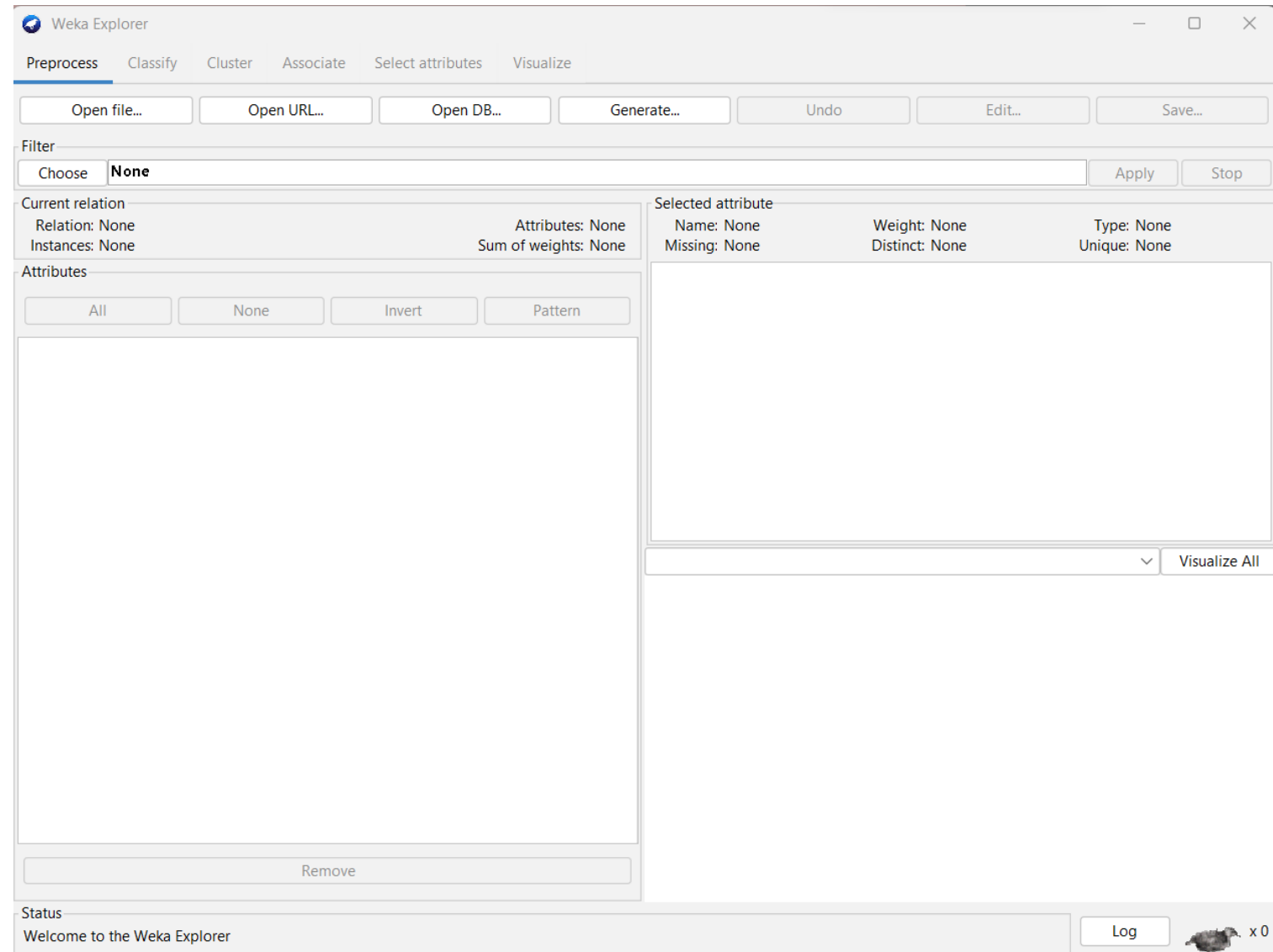
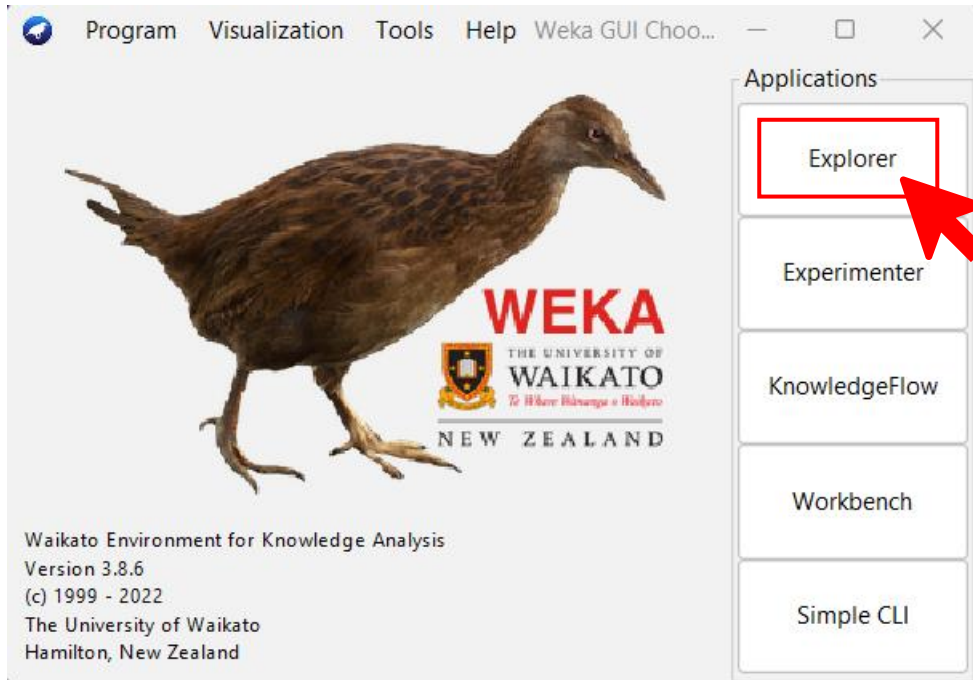


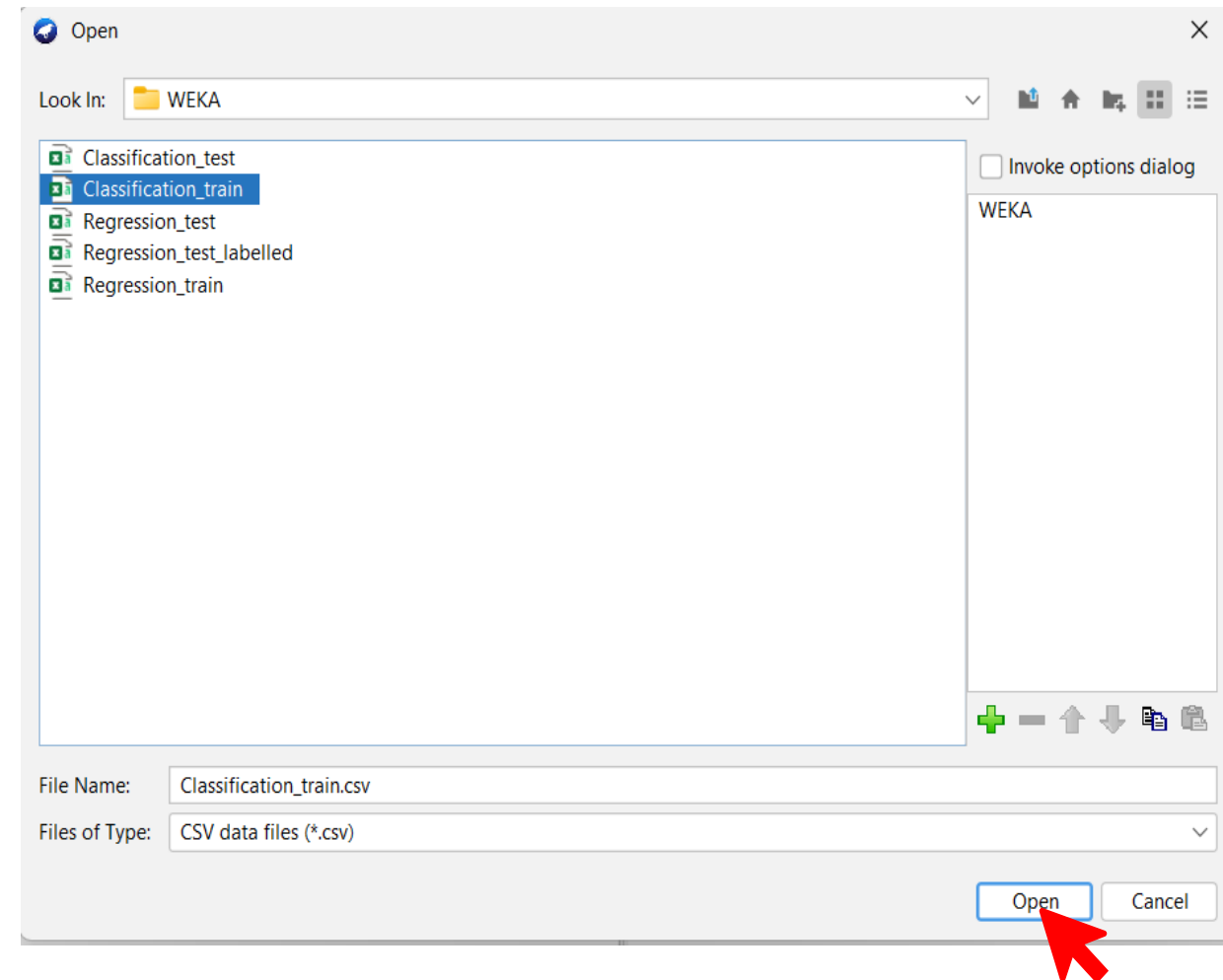
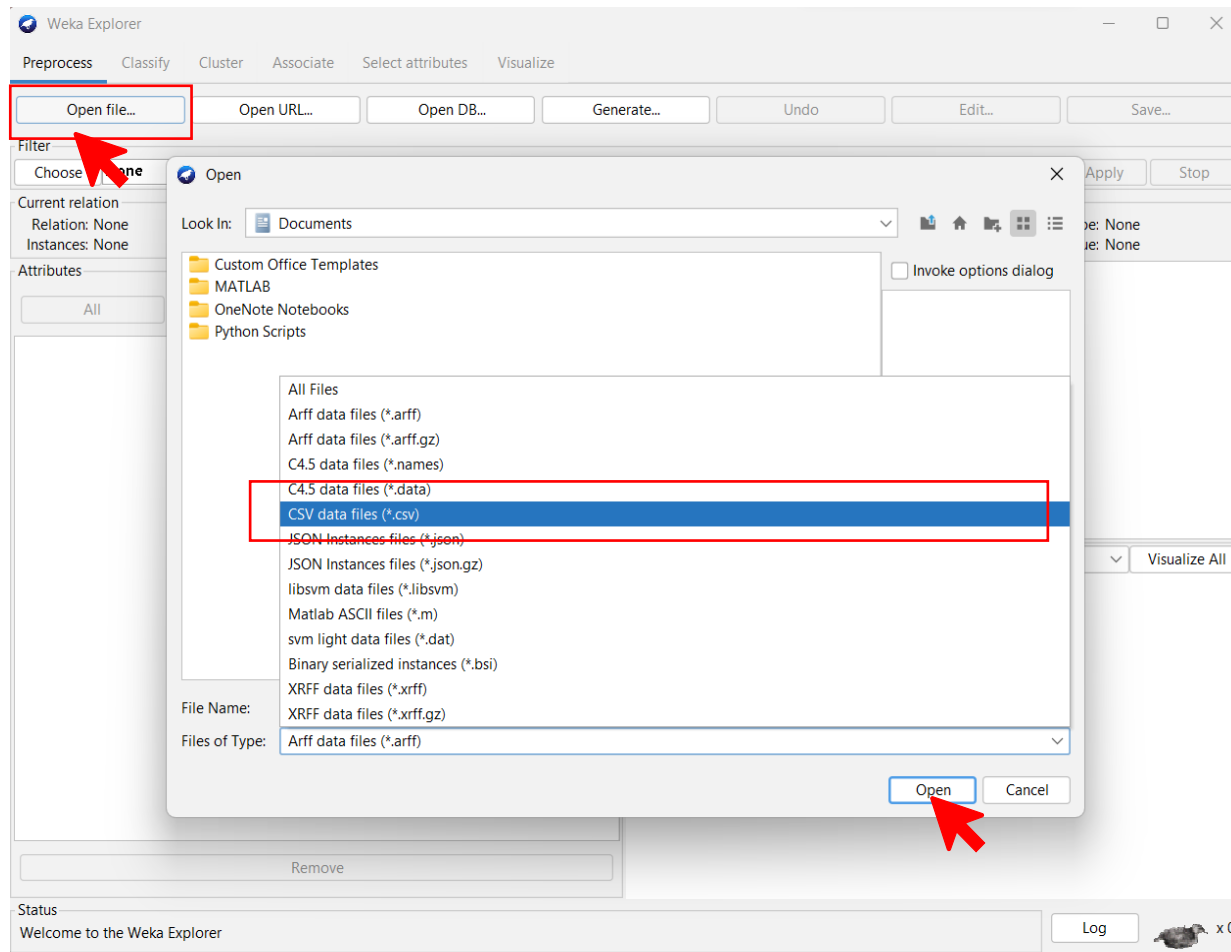
# **Classification using WEKA**

**Drug resistance class prediction using  
Support Vector Machine**

# WEKA GUI



# Loading Data



# Dataset for Classification

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: Classification\_train Instances: 342 Attributes: 55 Sum of weights: 342

Attributes: All None Invert Pattern

Selected attribute: Name: CATEGORY Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	S	139	139
2	M	203	203

Attributes list:

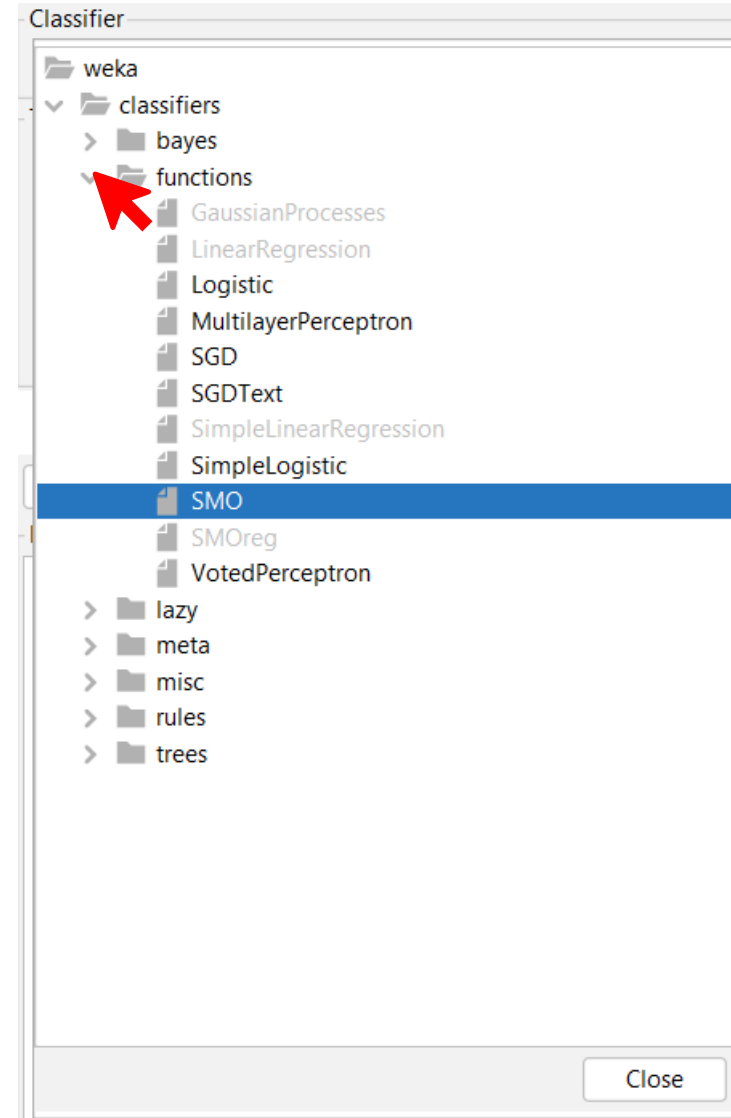
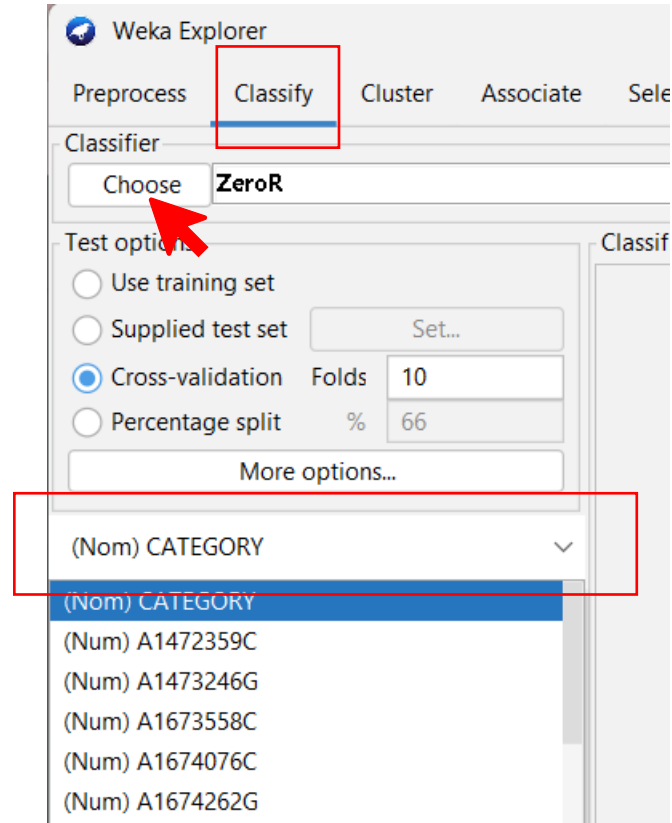
No.	Name
1	<input checked="" type="checkbox"/> CATEGORY
2	<input type="checkbox"/> A1472359C
3	<input type="checkbox"/> A1473246G

Viewer

Relation: Classification\_train

No.	1: CATEGORY Nominal	2: A1472359C Numeric	3: A1473246G Numeric	4: A1673558C Numeric	5: A1674076C Numeric	6: A1674262G Numeric	7: A4247033C Numeric	8: A4247429G Numeric	9: A4247847G Numeric	10: A4248003G Numeric	11: A4248319T Numeric	12: A44 Nur
1	S	0.0	0.0	0.0	0.1293	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	S	0.0	0.0	0.0	0.0866	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	S	0.0	0.0	0.0	0.1511	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	S	0.0	0.0	0.0	0.1223	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	S	0.0	0.0	0.0	0.1148	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	S	0.0	0.0	0.0	0.1375	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	S	0.0	0.0	0.0	0.1237	0.0	0.0	0.0	0.0	0.0	0.0	0.0

# Choosing Algorithm



# Algorithm parameters

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVectorPolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-

Test options Classifier output

weka.gui.GenericObjectEditor

weka.classifiers.functions.SMO

About

Implements John Platt's sequential minimal optimization algorithm for training a support vector classifier. [More](#) [Capabilities](#)

batchSize 100

buildCalibrationModels False

c 1.0

calibrator Choose **Logistic** -R 1.0E-8 -M -1 -num-decimal-pl

checksTurnedOff False

debug False

doNotCheckCapabilities False

epsilon 1.0E-12

filterType Normalize training data

kernel Choose **PolyKernel** -E 1.0 -C 250007

numDecimalPlaces 2

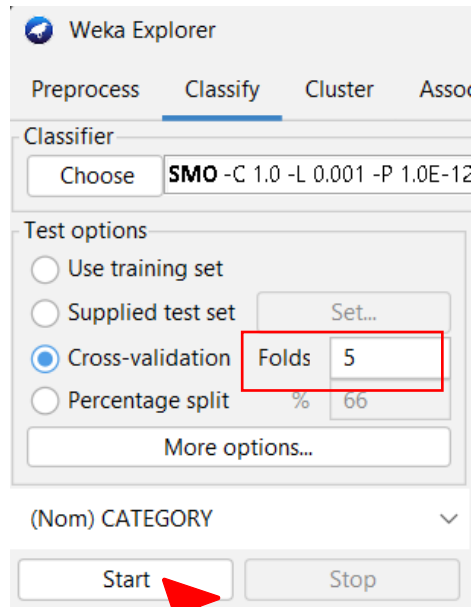
numFolds -1

randomSeed 1

toleranceParameter 0.001

Open... Save... OK Cancel

# 5-fold cross-validation



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      325           95.0292 %
Incorrectly Classified Instances    17           4.9708 %
Kappa statistic                    0.8989
Mean absolute error                 0.0497
Root mean squared error            0.223
Relative absolute error            10.2993 %
Root relative squared error        45.3907 %
Total Number of Instances         342

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1.000    0.084    0.891     1.000    0.942     0.904    0.958    0.891     S
                0.916    0.000    1.000     0.916    0.956     0.904    0.958    0.966     M
Weighted Avg.   0.950    0.034    0.956     0.950    0.951     0.904    0.958    0.936

=== Confusion Matrix ===

  a  b  <-- classified as
139  0 |   a = S
 17 186 |   b = M
```

**Confusion Matrix**

# Performance Metrics

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

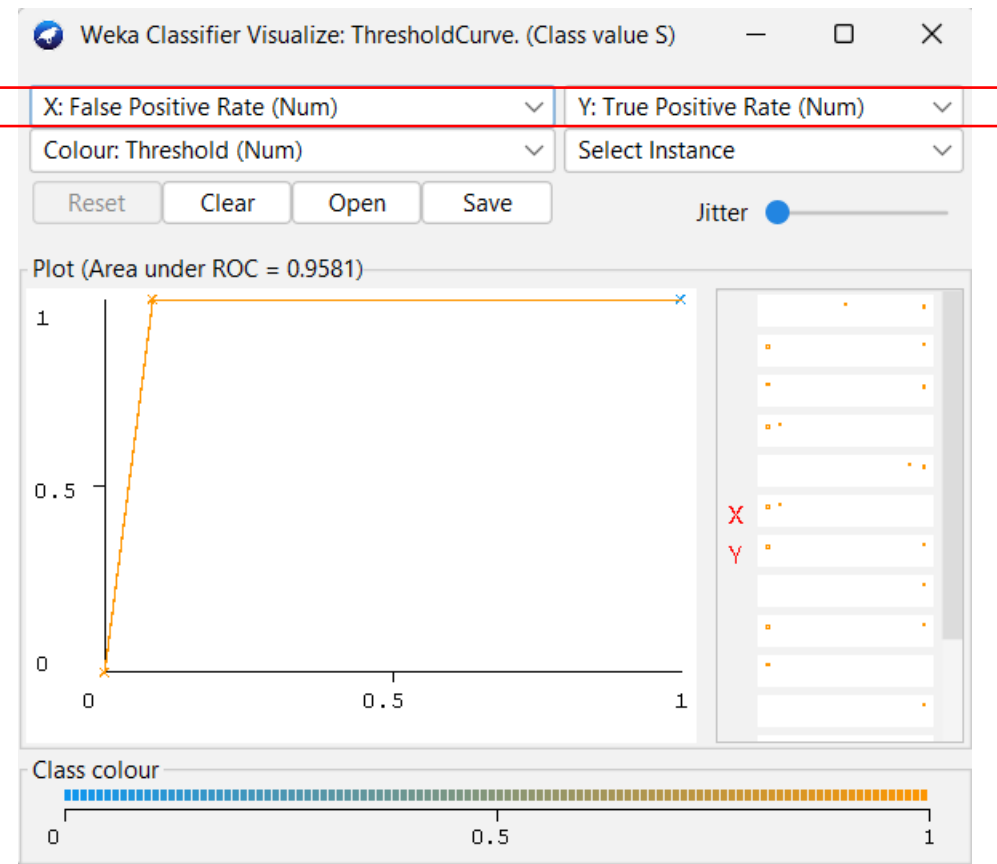
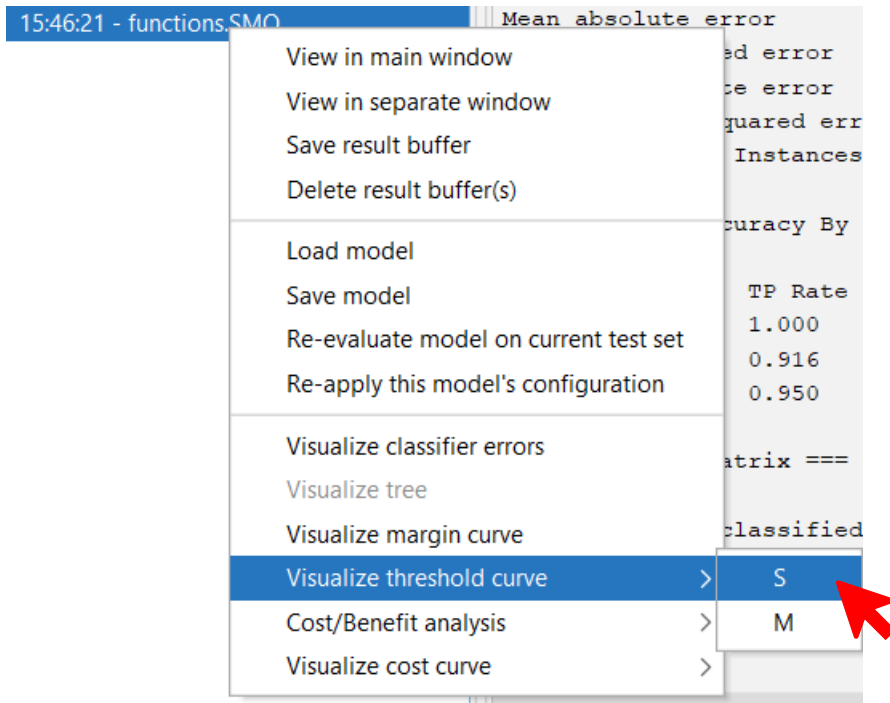
	S	M
S	139	0
M	17	186

Metrics	Formula
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$TP/(TP+FP)$
Recall/Sensitivity/True Positive Rate ( <b>TPR</b> )	$TP/(TP+FN)$
False Positive Rate (FPR)	$FP/(FP+TN)$
Specificity ( <b>1-FPR</b> )	$TN/(TN+FP)$
F1 score	$2TP/(2TP+FP+FN)$



# ROC Plot

## Right click on functions.SMO



# Model testing

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds

☐ Percentage split %

More options...

(Nom) CATEGORY

Start Stop

Result list (right-click for options)

15:28:18 - functions.SMO

Test Instances

Relation: None  
Instances: None

Attributes: None  
Sum of weights: None

Open file... Open URL...

Class No class

Close

Open

Look In: WEKA

- Classification\_test
- Classification\_train
- Regression\_test
- Regression\_test\_labelled
- Regression\_train

☐ Invoke options dialog

WEKA

File Name: Classification\_test.csv

Files of Type: CSV data files (\*.csv)

Open Cancel

Test Instances

Relation: Classification\_test  
Instances: ?

Attributes: 55  
Sum of weights: ?

Open file... Open URL...

Class (Nom) CATEGORY

Close

- No class
- (Nom) CATEGORY
- (Num) A1472359C
- (Num) A1473246G
- (Num) A1673558C

# Test Performance

## Right click on functions.SMO

Result list (right-click for options)

15:28:18 - function.SMO

- View in main window
- View in separate window
- Save result buffer
- Delete result buffer(s)
- Load model
- Save model
- Re-evaluate model on current test set
- Re-apply this model's configuration
- Visualize classifier errors
- Visualize tree
- Visualize margin curve
- Visualize threshold curve
- Cost/Benefit analysis
- Visualize cost curve

=== Re-evaluation on test set ===

User supplied test set

Relation: Classification\_test

Instances: unknown (yet). Reading incrementally

Attributes: 55

=== Summary ===

Correctly Classified Instances 83 98.8095 %

Incorrectly Classified Instances 1 1.1905 %

Kappa statistic 0.9754

Mean absolute error 0.0119

Root mean squared error 0.1091

Total Number of Instances 84

=== Detailed Accuracy By Class ===

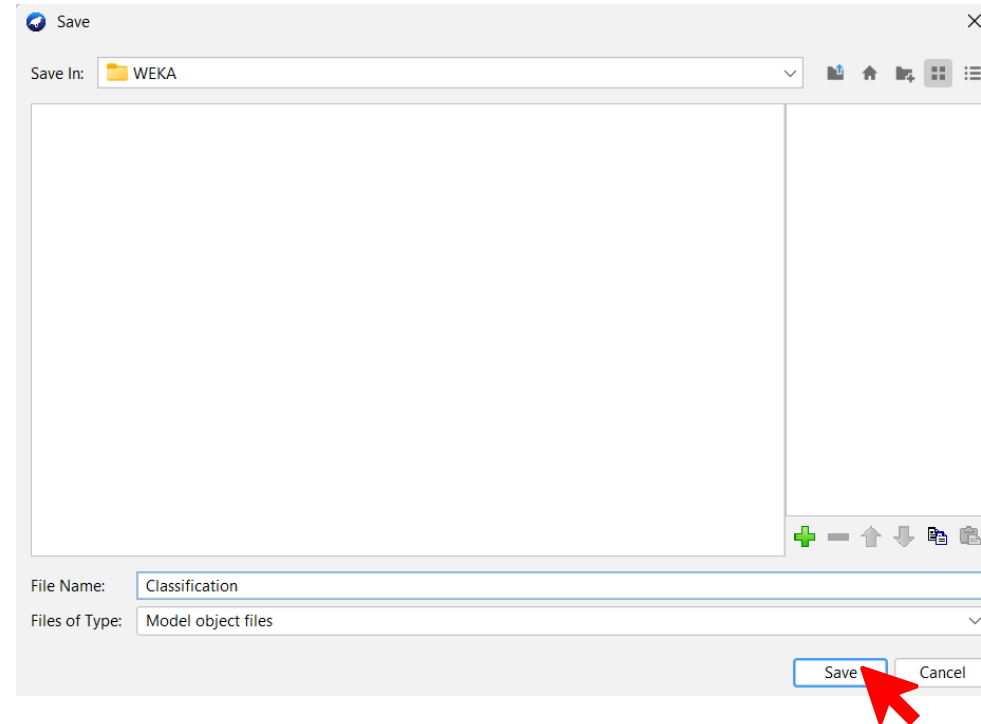
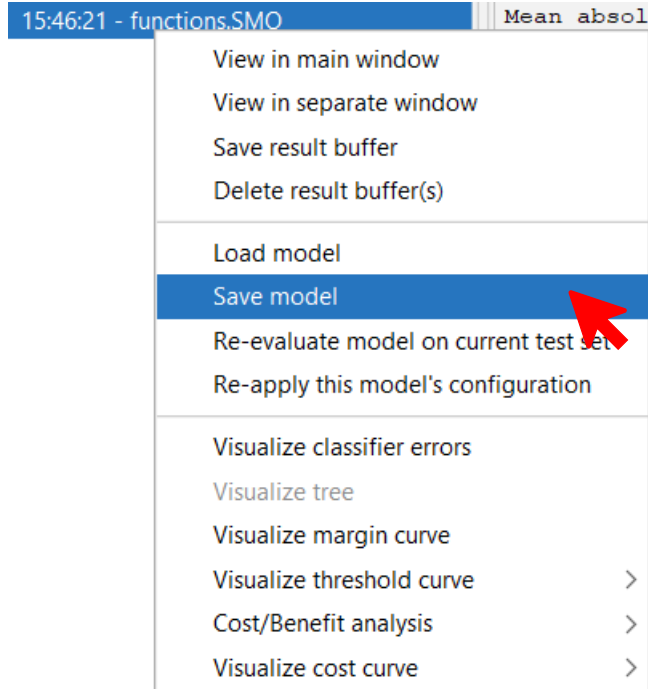
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.020	0.971	1.000	0.986	0.976	0.990	0.971	S
	0.980	0.000	1.000	0.980	0.990	0.976	0.990	0.992	M
Weighted Avg.	0.988	0.008	0.988	0.988	0.988	0.976	0.990	0.984	

=== Confusion Matrix ===

```
a b  <-- classified as
34 0 | a = S
 1 49 | b = M
```

# Saving Model

## Right click on functions.SMO



# Loading Model

## Right click

