

Data Mining using UI-Based Tools (WEKA, KNIME, Orange)

Tools Used

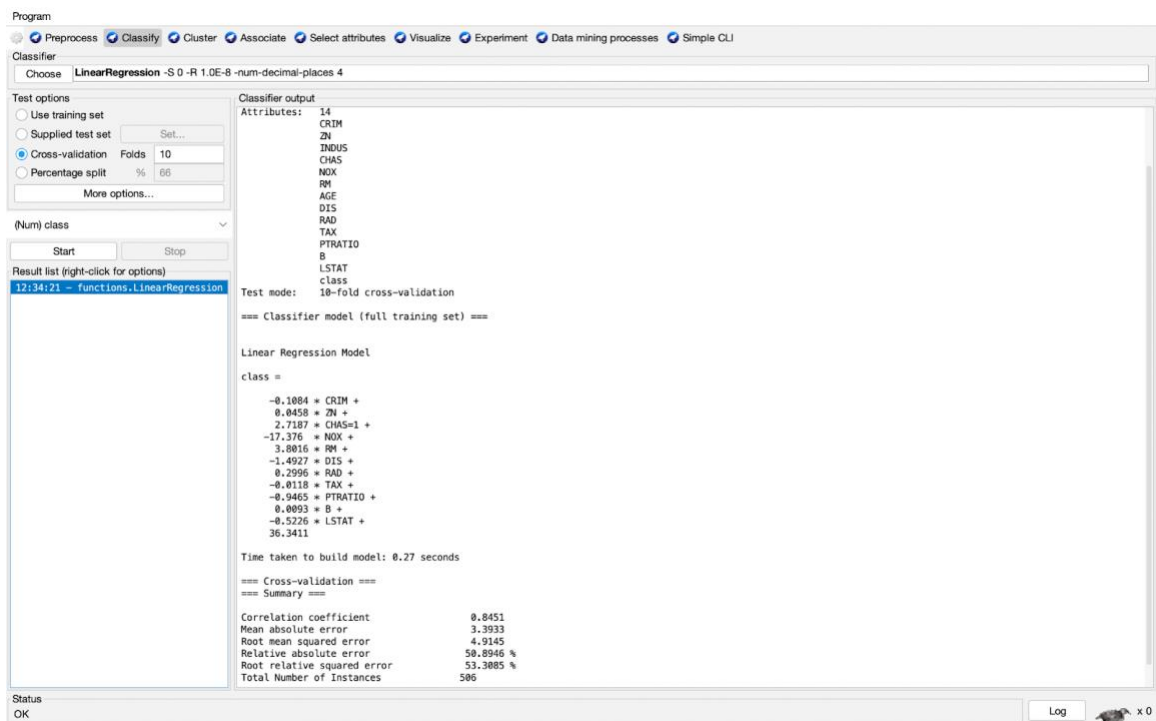
- WEKA – Linear Regression, Neural Network, Decision Tree
- KNIME – Regression and Clustering
- Orange – (Optional for visualization)

Q1. Linear Regression in WEKA

Built a linear regression equation with 11 terms.

Variables AGE and INDUS were omitted automatically by WEKA due to low predictive power.

Each term is a product of a coefficient and an independent explanatory variable.



Q2. Neural Network (Multilayer Perceptron in WEKA)

Tested multiple learning rate and momentum combinations.

Best performance achieved at learning rate = 0.1 and momentum = 0.1:

Learning Rate

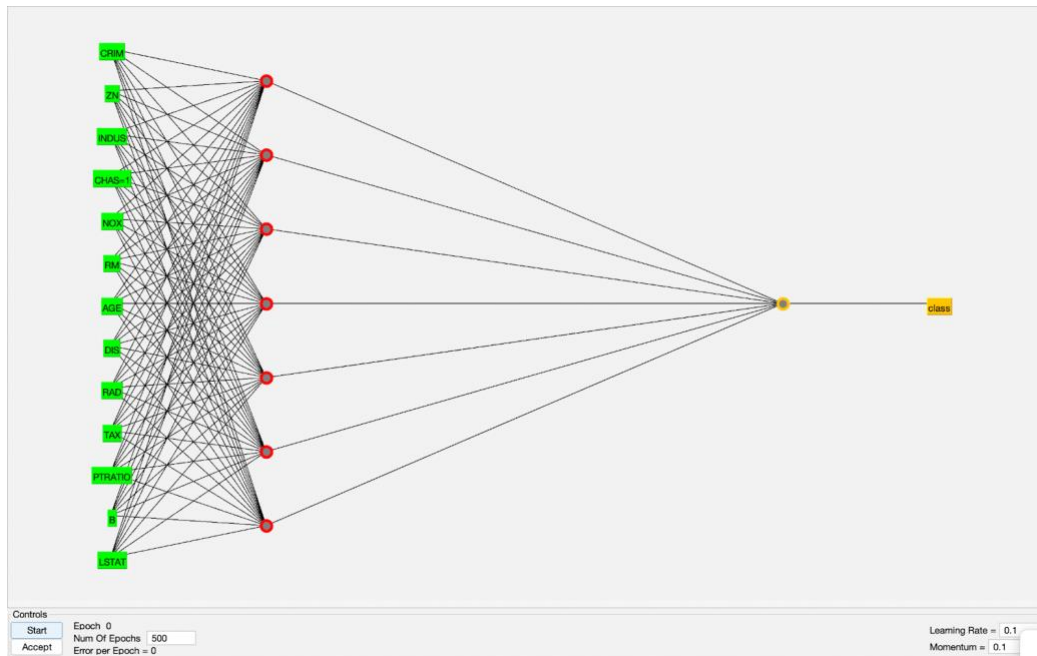
Momentum

RMSE

0.1

0.1

3.8654



Preprocess Classify Cluster Associate Select attributes Visualize

Classifier
Choose **MultilayerPerceptron -L 0.3 -M 0.3 -N 500 -V 0 -S 0 -E 20 -H a**

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
 More options...

(Num) class

Result list (right-click for options)
 14:42:36 - functions.LinearRegression
 14:57:29 - functions.MultilayerPerceptron
 14:58:06 - functions.MultilayerPerceptron
 14:58:23 - functions.MultilayerPerceptron
 14:58:46 - functions.MultilayerPerceptron
 14:59:06 - functions.MultilayerPerceptron
 14:59:23 - functions.MultilayerPerceptron
 14:59:38 - functions.MultilayerPerceptron
 14:59:53 - functions.MultilayerPerceptron
 15:00:06 - functions.MultilayerPerceptron

Classifier output

Inputs Weights
 Threshold -0.551932617735081
 Attrib CRIM 0.5551741116251884
 Attrib ZN 0.64698737681539414
 Attrib INDUS 0.27628778356948686
 Attrib CHAS-1 0.85994793358179698
 Attrib NOX -0.117538077439051
 Attrib RM -0.40787635806180696
 Attrib AGE -0.408529831291626
 Attrib DIS 0.25627422635383884
 Attrib RAD -0.09151853933382566
 Attrib TAX 0.022608143148336863
 Attrib PTRATIO -0.27592538912975264
 Attrib B -1.423398546528519
 Attrib LSTAT -0.25423852175891875

Sigmoid Node 7
 Inputs Weights
 Threshold 4.147811238849256
 Attrib CRIM 0.6378534437829111
 Attrib ZN -0.528857895873482
 Attrib INDUS -0.5231231525286125
 Attrib CHAS-1 -0.05784787265878687
 Attrib NOX 0.2821673142128951
 Attrib RM -3.3537611168595247
 Attrib AGE 0.0717791689627983
 Attrib DIS 0.726626148886923
 Attrib RAD -1.3815176456992388
 Attrib TAX -0.6087797185145497
 Attrib PTRATIO -0.0387874208948977278
 Attrib B -0.22664549958375805
 Attrib LSTAT 3.9395048891278877

Class
 Input
 Node 0

Time taken to build model: 0.63 seconds

=== Cross-validation ===
 === Summary ===
 Correlation coefficient 0.9897
 Mean absolute error 2.5655
 Root mean squared error 3.8654
 Relative absolute error 38.4792 %
 Root relative squared error 41.9288 %
 Total Number of Instances 506

Status
 OK

Log x 0

Q3. Linear Regression in KNIME

Constructed a regression model to predict `num_rings`.

Equation:

$$\begin{aligned} \text{num_rings} = & 3.8946 - 0.8249 \times (\text{sex} = \text{I}) + 0.0577 \times (\text{sex} = \text{M}) - 0.4583 \times \text{length} \\ & + 11.0751 \times \text{diameter} + 10.7615 \times \text{height} + 8.9754 \times \text{whole_weight} \\ & - 19.7869 \times \text{shucked_weight} - 10.5818 \times \text{viscera_weight} + 8.7418 \times \text{shell_weight} \end{aligned}$$

File

Statistics on Linear Regression

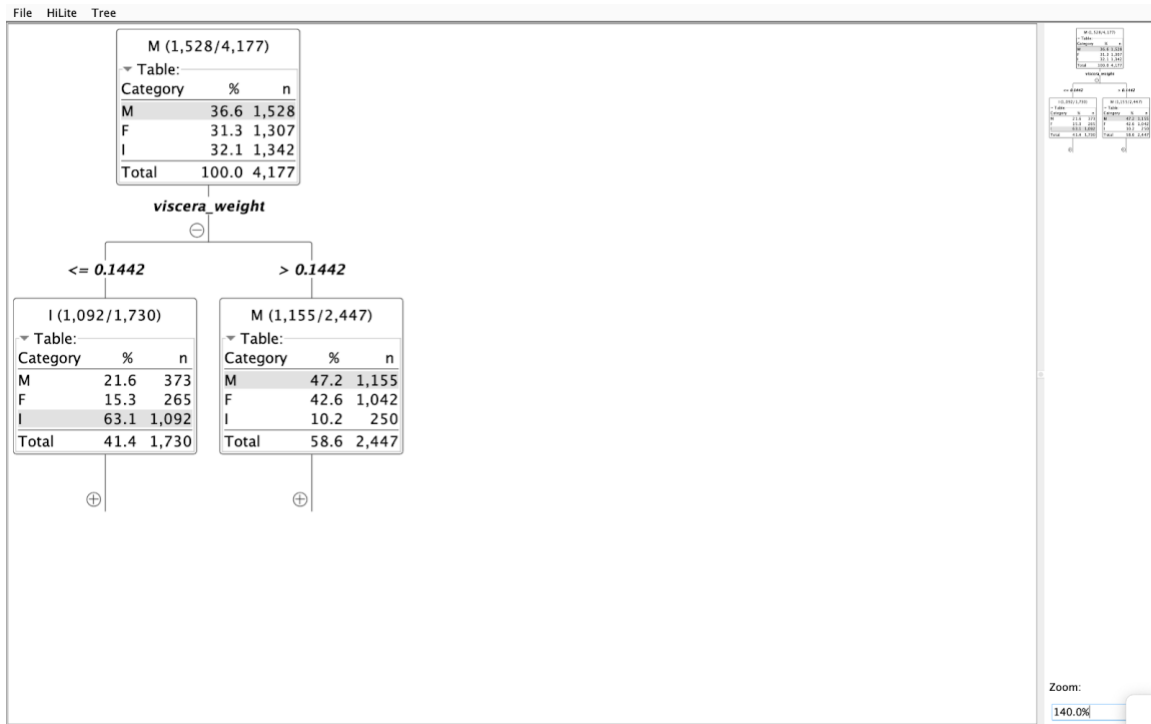
Variable	Coeff.	Std. Err.	t-value	P> t
sex=I	-0.8249	0.1024	-8.0558	1.11E-15
sex=M	0.0577	0.0833	0.6925	0.4887
length	-0.4583	1.8091	-0.2533	0.8
diameter	11.0751	2.2273	4.9725	6.88E-7
height	10.7615	1.5362	7.0053	2.86E-12
whole_weight	8.9754	0.7254	12.373	0.0
shucked_weight	-19.7869	0.8174	-24.2086	0.0
viscera_weight	-10.5818	1.2937	-8.1792	4.44E-16
shell_weight	8.7418	1.1247	7.7723	9.55E-15
Intercept	3.8946	0.2916	13.3576	0.0

R-Squared: 0.5379
Adjusted R-Squared: 0.5369

Q4. Decision Tree Learner (WEKA)

A decision tree was trained using WEKA's Decision Tree Learner.

Tree structure visualization can be inserted below.



Q5. Clustering with KNIME

Performed clustering using 4 attributes: length, diameter, height, num_rings.

Six clusters were identified with the following data point counts:

Cluster	Data Points
C1	418
C2	386
C3	1,087
C4	936
C5	1
C6	1,349

Q6. Simple Linear Regression in KNIME

Built a simple regression model using only length, diameter, and height.

Equation:

$$\text{num_rings} = 2.83648 - 11.9327 \times \text{length} + 25.7661 \times \text{diameter} + 20.3582 \times \text{height}$$

Info

4 instances (no missing data)
1 feature
No target variable.
1 meta attribute

Variables

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

Selection

☒ Select full rows

name

coef

1 intercept

2.83648

2 length

-11.9327

3 diameter

25.7661

4 height

20.3582

Restore Original Order

☒ Send Automatically

44