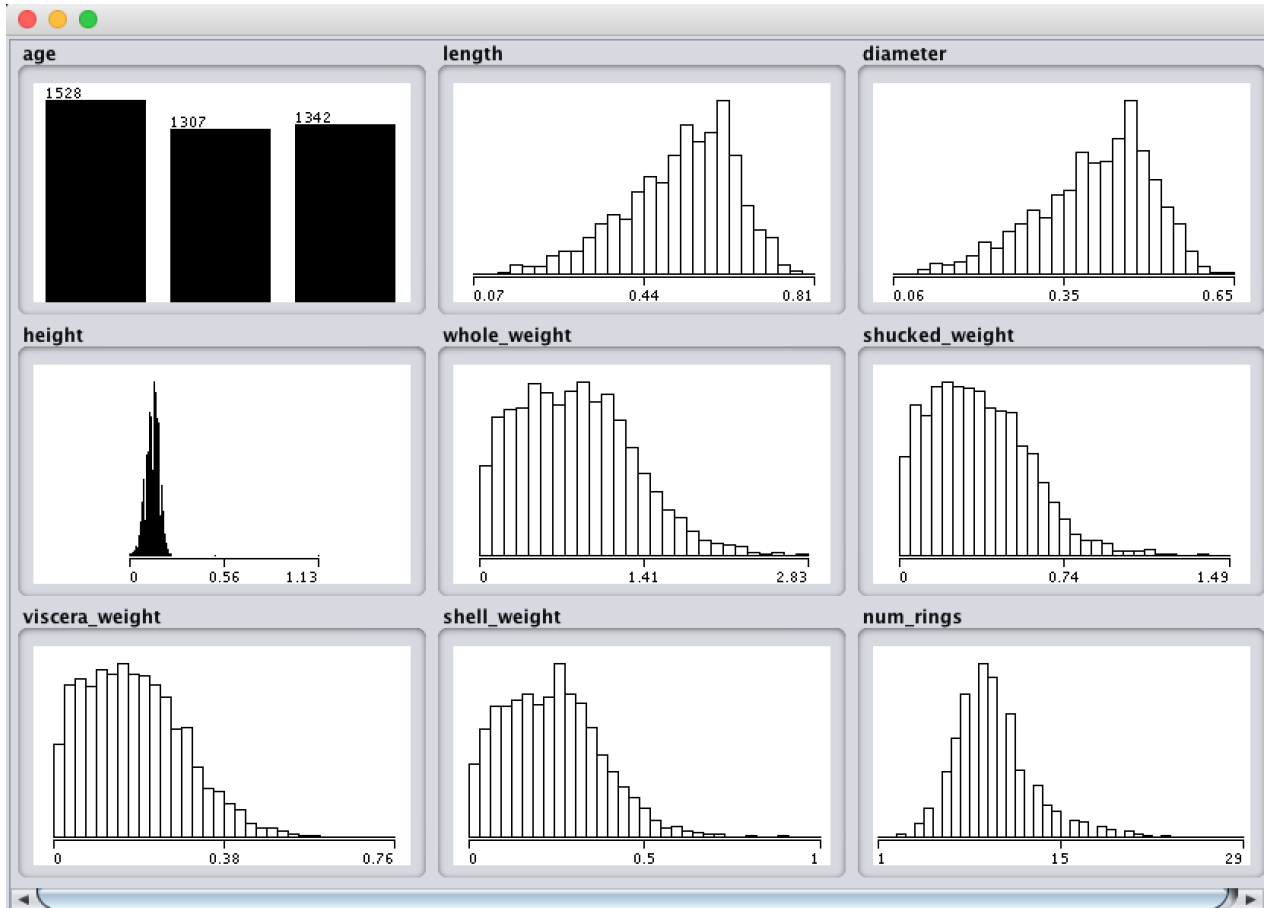


## 1. Weka:

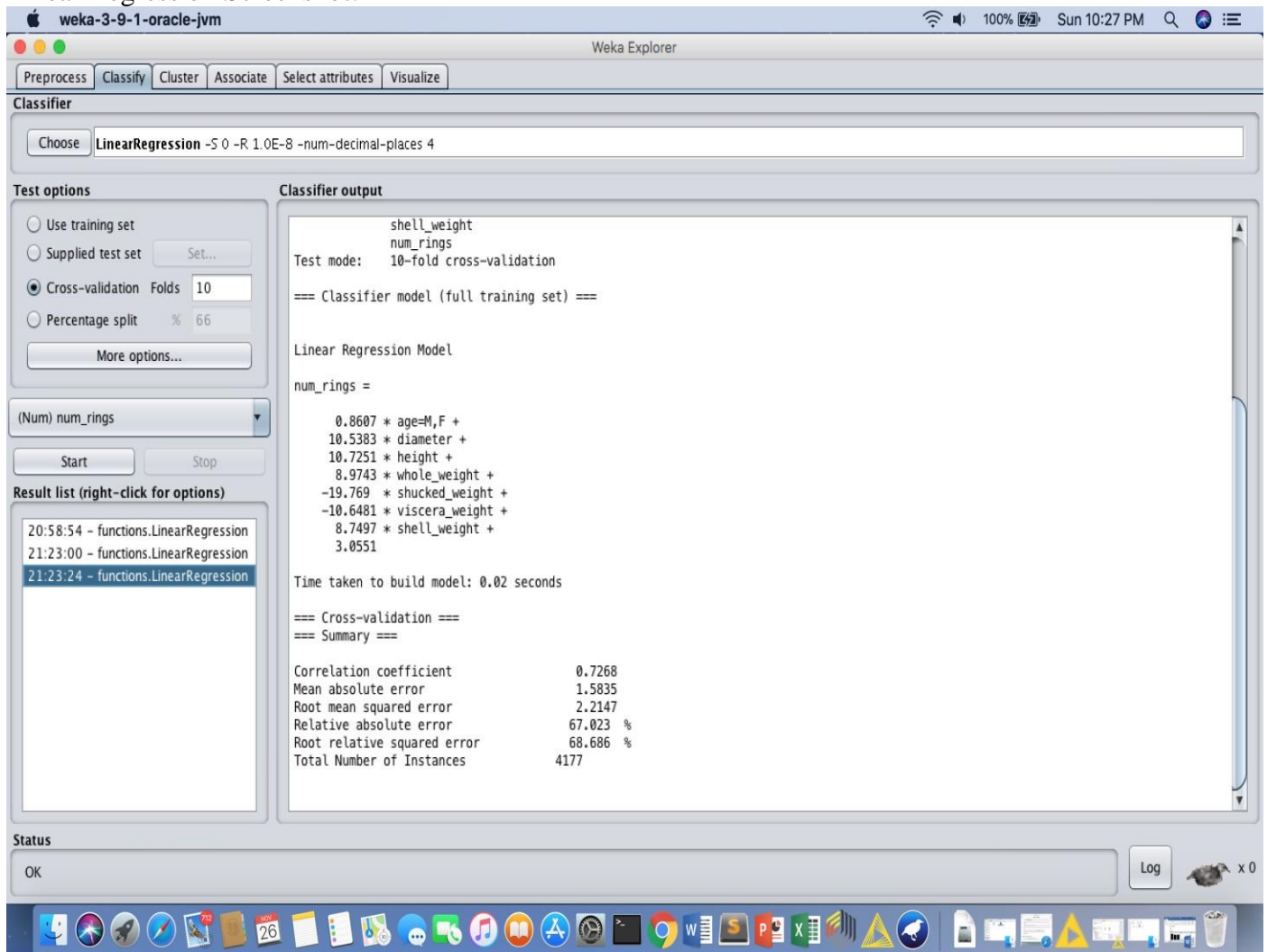
Distributions Screenshot:



1. Mean absolute error is 1.5835 (Which can be seen in Linear Regression Screenshot below).

Mean absolute error (MAE) is a measure of difference between two continuous variables. As the name suggests, the mean absolute error is an average of the absolute errors  $|e_i| = |y_i - x_i|$ , where  $y_i$  is the prediction and  $x_i$  the true value. Note that alternative formulations may include relative frequencies as weight factors. The mean absolute error uses the same scale as the data being measured. This is known as a scale-dependent accuracy measure and therefore cannot be used to make comparisons between series using different scales

## Linear Regression Screenshot:

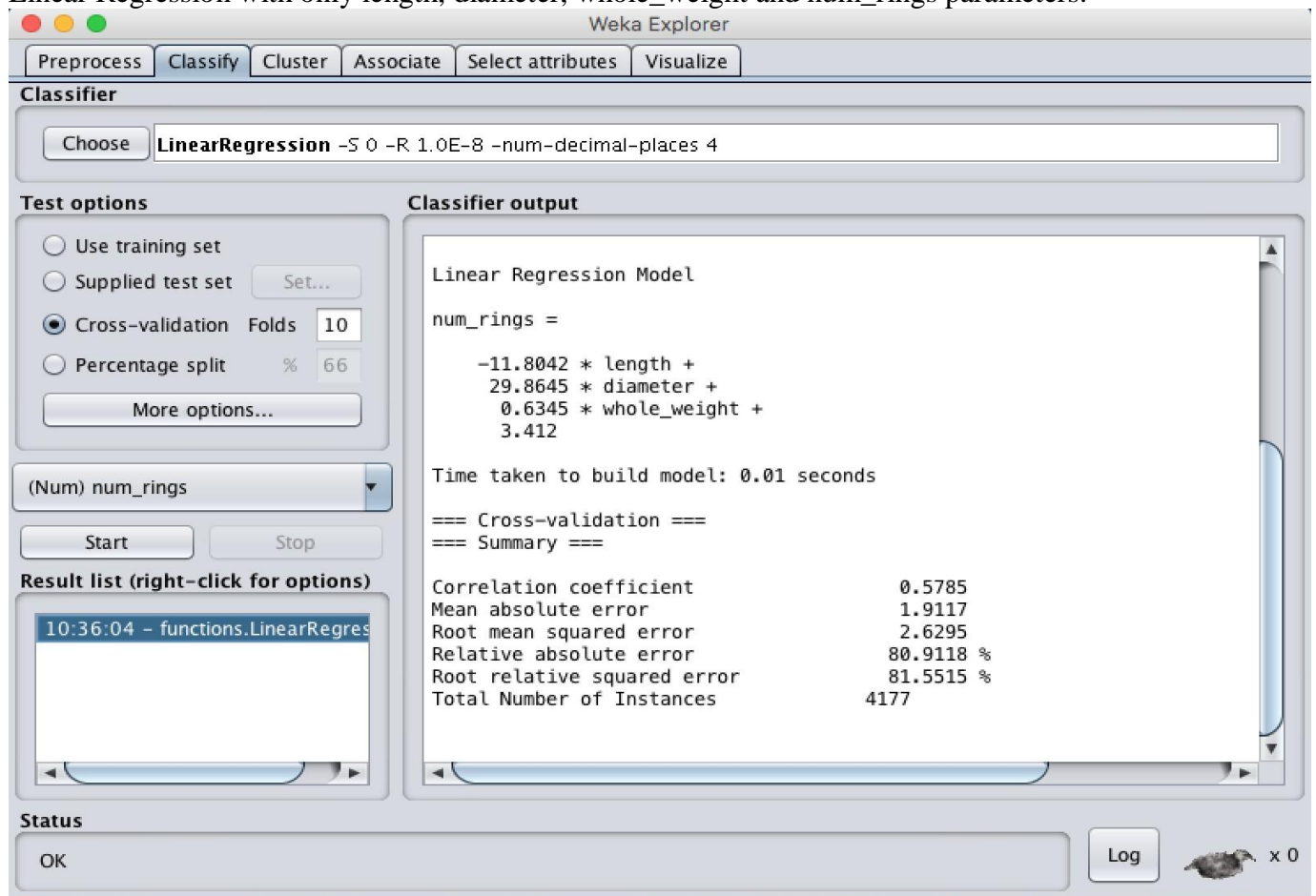


2. **Equation:**  $\text{num\_rings} = f(\text{age}\{M,F\}, \text{length}, \text{diameter}, \text{height}, \text{whole\_weight}, \text{shucked\_weight}, \text{viscera\_weight}, \text{shell\_weight}) =$

$$\begin{aligned}
 &0.8607 * \text{age}=M,F + \\
 &10.5383 * \text{diameter} + \\
 &10.7251 * \text{height} + \\
 &8.9743 * \text{whole\_weight} + \\
 &-19.769 * \text{shucked\_weight} + \\
 &-10.6481 * \text{viscera\_weight} + \\
 &8.7497 * \text{shell\_weight} + \\
 &3.0551
 \end{aligned}$$

3.0551 is the Intercept

Linear Regression with only length, diameter, whole\_weight and num\_rings parameters:



3. **Equation:**  $\text{num\_rings} = f(\text{length}, \text{diameter}, \text{whole\_weight}) =$

$$\begin{aligned}
 & -11.8042 * \text{length} + \\
 & 29.8645 * \text{diameter} + \\
 & 0.6345 * \text{whole\_weight} + \\
 & 3.412
 \end{aligned}$$

3.412 is the Intercept

## 2. Knime:

Linear Regression Screenshot:

The screenshot shows the KNIME Analytics Platform interface. The main workspace displays a workflow with two nodes: 'ARFF Reader' (Node 1) and 'Linear Regression Learner' (Node 2). The 'Linear Regression Learner' node is selected, and its 'Dialog Options' are visible on the right. The 'Target' field is set to 'length'. The 'Node Repository' on the left shows the 'Linear Regression Learner' node under the 'Statistics' category. The 'Workflow Coach' shows the 'Linear Regression Learner' node as the recommended node. The 'Console' window at the bottom right displays the following text:

```

*****
*** Welcome to the KNIME Analytics Platform v3.4.2.v201711100944 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/prerana/knime-workspace/.metadata/knime/knime.log
WARN ARFF Reader 2:1 No source location provided! Please enter a valid
  
```

## Result:

The screenshot shows the 'Linear Regression Result View' window in the KNIME Analytics Platform. The window displays the following statistics:

Multiple R-Squared: 0.5379  
Adjusted R-Squared: 0.5369

Variable	Coeff.	Std. Err.	t-value	P> t
age=I	-0.8249	0.1024	-8.0558	1.11E-15
age=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

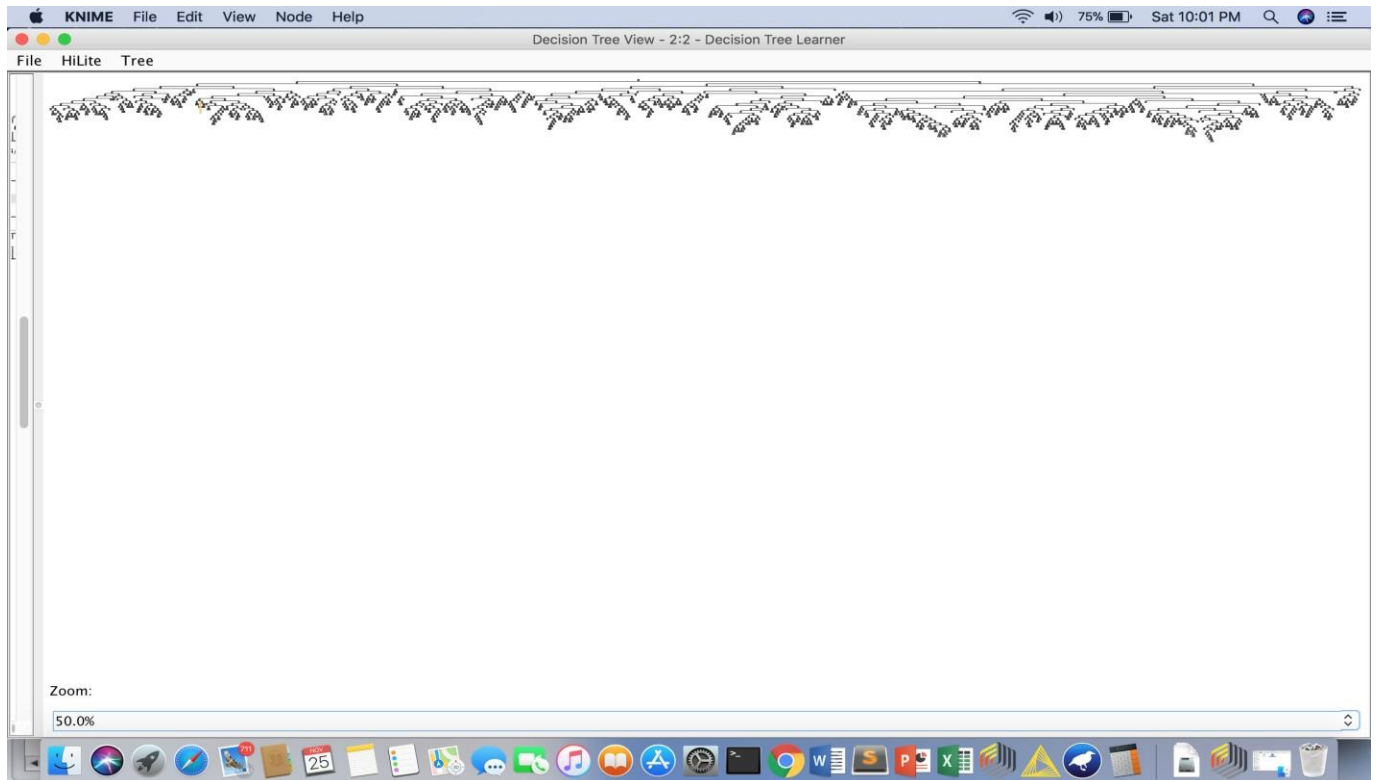
1. **Equation:**  $\text{num\_rings} = f(\text{age}\{I\}, \text{age}\{M\}, \text{length}, \text{diameter}, \text{height}, \text{whole\_weight}, \text{shucked\_weight}, \text{viscera\_weight}, \text{shell\_weight}) =$

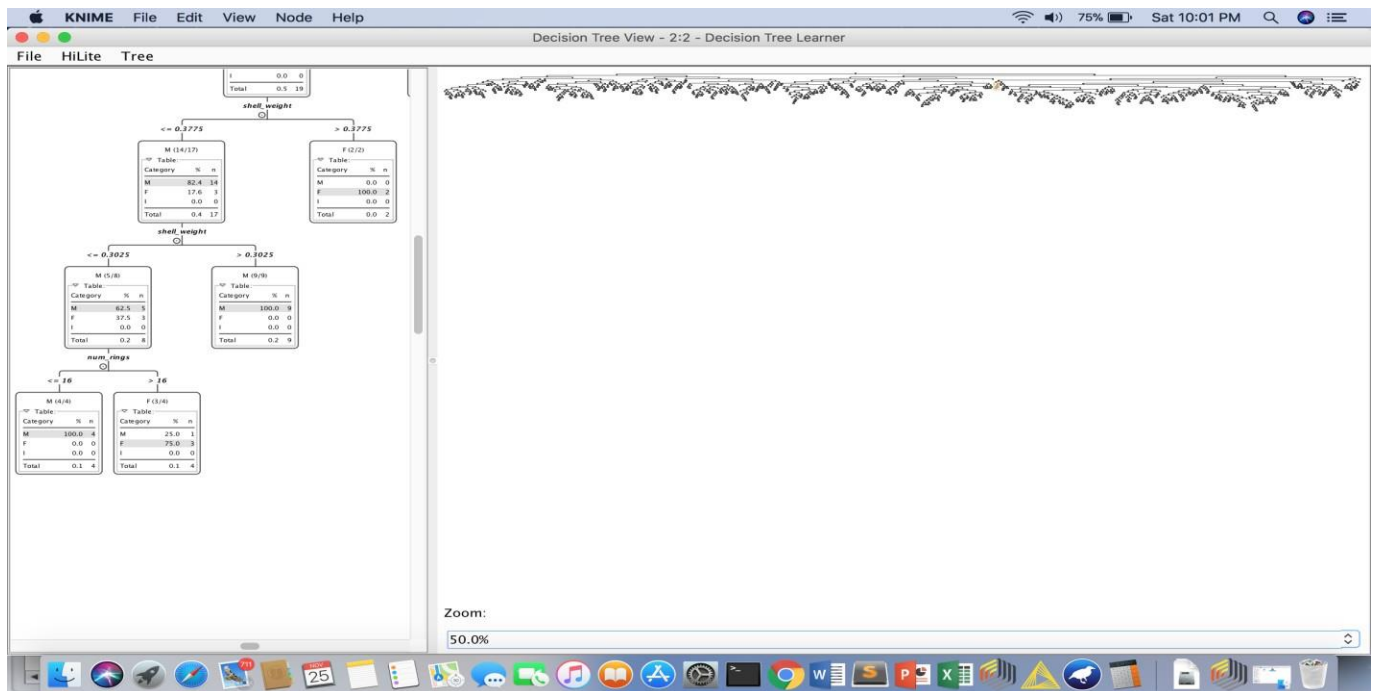
$$\begin{aligned} & -0.8249 * \text{age}=I + \\ & 0.0577 * \text{age}=M + \\ & -0.4583 * \text{length} + \\ & 11.0751 * \text{diameter} + \\ & 10.7615 * \text{height} + \\ & 8.9754 * \text{whole\_weight} + \\ & -19.7869 * \text{shucked\_weight} + \\ & -10.5818 * \text{viscera\_weight} + \\ & 8.7418 * \text{shell\_weight} + \\ & 3.8946 \end{aligned}$$

3.8946 is the Intercept

Parameters which have similar coefficients are: height, whole\_weight, shucked\_weight, viscera\_weight, shell\_weight. Length parameter is not shown in Weka.

## 2. Decision Tree Screenshots:

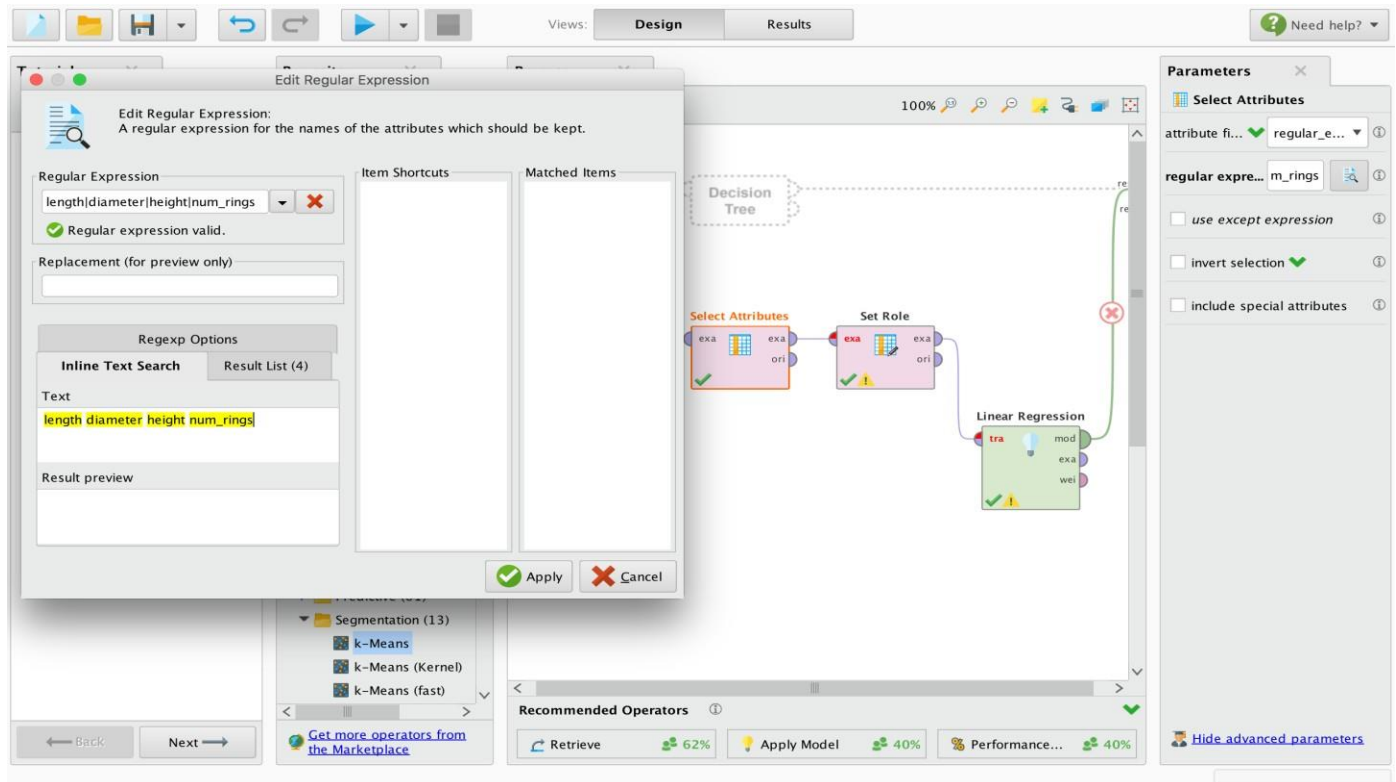




### 3. Rapid Miner:

**Regular Expression:** length|diameter|height|num\_rings

Screenshot:





## 1. kMeans clustering (6 clusters):

The screenshot shows the RapidMiner Studio interface with the 'Design' view selected. The process flow includes 'Retrieve Titanic Training', 'Decision Tree', 'Read ARFF', 'Select Attributes', and 'Clustering'. The 'Clustering' operator is highlighted, and its parameters are shown on the right:

- Clustering (k-Means)**
  - ☒ add cluster attribute
  - ☐ add as label
  - ☐ remove unlabeled
  - k: 6
  - max runs: 10
  - ☐ determine good start values
  - measure ty...: BregmanDI...
  - divergence: SquaredEu...

The 'Help' panel on the right provides a synopsis of the K-Means operator: "This operator performs clustering using the k-means algorithm. Clustering is concerned with grouping objects together that are similar to each other."

## Result:

The screenshot shows the 'Results' view of the RapidMiner Studio. The 'Cluster Model (Clustering)' operator is selected, and the 'ExampleSet (Clustering)' is displayed. The table shows 20 rows of data, each with a 'Row No.', 'id', 'cluster', and four numerical attributes: 'length', 'diameter', 'height', and 'num\_rings'.

Row No.	id	cluster	length	diameter	height	num_rings
1	1	cluster_2	0.455	0.365	0.095	15
2	2	cluster_5	0.350	0.265	0.090	7
3	3	cluster_3	0.530	0.420	0.135	9
4	4	cluster_0	0.440	0.365	0.125	10
5	5	cluster_5	0.330	0.255	0.080	7
6	6	cluster_3	0.425	0.300	0.095	8
7	7	cluster_4	0.530	0.415	0.150	20
8	8	cluster_2	0.545	0.425	0.125	16
9	9	cluster_3	0.475	0.370	0.125	9
10	10	cluster_4	0.550	0.440	0.150	19
11	11	cluster_2	0.525	0.380	0.140	14
12	12	cluster_0	0.430	0.350	0.110	10
13	13	cluster_1	0.490	0.380	0.135	11
14	14	cluster_0	0.535	0.405	0.145	10
15	15	cluster_0	0.470	0.355	0.100	10
16	16	cluster_1	0.500	0.400	0.130	12
17	17	cluster_5	0.355	0.280	0.085	7
18	18	cluster_0	0.440	0.340	0.100	10
19	19	cluster_5	0.365	0.295	0.080	7
20	20	cluster_3	0.450	0.320	0.100	9

Number of data points in each cluster: Cluster 0: 634, Cluster 1: 754 Cluster 2: 499, Cluster 3: 1257, Cluster 4: 194, Cluster 5: 839. Total: 4177

Screenshot:

The screenshot displays the RapidMiner Studio interface with the 'Cluster Model' process results. The main window shows the 'Cluster Model' description, listing the number of items in each cluster and the total count.

**Cluster Model Description:**

- 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

The interface includes a 'Tutorials' panel on the left, a 'Repository' panel on the right showing the data source, and a 'Views' bar at the top with 'Design' and 'Results' tabs.

## 2. Linear Regression

The screenshot displays the RapidMiner Studio interface with the 'Linear Regression' process setup. The main window shows the 'Process' diagram, which includes the 'Read ARFF', 'Select Attributes', 'Set Role', and 'Linear Regression' operators. The 'Help' panel on the right provides detailed information about the 'Linear Regression' operator.

**Linear Regression Help:**

**Tags:** Supervised, Classification, Regression, Model, Least squares, Ordinary Ridge, OLS, Glim, Generalized, Functions

**Synopsis**  
This operator calculates a linear regression model from the input ExampleSet.  
[Jump to Tutorial Process](#)

**Description**  
Regression is a technique used for numerical prediction. Regression is a statistical measure that attempts to determine the strength of the relationship between one dependent variable ( i.e. the label attribute) and a series of other changing variables (regular attributes). Just like Classification is used for predicting categorical labels, Regression is used for predicting a continuous value. For example, we may wish to predict the salary of university graduates with 5 years of work experience, or the potential sales of a new

The interface includes a 'Tutorials' panel on the left, a 'Repository' panel on the right showing the data source, and a 'Views' bar at the top with 'Design' and 'Results' tabs.



Result:

The screenshot shows the Orange3 data mining software interface. The main window displays the results of a Linear Regression model. The top toolbar includes icons for file operations, a 'Views' dropdown set to 'Results', and a 'Need help?' button. The left sidebar contains a 'Tutorials' panel with a 'Modeling' section and a 'View All' link. The central workspace shows a 'Result History' panel with a 'Data' icon, a 'Description' icon, and an 'Annotations' icon. The main area displays a table of regression results for the 'LinearRegression (Linear Regression)' model. The table has columns for Attribute, Coefficient, Std. Error, Std. Coeffi..., Tolerance, t-Stat, p-Value, and Code. The right sidebar shows a 'Repository' panel with a tree view of data sources, including 'Samples', 'DB', 'Local Repository (prerana)', and 'Cloud Repository (disconne)'. The 'Local Repository' is expanded, showing 'data (prerana)', 'shells (prerana - v1, 1)', 'processes (prerana)', 'shellIII (prerana - v1, 11/2)', and 'shells (prerana - v1, 11/2)'.

Attribute	Coefficient	Std. Error	Std. Coeffi...	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	****

**Equation:** Num\_rings = f(length, diameter, height) =  
-11.933 \* length +  
25.766 \* diameter +  
20.358 \* height +  
2.836

Intercept = 2.836