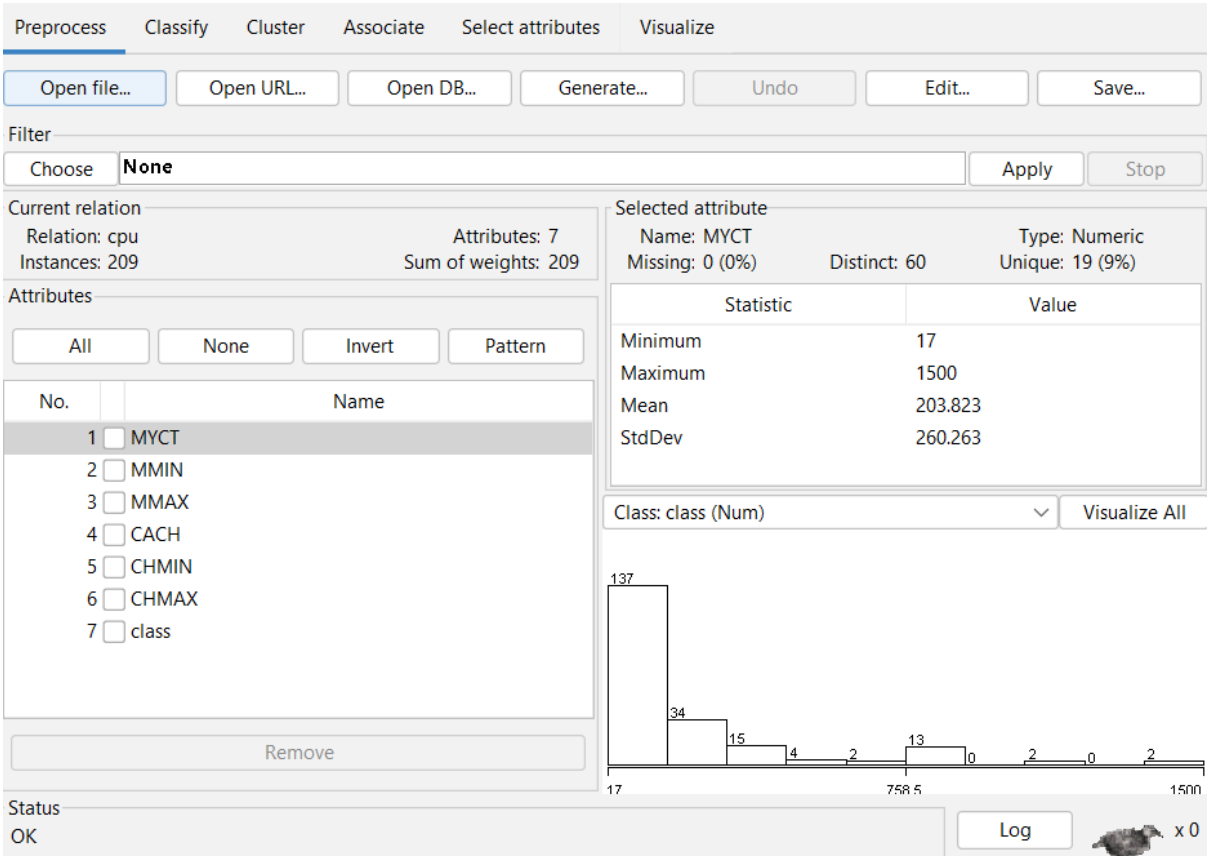
AI _ ML lab

Exploration of WEKA tool

Experiment 1:

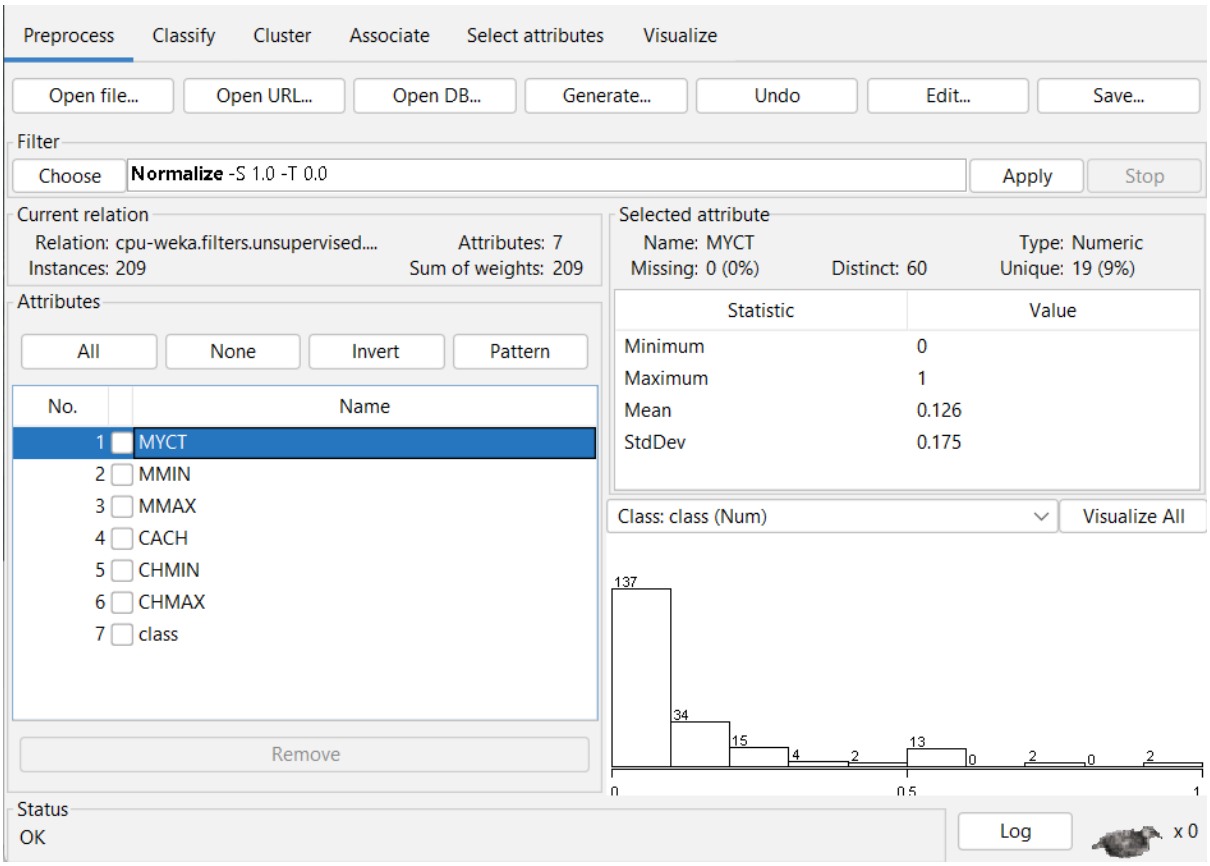
Exploration of WEKA tool for Regression task.

- **Open WEKA**:
  - Launch WEKA and choose the "Explorer" option.
- **Load Dataset**:
  - Click on the "Preprocess" tab.
  - Load a dataset (e.g., cpu.arff from the data folder in WEKA).
  -



Normalize

- For building any regression model first we have to need to do some data scaling beacause you're going to see their each variable has different minimum, maximum,mean & StdDev value .
- For this we have to click " choose"
- Then " unsupervised"
- Then "Normalize"



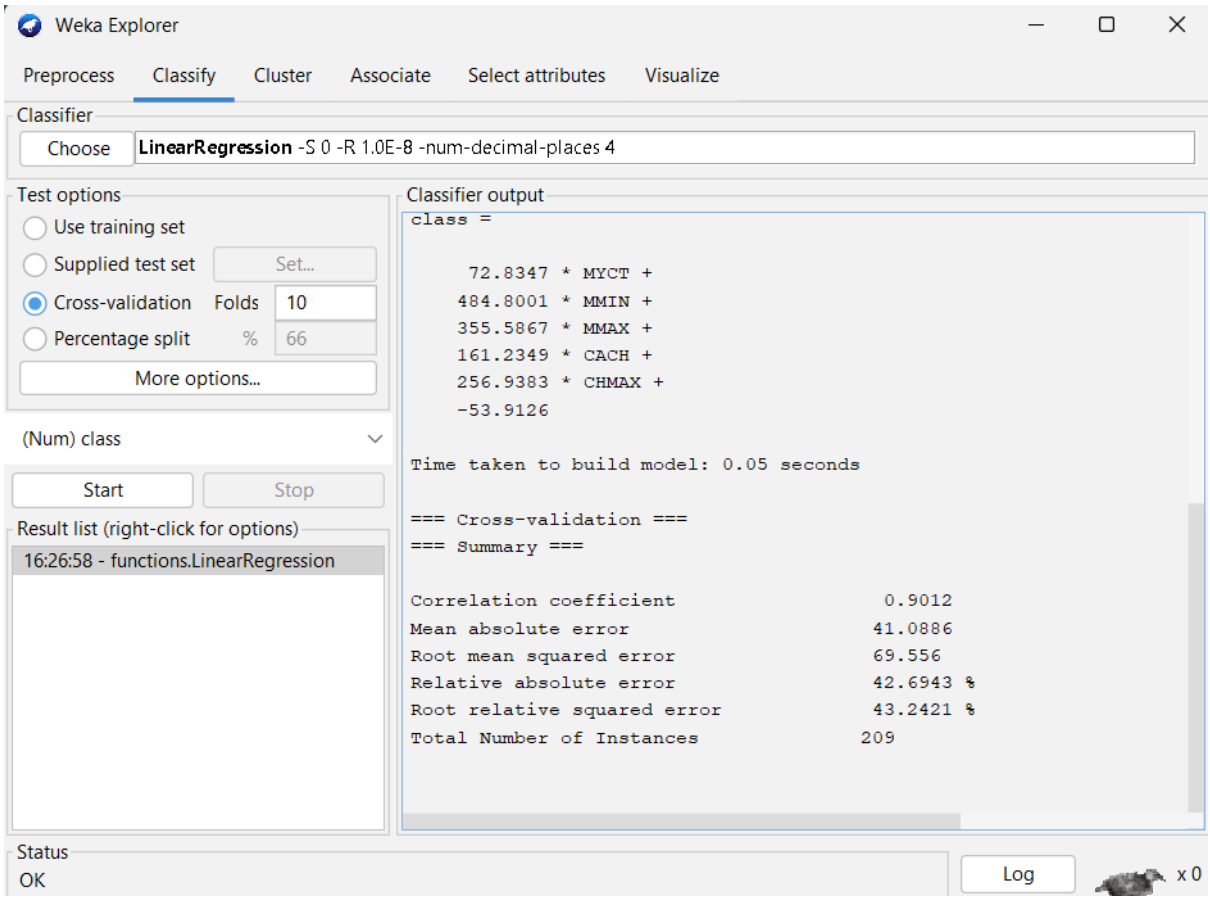
Now all the max and minimum value are 1 & 0.
Now it is comparable.

Select Regression Algorithm:

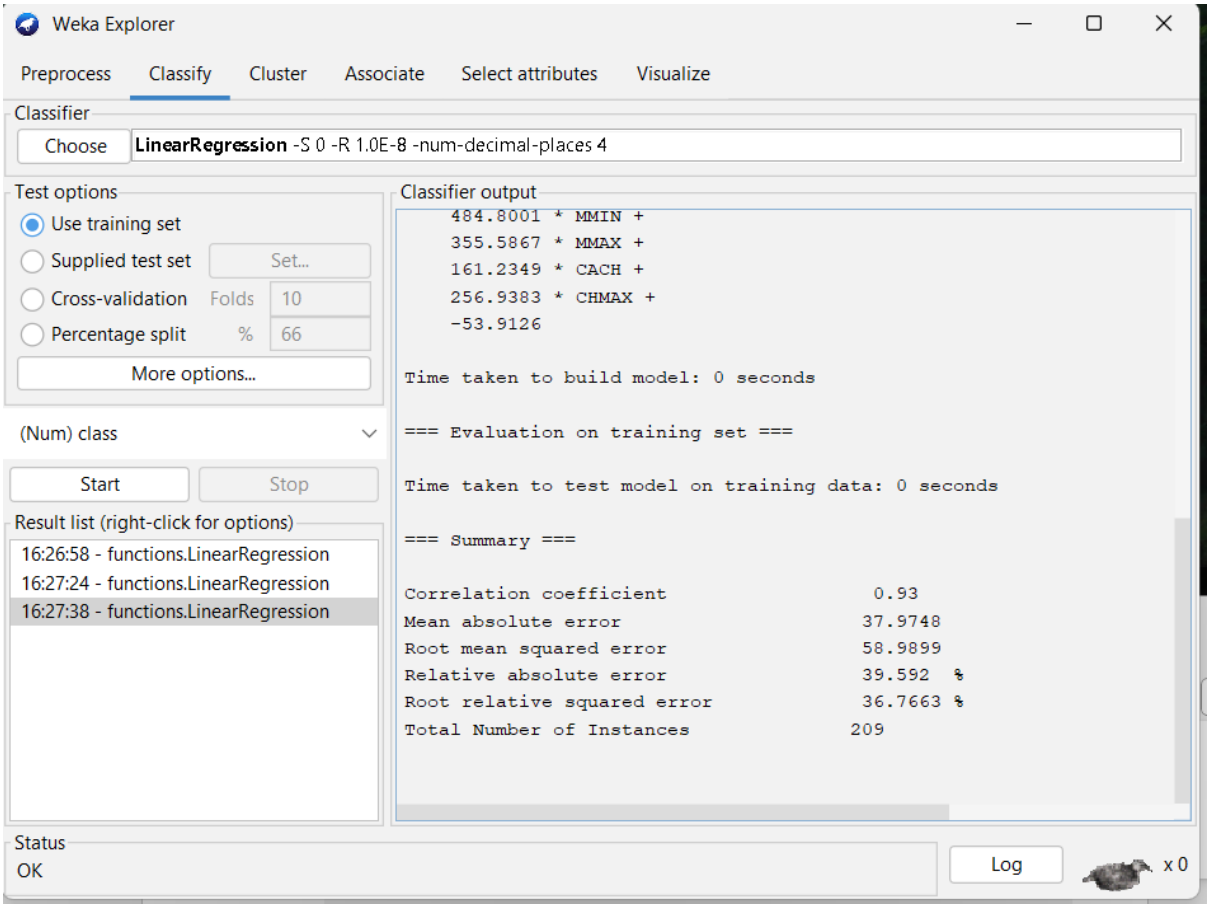
- Navigate to the "Classify" tab.
- Set the "Classifier" to a regression algorithm like LinearRegression.

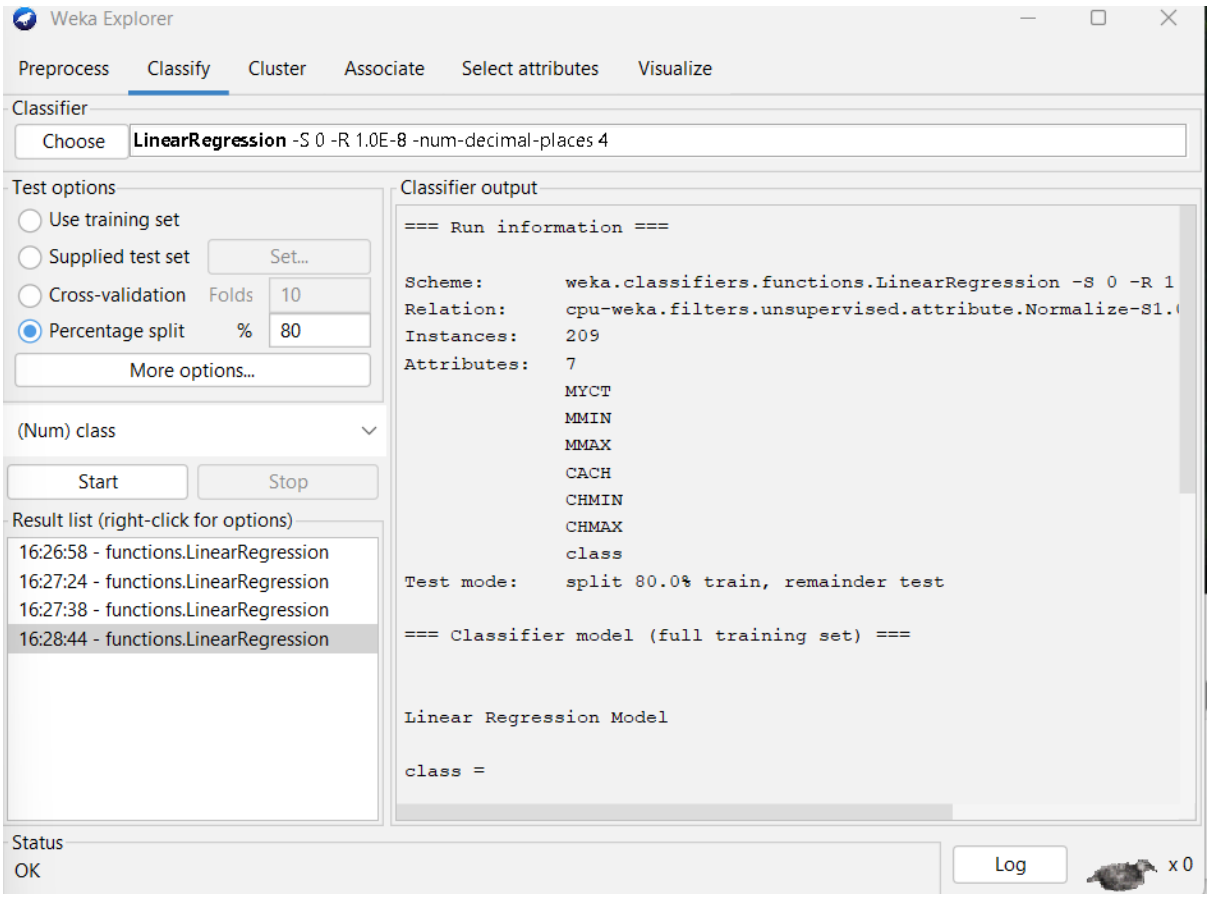
- For cross – validation no.of fold to be 10 (we are going to splitting the data into 10 segments and using 9 segements in order to build a model and 1 left segement is use for prediction . . we will iterated the same model 10 times.)

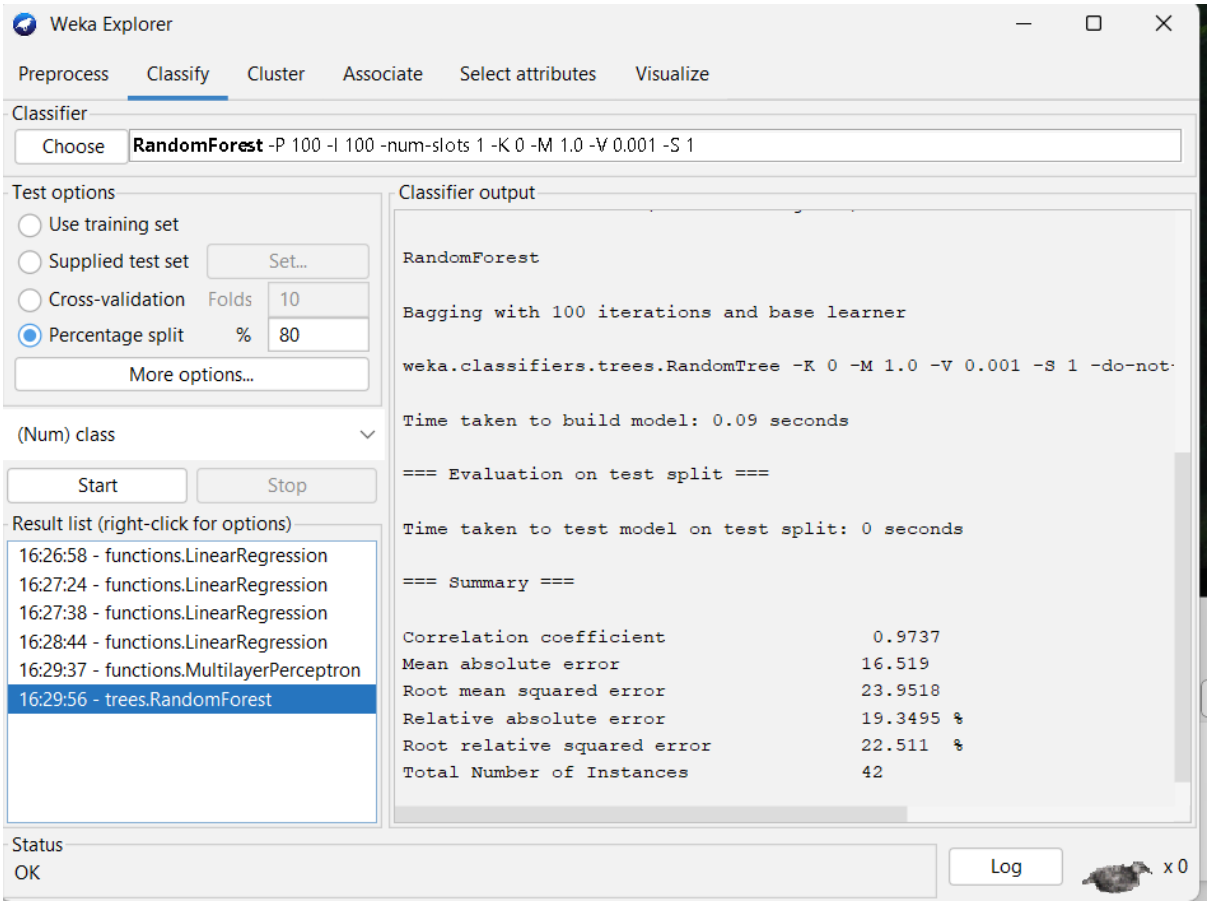
- **Setup and Run**:
  - Choose the target attribute (dependent variable) under the "Class" dropdown.
  - Click "Start" to run the regression.

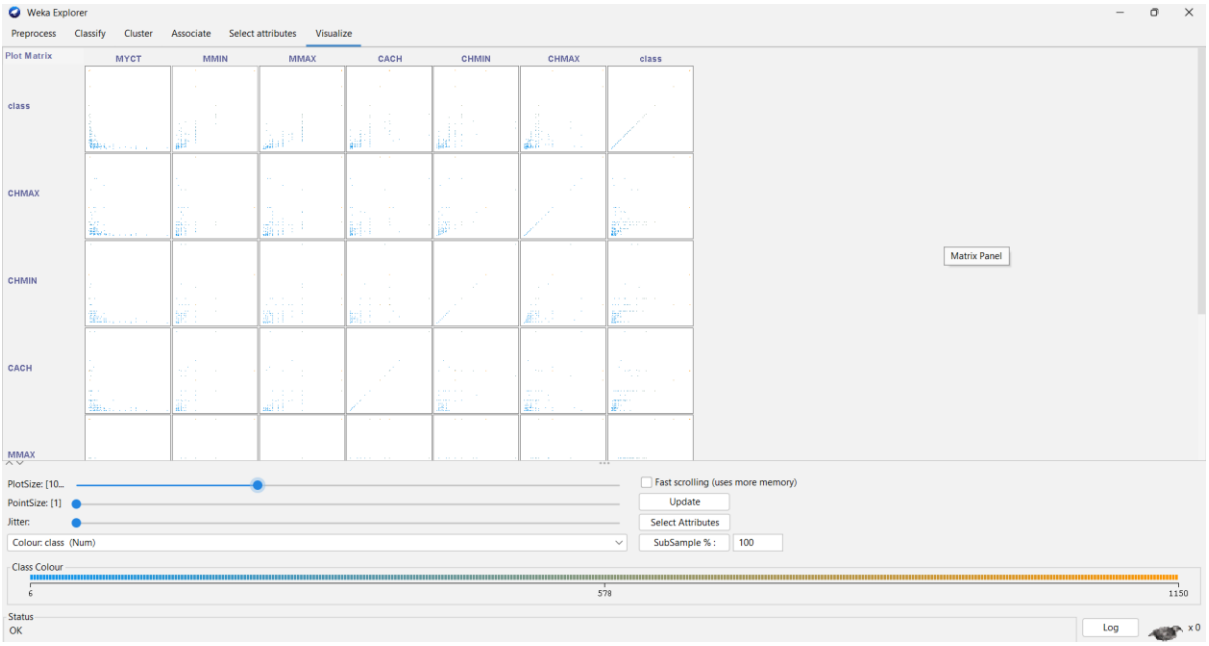


```
Linear Regression Model

class =

      72.8347 * MYCT +
     484.8001 * MMIN +
     355.5867 * MMAX +
     161.2349 * CACH +
     256.9383 * CHMAX +
     -53.9126

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correlation coefficient                  0.93
Mean absolute error                     37.9748
Root mean squared error                 58.9899
```


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...
Classifier output
484.8001 * MMIN +
355.5867 * MMAX +
161.2349 * CACH +
256.9383 * CHMAX +
-53.9126
Time taken to build model: 0 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds
=== Summary ===
Correlation coefficient 0.93
Mean absolute error 37.9748
Root mean squared error 58.9899
Relative absolute error 39.592 %
Root relative squared error 36.7663 %
Total Number of Instances 209
(Num) class
Start
Stop
Result list (right-click for options)
16:26:58 - functions.LinearRegression
16:27:24 - functions.LinearRegression
16:27:38 - functions.LinearRegression
Status
OK
Log
x 0


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
%
80
More options...
Classifier output
=== Run information ===
Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1
Relation: cpu-weka.filters.unsupervised.attribute.Normalize-S1.
Instances: 209
Attributes: 7
MYCT
MMIN
MMAX
CACH
CHMIN
CHMAX
class
Test mode: split 80.0% train, remainder test
=== Classifier model (full training set) ===
Linear Regression Model
class =
(Num) class
Start
Stop
Result list (right-click for options)
16:26:58 - functions.LinearRegression
16:27:24 - functions.LinearRegression
16:27:38 - functions.LinearRegression
16:28:44 - functions.LinearRegression
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
80
More options...
Classifier output
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds
=== Summary ===
Correlation coefficient 0.9737
Mean absolute error 16.519
Root mean squared error 23.9518
Relative absolute error 19.3495 %
Root relative squared error 22.511 %
Total Number of Instances 42
(Num) class
Start
Stop
Result list (right-click for options)
16:26:58 - functions.LinearRegression
16:27:24 - functions.LinearRegression
16:27:38 - functions.LinearRegression
16:28:44 - functions.LinearRegression
16:29:37 - functions.MultilayerPerceptron
16:29:56 - trees.RandomForest
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
MYCT
MMIN
MMAX
CACH
CHMIN
CHMAX
class
Matrix Panel
PlotSize: [10...
PointSize: [1]
Jitter:
Colour: class (Num)
Fast scrolling (uses more memory)
Update
Select Attributes
SubSample % :
100
Class Colour
6
578
1150
Status
OK
Log
x 0