HOMEWORK-5

[1]

Mean Absolute Error

Mean absolute error here is 1.5835

[Mean Absolute Error (MAE): MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.]

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

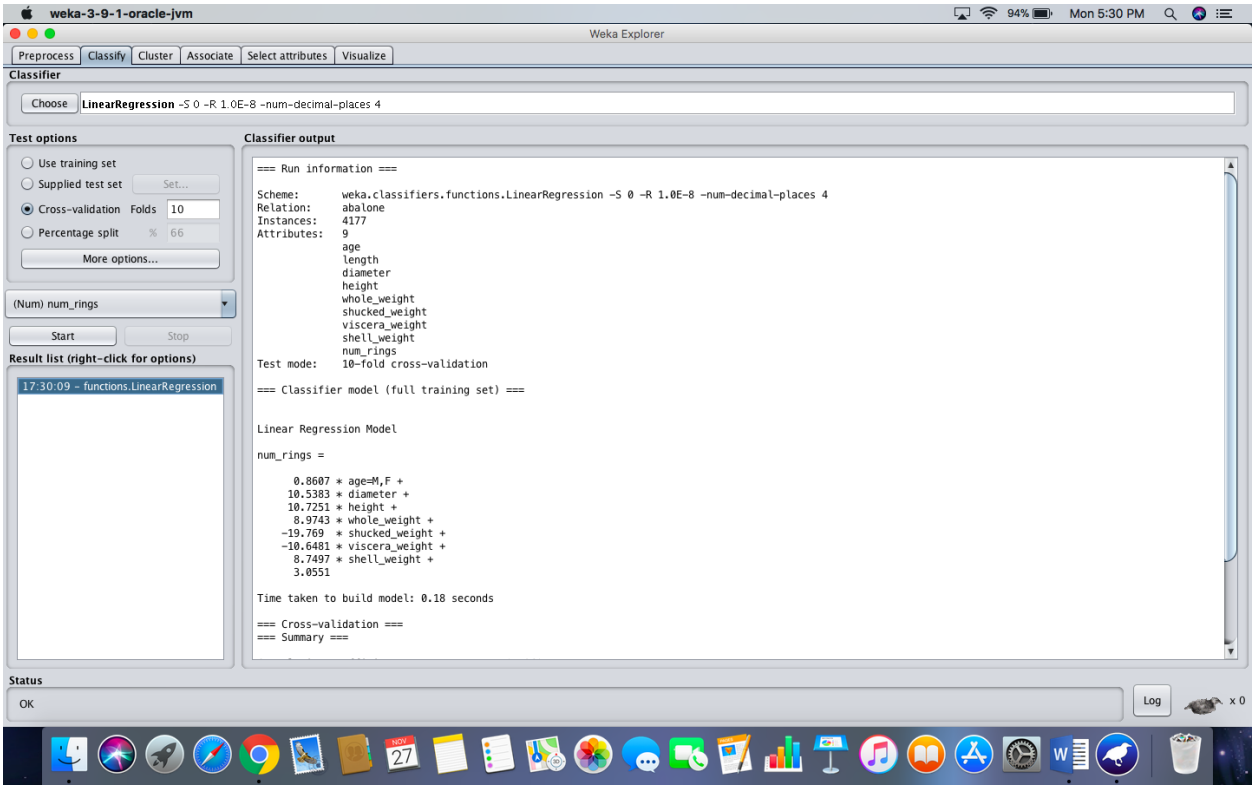
[2]

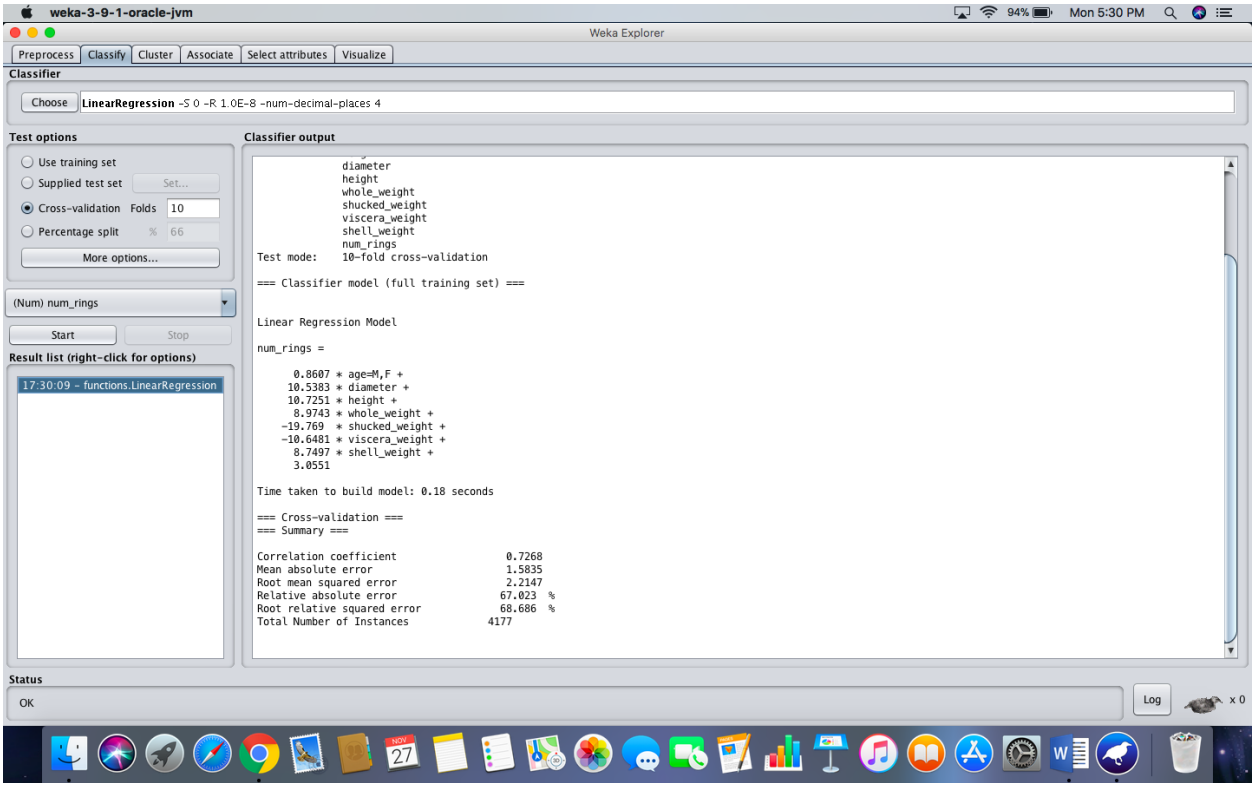
Num_rings

num_rings =

0.8607 * age=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

The co-efficient of the Length parameter is 0.

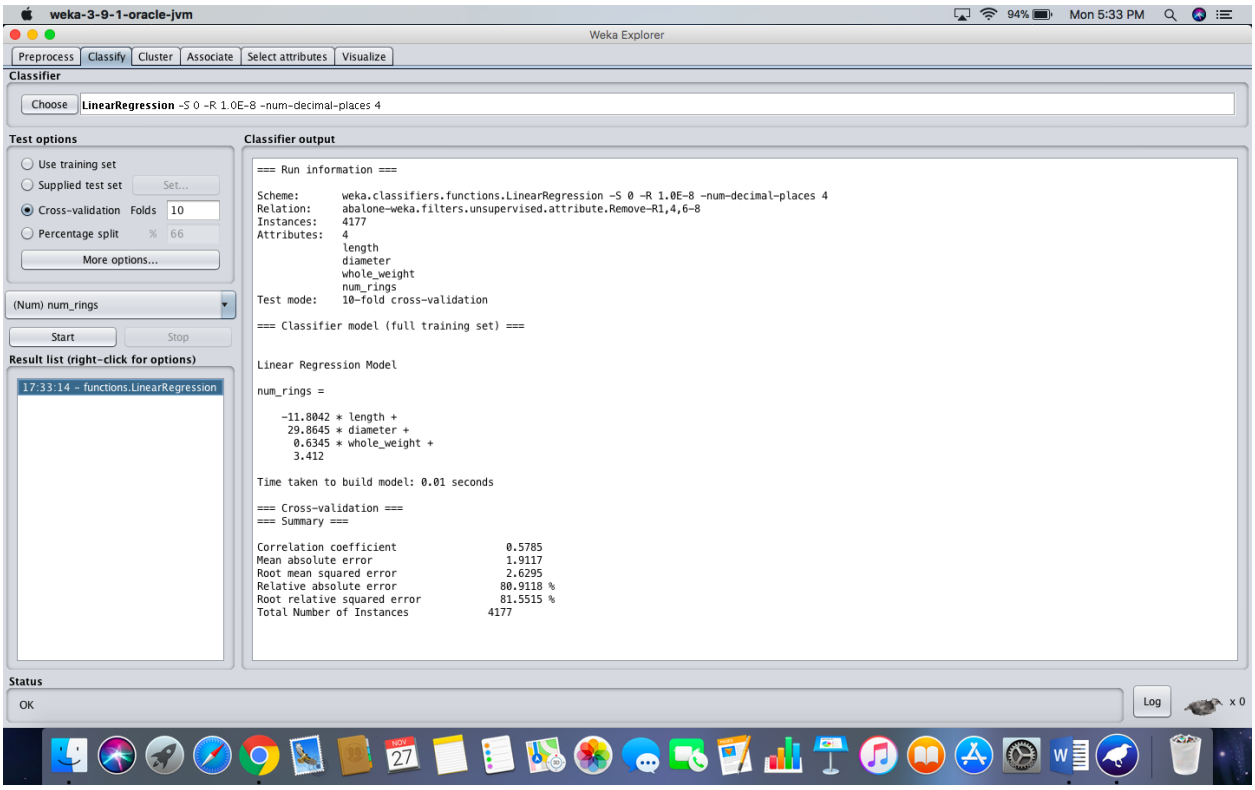

weka-3-9-1-oracle-jvm
Mon 5:30 PM
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 10
Percentage split % 66
More options...
(Num) num_rings
Start
Stop
Result list (right-click for options)
17:30:09 - functions.LinearRegression
Classifier output
=== Run information ===
Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation: abalone
Instances: 4177
Attributes: 9
age
length
diameter
height
whole_weight
shucked_weight
viscera_weight
shell_weight
num_rings
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
Linear Regression Model
num_rings =
0.8607 * age=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
Time taken to build model: 0.18 seconds
=== Cross-validation ===
=== Summary ===
Status
OK
Log
x 0


weka-3-9-1-oracle-jvm
Mon 5:30 PM
Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 10
Percentage split % 66
More options...
(Num) num_rings
Start
Stop
Result list (right-click for options)
17:30:09 - functions.LinearRegression
Classifier output
diameter
height
whole_weight
shucked_weight
viscera_weight
shell_weight
num_rings
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
Linear Regression Model
num_rings =
0.8607 * age=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
Time taken to build model: 0.18 seconds
=== Cross-validation ===
=== Summary ===
Correlation coefficient 0.7268
Mean absolute error 1.5835
Root mean squared error 2.2147
Relative absolute error 67.023 %
Root relative squared error 68.686 %
Total Number of Instances 4177
Status
OK
Log
x 0

[3]

num_rings =

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

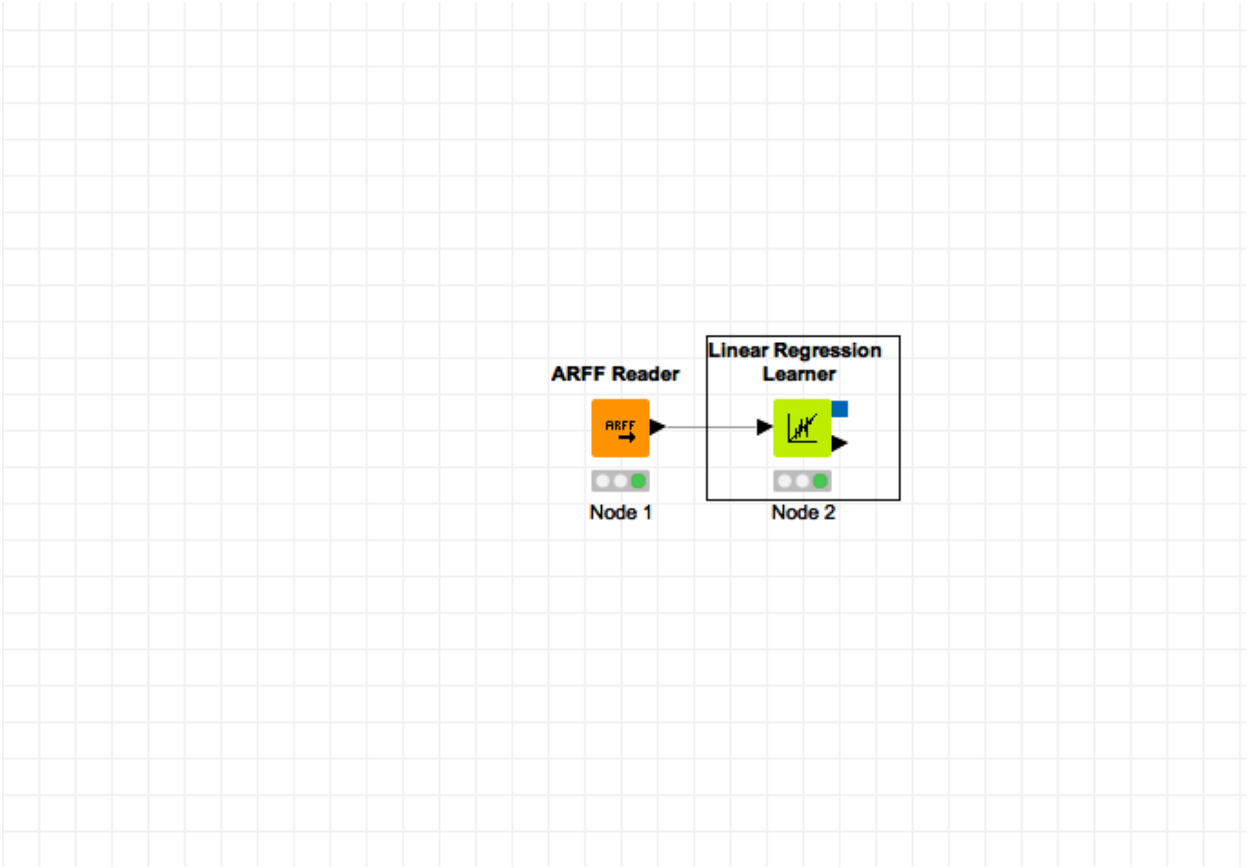


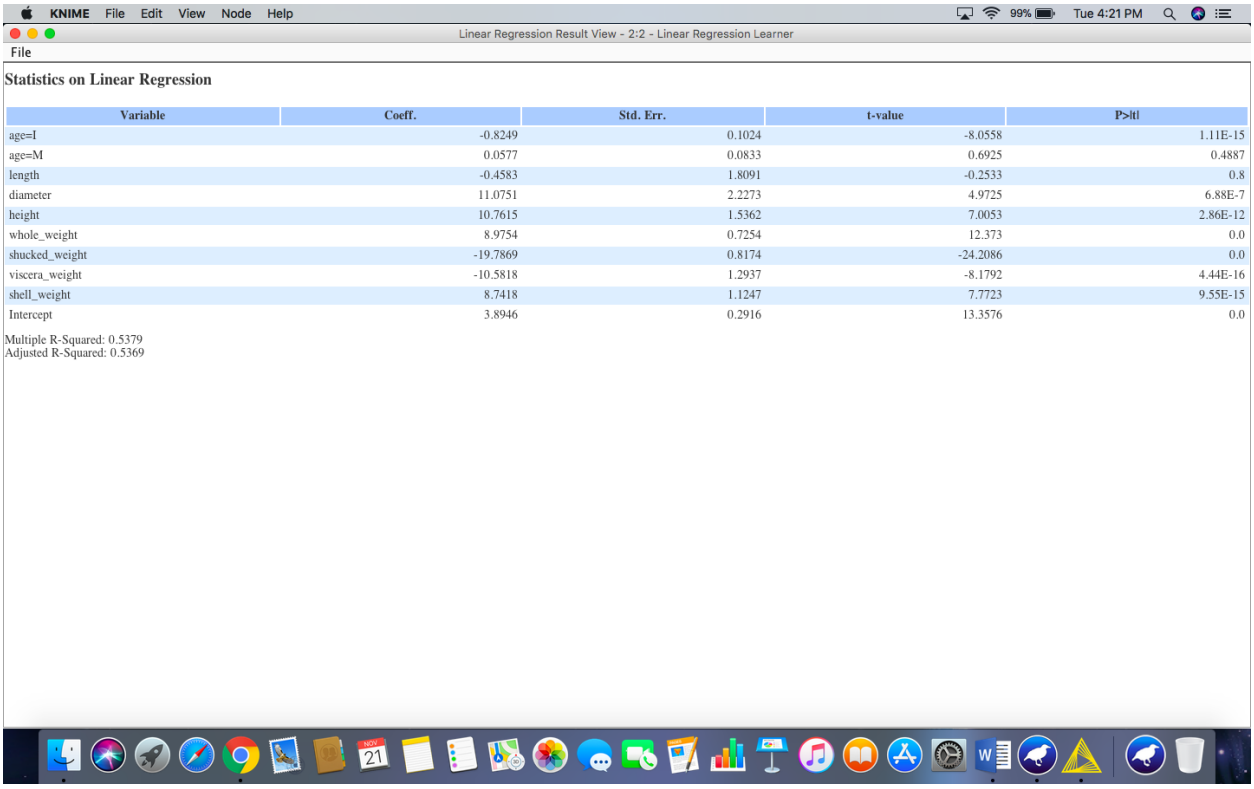
Mean absolute error here is 1.9117

[4]

Linear Equation :
Num_rings=

-0.8249 * age=I +
0.0577 * age=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 +
3.8946





Statistics on Linear Regression

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

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

Compare :

Weka -
0.8607 * age=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

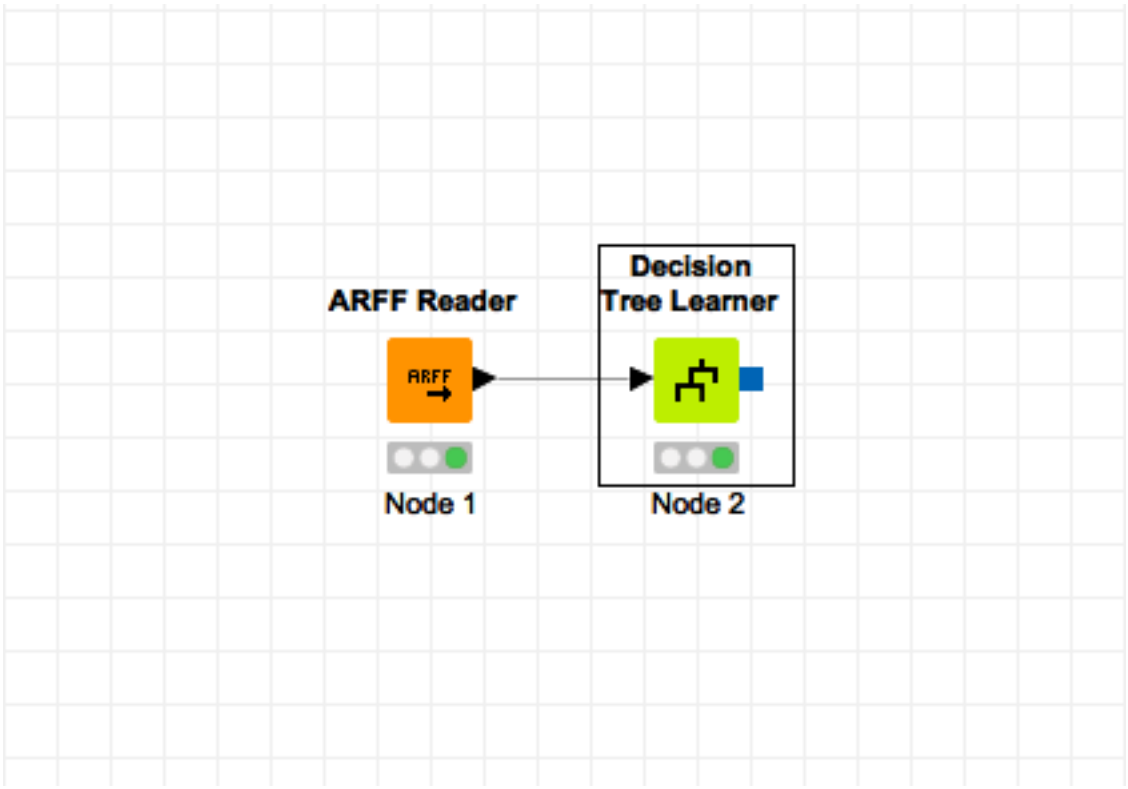
(length coefficient : 0)

KNIME -
-0.8249 * age=I +
0.0577 * age=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 +
3.8946

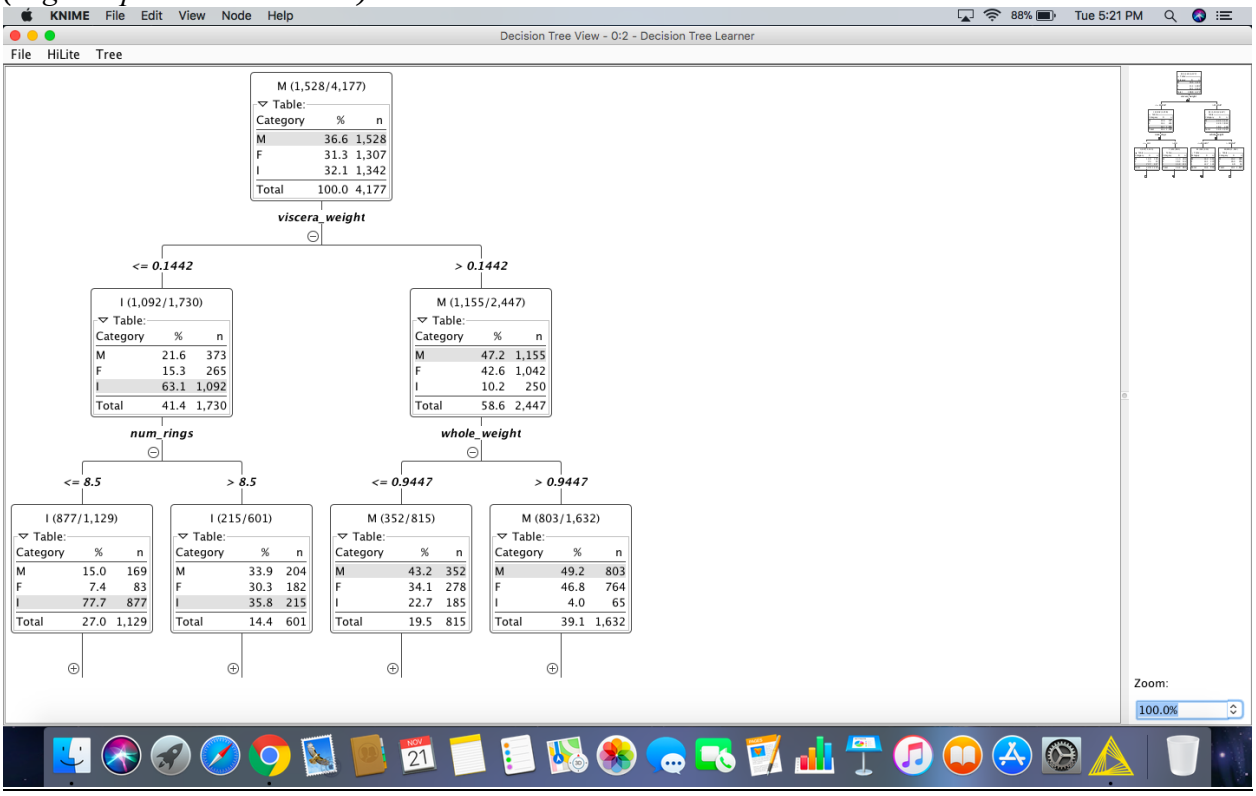
Similar coefficients : length, height, whole_weight, shucked_weight, viscera_weight, shell_weight

(diameter is slightly more than 0.5)

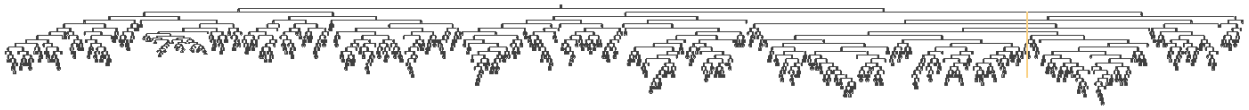
[5]

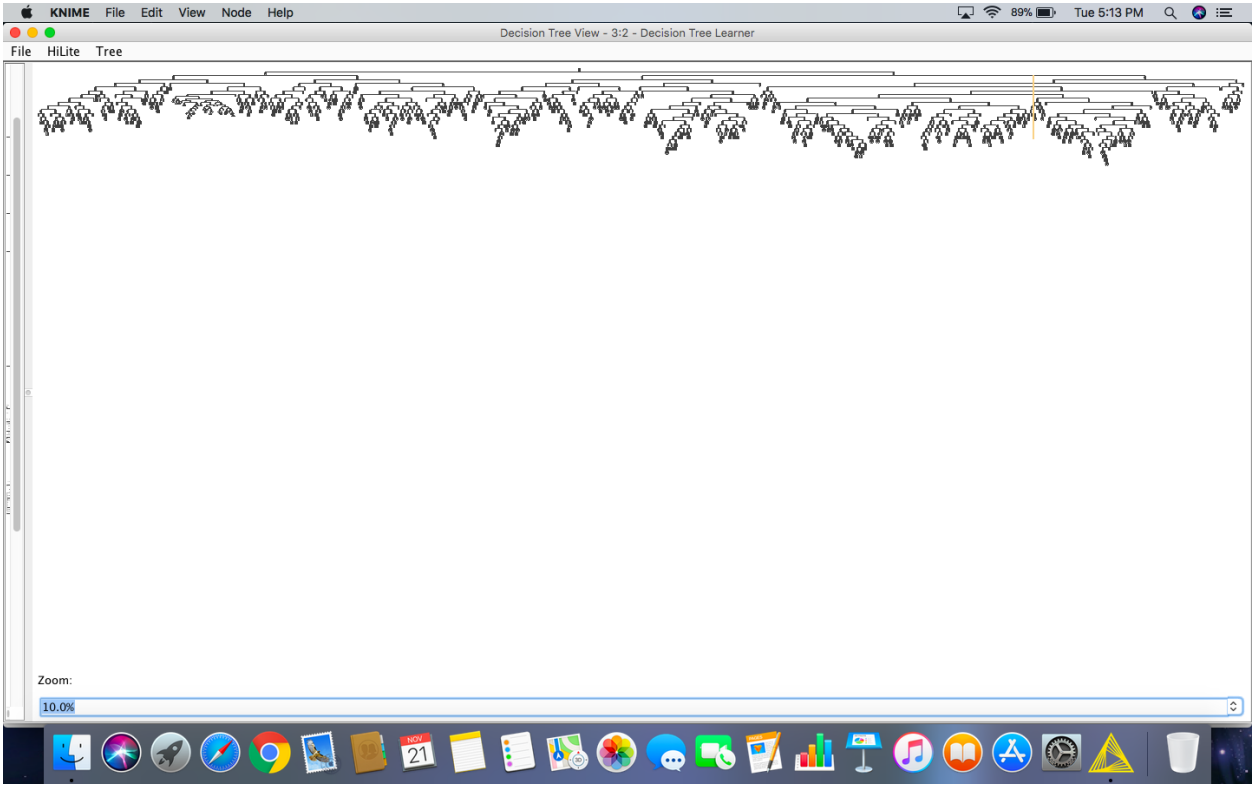


('Age' is predicted variable)



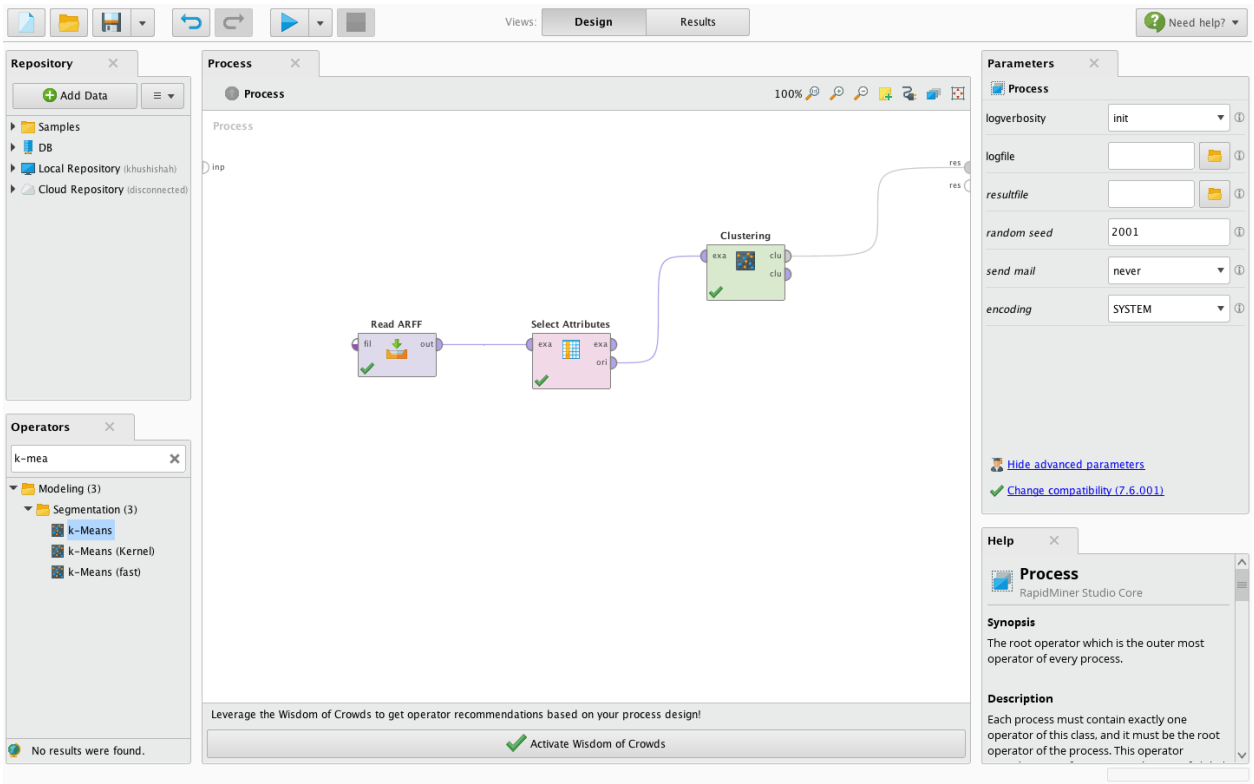
DECISION TREE

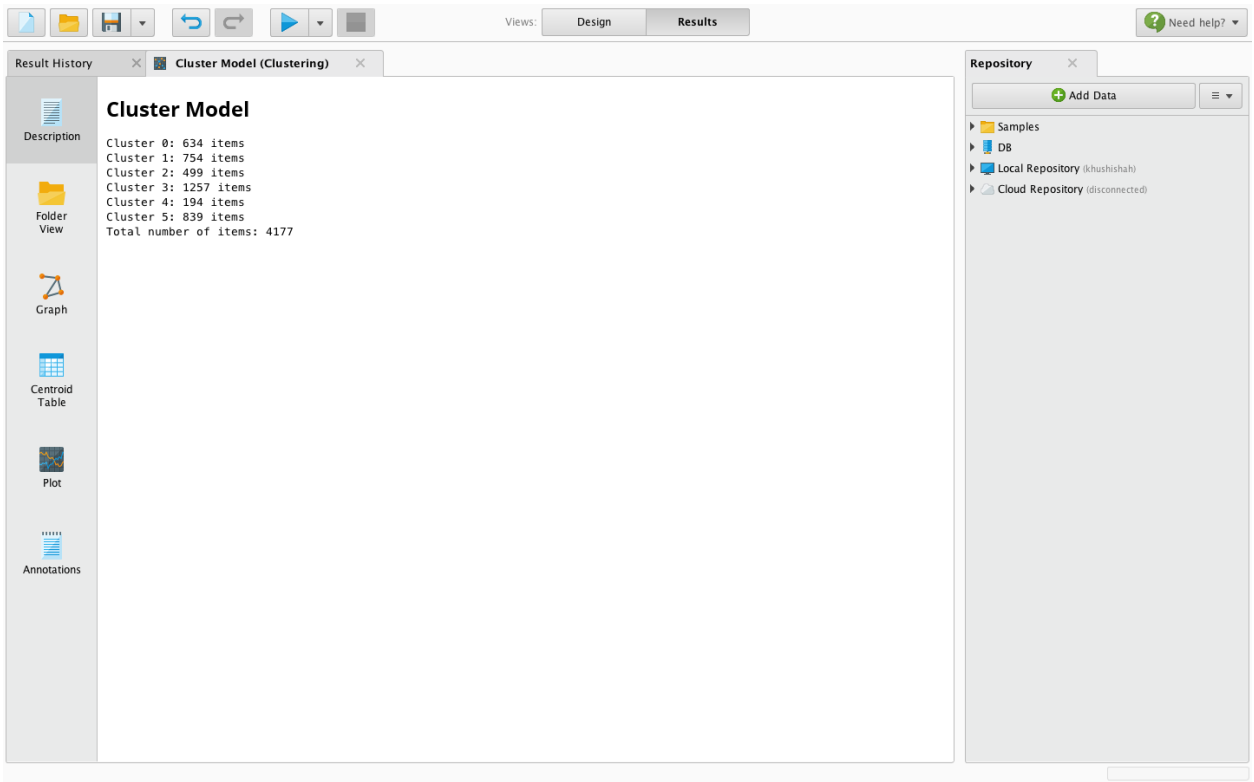




BONUS QUESTION

[1]

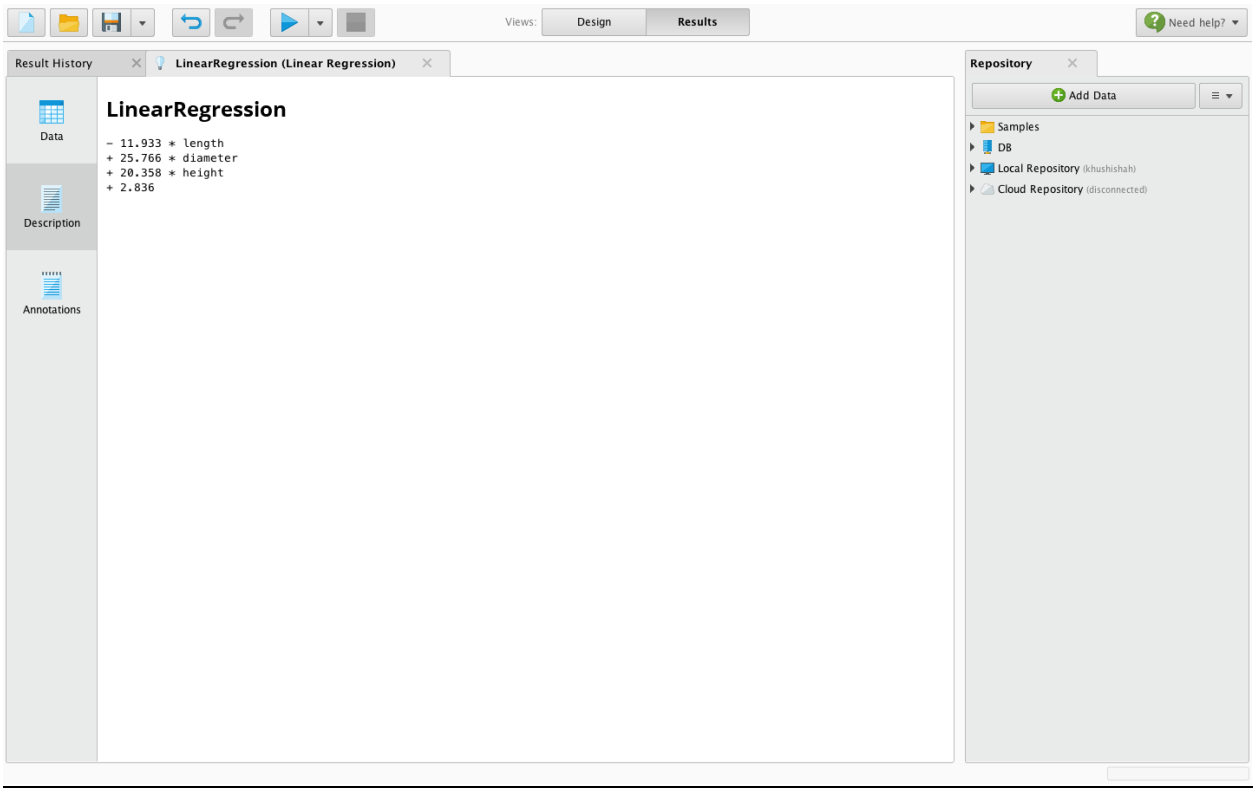




Cluster Model
Cluster 0: 634 items
Cluster 1: 754 items
Cluster 2: 499 items
Cluster 3: 1257 items
Cluster 4: 194 items
Cluster 5: 839 items
Total number of items: 4177

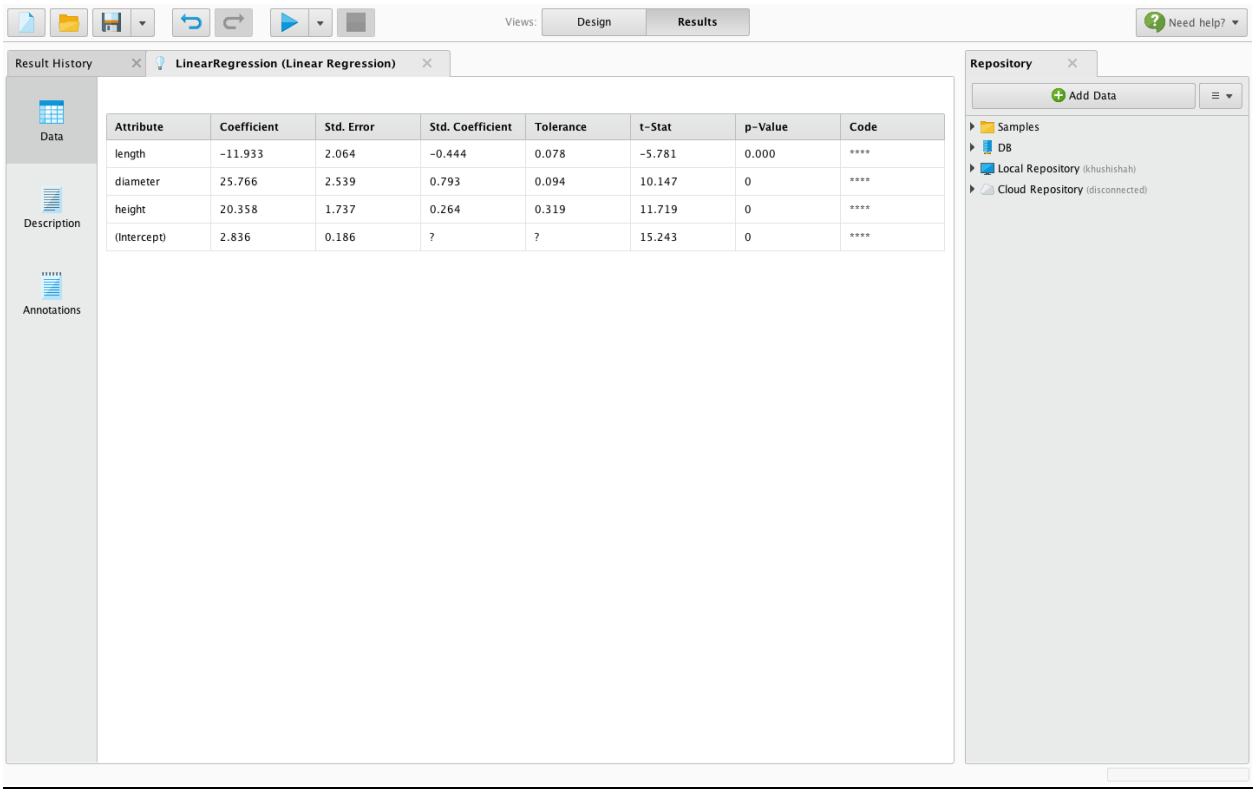
[2]

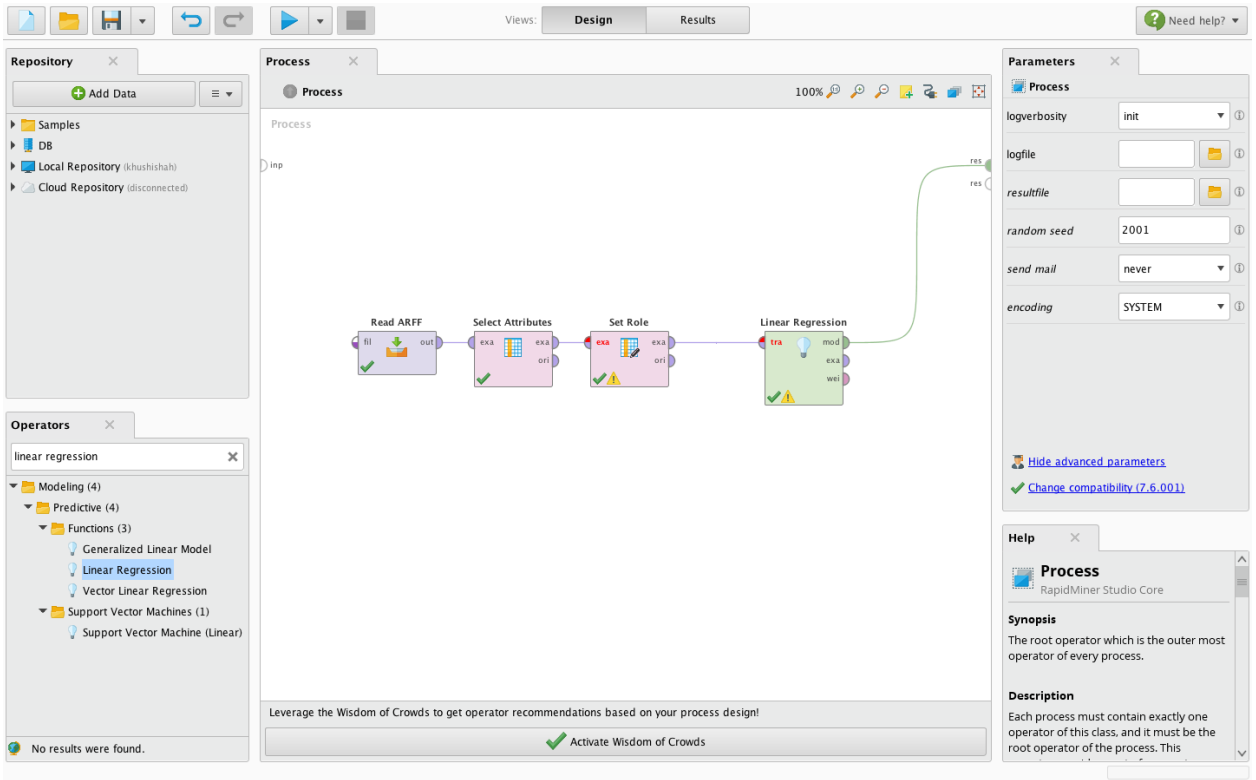
Linear Regression



LinearRegression

```
- 11.933 * length
+ 25.766 * diameter
+ 20.358 * height
+ 2.836
```


Views:
Design
Results
Need help?
Result History
LinearRegression (Linear Regression)
Data
Description
Annotations
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 | ****
Repository
Add Data
Samples
DB
Local Repository (khushishah)
Cloud Repository (disconnected)


Views:
Design
Results
Need help?
Repository
Add Data
Samples
DB
Local Repository (khushishah)
Cloud Repository (disconnected)
Operators
linear regression
Modeling (4)
Predictive (4)
Functions (3)
Generalized Linear Model
Linear Regression
Vector Linear Regression
Support Vector Machines (1)
Support Vector Machine (Linear)
No results were found.
Process
Process
100%
Read ARFF
Select Attributes
Set Role
Linear Regression
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Process
logverbosity
init
logfile
resultfile
random seed
2001
send mail
never
encoding
SYSTEM
Hide advanced parameters
Change compatibility (7.6.001)
Help
Process
RapidMiner Studio Core
Synopsis
The root operator which is the outer most operator of every process.
Description
Each process must contain exactly one operator of this class, and it must be the root operator of the process. This

Linear Regression Equation:

Num_rings = - 11.933 * length
+ 25.766 * diameter
+ 20.358 * height
+ 2.836