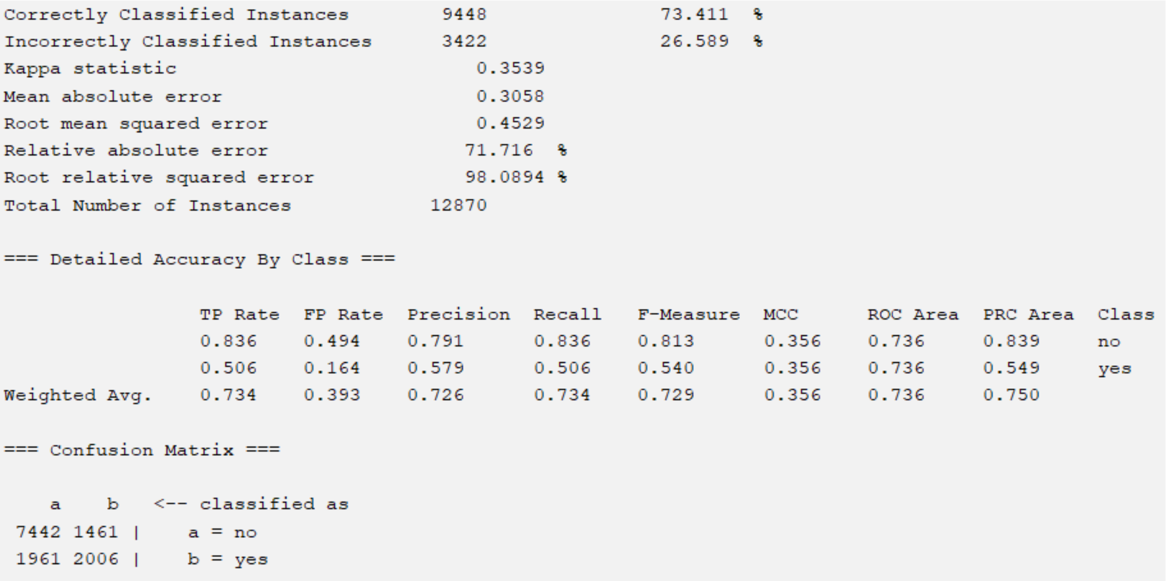
X: ID (Num)
Y: age (Num)
Z: job (Nom)
Colour: y (Nom)
Sample % 10.00 Random seed 1
Status
OK

Figure: Weka Explorer Naïve bayes classification

Conclusion & Discussion

Naïve Bayes classifier:

There are 12870 instances in the Naive Bayes model. In this case, there are 9448 correctly classified instances and 3422 incorrectly classified instances.

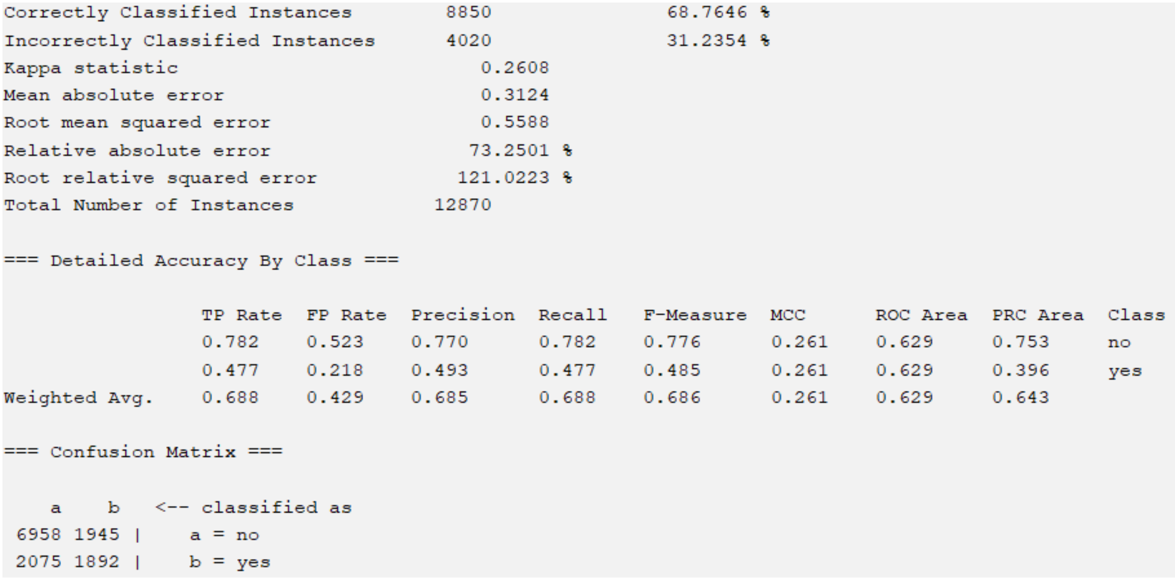


```
Correctly Classified Instances        9448               73.411  %
Incorrectly Classified Instances      3422               26.589  %
Kappa statistic                          0.3539
Mean absolute error                      0.3058
Root mean squared error                  0.4529
Relative absolute error                 71.716  %
Root relative squared error             98.0894 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.836    0.494    0.791      0.836    0.813      0.356    0.736     0.839     no
                 0.506    0.164    0.579      0.506    0.540      0.356    0.736     0.549     yes
Weighted Avg.    0.734    0.393    0.726      0.734    0.729      0.356    0.736     0.750

=== Confusion Matrix ===

    a    b   <-- classified as
 7442 1461 |    a = no
 1961 2006 |    b = yes
```

Figure: Naive Bayes model accuracy with confusion matrix

K-Nearest Neighbour Classification:

In the KNN model, there are 12970 occurrences. In this situation, there are 8850 cases that are correctly classified and 4020 instances that are incorrectly classified.

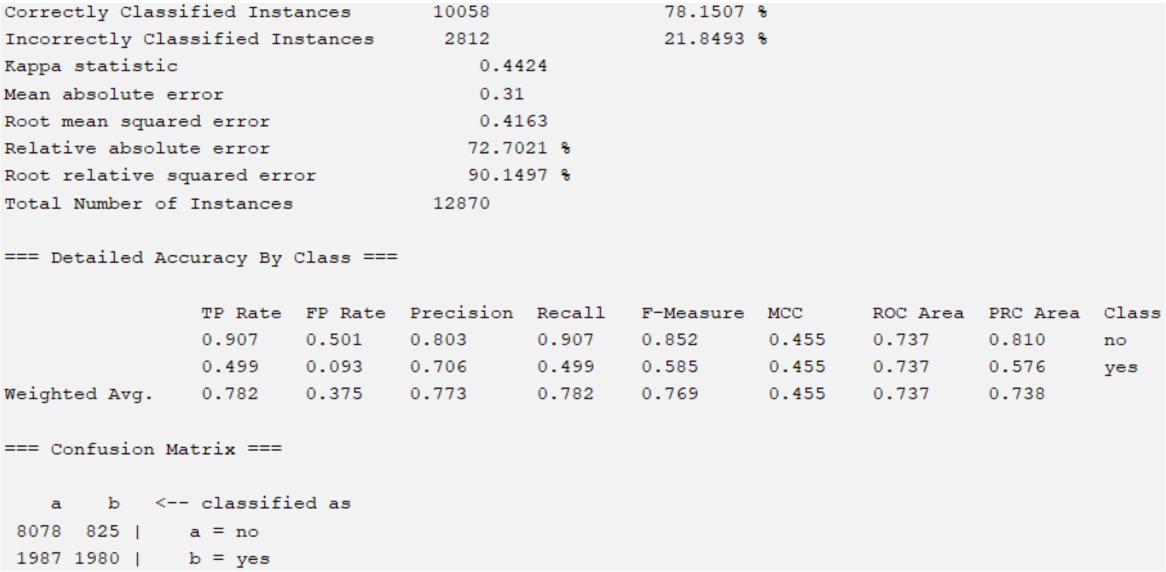


```
Correctly Classified Instances        8850               68.7646 %
Incorrectly Classified Instances      4020               31.2354 %
Kappa statistic                          0.2608
Mean absolute error                      0.3124
Root mean squared error                  0.5588
Relative absolute error                 73.2501 %
Root relative squared error            121.0223 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.782    0.523    0.770      0.782    0.776      0.261    0.629     0.753     no
                 0.477    0.218    0.493      0.477    0.485      0.261    0.629     0.396     yes
Weighted Avg.    0.688    0.429    0.685      0.688    0.686      0.261    0.629     0.643

=== Confusion Matrix ===

    a    b   <-- classified as
 6958 1945 |    a = no
 2075 1892 |    b = yes
```

Figure: Naive KNN accuracy with confusion matrix

Decision Trees:

There are 12870 instances in the Decision Tree. There are 10058 instances that are correctly classified and 2812 instances that are incorrectly classified in this circumstance.



```
Correctly Classified Instances       10058               78.1507 %
Incorrectly Classified Instances      2812               21.8493 %
Kappa statistic                          0.4424
Mean absolute error                      0.31
Root mean squared error                  0.4163
Relative absolute error                 72.7021 %
Root relative squared error             90.1497 %
Total Number of Instances            12870

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.907    0.501    0.803      0.907    0.852      0.455    0.737     0.810     no
                 0.499    0.093    0.706      0.499    0.585      0.455    0.737     0.576     yes
Weighted Avg.    0.782    0.375    0.773      0.782    0.769      0.455    0.737     0.738

=== Confusion Matrix ===

    a    b   <-- classified as
 8078  825 |    a = no
 1987 1980 |    b = yes
```

Figure: Decision Tree model accuracy with confusion matrix

As we can see, the percentage of cases of Naive Bayes that are successfully categorized is 73.41%, the percentage of instances of KNN that are correctly classified is 68.76%, and the percentage of instances of a Decision tree that is correctly classified is 78.15%. We may infer that the Decision Tree performs better in this dataset since it correctly classifies more instances than Naive Bayes and KNN.

Here the purpose of this report was to identify an appropriate classifier for bank marketing response prediction that will be able to categorize the bank marketing response as correctly as possible and be able to forecast the class from the test data set. Following the application of three different classifiers—KNN, naive Bayes, and decision tree the dataset's best classifier with a 78.15% accuracy rate is the decision tree classifier. Then, a training set was taken from the original dataset to create a machine learning model. The model was tested using a prepared test dataset, and the results showed that the model's accuracy was 73.83% for the prepared test dataset. Creating training and the testing dataset is an important concept in data science as it is used to improve generalization and minimize overfitting. This also helps to give an unbiased evaluation of the accuracy of the model itself.