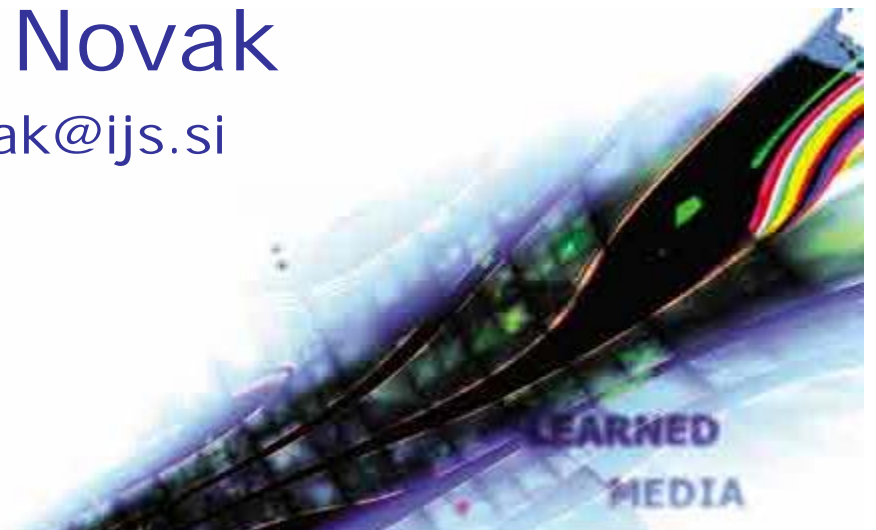


Numeric prediction in Weka

11.11.2008

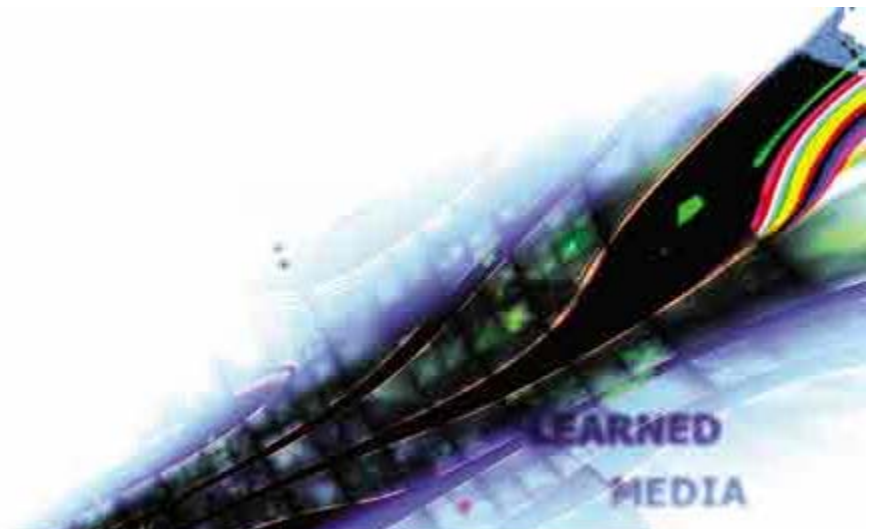
Petra Kralj Novak

Petra.Kralj.Novak@ijs.si



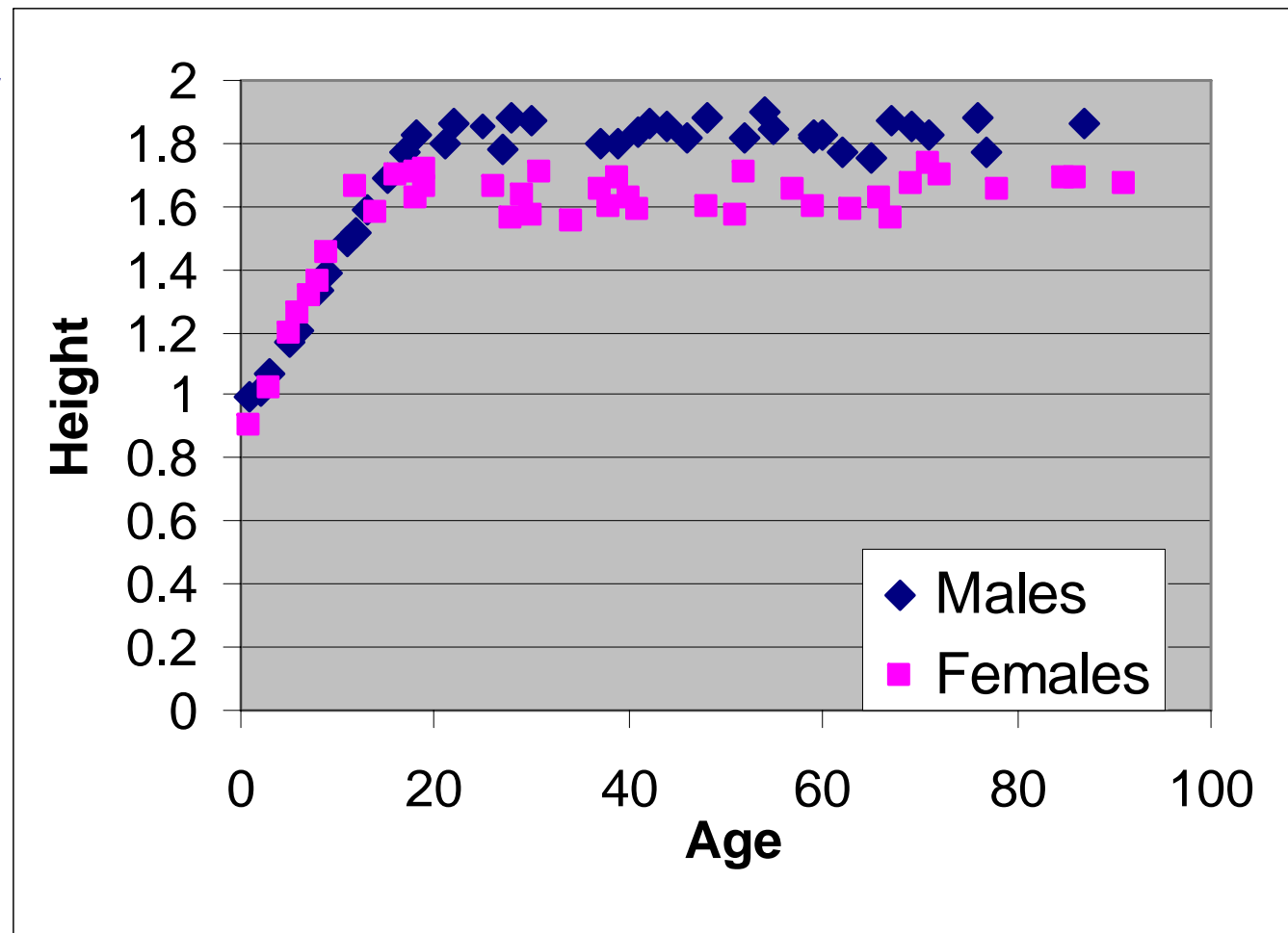
Numeric prediction models

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



regressionAgeHeight.csv

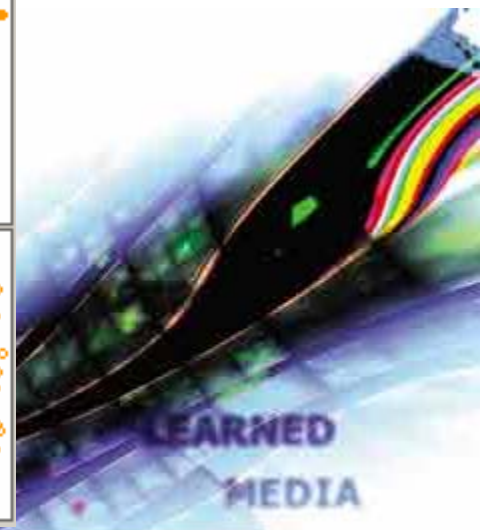
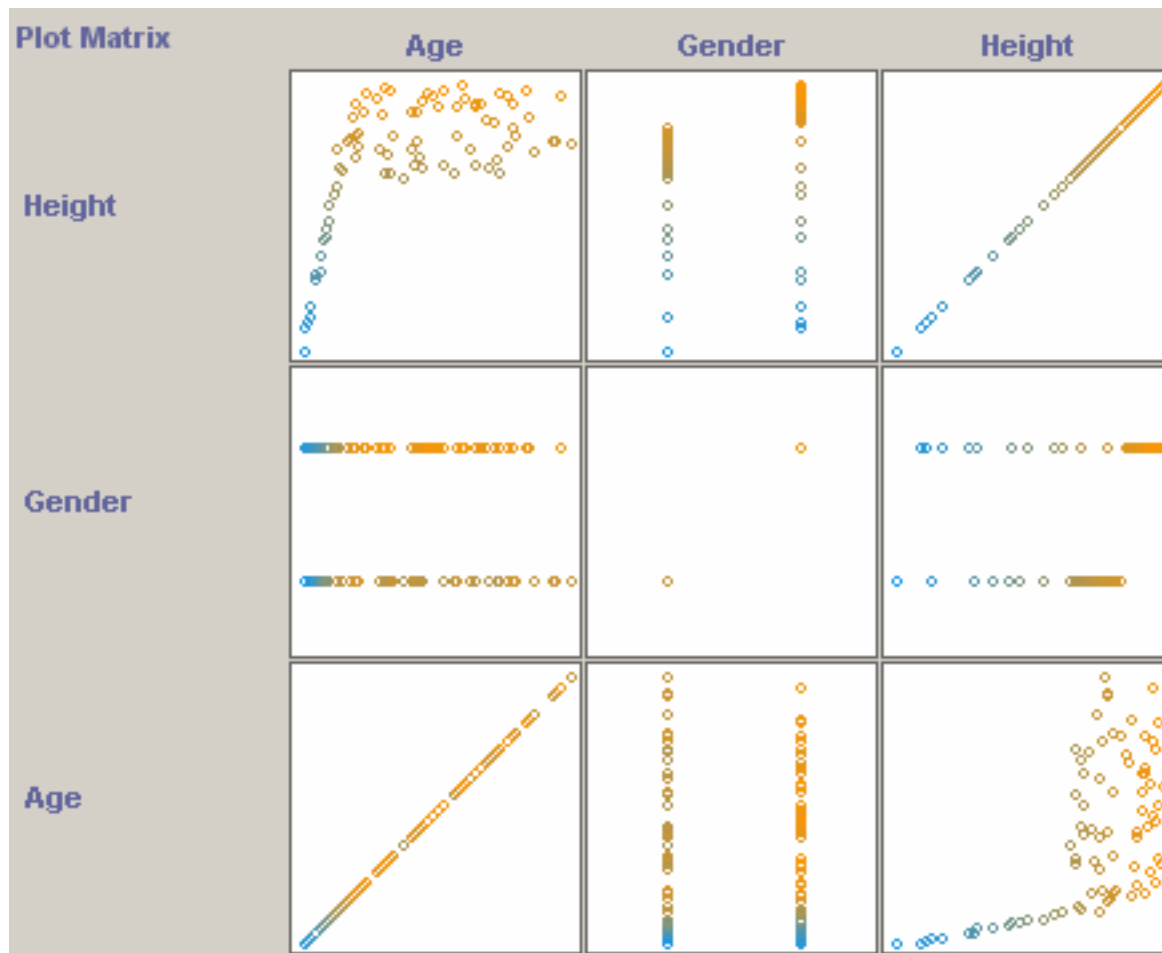
- Data about 80 people:
- Age
- Gender
- Height



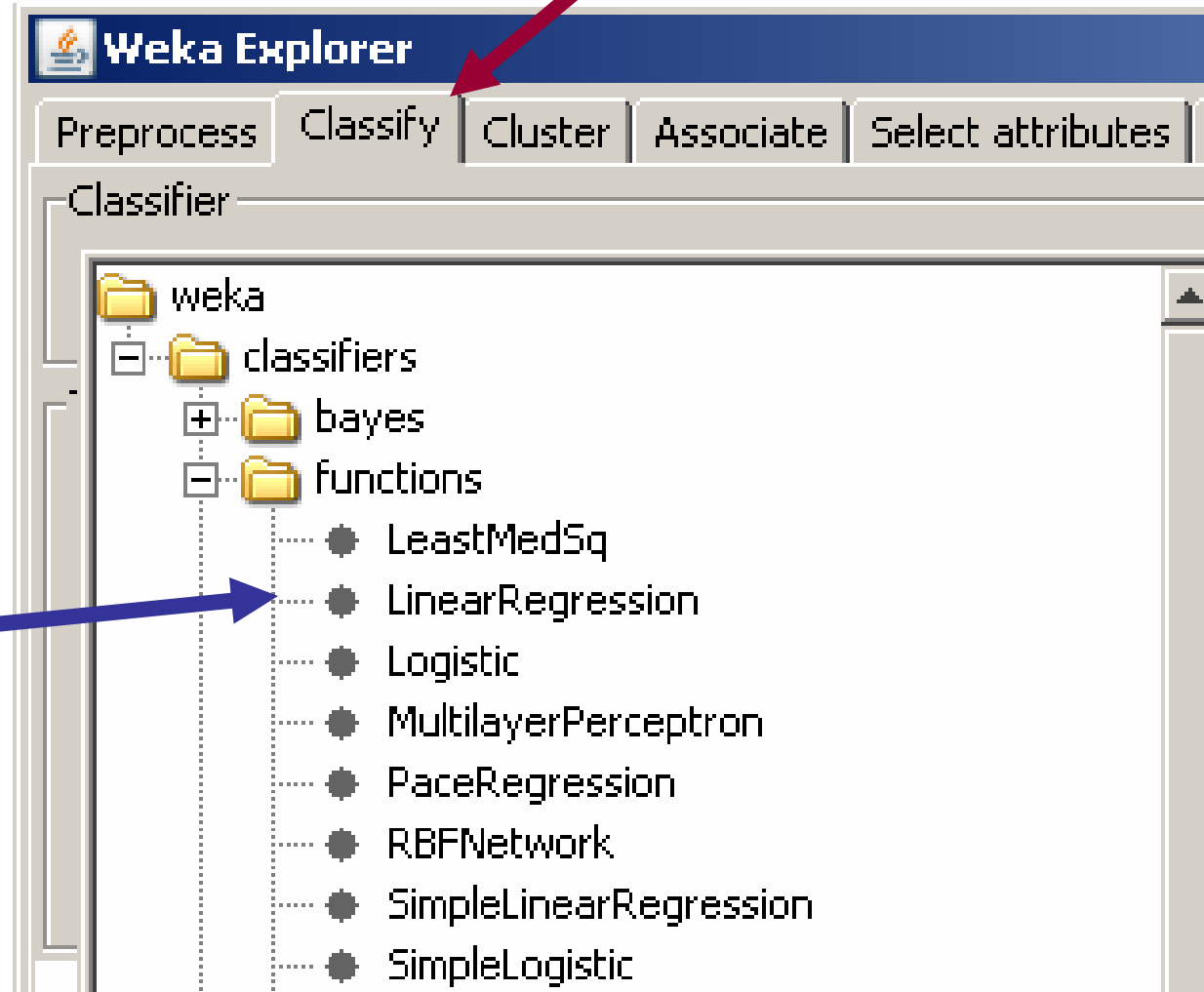
Filename:

The screenshot shows the Weka Explorer application window. At the top, there are menu tabs: 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' and 'Instances: 80' with 'Attributes: 3'. The 'Attributes' section has 'All', 'None', and 'Invert' buttons, and a table listing attributes: Age, Gender, and Height (selected). A 'Remove' button is at the bottom of this section. The 'Selected attribute' section shows 'Name: Height', 'Type: Numeric', 'Missing: 0 (0%)', 'Distinct: 80', and 'Unique: 80 (100%)'. It includes a table of statistics: Minimum (0.902), Maximum (1.895), Mean (1.628), and StdDev (0.236). Below this is a dropdown for 'Class: Height (Num)' and a 'Visualize All' button. A histogram shows the distribution of 'Height' values with bars at 0.9 (count 5), 1.4 (count 4), and 1.9 (count 35). The status bar at the bottom shows 'Status OK', a 'Log' button, and a small icon with '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

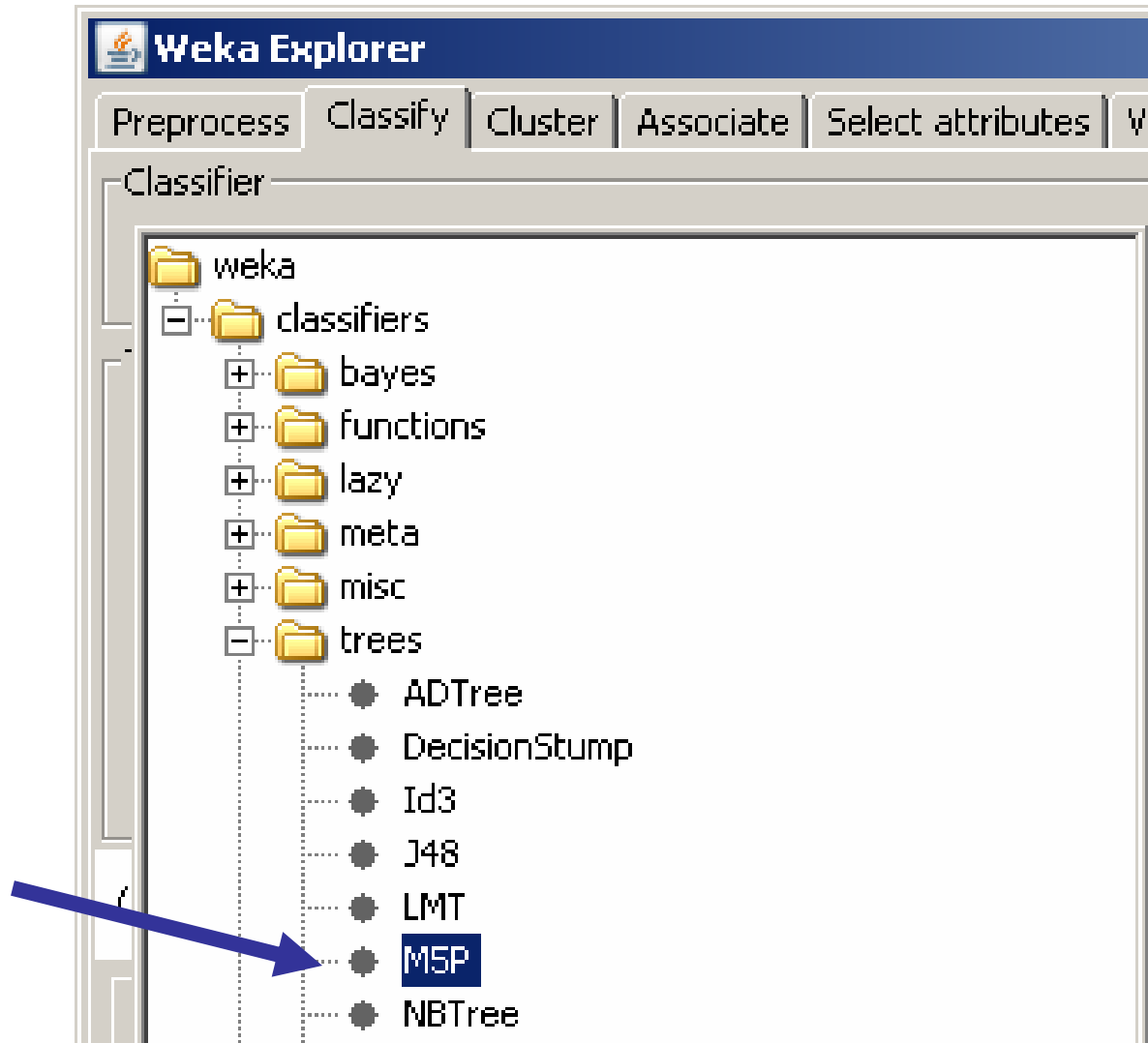
Time taken to build model: 0 seconds

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

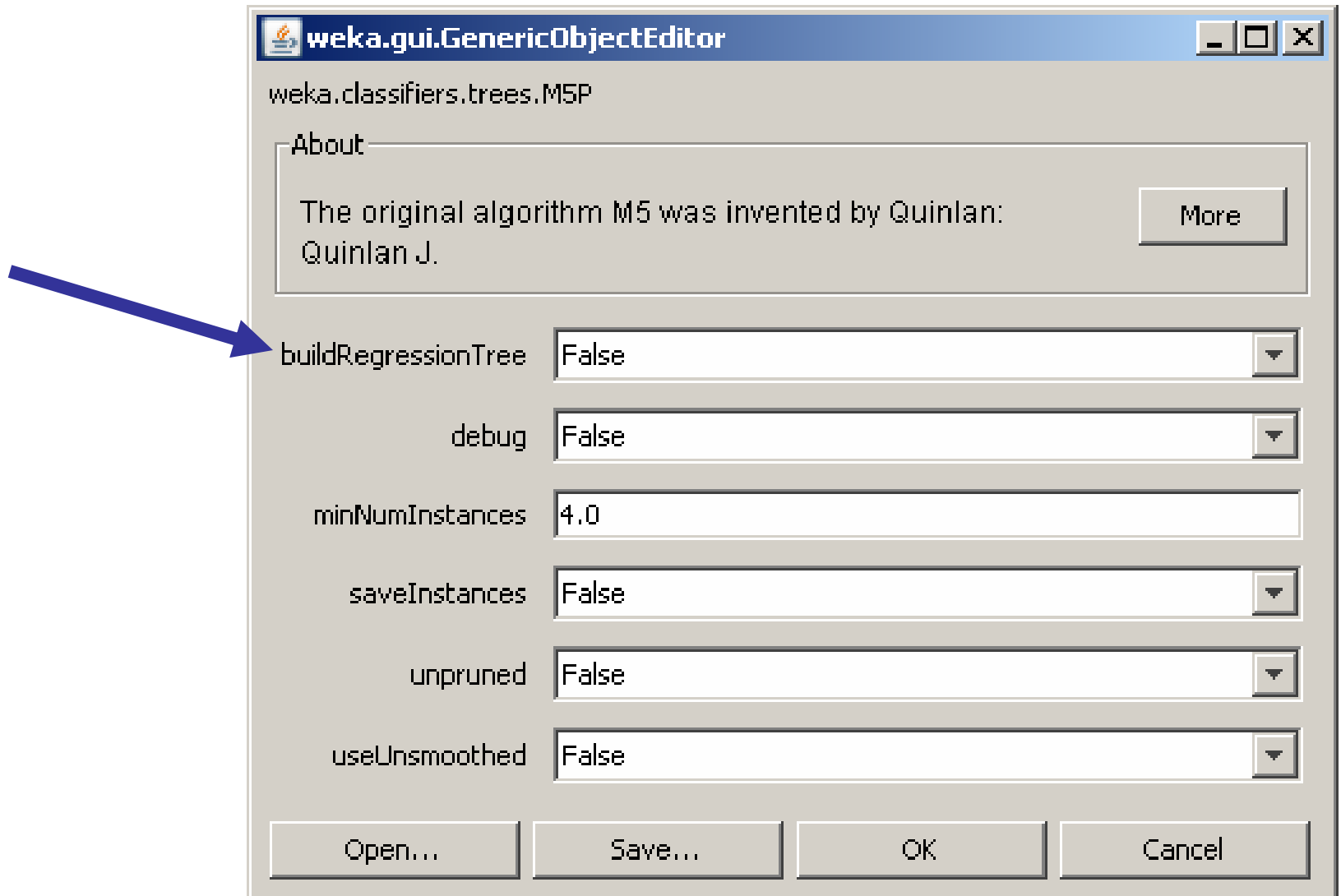
Correlation coefficient           0.6204
Mean absolute error              0.142
Root mean squared error         0.1844
Relative absolute error        80.1623 %
Root relative squared error    77.2023 %
Total Number of Instances      80
  
```

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 overlapping 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 selected 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.

Weka Explorer

Preprocess Classify Cluster

Classifier

- weka
 - classifiers
 - bayes
 - functions
 - lazy
 - IB1
 - IBk**
 - KStar
 - LBR
 - LWL
 - meta
 - misc
 - trees
 - rules

weka.gui.GenericObjectEditor

weka.classifiers.lazy.IBk

About

K-nearest neighbours classifier. More

KNN 3

crossValidate False

debug False

distanceWeighting No distance weighting

meanSquared False

noNormalization False

windowSize 0

Open... Save... OK Cancel

DEPARTMENT OF KNOWLEDGE TECHNOLOGIES
Jozef Stefan Institute

Weka → classifiers → rules → ZeroR

