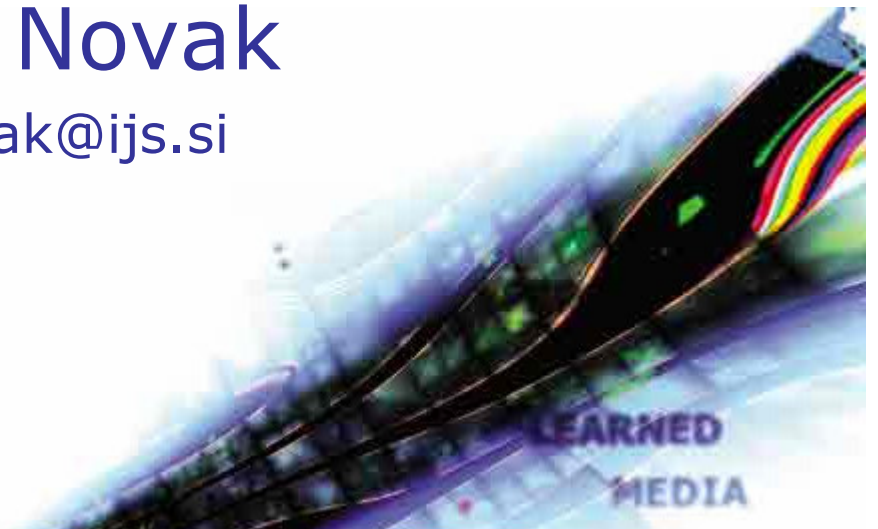


Numeric prediction in Weka

24.11.2009

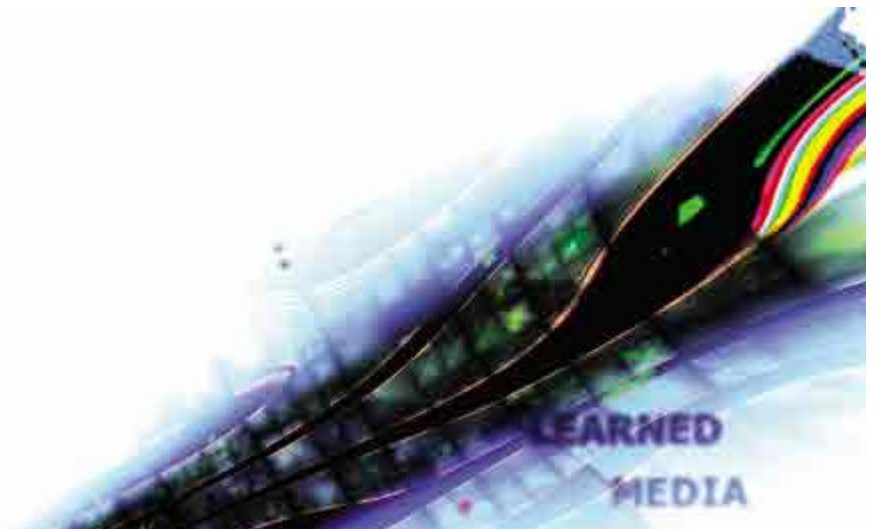
Petra Kralj Novak

Petra.Kralj.Novak@ijs.si



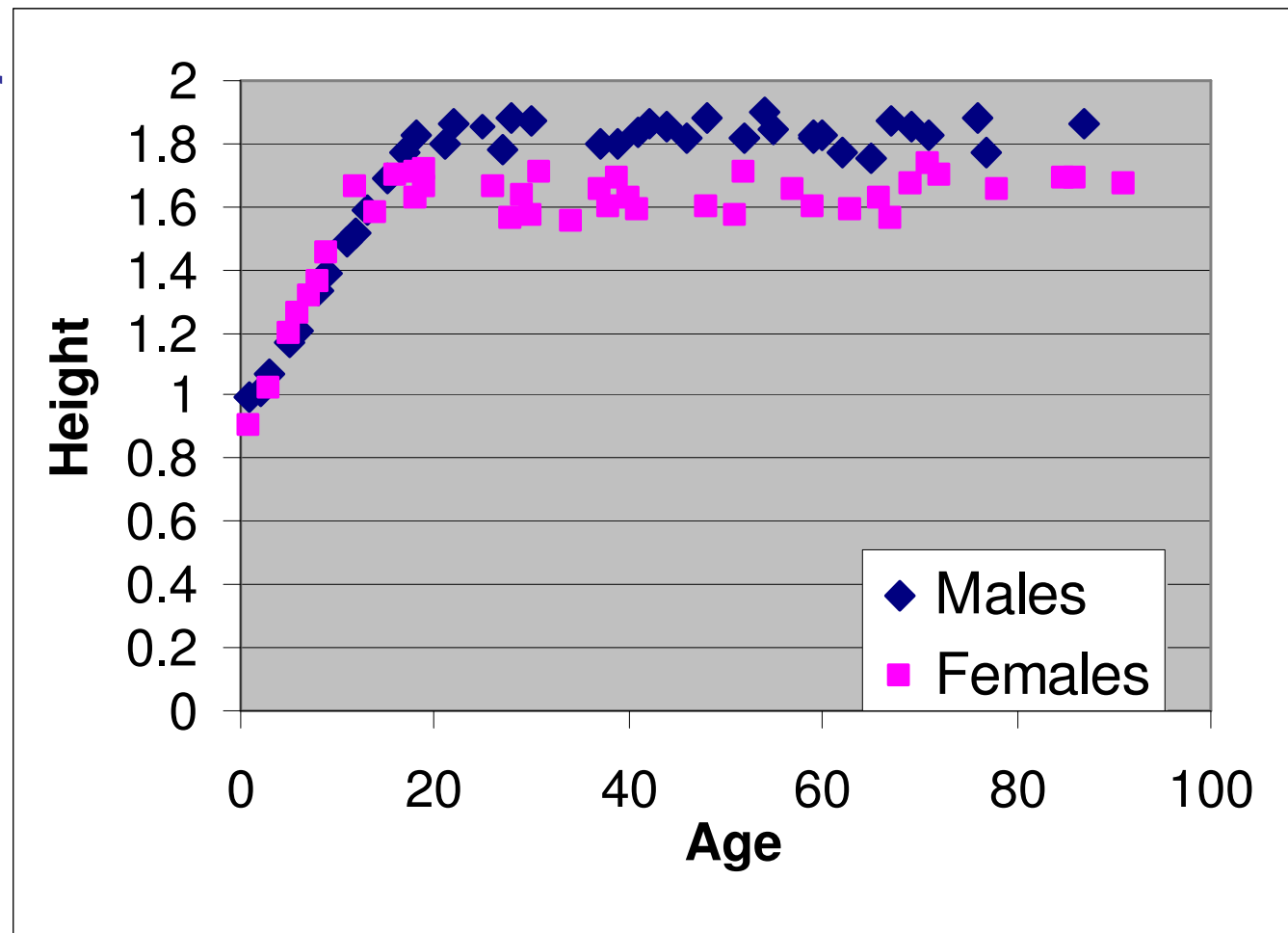
Numeric prediction models

- LinearRegression
- M5P Regression and model trees
- KNN
- Baseline predictor



regressionAgeHeight.csv

- Data about 80 people:
- Age
- Gender
- Height



Filename:

The screenshot shows the Weka Explore application window. At the top, there are tabs for 'Preprocess', 'Classify', 'Cluster', 'Associate', 'Select attributes', and 'Visualize'. Below these are buttons for 'Open file...', 'Open URL...', 'Open DB...', 'Undo', 'Edit...', and 'Save...'. A 'Filter' section contains a 'Choose' button, a dropdown menu set to 'None', and an 'Apply' button. The 'Current relation' section displays 'Relation: regressionAheHeight', 'Instances: 80', and 'Attributes: 3'. The 'Attributes' section has 'All', 'None', and 'Invert' buttons, and a table with columns 'No.' and 'Name'. The table lists three attributes: 'Age', 'Gender', and 'Height', with 'Height' selected. A 'Remove' button is at the bottom of this section. The 'Selected attribute' section shows 'Name: Height', 'Missing: 0 (0%)', 'Distinct: 80', and 'Type: Numeric Unique: 80 (100%)'. Below this is a table of statistics for the selected attribute.

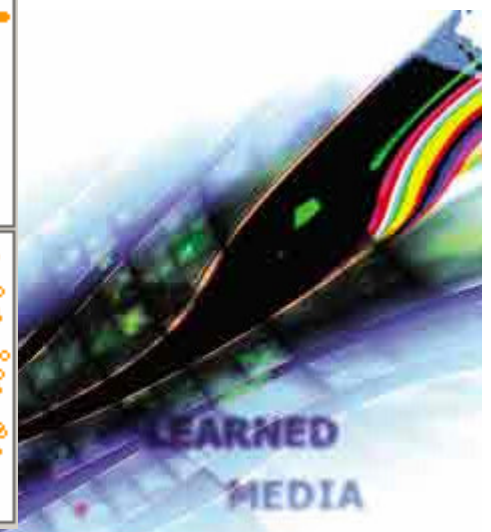
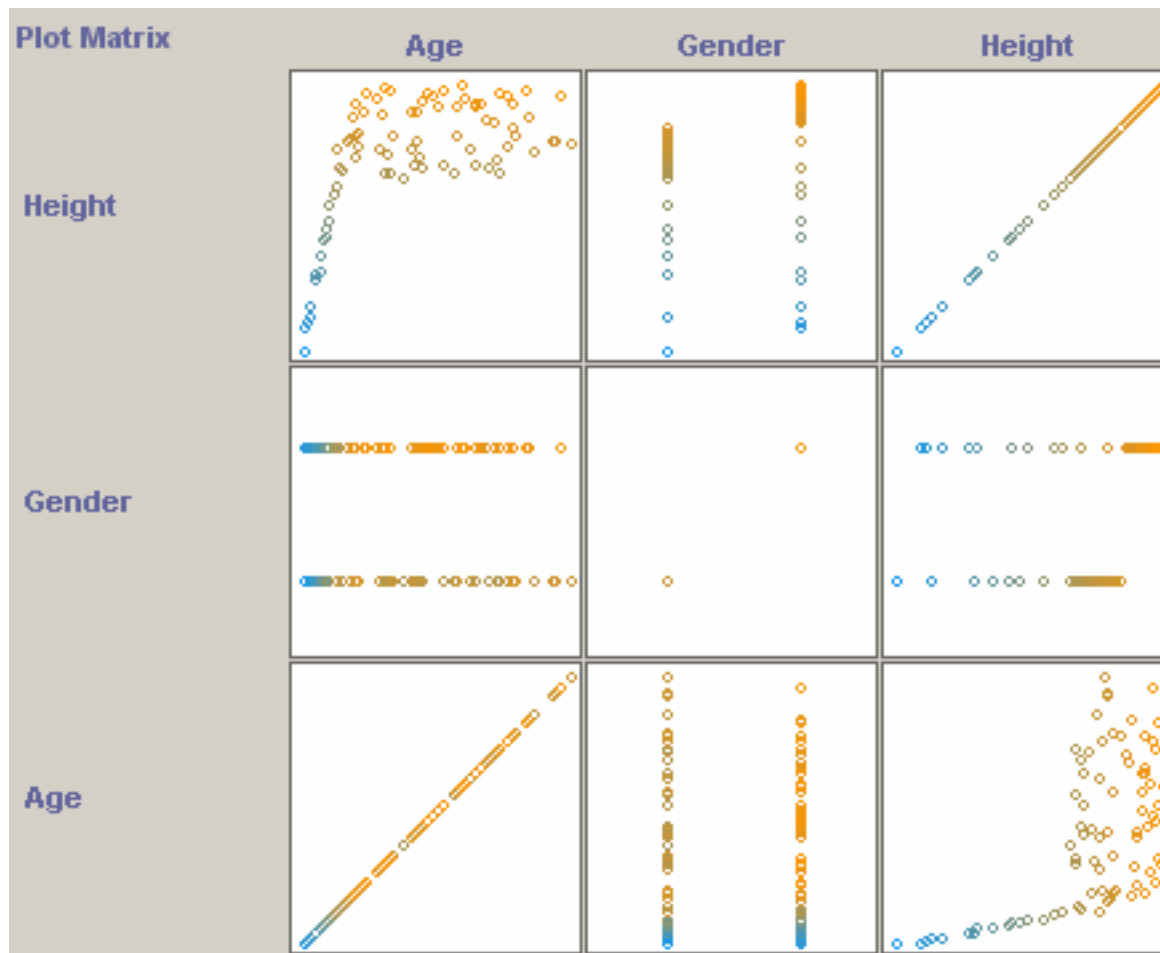
Statistic	Value
Minimum	0.902
Maximum	1.895
Mean	1.628
StdDev	0.236

The 'Class: Height (Num)' section features a dropdown menu and a 'Visualize All' button. Below is a histogram showing the distribution of the 'Height' attribute. The x-axis ranges from 0.9 to 1.9, and the y-axis shows frequencies of 5, 4, 6, 30, and 35 for different bins.

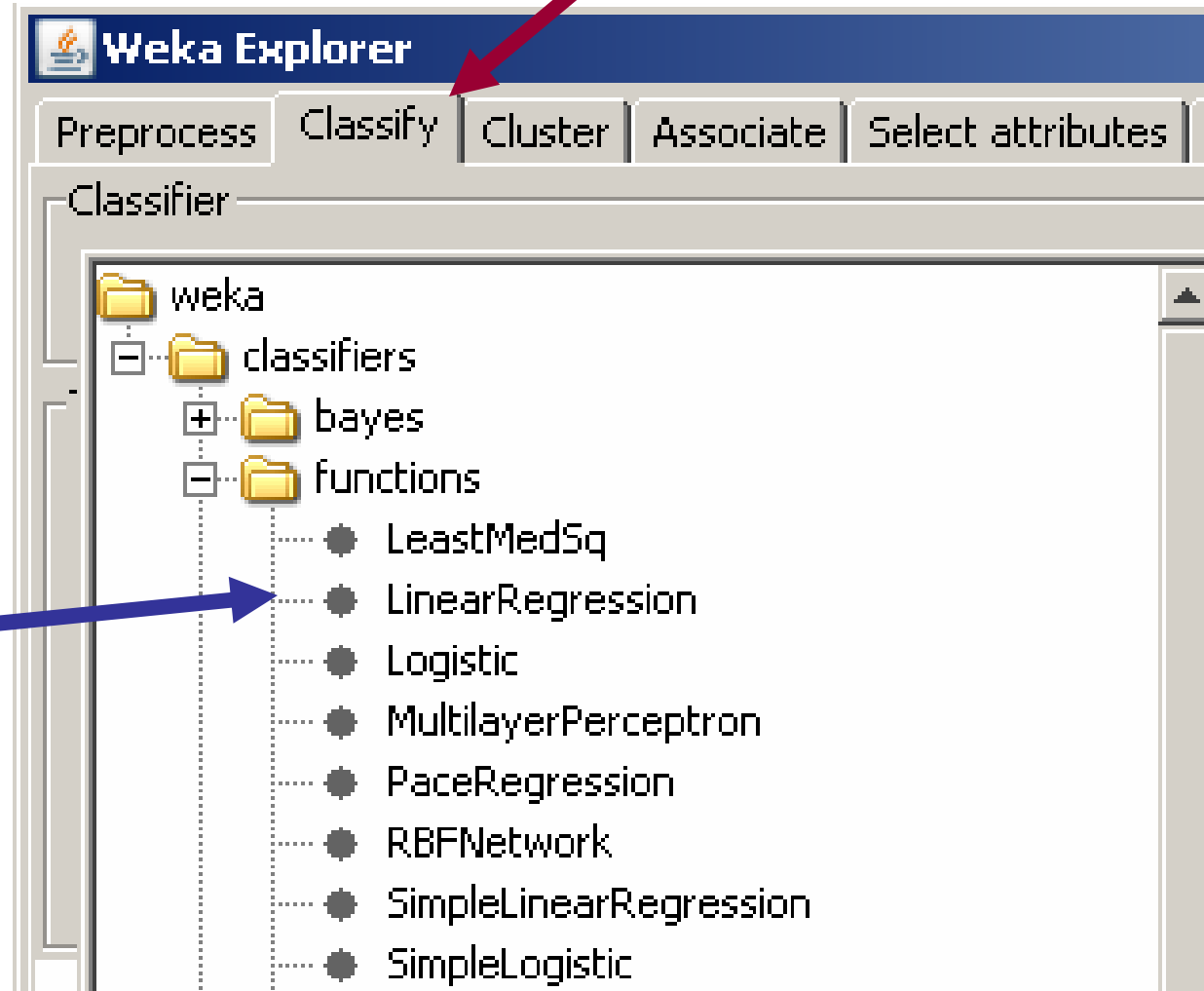
Status: OK

Log x 0

Visualization in Weka



Weka → classifiers → functions → LinearRegression



Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier: Choose **MSP -M 4.0**

Test options:

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Num) Height

Result list (right-click for options):

- 15:24:52 - trees.M5P
- 15:58:13 - functions.LinearRegression**
- 16:03:32 - functions.LinearRegression

Classifier output:

```

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

Linear Regression Model


Height =

      0.0056 * Age +
      0.1292 * Gender=M +
      1.3506

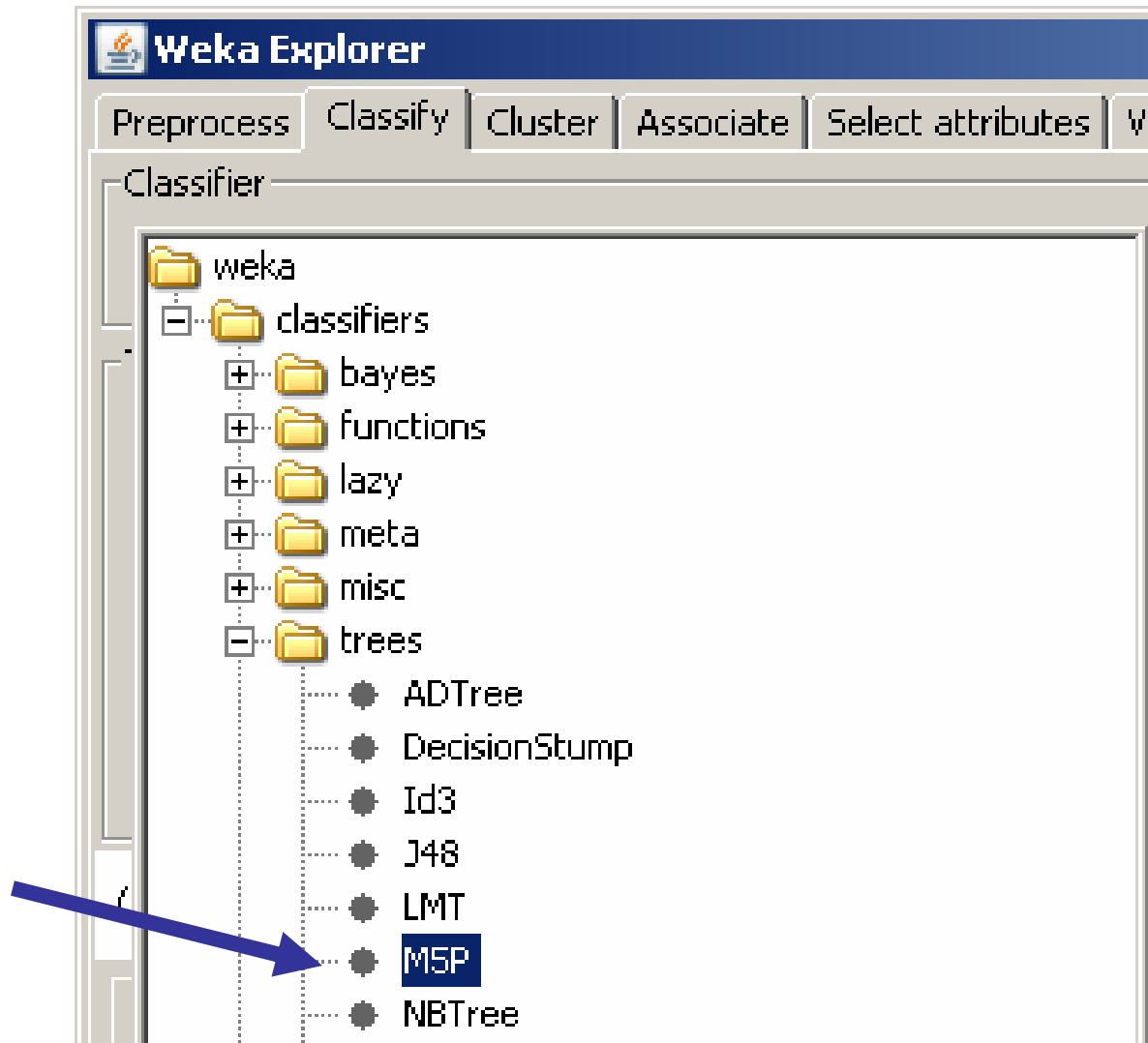
Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

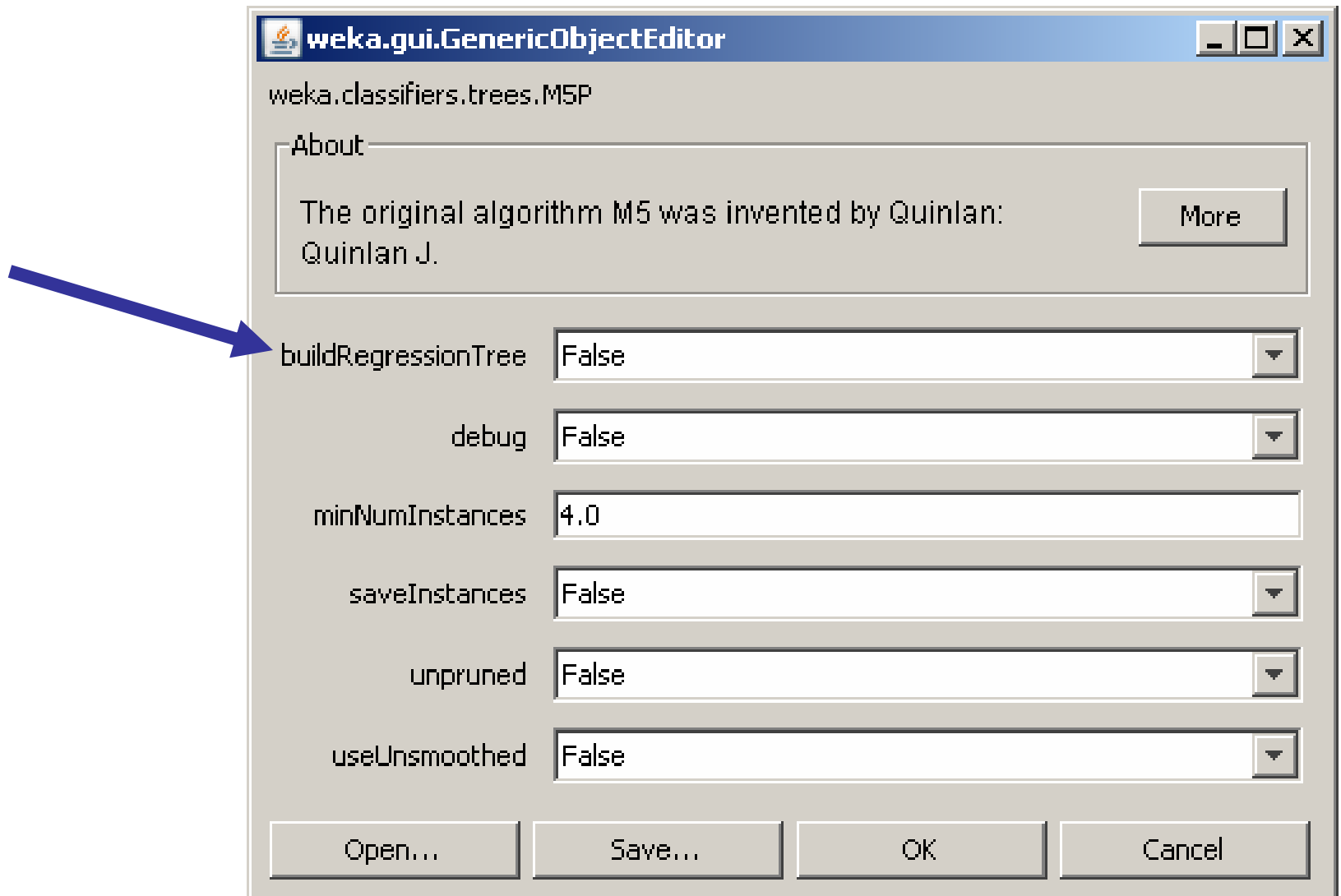
Correlation coefficient          0.6204
Mean absolute error             0.142
Root mean squared error        0.1844
Relative absolute error        80.1623 %
Root relative squared error    77.2023 %
Total Number of Instances      80
  
```

Status: OK  x 0

Weka → classifiers → trees → M5P



builtRegressionTree = True → regression tree
builtRegressionTree = False → model tree



kNN: weka → classifiers → lazy
→ IBk

The image shows two windows from the Weka software. The left window, titled 'Weka Explorer', displays a tree view of the classifier hierarchy. A blue arrow points to the 'IBk' classifier under the 'lazy' folder. The right window, titled 'weka.gui.GenericObjectEditor', shows the configuration for the 'weka.classifiers.lazy.IBk' classifier. A blue arrow points to the 'KNN' field, which is set to '3'. Other configuration options include 'crossValidate' (False), 'debug' (False), 'distanceWeighting' (No distance weighting), 'meanSquared' (False), 'noNormalization' (False), and 'windowSize' (0). Buttons for 'Open...', 'Save...', 'OK', and 'Cancel' are visible at the bottom.

Property	Value
KNN	3
crossValidate	False
debug	False
distanceWeighting	No distance weighting
meanSquared	False
noNormalization	False
windowSize	0

Weka → classifiers → rules → ZeroR

