

Build Regression Model Using WEKA

ISE4132 : AI Application System



AMIN AL (아민알)
Integrated System Engineering (ISE)
010-6853-6648
alaminanik@inha.ac.kr

WEKA Software Download



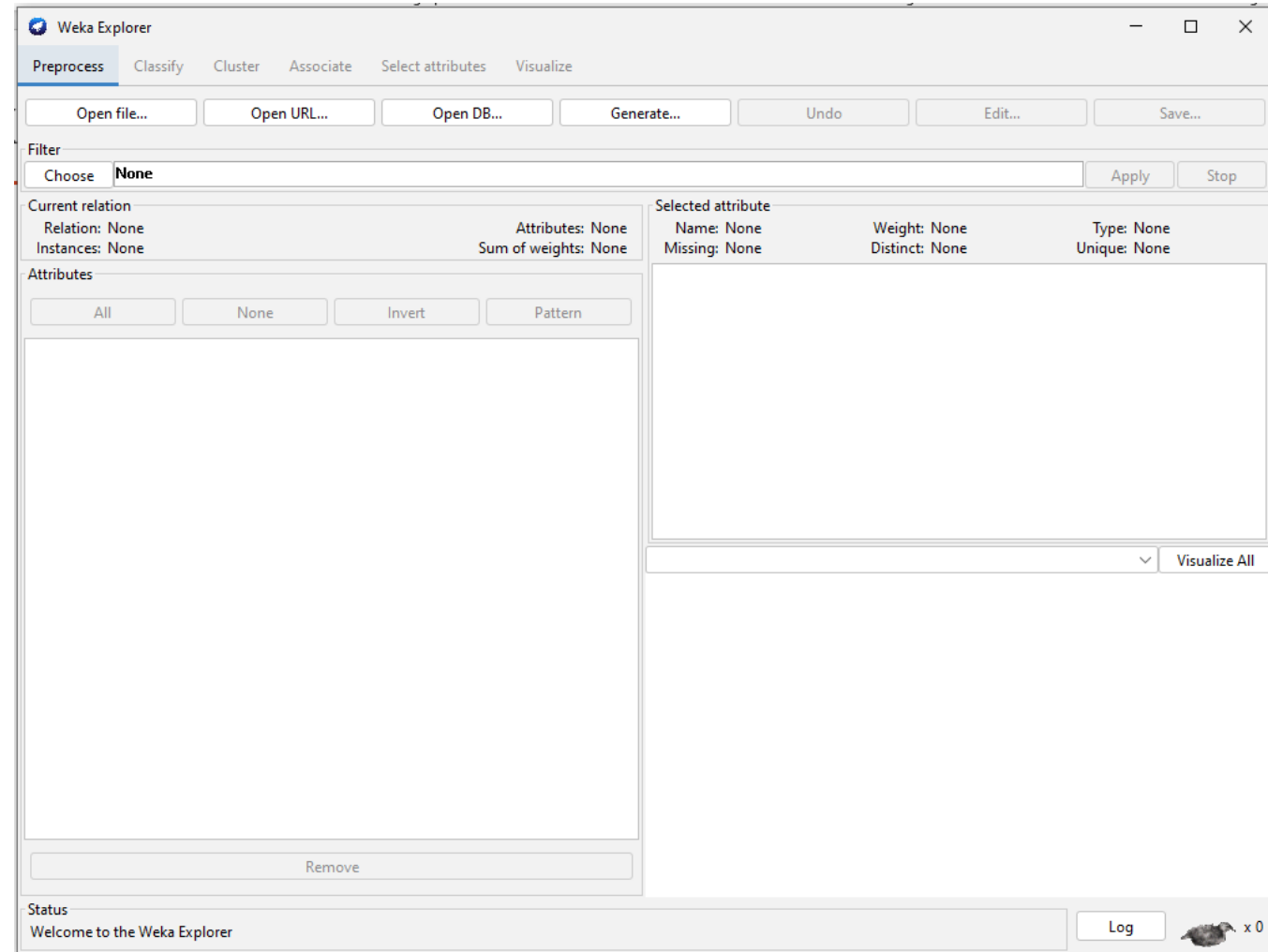
[Download Link](#)



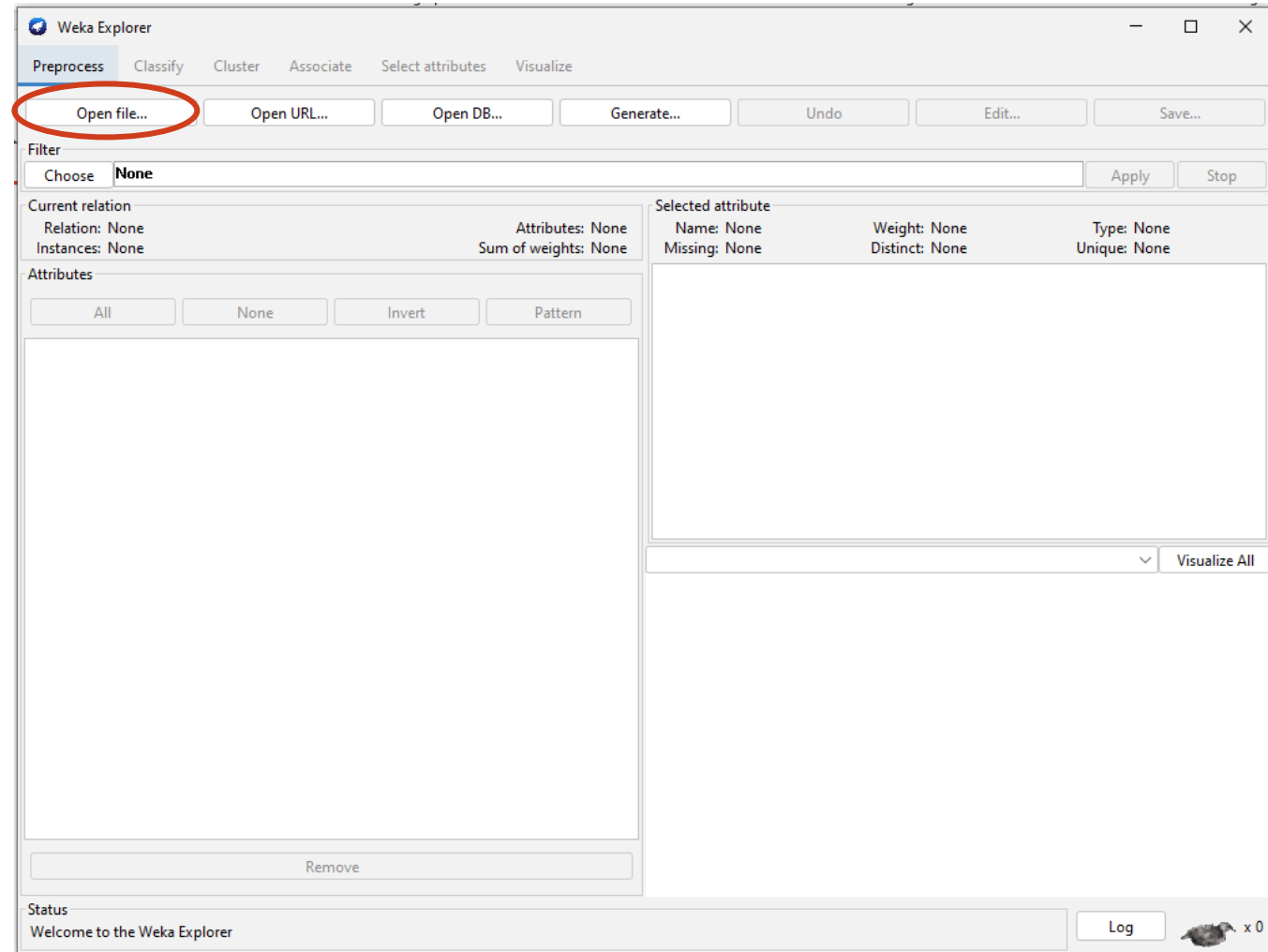
Selecting Dataset



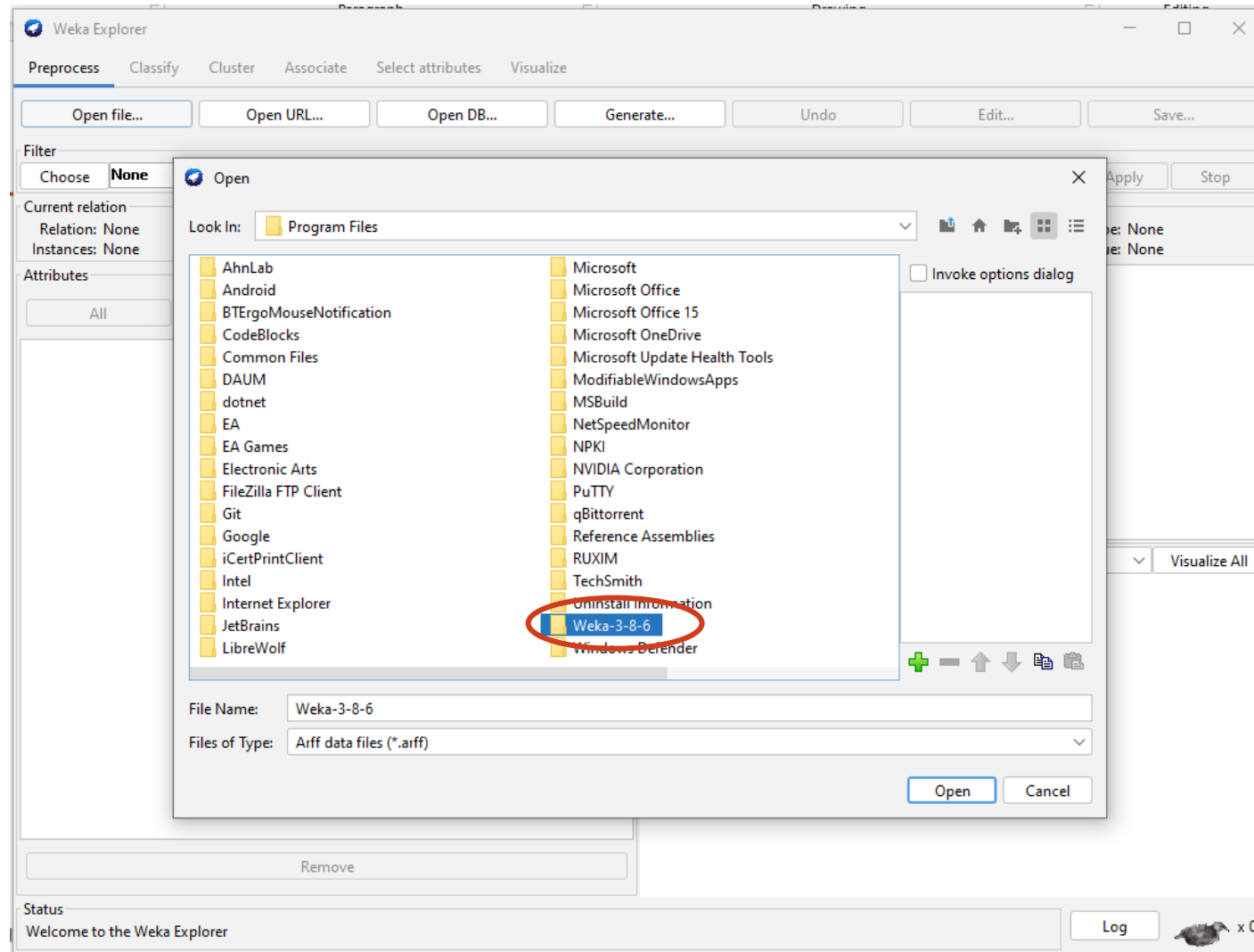
Selecting Dataset



Selecting Dataset



Selecting Dataset



Weka-3-8-6 > data > cpu

Selecting Dataset



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **None** Apply Stop

Current relation
Relation: cpu
Instances: 209
Attributes: 7
Sum of weights: 209

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> MYCT
2	<input type="checkbox"/> MMIN
3	<input type="checkbox"/> MMAX
4	<input type="checkbox"/> CACH
5	<input type="checkbox"/> CHMIN
6	<input type="checkbox"/> CHMAX
7	<input type="checkbox"/> class

Remove

Selected attribute
Name: MYCT
Missing: 0 (0%)
Distinct: 60
Type: Numeric
Unique: 19 (9%)

Statistic	Value
Minimum	17
Maximum	1500
Mean	203.823
StdDev	260.263

Class: class (Num) Visualize All

Status
OK

Log x 0

Min Max Normalization



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose **Normalize -S 1.0 -T 0.0** Apply Stop

Current relation
Relation: cpu
Instances: 209
Attributes: 7
Sum of weights: 209

Attributes

All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> MYCT
2	<input type="checkbox"/> MMIN
3	<input type="checkbox"/> MMAX
4	<input type="checkbox"/> CACH
5	<input type="checkbox"/> CHMIN
6	<input type="checkbox"/> CHMAX
7	<input type="checkbox"/> class

Remove

Status
OK

Selected attribute
Name: MYCT
Missing: 0 (0%)
Distinct: 60
Type: Numeric
Unique: 19 (9%)

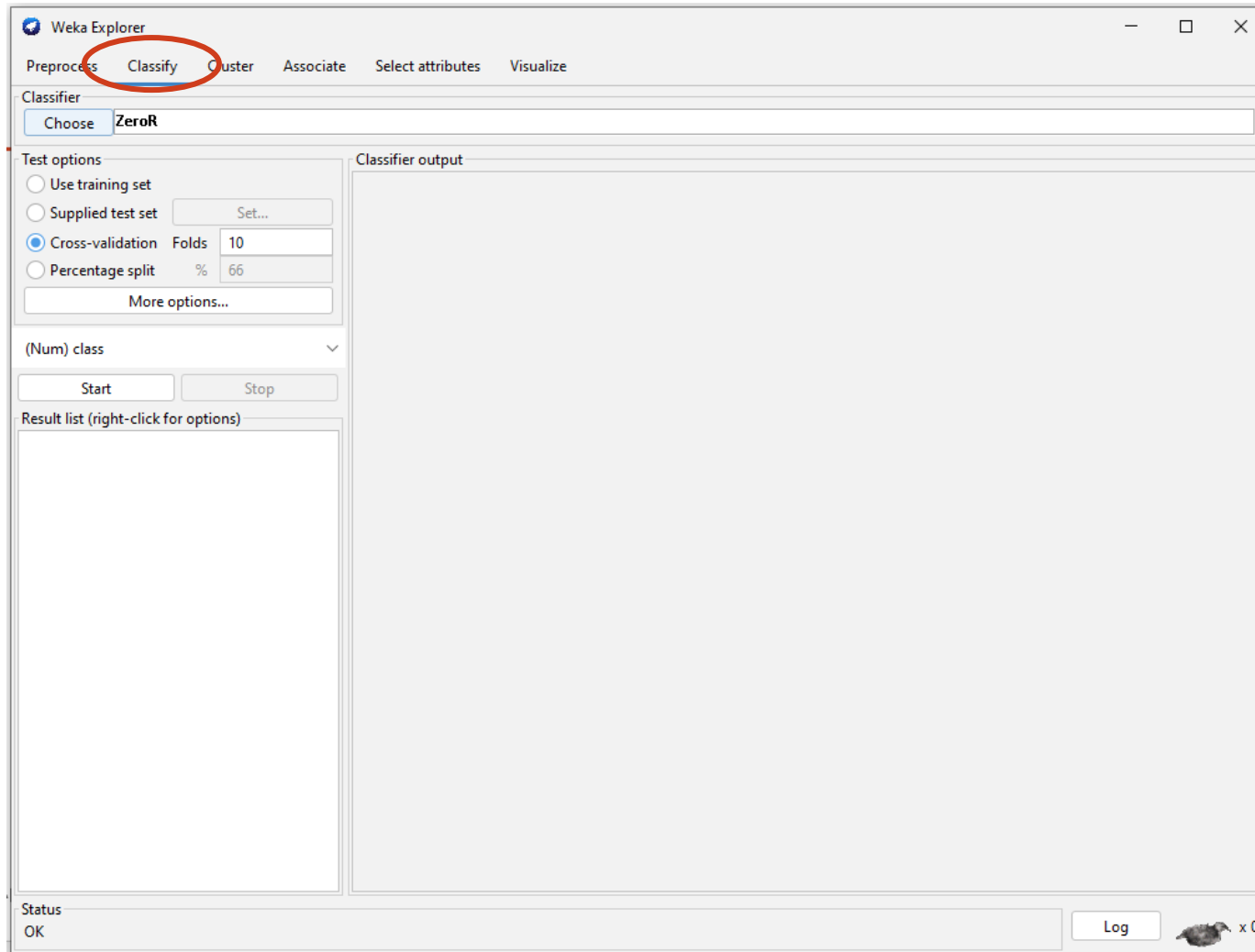
Statistic	Value
Minimum	17
Maximum	1500
Mean	203.823
StdDev	260.263

Class: class (Num) Visualize All

Log x 0

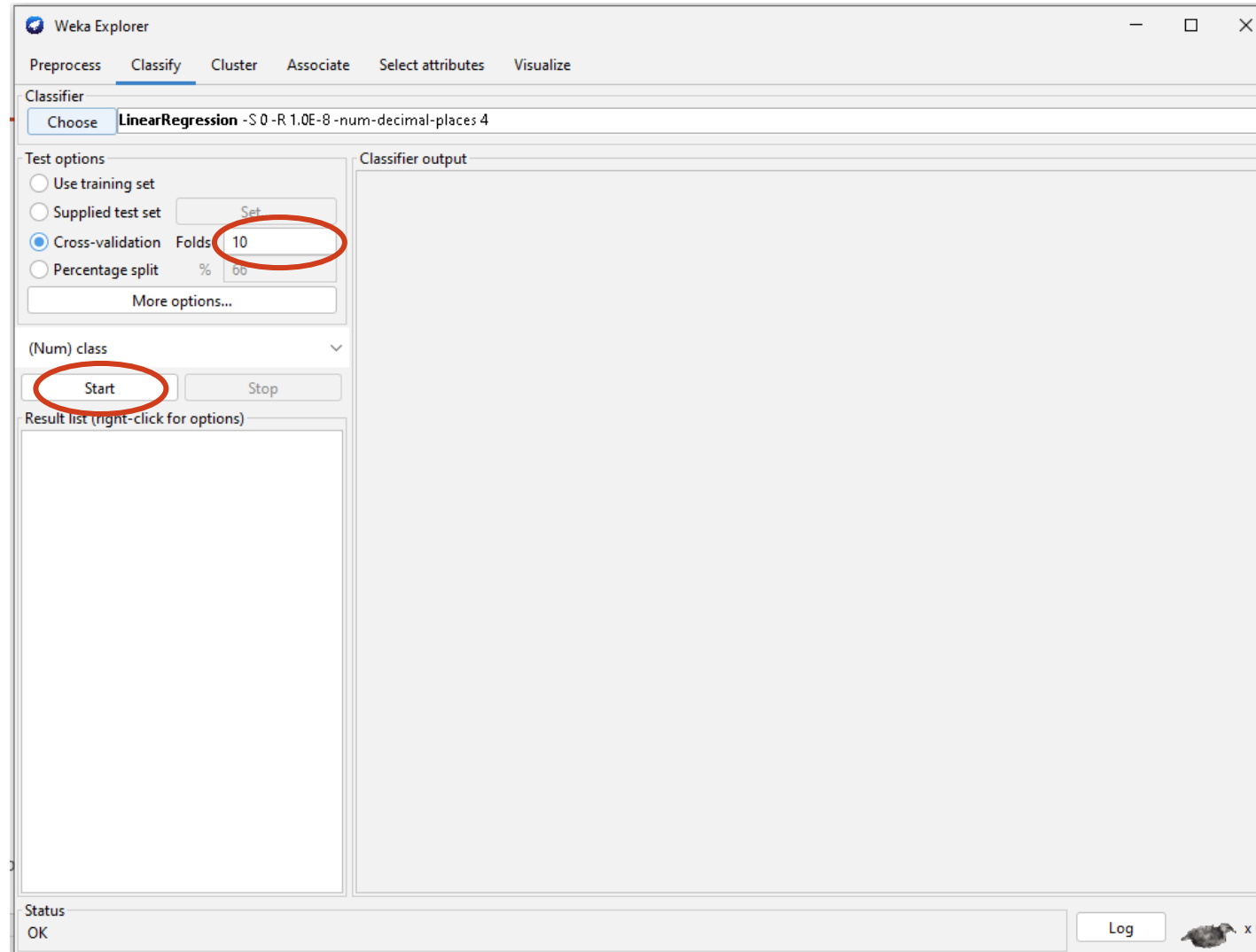
Filter > Choose > filters >
unsupervised > attribute >
Normalize > apply

Build ML Model



Choose > classifiers > function >
linear regression

Build ML Model Using LR



LR Model Output Using Cross Validation



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) class

Start Stop

Result list (right-click for options)

16:08:37 - functions.LinearRegression

Classifier output

MMA
CACH
CHMIN
CHMAX
class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Linear Regression Model

class =

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

Time taken to build model: 0.05 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9012
Mean absolute error	41.0886
Root mean squared error	69.556
Relative absolute error	42.6943 %
Root relative squared error	43.2421 %
Total Number of Instances	209

Status

OK

Log x 1

LR Model Output Using Training Set



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) class

Start Stop

Result list (right-click for options)

16:08:37 - functions.LinearRegression

Classifier output

MMAX
CACH
CHMIN
CHMAX
class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

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: 0.05 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9012
Mean absolute error	41.0886
Root mean squared error	69.556
Relative absolute error	42.6943 %
Root relative squared error	43.2421 %
Total Number of Instances	209

Status: OK

Log x 0

LR Model Output Using Training Set



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) class

Start Stop

Result list (right-click for options)

- 16:08:37 - functions.LinearRegression
- 16:12:30 - functions.LinearRegression**

Classifier output

CHMAX
class

Test mode: evaluate on training data

=== Classifier model (full training set) ===

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
Relative absolute error	39.592 %
Root relative squared error	36.7663 %
Total Number of Instances	209

Status
OK

Log x 1

LR Model Output Using Split



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...

(Num) class: (Num) class

Start Stop

Result list (right-click for options):

- 16:08:37 - functions.LinearRegression
- 16:12:30 - functions.LinearRegression

Classifier output:

CHMAX
class

Test mode: evaluate on training data

=== Classifier model (full training set) ===

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
Relative absolute error	39.592 %
Root relative squared error	36.7663 %
Total Number of Instances	209

Status: OK

Log x 0

LR Model Output Using Split



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...

(Num) class
Start Stop

Result list (right-click for options)
16:08:37 - functions.LinearRegression
16:12:30 - functions.LinearRegression
16:15:35 - functions.LinearRegression

Classifier output

```
CHMAX
class
Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

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 test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correlation coefficient          0.9186
Mean absolute error             36.6448
Root mean squared error        48.1821
Relative absolute error        42.9238 %
Root relative squared error    45.2838 %
Total Number of Instances      42
```

Status
OK

Log x 1

Build ML Model Using SVR



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **SMOreg** -C 1.0 -N 0 -I "weka.classifiers.functions.supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "weka.classifiers.functions.supportVector.PolyKern"

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds 10

☒ Percentage split % 80

(Num) class

Result list (right-click for options)

- 16:08:37 - functions.LinearRegression
- 16:12:30 - functions.LinearRegression
- 16:15:35 - functions.LinearRegression
- 19:51:57 - functions.SMOreg**

Classifier output

=== Classifier model (full training set) ===

SMOreg

weights (not support vectors):

- + 0.01 * (normalized) MYCT
- + 0.4321 * (normalized) MMIN
- + 0.1847 * (normalized) MMAX
- + 0.1175 * (normalized) CACH
- + 0.0973 * (normalized) CHMIN
- + 0.0235 * (normalized) CHMAX
- 0.0168

Number of kernel evaluations: 21945 (93.081% cached)

Time taken to build model: 0.04 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correlation coefficient	0.9378
Mean absolute error	22.0259
Root mean squared error	35.7805
Relative absolute error	25.7999 %
Root relative squared error	33.6281 %
Total Number of Instances	42

Status OK

x 0

Build ML Model Using RF



The screenshot shows the Weka Explorer interface. The 'Classifier' tab is active, and the 'RandomForest' classifier is selected in the left pane. The right pane displays the model configuration and evaluation results.

Classifier: supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "weka.classifiers.functions.supportVector.PolyKern

model (full training set) ==

support vectors):

- * (normalized) MYCT
- * (normalized) MMIN
- * (normalized) MMAX
- * (normalized) CACH
- * (normalized) CHMIN
- * (normalized) CHMAX

1 evaluations: 21945 (93.081% cached)

Build model: 0.04 seconds

on test split ==

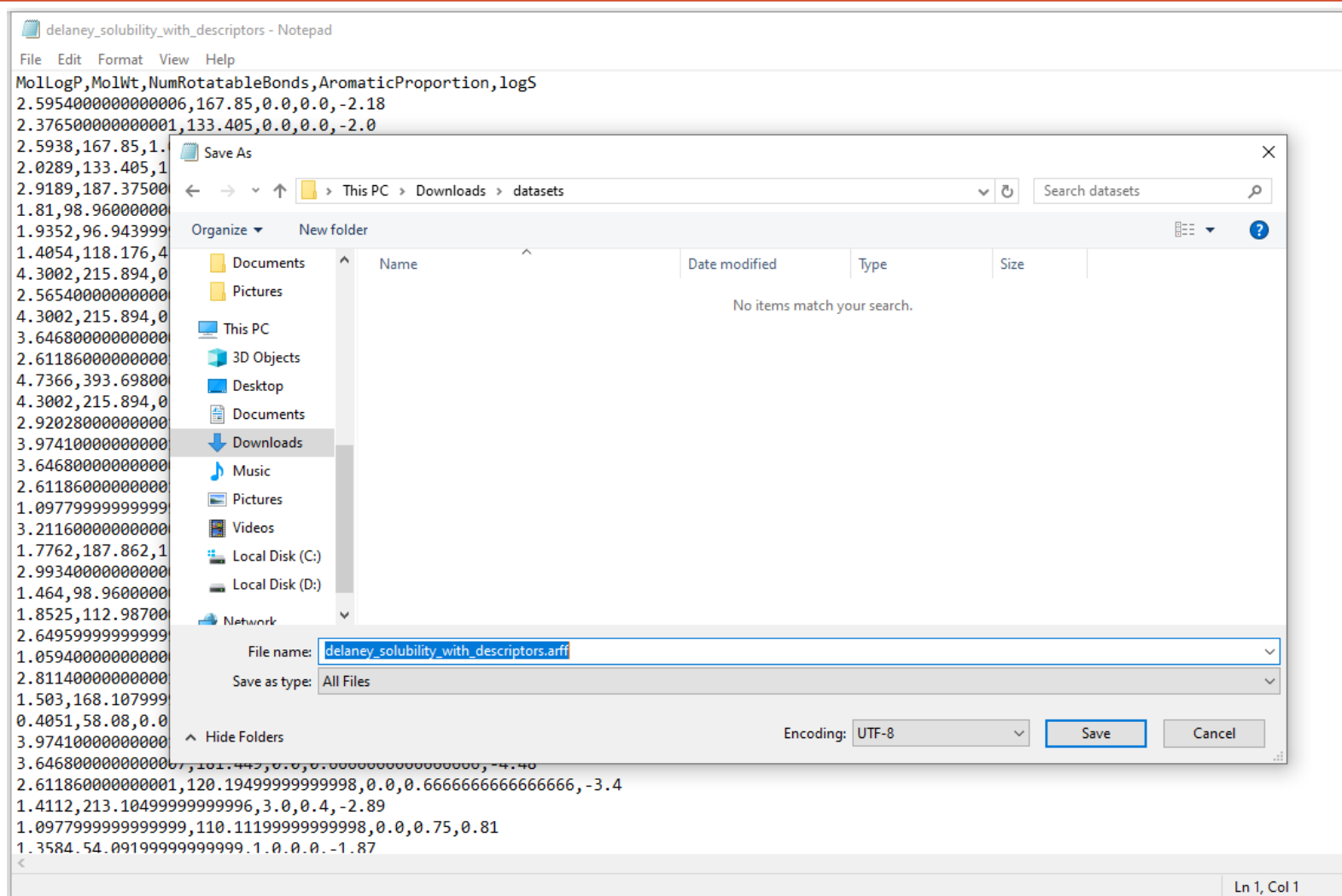
Test model on test split: 0 seconds

Efficient	0.9378
Error	22.0259
ed error	35.7805
te error	25.7999 %
Root relative squared error	33.6281 %
Total Number of Instances	42

Status: OK

Log x 0

Converting a Dataset



txt to arff

Modify to Import The Dataset into WEKA



@relation 'delaney'

@attribute MolLogP numeric

@attribute MolWt numeric

@attribute NumRotatableBonds numeric

@attribute AromaticProportion numeric

@attribute logS numeric

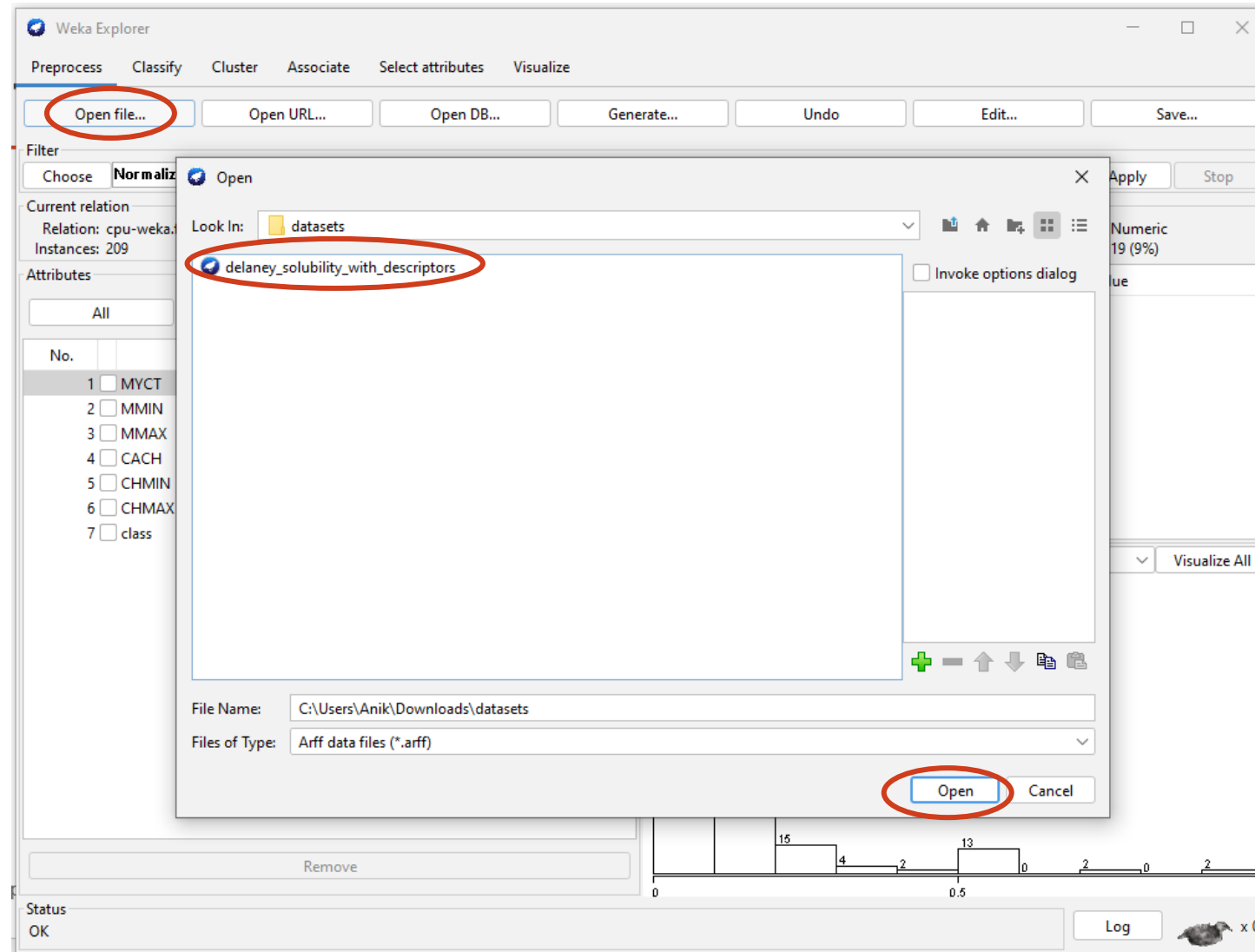
@data

```
*delaney_solubility_with_descriptors - Notepad
File Edit Format View Help
@relation 'delaney'

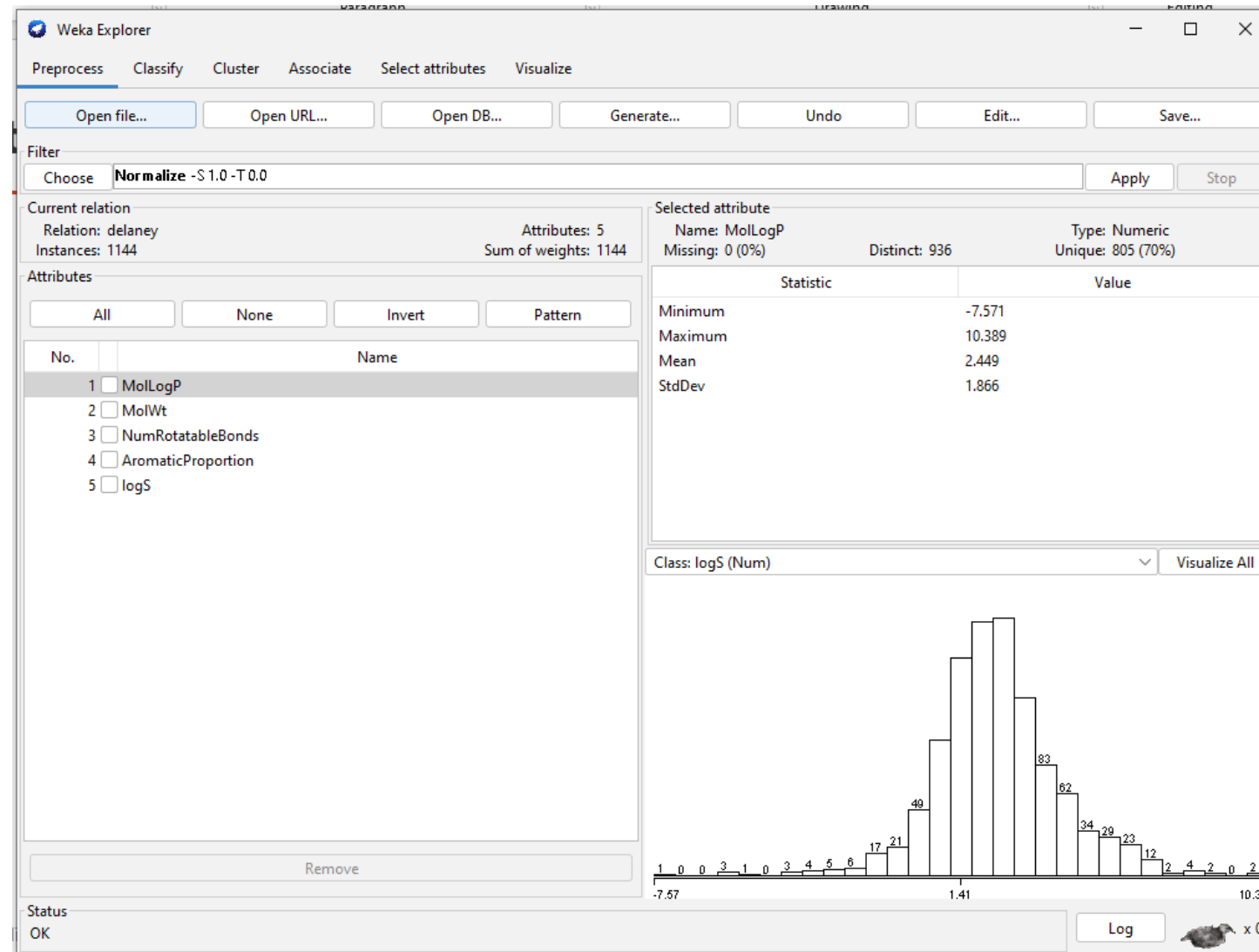
@attribute MolLogP numeric
@attribute MolWt numeric
@attribute NumRotatableBonds numeric
@attribute AromaticProportion numeric
@attribute logS numeric

@data
2.5954000000000006,167.85,0.0,0.0,-2.18
2.3765000000000001,133.405,0.0,0.0,-2.0
2.5938,167.85,1.0,0.0,-1.74
2.0289,133.405,1.0,0.0,-1.48
2.9189,187.37500000000003,1.0,0.0,-3.04
1.81,98.96000000000001,0.0,0.0,-1.29
1.9352,96.94399999999999,0.0,0.0,-1.64
1.4054,118.176,4.0,0.0,-0.43
4.3002,215.894,0.0,0.6,-4.57
2.5654000000000003,132.20599999999996,0.0,0.6,-4.37|
4.3002,215.894,0.0,0.6,-4.63
3.6468000000000007,181.44899999999998,0.0,0.6666666666666666
2.6118600000000001,120.195,0.0,0.6666666666666666,-3.2
4.7366,393.69800000000004,0.0,0.6,-6.98
```

Open the New Dataset



Open the New Dataset



Class Activity



1. Apply Standardize filter (see changes on mean and StdDev)
2. Apply Linear Regression classifier using split 80% and check result
3. Change model parameter (attributeSelectionMethod > No attribute selection)
4. Try cross validation
5. Try Random Forest using cross validation