***Mean Absolute Error**:*

Mean Absolute Error is a model evaluation metric used with regression models.

The mean absolute error of a model with respect to a test set is the mean of the absolute values of the individual prediction errors on over all instances in the test set.

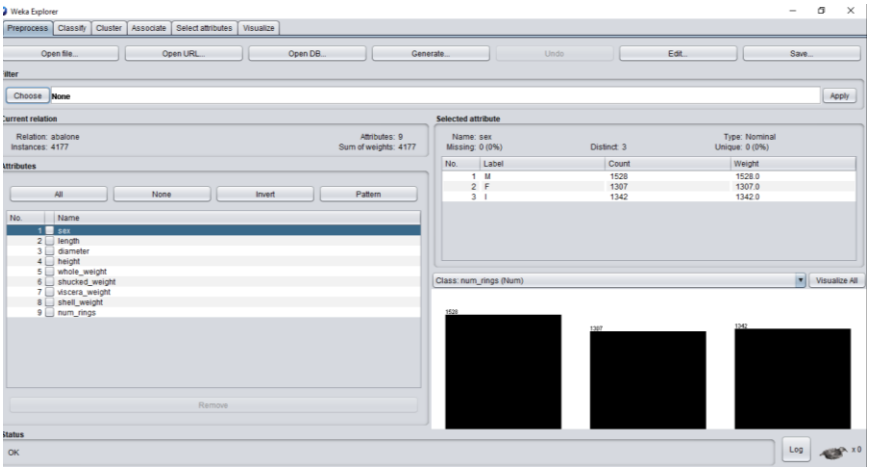
Each prediction error is the difference between the true value and the predicted value for the instance.

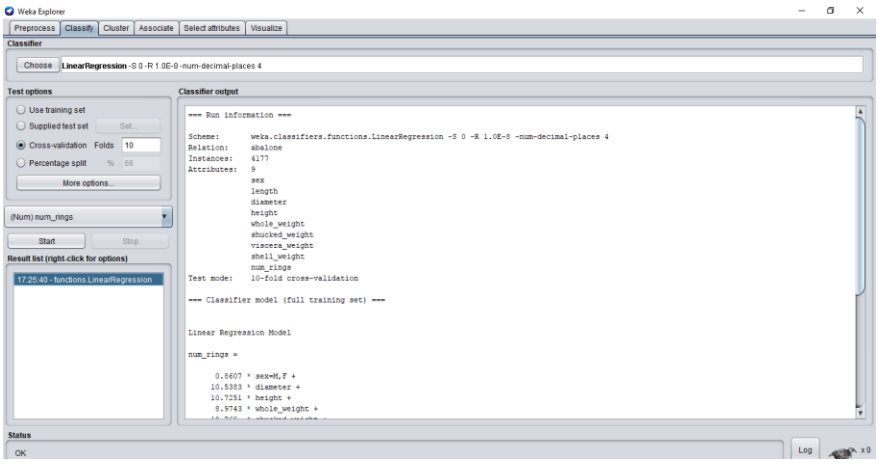
$$\mathbf{MAE} = \frac{\sum_{i=1}^{n} \boldsymbol{abs}(\mathbf{yi} - \boldsymbol{\lambda}(\mathbf{xi}))}{\boldsymbol{n}}$$

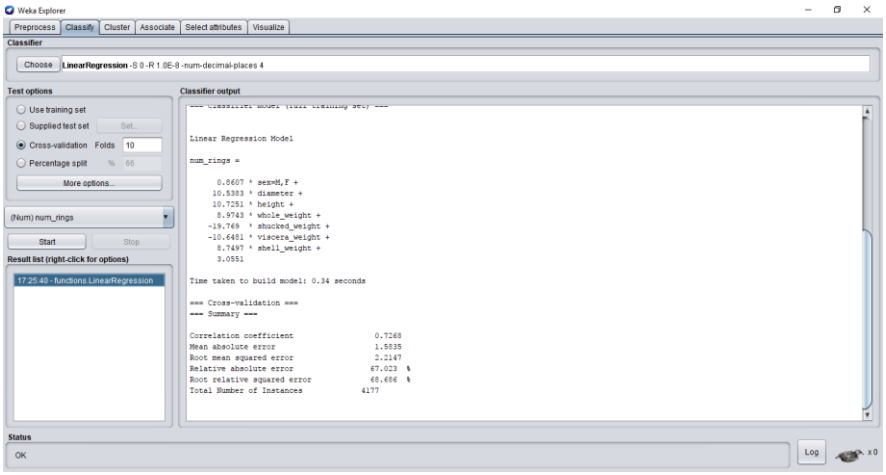
where y_i is the true target value for test instance x_i, $\lambda(x_i)$ is the predicted target value for test instance x_i, and n is the number of test instances.

WEKA: 1

All Parameters Equation and Analysis with linear regression:







Equation:

num_rings =

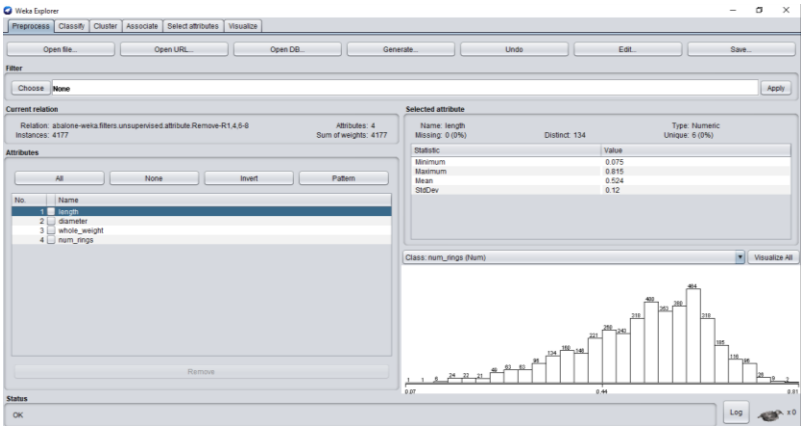
0.8607 * sex=M, F + 10.5383 * diameter + 10.7251 * height + 8.9743 * whole_weight + (-19.769) * shucked_weight + (-10.6481) * viscera_weight + 8.7497 * shell_weight + 3.0551

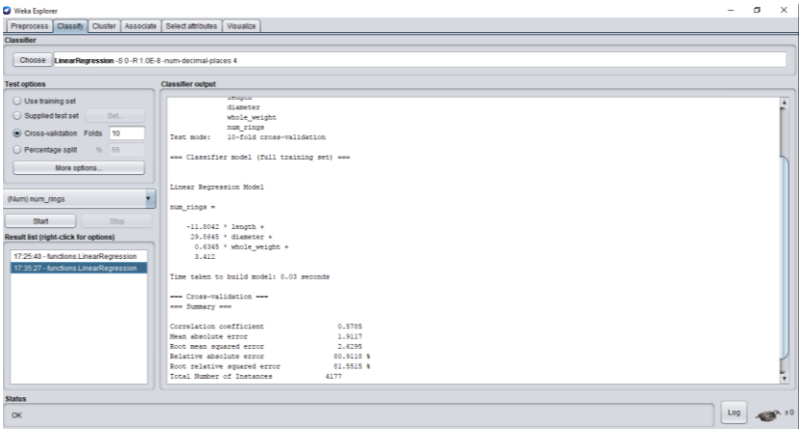
(Here the co-efficient for **sex=I** and co-efficient for **length** is equal to **zero** for both hence they don't appear in the equation, also the last term **3.0551** is the **intercept** for the linear model)

WEKA: 2

Weka 3 Parameters Equation and Analysis with linear regression:

(Removing all parameters i.e., measurements except: length, diameter, whole_weight, num_rings)





Equation :

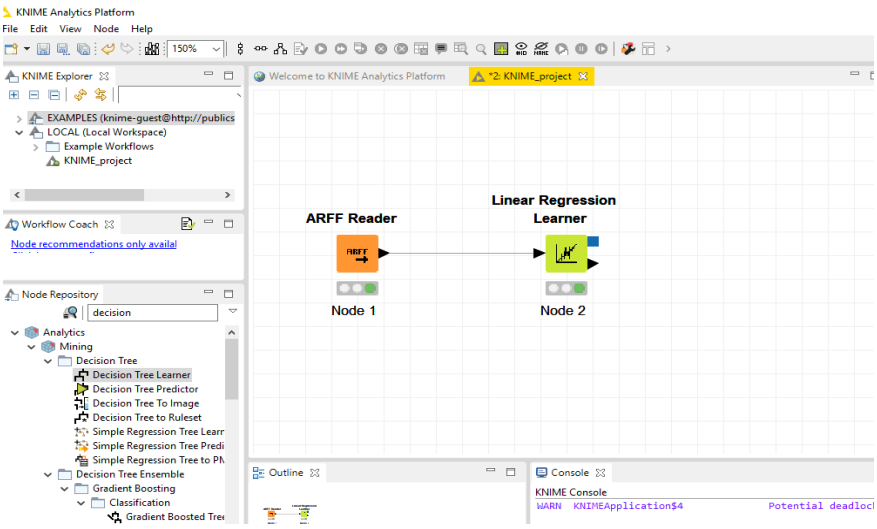
num_rings =

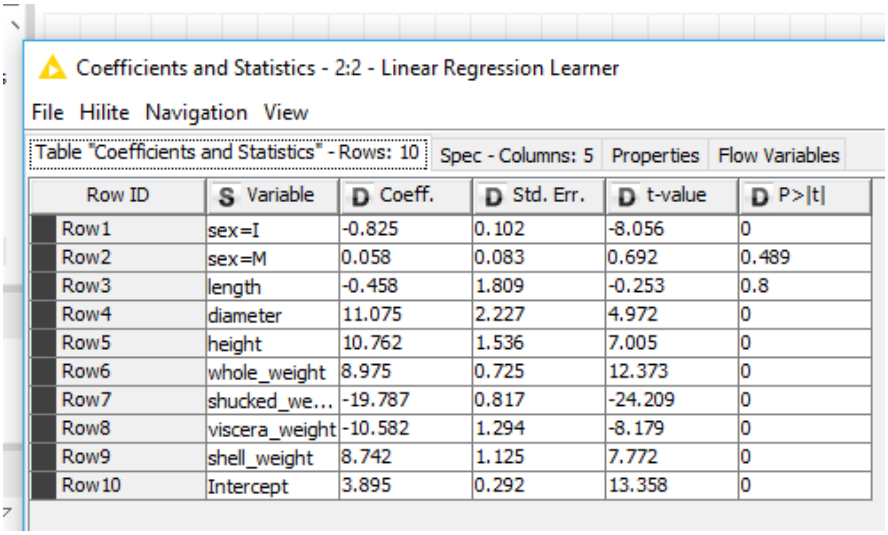
-11.8042 * length + 29.8645 * diameter + 0.6345 * whole_weight + 3.412

(All three parameters or attributes appear in the equation **none** of them have **zero** co-efficient, also the **intercept** for the linear model is **3.412**)

KNIME: 1

Using KNIME to perform linear regression on **all 9 parameters**: Following screengrab shows 2 nodes: One for **AARF reader** and another for **Linear Regression Learner**. The **green** dot below both nodes show that they are **executed**. Also, the pictures show the co-efficient of all parameters.





Coefficients and Statistics - 2:2 - Linear Regression Learner

File Hilite Navigation View

Table "Coefficients and Statistics" - Rows: 10 | Spec - Columns: 5 | Properties | Flow Variables

Row ID	Variable	Coeff.	Std. Err.	t-value	P>|t|
Row1	sex=I	-0.825	0.102	-8.056	0
Row2	sex=M	0.058	0.083	0.692	0.489
Row3	length	-0.458	1.809	-0.253	0.8
Row4	diameter	11.075	2.227	4.972	0
Row5	height	10.762	1.536	7.005	0
Row6	whole_weight	8.975	0.725	12.373	0
Row7	shucked_we...	-19.787	0.817	-24.209	0
Row8	viscera_weight	-10.582	1.294	-8.179	0
Row9	shell_weight	8.742	1.125	7.772	0
Row10	Intercept	3.895	0.292	13.358	0

Equation:

num_rings =

(-0.825) * sex=I + 0.058 * sex = M + (-0.458) * length + 11.075 * diameter + 10.762 * height + 8.975 * whole_weight + (-19.787) * shucked_weight + (-10.582) * viscera_weight + 8.742 * shell_weight + 3.895 (Intercept)

KNIME: 2

Comparison:

Parameter that have similar co-efficient in the window of (± 0.5)

Analysis Table below:

Attributes / Parameters	Analyzing WEKA and KNIME co-efficient	Difference	Parameters with similar co-efficient (within 0.5 window)
sex	All 3 I,M,F	Around 0.8	No
length	0 – (-0.458)	0.458	Yes
diameter	11.075 -10.5383	0.5367	0.5367 > 0.50
height	10.762 - 10.7251	0.0369	Yes
whole_weight	8.975-8.9743	0.0007	Yes
shucked_weight	-19.769 - (-19.787)	0.018	Yes
viscera_weight	-10.582- (-10.6481)	0.0661	Yes
shell_weight	8.7497 – 8.742	0.0077	Yes

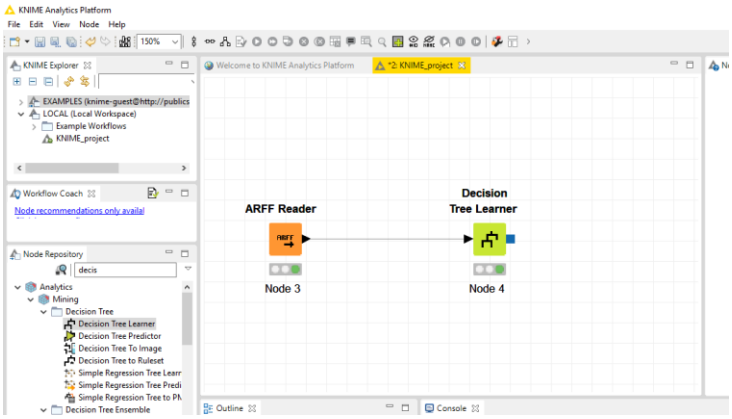
Here, **sex = I, M, F co-efficient all differ by almost 0.8** in WEKA and KNIME and the **diameter** parameter is **negligibly greater** than 0.5 in KNIME, so I am not considering it to be similar because of above reason in the table.

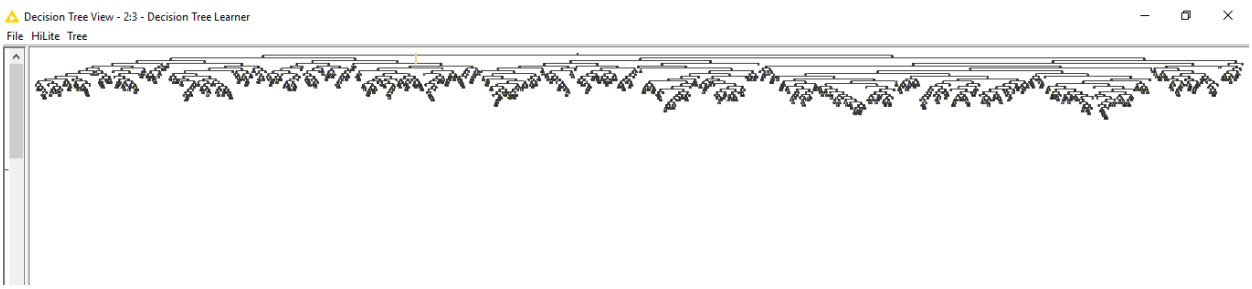
The parameters that have similar co-efficient are:

length, height, whole_weight, shucked_weight, viscera_weight, shell_weight (Total 6)

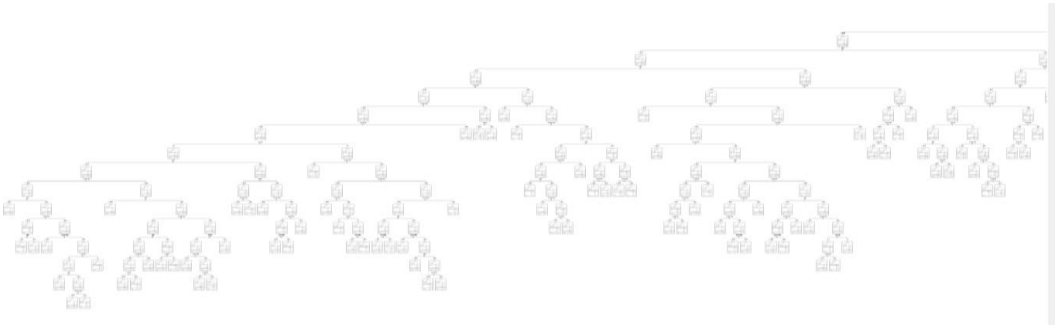
KNIME: 3

Setting up a **"Decision Tree Learner"** predictor, where "**sex**" is the predicted variable: Below is the screenshot for the same:

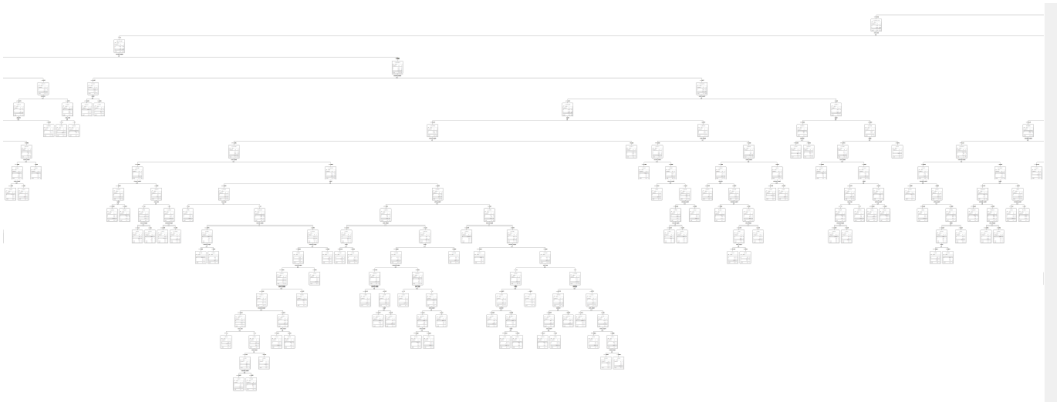




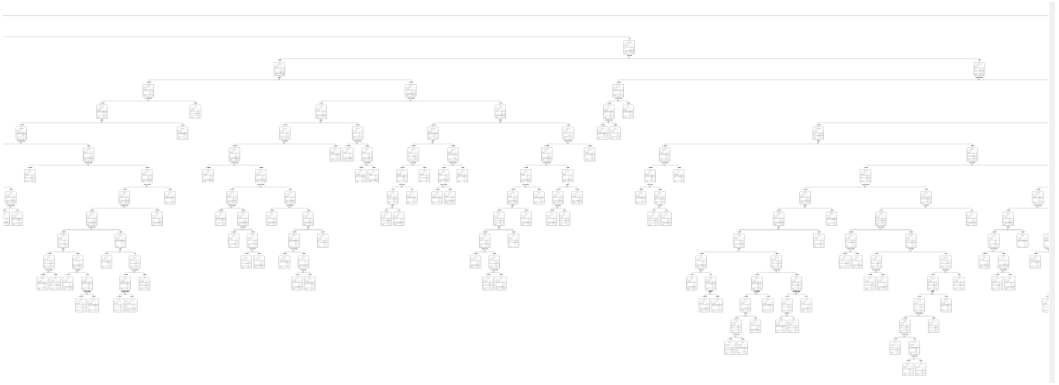
Below is a clearer decision tree image starting from left to right in different parts:



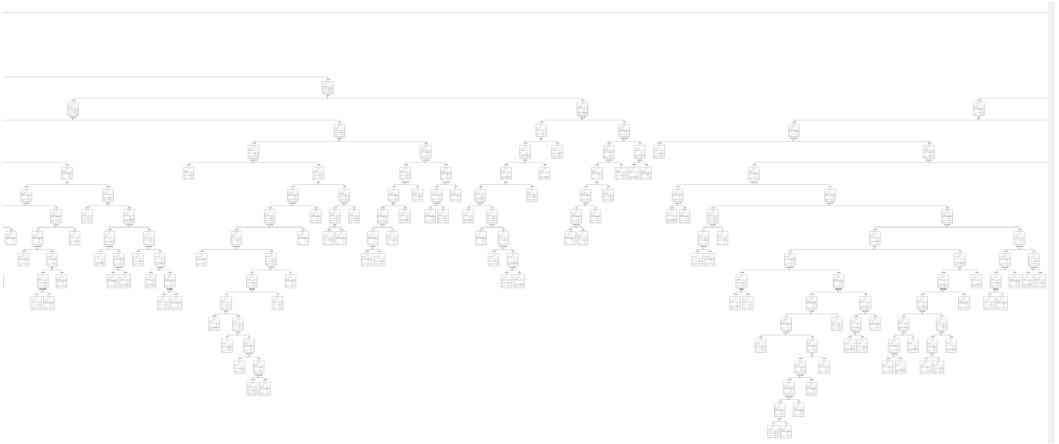
1



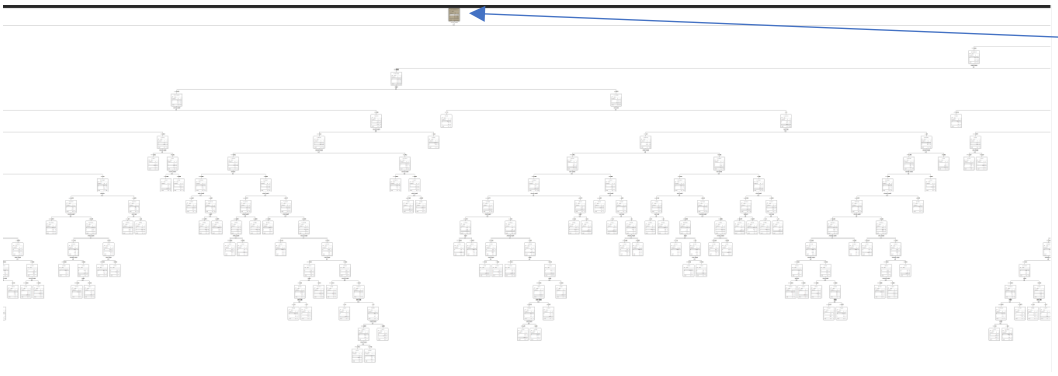
2



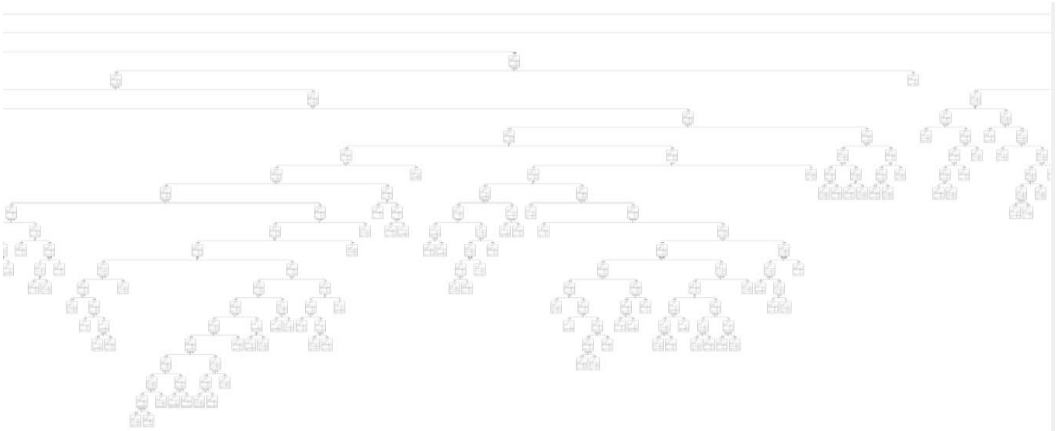
3



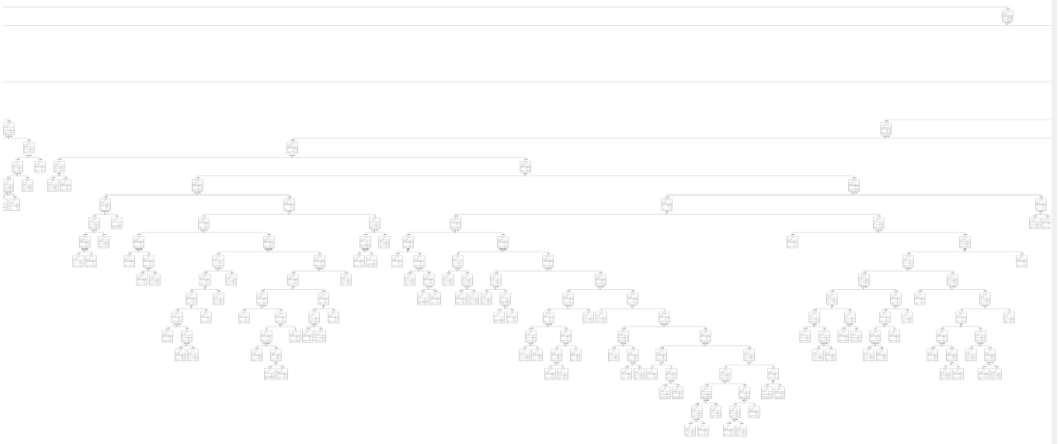
4



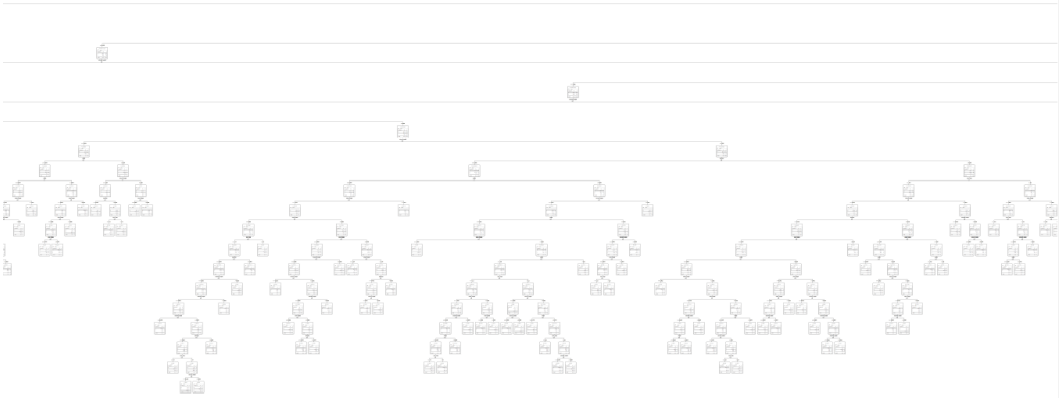
5



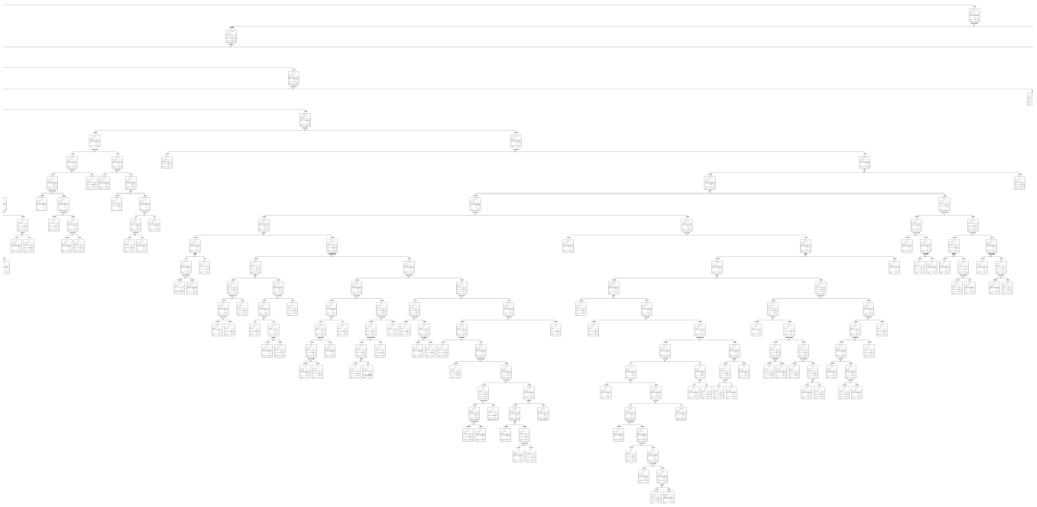
6



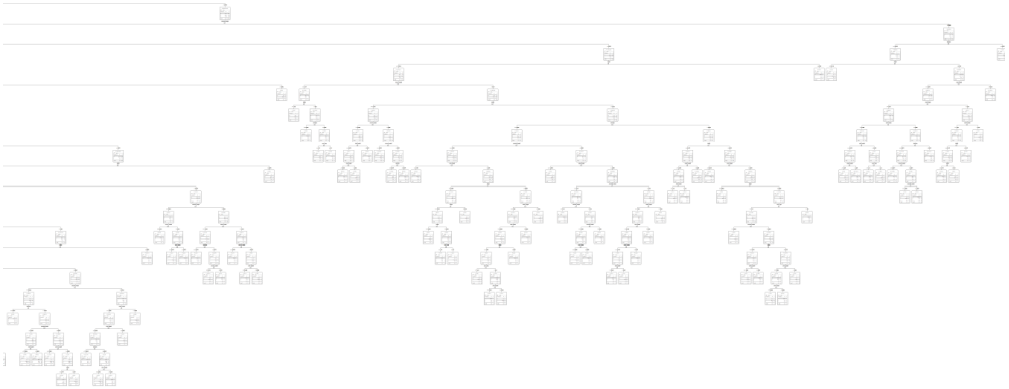
7



8

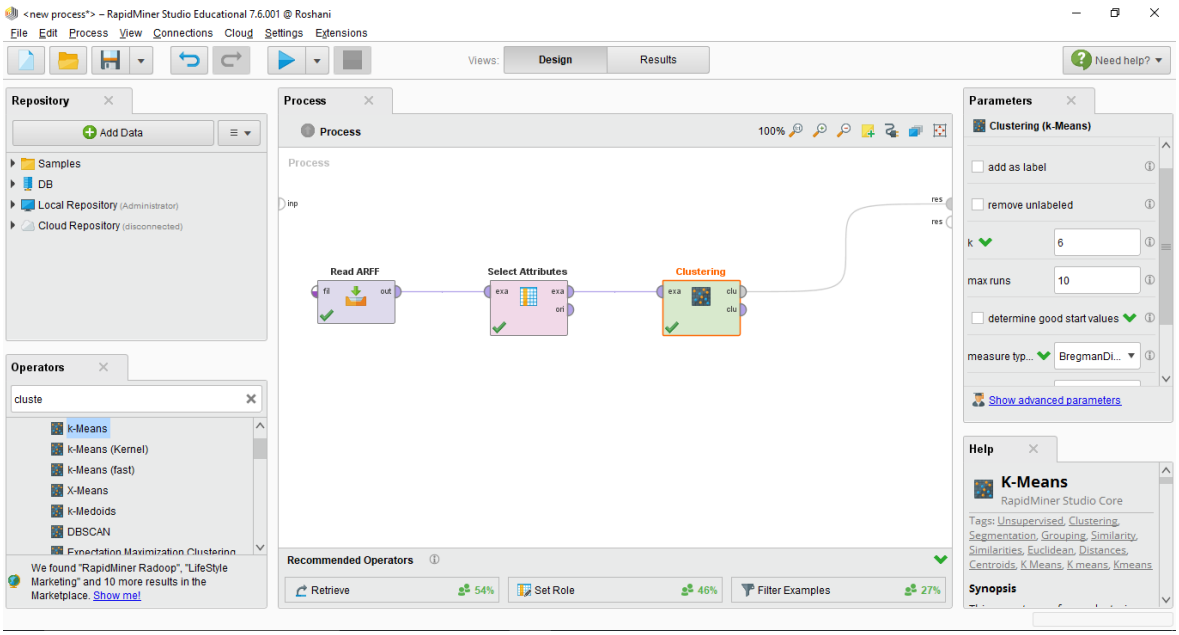


9



10

RAPIDMINER: 1

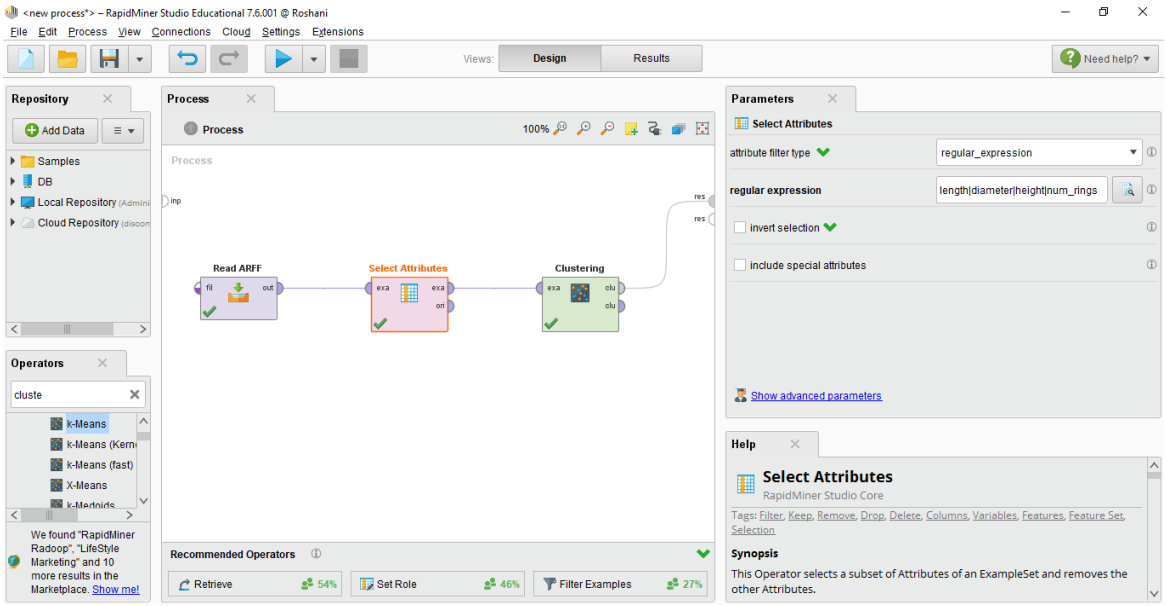


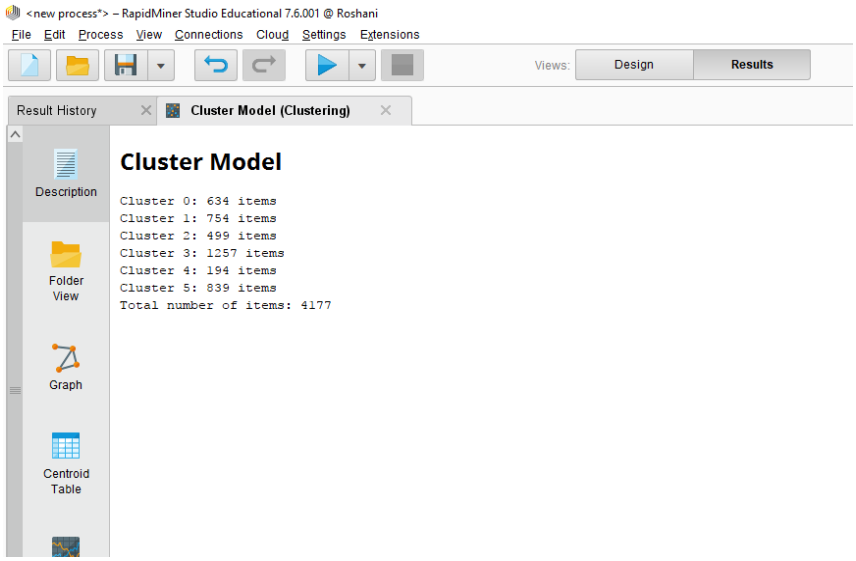
The above picture shows That the three nodes are arranged such that:

Read ARFF node: Is used to read the dataset file shells.arff

Select Attributes node: (From the picture below) I have selected attributes that are mentioned by sing the following regular expression **"length | diameter | height | num_rings"** that selects only the required attributes.

Clustering (k-Means) node: k is set to **6** which gives 6 clusters of data points based on their similarity of attributes.





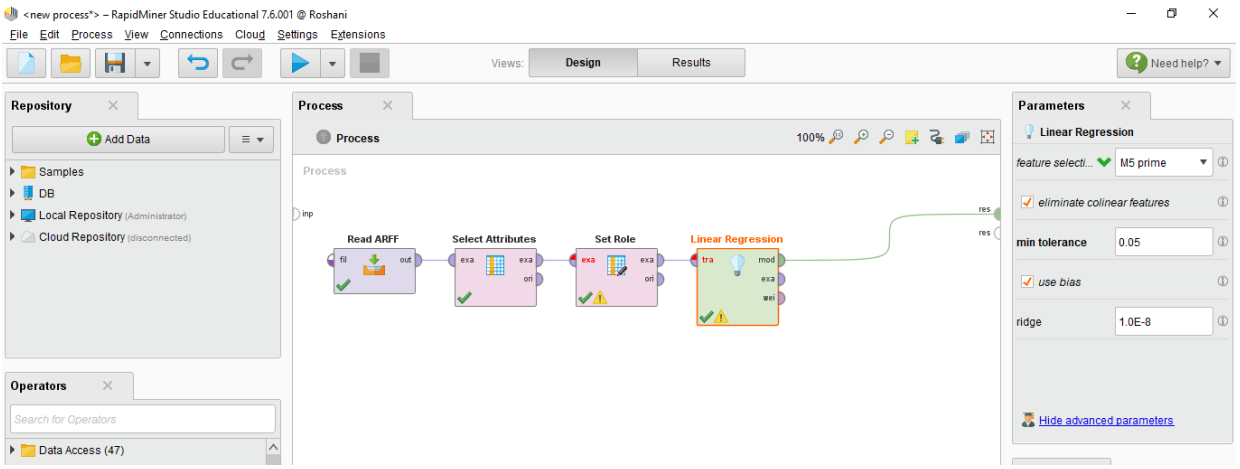
The above picture shows that 6 clusters are formed, and each cluster contains specific number of data points they are as below:

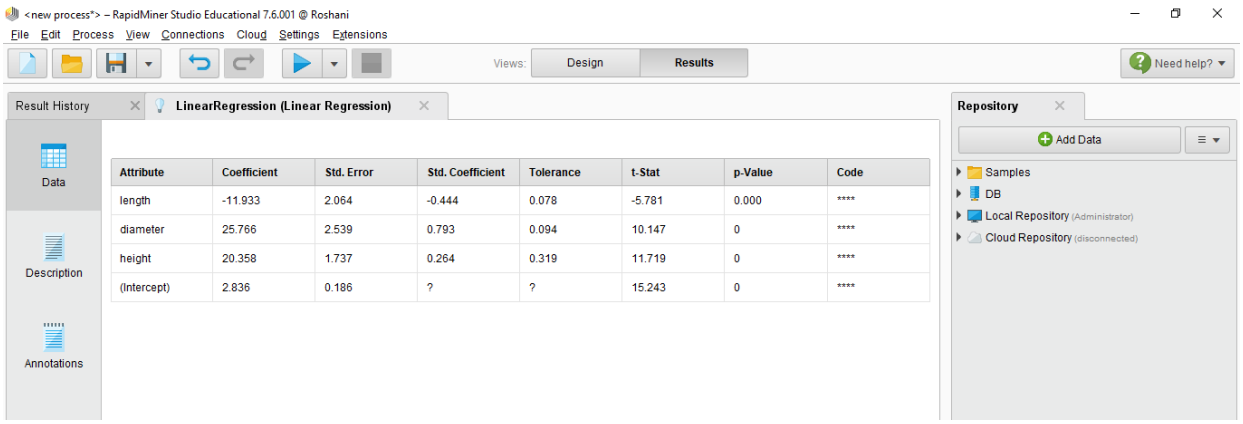
Cluster 0: 634 items
Cluster 1: 754 items
Cluster 2: 499 items
Cluster 3: 1257 items
Cluster 4: 194 items
Cluster 5: 839 items
Total number of items: 4177

Linear regression in Rapid miner:

Picture below shows the node arrangement:

Steps: Reading the dataset file ---> Selecting the attributes ---> set role for num_rings as label (so that predictor knows which attribute to predict) ---> use linear regression to predict num_rings





Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
length	-11.933	2.064	-0.444	0.078	-5.781	0.000	****
diameter	25.766	2.539	0.793	0.094	10.147	0	****
height	20.358	1.737	0.264	0.319	11.719	0	****
(Intercept)	2.836	0.186	?	?	15.243	0	****

Following is the equation for predicting num_rings through linear regression model:

num_rings = (-11.933) * length + 25.766 * diameter + 20.358 * height + 2.836 (intercept)