

Build Classification Model Using WEKA

ISE4132 : AI Application System

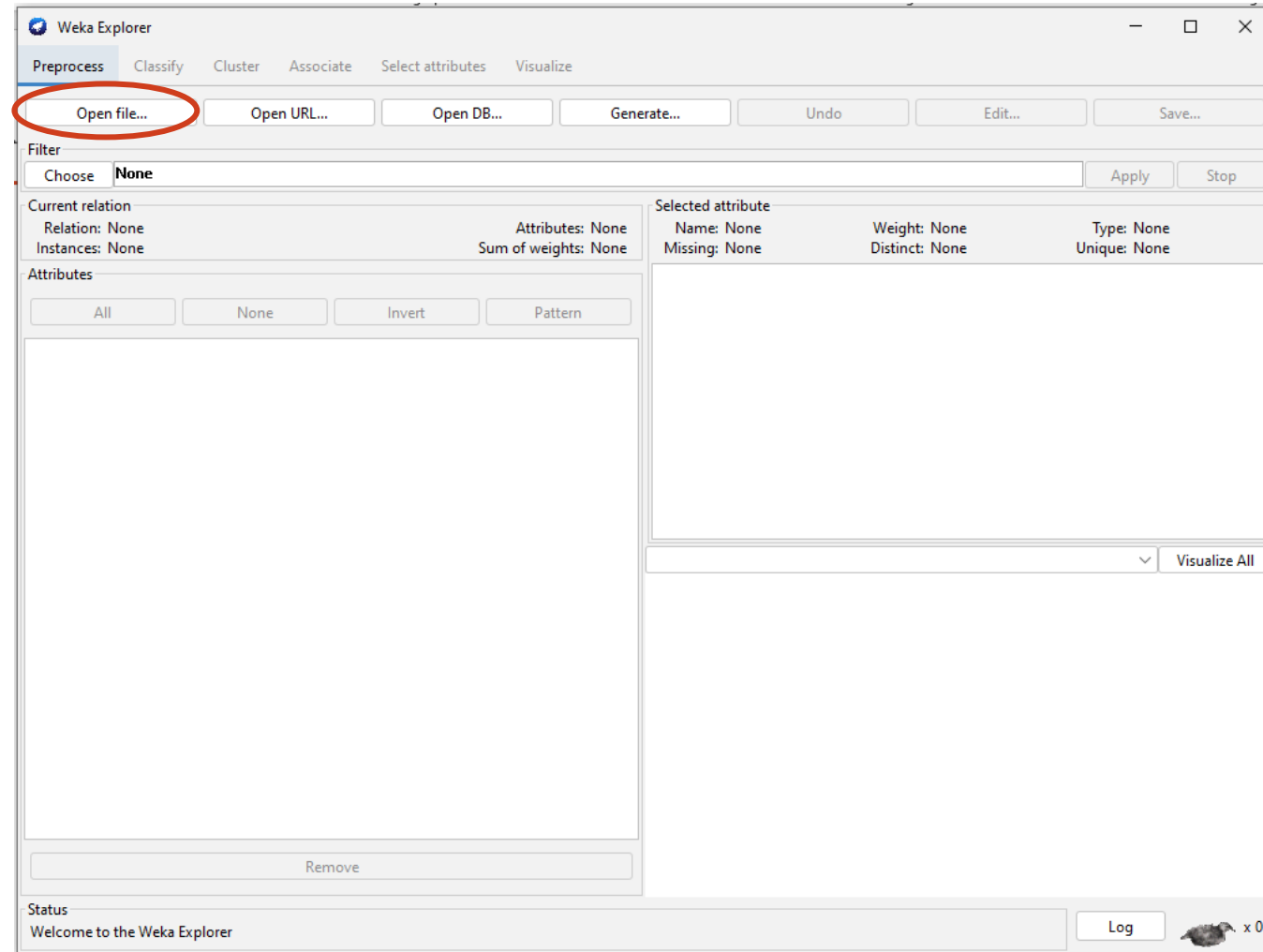


AMIN AL (아민알)
Integrated System Engineering (ISE)
010-6853-6648
alaminanik@inha.ac.kr

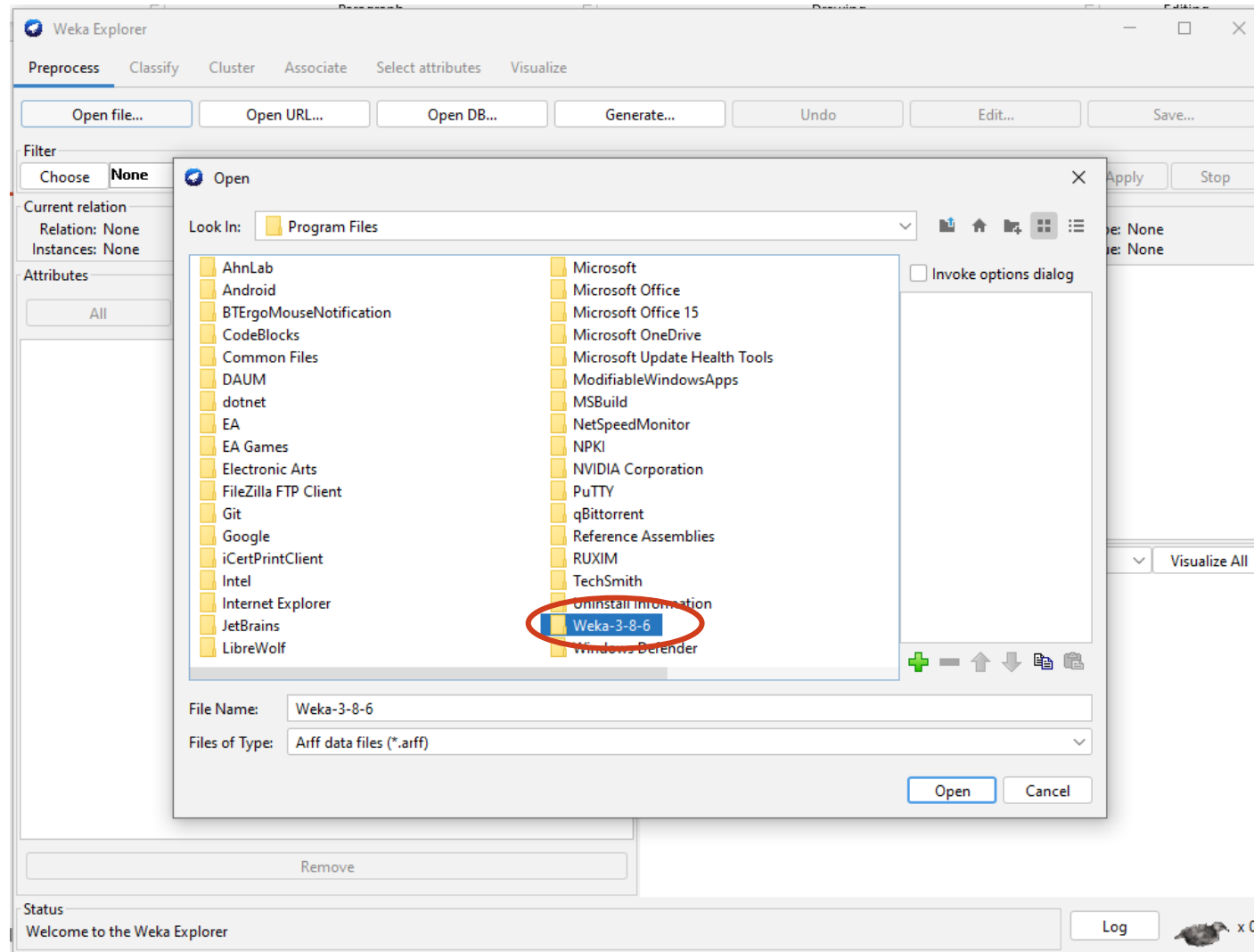
Selecting Dataset



Selecting Dataset



Selecting Dataset



Weka-3-8-6 > data > breast-cancer

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: breast-cancer
Instances: 286
Attributes: 10
Sum of weights: 286

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> age
2	<input type="checkbox"/> menopause
3	<input type="checkbox"/> tumor-size
4	<input type="checkbox"/> inv-nodes
5	<input type="checkbox"/> node-caps
6	<input type="checkbox"/> deg-malig
7	<input type="checkbox"/> breast
8	<input type="checkbox"/> breast-quad
9	<input type="checkbox"/> irradiat
10	<input type="checkbox"/> Class

Remove

Status: OK

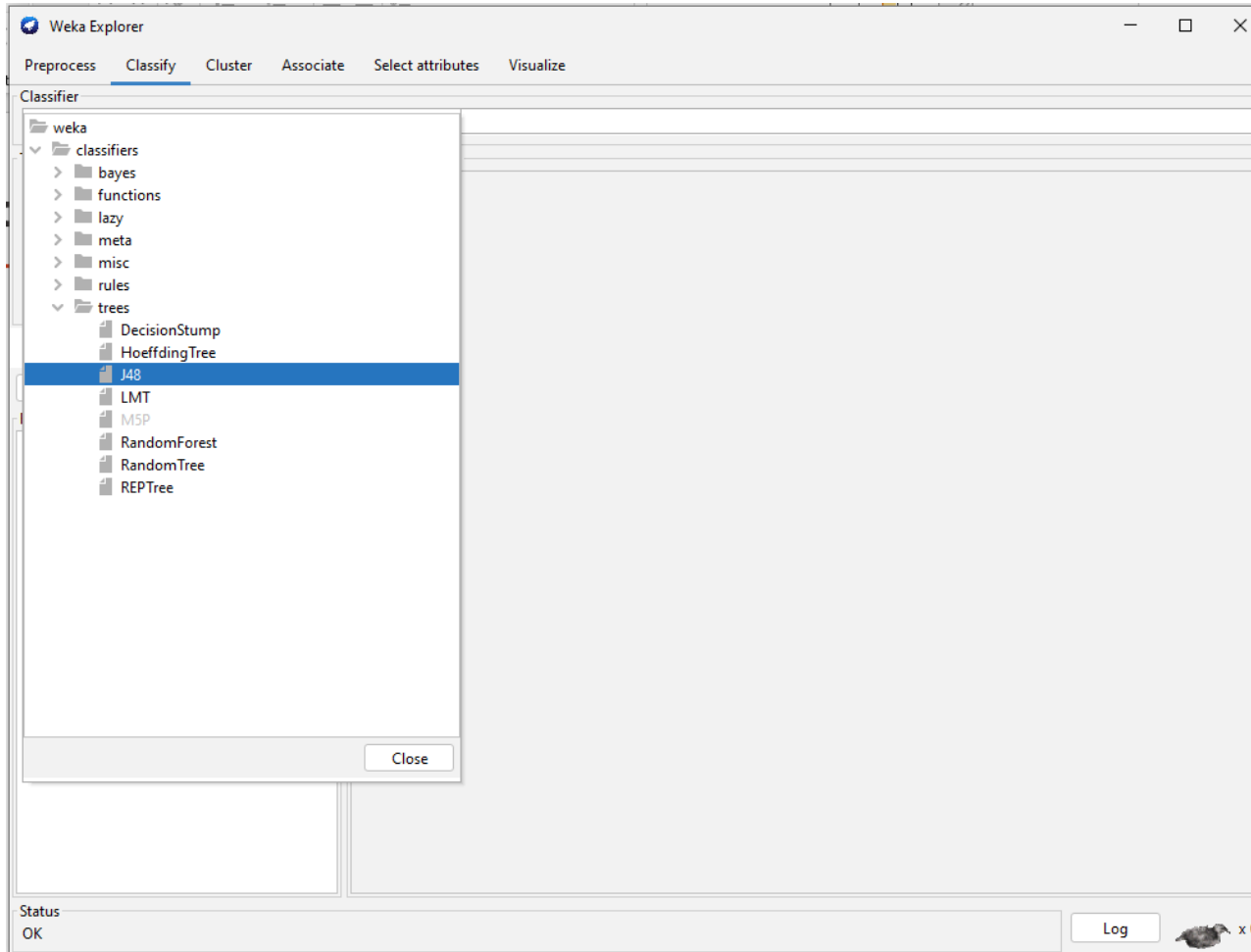
Selected attribute
Name: age
Missing: 0 (0%)
Distinct: 6
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	10-19	0	0
2	20-29	1	1
3	30-39	36	36
4	40-49	90	90
5	50-59	96	96
6	60-69	57	57
7	70-79	6	6
8	80-89	0	0
9	90-99	0	0

Class: Class (Nom) Visualize All

Log x 0

Select Classifiers



Classifiers: trees > J48

J48 is a machine learning decision tree classification algorithm based on Iterative Dichotomiser 3. It is very helpful in examine the data categorically and continuously.

Build ML Model Using J48 (Use Training Set)



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☒ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

09:52:09 - trees.J48

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	217	75.8741 %
Incorrectly Classified Instances	69	24.1259 %
Kappa statistic	0.2899	
Mean absolute error	0.3658	
Root mean squared error	0.4269	
Relative absolute error	87.4491 %	
Root relative squared error	93.4017 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.965	0.729	0.758	0.965	0.849	0.352	0.639	0.767	no-recurrence-events
	0.271	0.035	0.767	0.271	0.400	0.352	0.639	0.461	recurrence-events
Weighted Avg.	0.759	0.523	0.760	0.759	0.716	0.352	0.639	0.676	

=== Confusion Matrix ===

a	b	<-- classified as
194	7	a = no-recurrence-events
62	23	b = recurrence-events

Status

OK

Log x 1

Build ML Model Using J48 (Use Cross-Validation)



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier
Choose J48 -C 0.25 -M 2

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 60
More options...

(Nom) Class
Start Stop

Result list (right-click for options)
09:52:09 - trees.J48
12:24:01 - trees.J48

Classifier output

Number of Leaves : 4
Size of the tree : 6
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances	216	75.5245 %
Incorrectly Classified Instances	70	24.4755 %
Kappa statistic	0.2826	
Mean absolute error	0.3676	
Root mean squared error	0.4324	
Relative absolute error	87.8635 %	
Root relative squared error	94.6093 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.960	0.729	0.757	0.960	0.846	0.339	0.584	0.736	no-recurrence-events
	0.271	0.040	0.742	0.271	0.397	0.339	0.584	0.436	recurrence-events
Weighted Avg.	0.755	0.524	0.752	0.755	0.713	0.339	0.584	0.647	

=== Confusion Matrix ===

```
a  b  <-- classified as
193  8 | a = no-recurrence-events
 62 23 | b = recurrence-events
```

Status
OK

Log x 1

Build ML Model Using J48 (Use Split)



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier
Choose **J48 -C 0.25 -M 2**

Test options
☐ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☒ Percentage split % 80
More options...

(Nom) Class
Start Stop

Result list (right-click for options)
09:52:09 - trees.J48
12:24:01 - trees.J48
12:27:04 - trees.J48

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	41	71.9298 %
Incorrectly Classified Instances	16	28.0702 %
Kappa statistic	0.2995	
Mean absolute error	0.3707	
Root mean squared error	0.4619	
Relative absolute error	83.7729 %	
Root relative squared error	94.2422 %	
Total Number of Instances	57	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.972	0.714	0.700	0.972	0.814	0.379	0.628	0.711	no-recurrence-events
	0.286	0.028	0.857	0.286	0.429	0.379	0.628	0.522	recurrence-events
Weighted Avg.	0.719	0.461	0.758	0.719	0.672	0.379	0.628	0.641	

=== Confusion Matrix ===

```
a b <-- classified as
35 1 | a = no-recurrence-events
15 6 | b = recurrence-events
```

Status
OK

Log x 1

Build ML Model Using Random Forest



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...

(Nom) Class

Start Stop

Result list (right-click for options)

- 09:52:09 - trees.J48
- 12:24:01 - trees.J48
- 12:27:04 - trees.J48
- 12:31:43 - trees.RandomForest

Classifier output

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

Time taken to build model: 0.08 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances	280	97.9021 %
Incorrectly Classified Instances	6	2.0979 %
Kappa statistic	0.9498	
Mean absolute error	0.1407	
Root mean squared error	0.1869	
Relative absolute error	33.6442 %	
Root relative squared error	40.8903 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.985	0.035	0.985	0.985	0.985	0.950	0.998	0.999	no-recurrence-events
	0.965	0.015	0.965	0.965	0.965	0.950	0.998	0.995	recurrence-events
Weighted Avg.	0.979	0.029	0.979	0.979	0.979	0.950	0.998	0.998	

=== Confusion Matrix ===

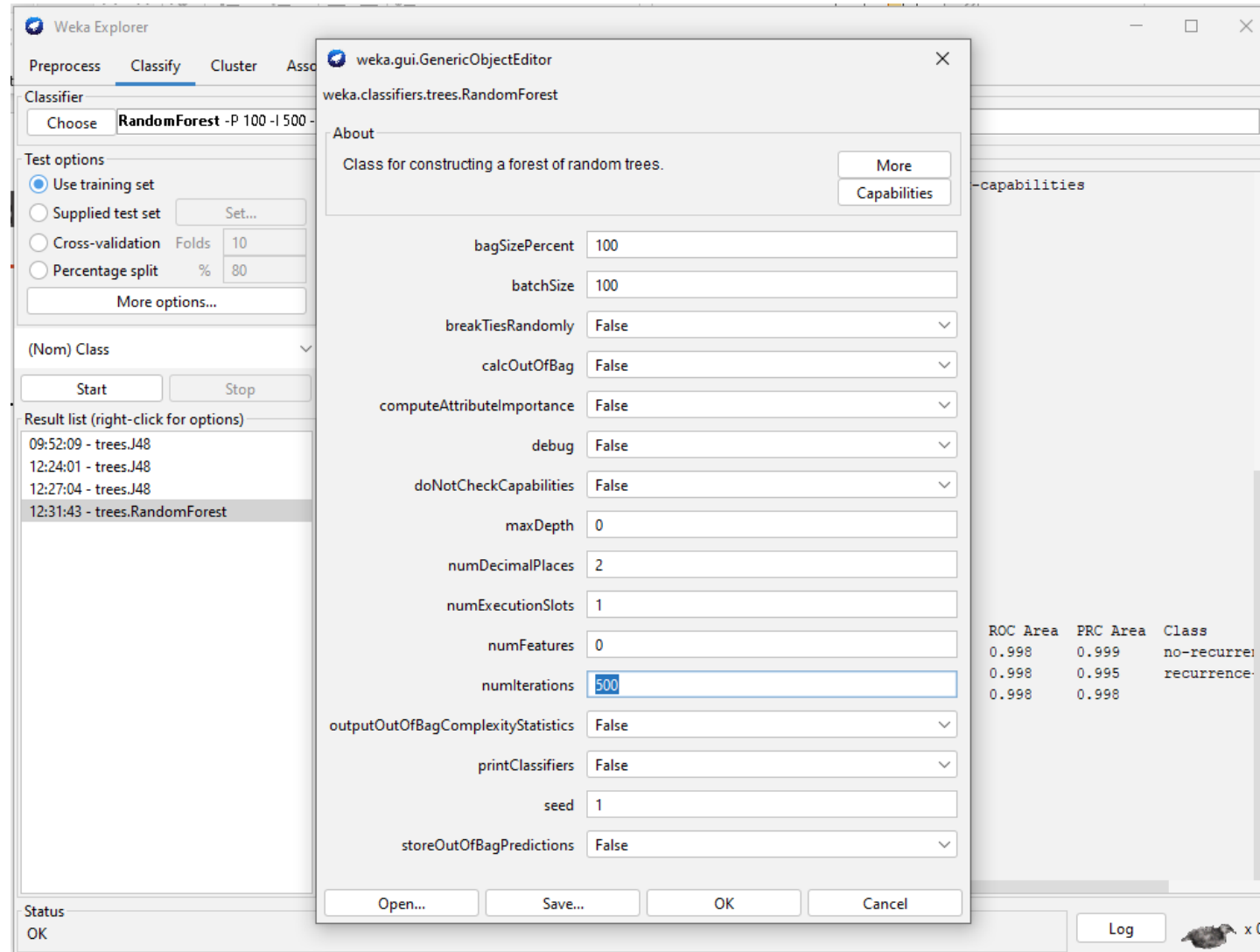
a	b	<-- classified as
198	3	a = no-recurrence-events
3	82	b = recurrence-events

Status OK

Log x 1

Now try with cross-validation and percentage split. Then check the results.

Configure ML Model



Use Multilayer Perception



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

weka

- classifiers
 - bayes
 - functions
 - GaussianProcesses
 - LinearRegression
 - Logistic
 - MultilayerPerceptron**
 - SGD
 - SGDText
 - SimpleLinearRegression
 - SimpleLogistic
 - SMO
 - SMOreg
 - VotedPerceptron
 - lazy
 - meta
 - misc
 - rules
 - trees

E 20 -H a

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

uild model: 0.14 seconds

on test split ==

est model on test split: 0.01 seconds

ified Instances 39 68.4211 %

ssified Instances 18 31.5789 %

0.2467

error 0.3704

ed error 0.4721

te error 83.709 %

guared error 96.3084 %

Instances 57

uracy By Class ==

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.889	0.667	0.696	0.889	0.780	0.272	0.673	0.780	no-recurre
0.333	0.111	0.636	0.333	0.437	0.272	0.673	0.597	recurrence-
0.684	0.462	0.674	0.684	0.654	0.272	0.673	0.712	

atrix ==

classified as

32 4 | a = no-recurrence-events

14 7 | b = recurrence-events

Close

Status OK

Log x 0

Use Support Vector Machine



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

- weka
 - classifiers
 - bayes
 - functions
 - GaussianProcesses
 - LinearRegression
 - Logistic
 - MultilayerPerceptron
 - SGD
 - SGDText
 - SimpleLinearRegression
 - SimpleLogistic
 - SMO**
 - SMOreg
 - VotedPerceptron
 - lazy
 - meta
 - misc
 - rules
 - trees

weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E

Build model: 0.14 seconds

on test split ==

Test model on test split: 0.01 seconds

Classified Instances	39	68.4211 %
Unclassified Instances	18	31.5789 %
Error	0.2467	
Reduced error	0.3704	
Test error	0.4721	
Test error	83.709 %	
Squared error	96.3084 %	
Instances	57	

Accuracy By Class ==

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.889	0.667	0.696	0.889	0.780	0.272	0.673	0.780	no-recurrence-events
0.333	0.111	0.636	0.333	0.437	0.272	0.673	0.597	recurrence-events
0.684	0.462	0.674	0.684	0.654	0.272	0.673	0.712	

Confusion Matrix ==

Classified as

32	4	a = no-recurrence-events
14	7	b = recurrence-events

Status
OK

Log x 0

Class Activity



1. Download csv file from iclass
2. Convert the file for weka dataset
3. Load the dataset into Weka
4. Normalize the data
5. Use J48 model with all test options and check the result.
6. Use Random Forest model with all test options and check the result.
7. Use Multilayer Perception model with all test options and check the result.
8. Use SVO model with all test options and check the result.

