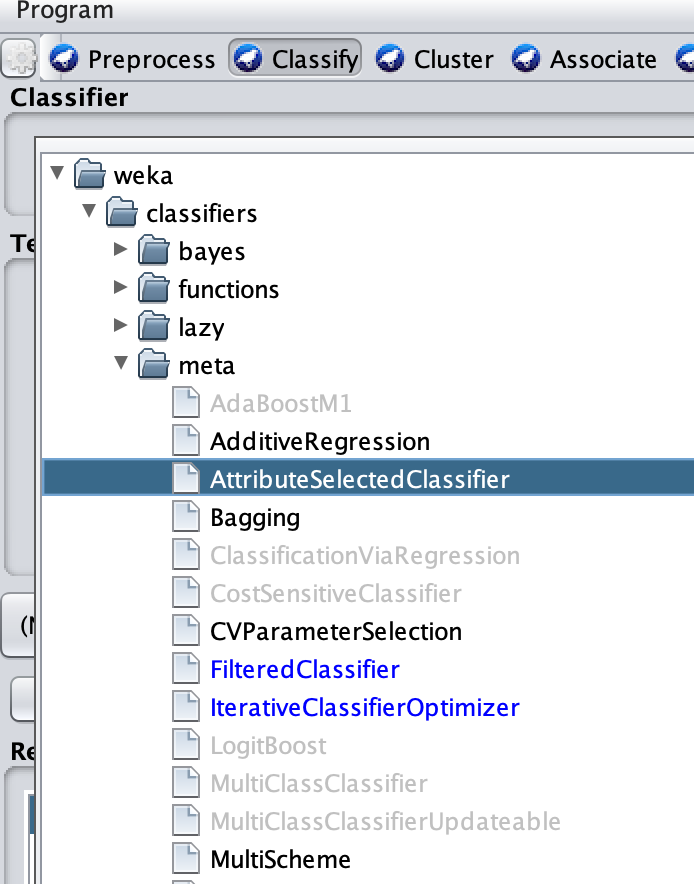
# Modeling Using WEKA:

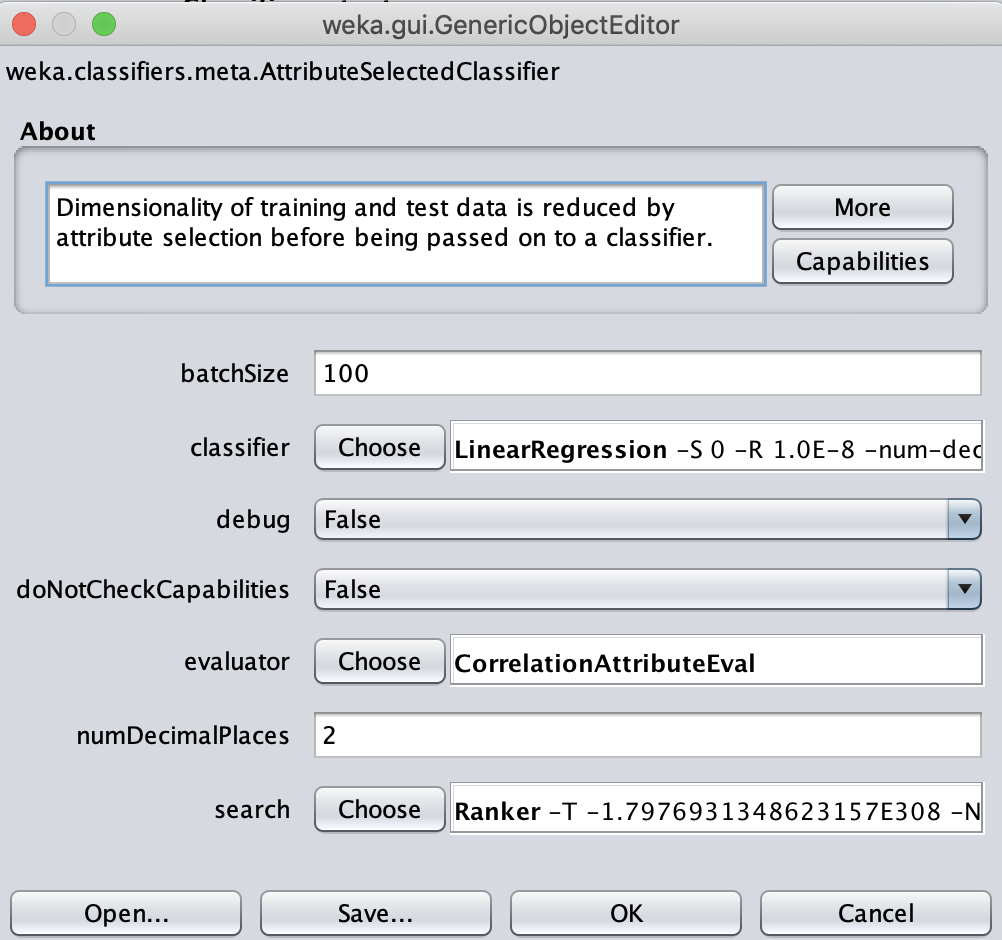
Data is first preprocessed using python notebook. Missing data were linearly interpolated and outliers were removed over certain percentiles for some variables. WEKA is used to build some models as below:

## Linear Regression Model:

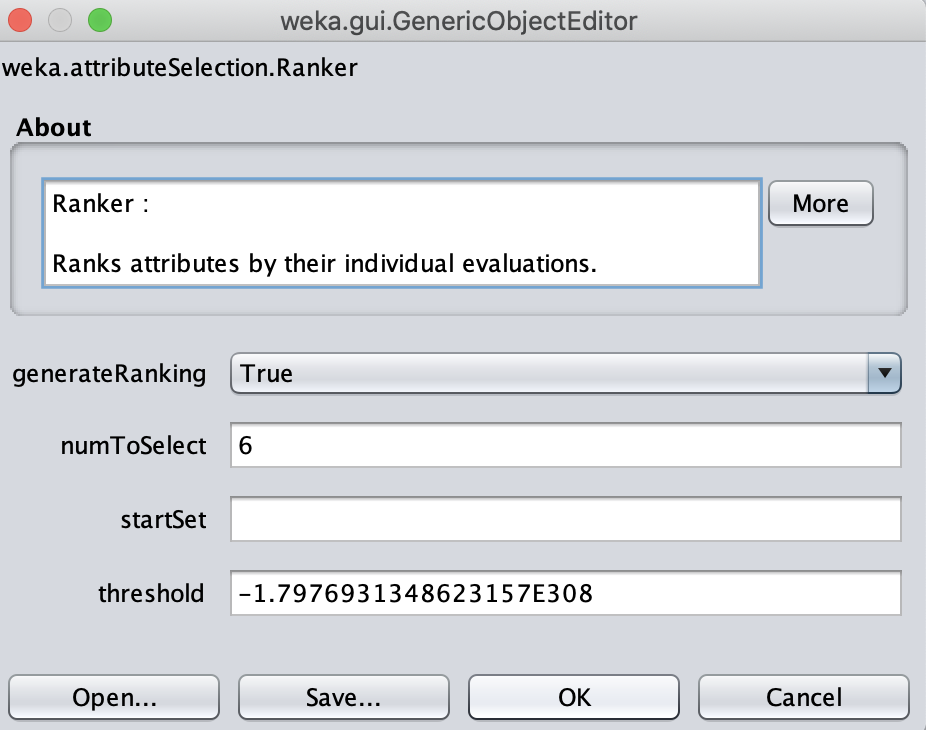
* Open WEKA workbench, and open the normalized data file (normalized\_dataset.csv).
* On the “classify” tab, click on the “choose” button and select “AttributeSelectedClassifier” under the “weka -> classifiers -> meta” group.



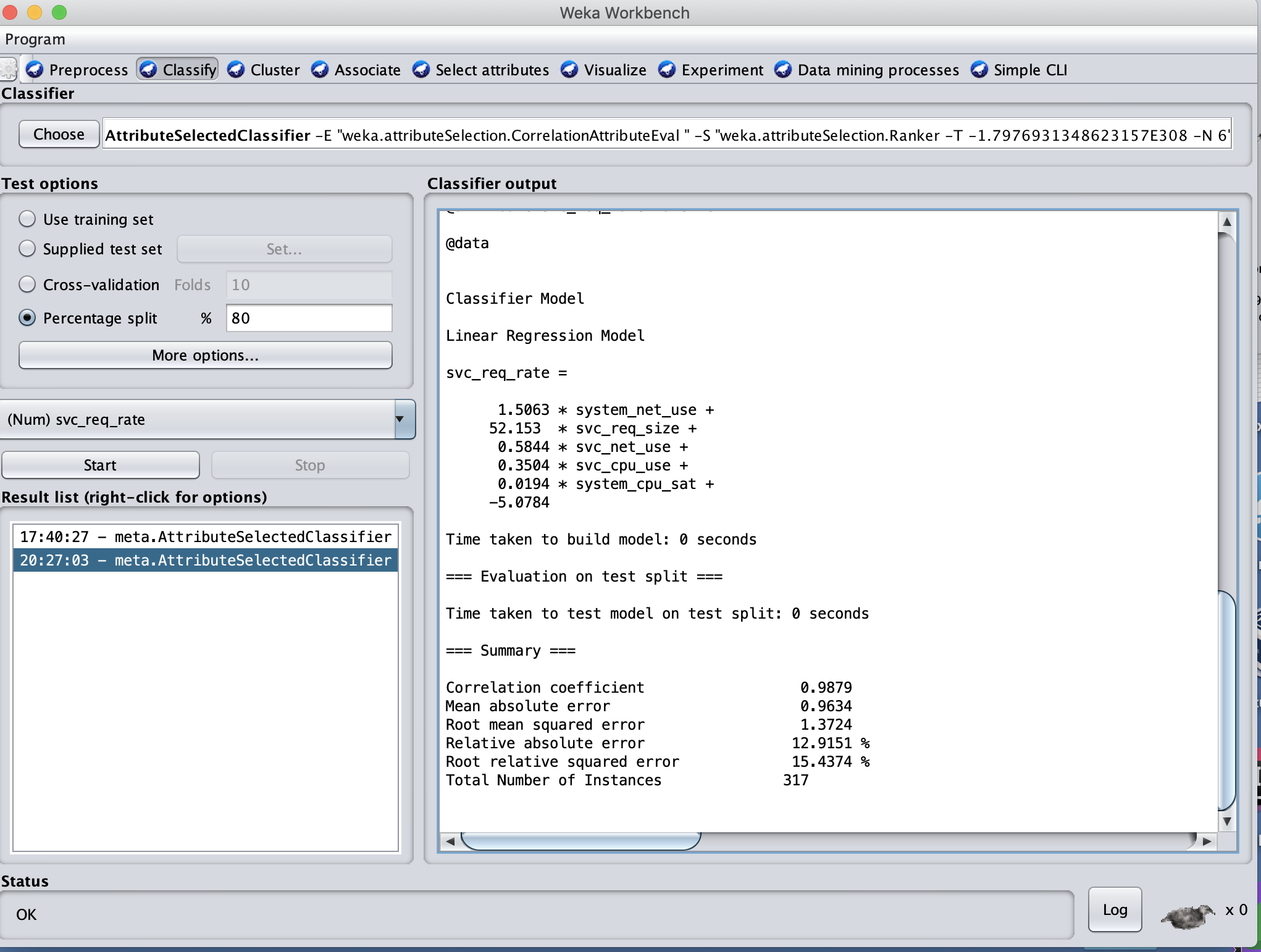
* Click on the “classifier” field to open the properties window.



* As shown above, choose “Linear Regression” as the classifier, “CorrelationAttributesEval” as the evaluator and “Ranker” as the search method
* Click on the “Ranker” search method field to open the properties window



* Choose number of 6 maximum attributes to select. Leave the other fields as they are.
* Click on “OK” button to close all properties window
* Click on the “Start” button on the main classify window to run the model. Use 80 as the percentage split value.



Result:

=== Run information ===

Scheme: weka.classifiers.meta.AttributeSelectedClassifier -E "weka.attributeSelection.CorrelationAttributeEval " -S "weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N 6" -W weka.classifiers.functions.LinearRegression -- -S 0 -R 1.0E-8 -num-decimal-places 4

Relation: normalized\_dataset

Instances: 1584

Attributes: 12

ltcy

svc\_cpu\_use

svc\_cpu\_thr

svc\_net\_use

svc\_disk\_use

system\_cpu\_use

system\_cpu\_sat

system\_net\_use

svc\_req\_size

svc\_resp\_size

svc\_pods

svc\_req\_rate

Test mode: split 80.0% train, remainder test

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

AttributeSelectedClassifier:

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (numeric): 12 svc\_req\_rate):

Correlation Ranking Filter

Ranked attributes:

0.985 8 system\_net\_use

0.855 9 svc\_req\_size

0.749 4 svc\_net\_use

0.744 2 svc\_cpu\_use

0.731 6 system\_cpu\_use

0.689 7 system\_cpu\_sat

Selected attributes: 8,9,4,2,6,7 : 6

Header of reduced data:

@relation 'normalized\_dataset-weka.filters.unsupervised.attribute.Remove-V-R8-9,4,2,6-7,12'

@attribute system\_net\_use numeric

@attribute svc\_req\_size numeric

@attribute svc\_net\_use numeric

@attribute svc\_cpu\_use numeric

@attribute system\_cpu\_use numeric

@attribute system\_cpu\_sat numeric

@attribute svc\_req\_rate numeric

@data

Classifier Model

Linear Regression Model

svc\_req\_rate =

1.5063 \* system\_net\_use +

52.153 \* svc\_req\_size +

0.5844 \* svc\_net\_use +

0.3504 \* svc\_cpu\_use +

0.0194 \* system\_cpu\_sat +

-5.0784

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correlation coefficient **0.9879**

Mean absolute error 0.9634

Root mean squared error 1.3724

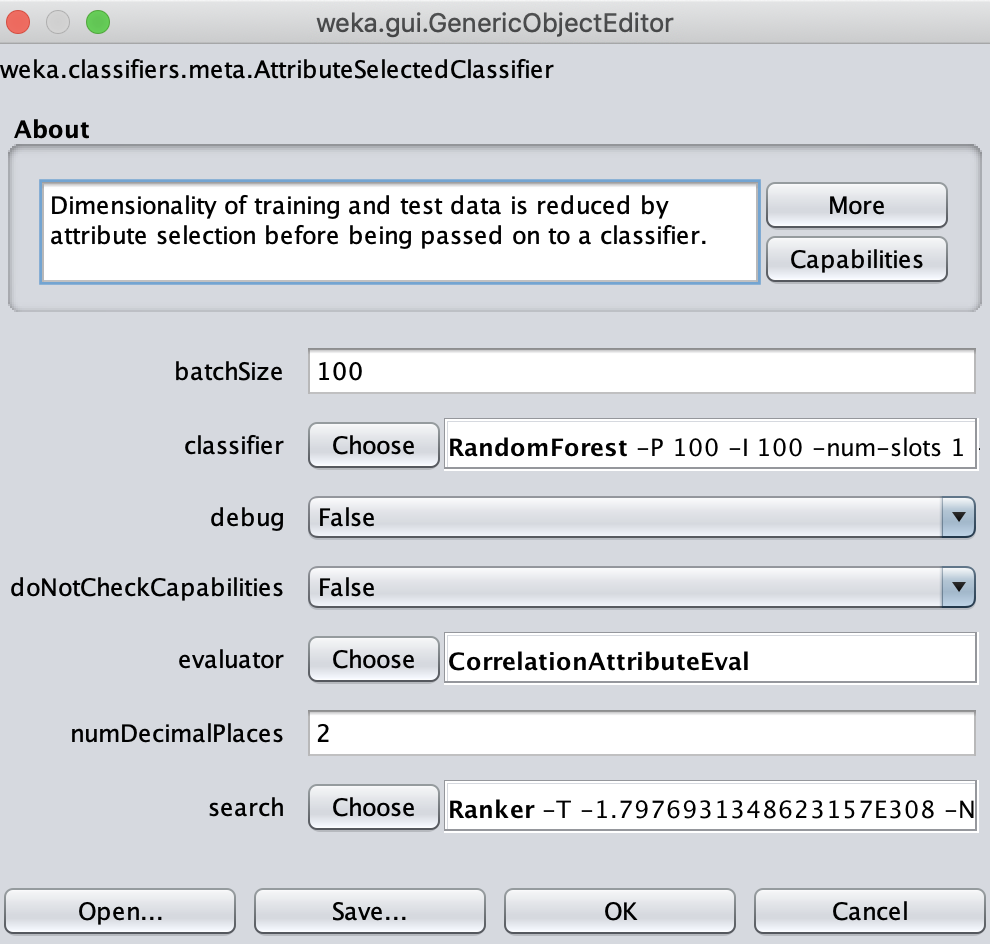
Relative absolute error **12.9151 %**

Root relative squared error **15.4374 %**

Total Number of Instances 317

## Random Forest:

All selection above on WEKA, just click on the “choose” field button on the “classify” tab-window, and change the classifier to be “RandomForest”



Result

=== Run information ===

Scheme: weka.classifiers.meta.AttributeSelectedClassifier -E "weka.attributeSelection.CorrelationAttributeEval " -S "weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N 6" -W weka.classifiers.trees.RandomForest -- -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: normalized\_dataset

Instances: 1584

Attributes: 12

ltcy

svc\_cpu\_use

svc\_cpu\_thr

svc\_net\_use

svc\_disk\_use

system\_cpu\_use

system\_cpu\_sat

system\_net\_use

svc\_req\_size

svc\_resp\_size

svc\_pods

svc\_req\_rate

Test mode: split 80.0% train, remainder test

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

AttributeSelectedClassifier:

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (numeric): 12 svc\_req\_rate):

Correlation Ranking Filter

Ranked attributes:

0.985 8 system\_net\_use

0.855 9 svc\_req\_size

0.749 4 svc\_net\_use

0.744 2 svc\_cpu\_use

0.731 6 system\_cpu\_use

0.689 7 system\_cpu\_sat

Selected attributes: 8,9,4,2,6,7 : 6

Header of reduced data:

@relation 'normalized\_dataset-weka.filters.unsupervised.attribute.Remove-V-R8-9,4,2,6-7,12'

@attribute system\_net\_use numeric

@attribute svc\_req\_size numeric

@attribute svc\_net\_use numeric

@attribute svc\_cpu\_use numeric

@attribute system\_cpu\_use numeric

@attribute system\_cpu\_sat numeric

@attribute svc\_req\_rate numeric

@data

Classifier Model

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.53 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correlation coefficient 0.9939

Mean absolute error 0.6826

Root mean squared error 0.9825

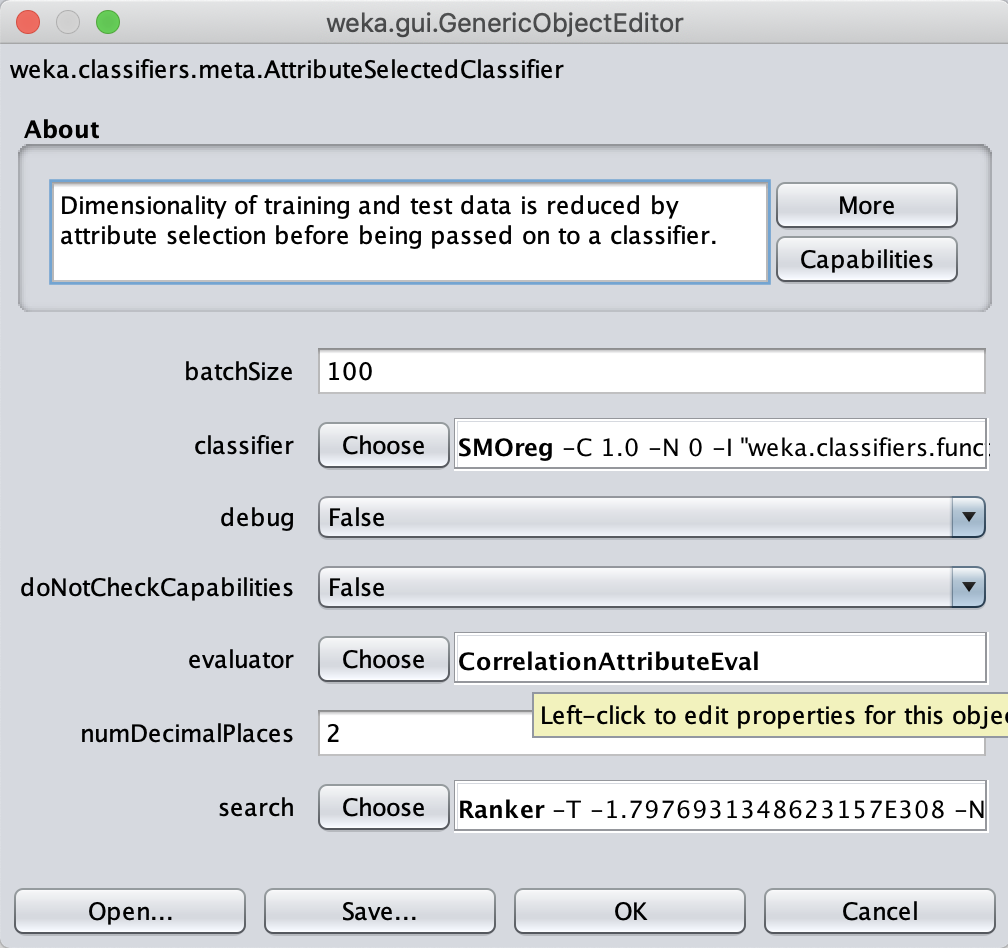
Relative absolute error 9.1502 %

Root relative squared error 11.052 %

Total Number of Instances 317

Support Vector Regression

All selection above are same on WEKA, just click on the “choose” field button on the “classify” tab-window, and change the classifier to be “SMOreg” under “weka.classifiers.functions” group.



Result

=== Run information ===

Scheme: weka.classifiers.meta.AttributeSelectedClassifier -E "weka.attributeSelection.CorrelationAttributeEval " -S "weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N 6" -W weka.classifiers.functions.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.PolyKernel -E 1.0 -C 250007"

Relation: normalized\_dataset

Instances: 1584

Attributes: 12

ltcy

svc\_cpu\_use

svc\_cpu\_thr

svc\_net\_use

svc\_disk\_use

system\_cpu\_use

system\_cpu\_sat

system\_net\_use

svc\_req\_size

svc\_resp\_size

svc\_pods

svc\_req\_rate

Test mode: split 80.0% train, remainder test

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

AttributeSelectedClassifier:

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (numeric): 12 svc\_req\_rate):

Correlation Ranking Filter

Ranked attributes:

0.985 8 system\_net\_use

0.855 9 svc\_req\_size

0.749 4 svc\_net\_use

0.744 2 svc\_cpu\_use

0.731 6 system\_cpu\_use

0.689 7 system\_cpu\_sat

Selected attributes: 8,9,4,2,6,7 : 6

Header of reduced data:

@relation 'normalized\_dataset-weka.filters.unsupervised.attribute.Remove-V-R8-9,4,2,6-7,12'

@attribute system\_net\_use numeric

@attribute svc\_req\_size numeric

@attribute svc\_net\_use numeric

@attribute svc\_cpu\_use numeric

@attribute system\_cpu\_use numeric

@attribute system\_cpu\_sat numeric

@attribute svc\_req\_rate numeric

@data

Classifier Model

SMOreg

weights (not support vectors):

+ 0.9686 \* (normalized) system\_net\_use

+ 0.0749 \* (normalized) svc\_req\_size

+ 0.0004 \* (normalized) svc\_net\_use

- 0.0052 \* (normalized) svc\_cpu\_use

- 0.0034 \* (normalized) system\_cpu\_use

+ 0.023 \* (normalized) system\_cpu\_sat

- 0.0031

Number of kernel evaluations: 5570703 (80.172% cached)

Time taken to build model: 1.44 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correlation coefficient **0.9878**

Mean absolute error 0.9583

Root mean squared error 1.3875

Relative absolute error **12.8469 %**

Root relative squared error **15.6072 %**

Total Number of Instances 317