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

Tools Used

- WEKA – Linear Regression, Neural Network, Decision Tree

- KNIME – Regression and Clustering

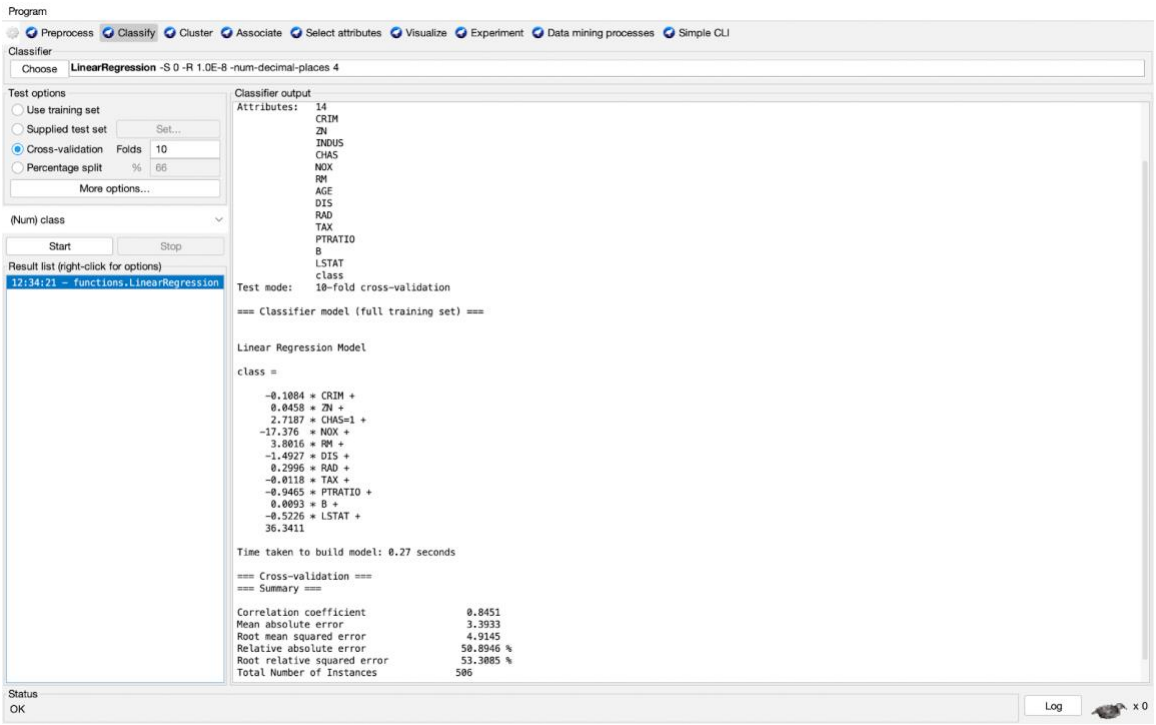
- Orange – (Optional for visualization)

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.



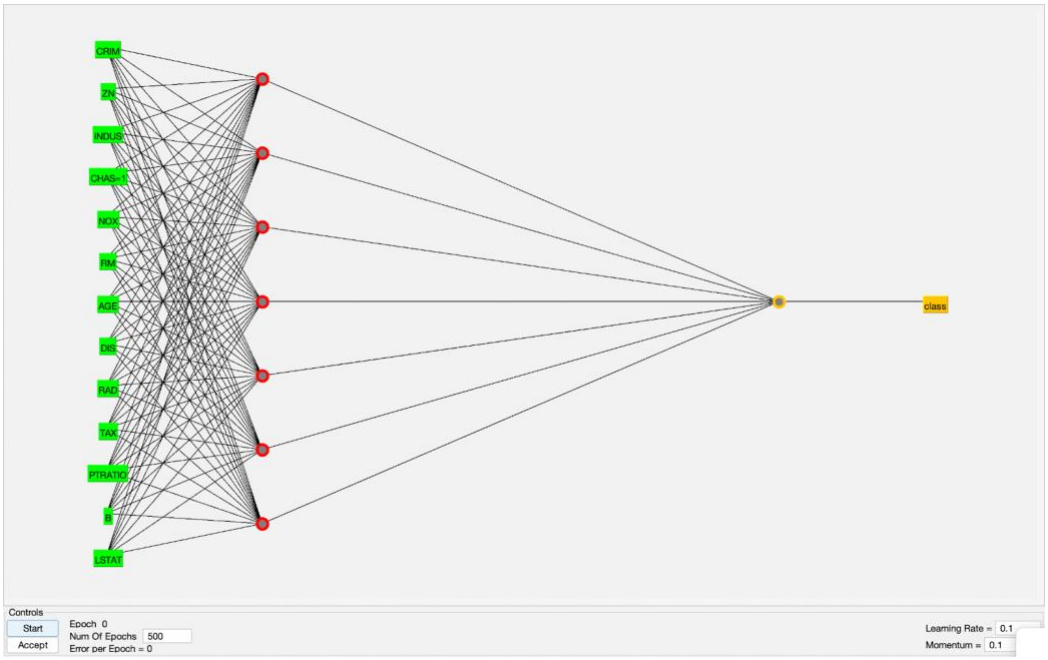
Q2. Neural Network (Multilayer Perceptron in WEKA)

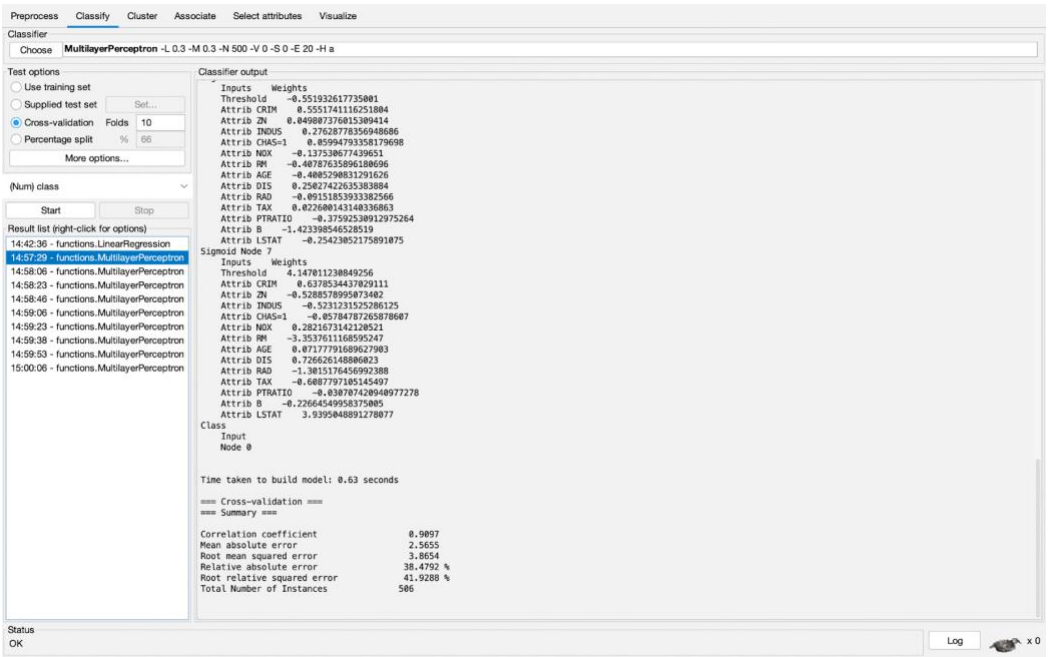
Tested multiple learning rate and momentum combinations.

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

Learning Rate	Momentum	RMSE

0.1	0.1	3.8654





Q3. Linear Regression in KNIME

Constructed a regression model to predict `num_rings`.

Equation:

num_rings = 3.8946 − 0.8249 × (sex = I) + 0.0577 × (sex = M) − 0.4583 × length
+ 11.0751 × diameter + 10.7615 × height + 8.9754 × whole_weight
− 19.7869 × shucked_weight − 10.5818 × viscera_weight + 8.7418 × shell_weight

File

Statistics on Linear Regression

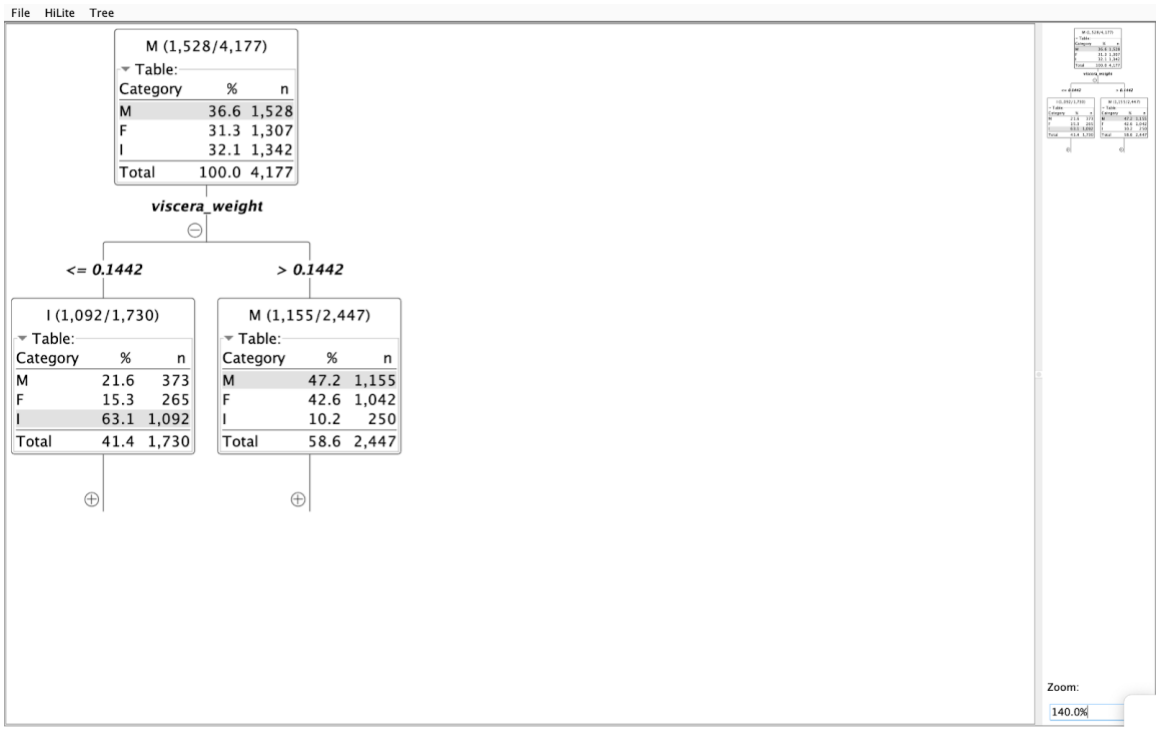
Variable	Coeff.	Std. Err.	t-value	P>|t|
sex=I	-0.8249	0.1024	-8.0558	1.11E-15
sex=M	0.0577	0.0833	0.6925	0.4887
length	-0.4583	1.8091	-0.2533	0.8
diameter	11.0751	2.2273	4.9725	6.88E-7
height	10.7615	1.5362	7.0053	2.86E-12
whole_weight	8.9754	0.7254	12.373	0.0
shucked_weight	-19.7869	0.8174	-24.2086	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:

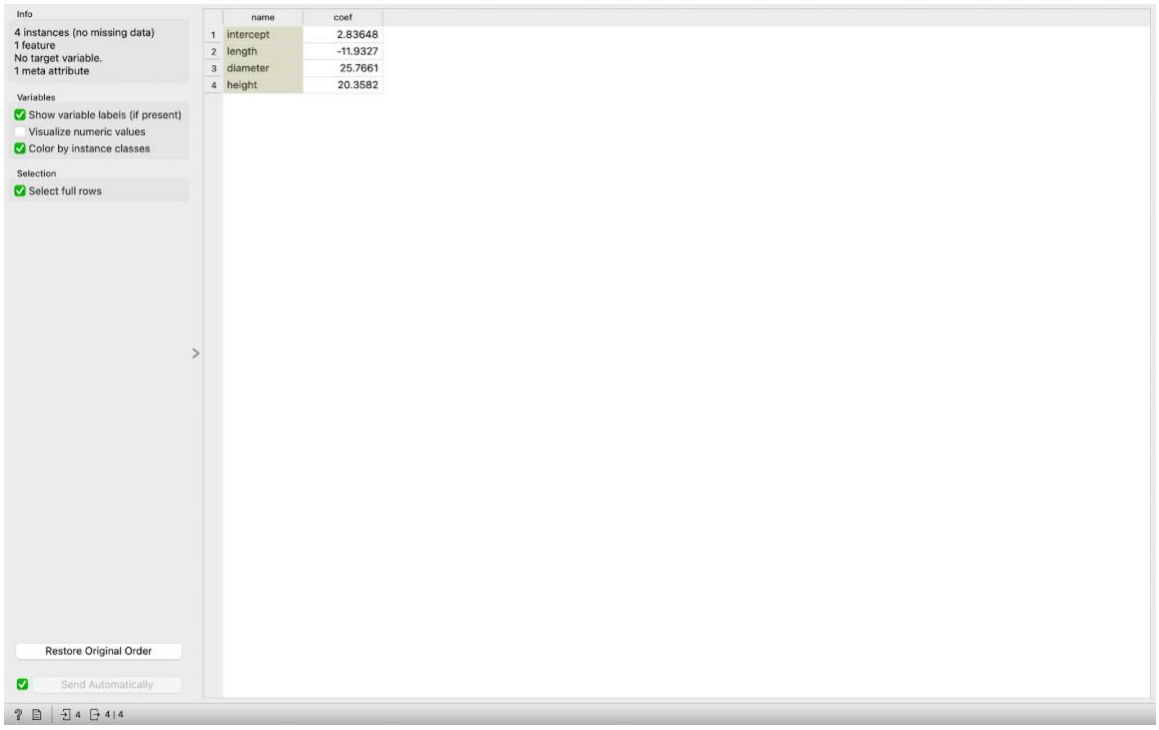
Cluster	Data Points
C1	418
C2	386
C3	1,087
C4	936
C5	1
C6	1,349

Q6. Simple Linear Regression in KNIME

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

Equation:

num_rings = 2.83648 - 11.9327 × length + 25.7661 × diameter + 20.3582 × 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
Restore Original Order
Send Automatically
4
4 | 4
name
coef
1 intercept 2.83648
2 length -11.9327
3 diameter 25.7661
4 height 20.3582