



PHÂN LỚP VỚI CÂY QUYẾT ĐỊNH

Giảng viên: Nguyễn Tu Trung
BM HTTT, Khoa CNTT, Trường ĐH Thủy Lợi

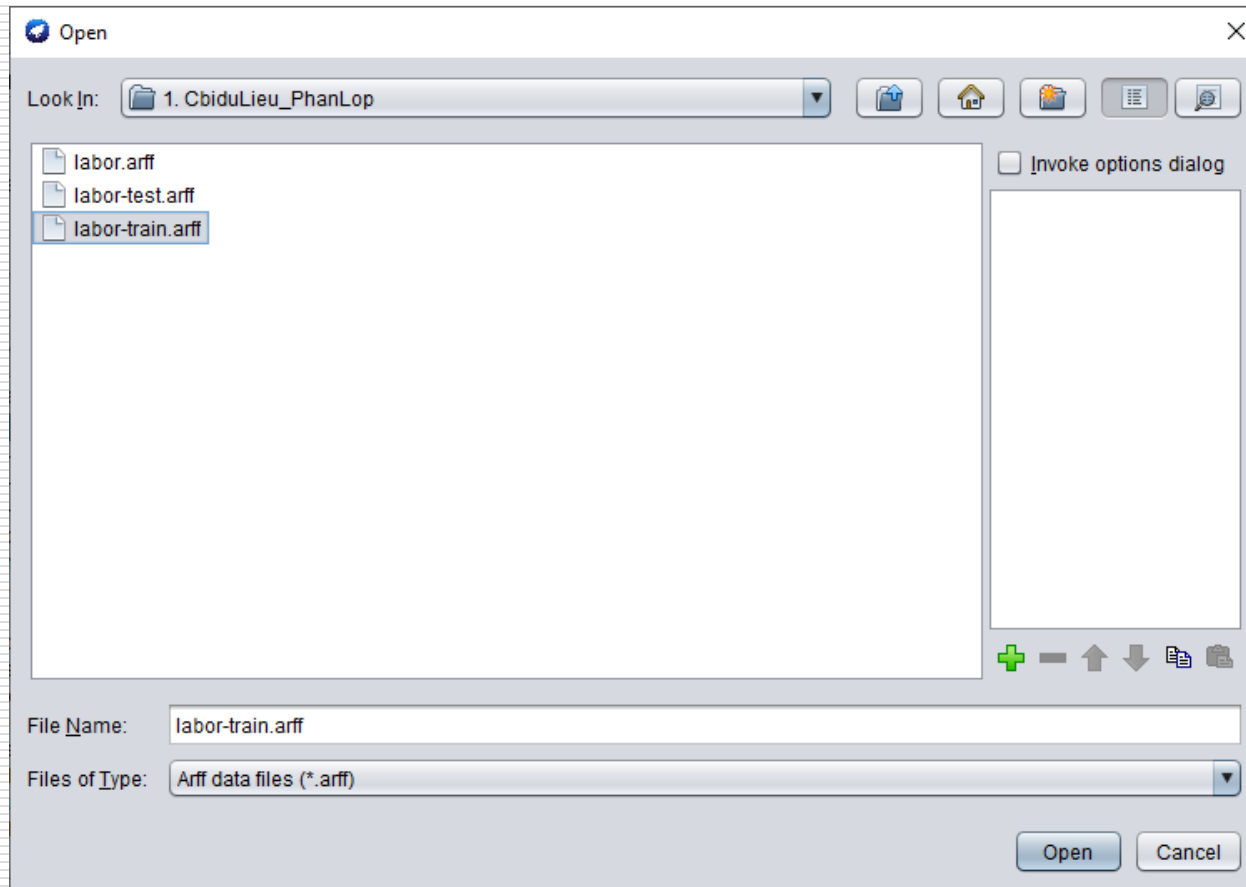
Hà Nội, 2021

Nội dung

- ❖ Mở file labor-train.arff
- ❖ Thiết lập tham số
- ❖ Huấn luyện
- ❖ View cây quyết định
- ❖ Kiểm thử
- ❖ Save và load model

Mở file labor-train.arff

- ❖ Trong Tab Preprocessing, chọn Open File
- ❖ Chọn file labor-train.arff đã chuẩn bị



Mở file labor-train.arff

❖ Kết quả:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

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

Filter: Choose **Resample -S 1 -Z 20.0** [Apply] [Stop]

Current relation
Relation: labor-neg-data-weka.filters.unsupervised insta...
Instances: 46
Attributes: 17
Sum of weights: 46

Attributes
[All] [None] [Invert] [Pattern]

No.	Name
1	<input checked="" type="checkbox"/> duration
2	<input type="checkbox"/> wage-increase-first-year
3	<input type="checkbox"/> wage-increase-second-year
4	<input type="checkbox"/> wage-increase-third-year
5	<input type="checkbox"/> cost-of-living-adjustment
6	<input type="checkbox"/> working-hours
7	<input type="checkbox"/> pension
8	<input type="checkbox"/> standby-pay
9	<input type="checkbox"/> shift-differential
10	<input type="checkbox"/> education-allowance
11	<input type="checkbox"/> statutory-holidays
12	<input type="checkbox"/> vacation
13	<input type="checkbox"/> longterm-disability-assistance
14	<input type="checkbox"/> contribution-to-dental-plan
15	<input type="checkbox"/> bereavement-assistance
16	<input type="checkbox"/> contribution-to-health-plan
17	<input type="checkbox"/> class

[Remove]

Selected attribute
Name: duration
Missing: 0 (0%)
Distinct: 3
Type: Numeric
Unique: 0 (0%)

Statistic	Value
Minimum	1
Maximum	3
Mean	2.13
StdDev	0.687

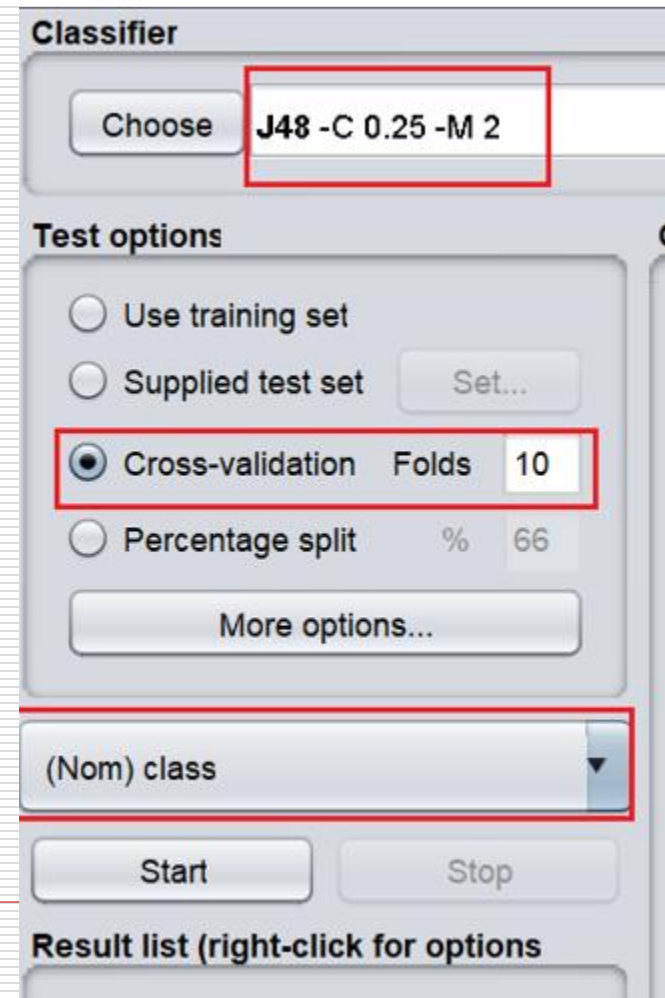
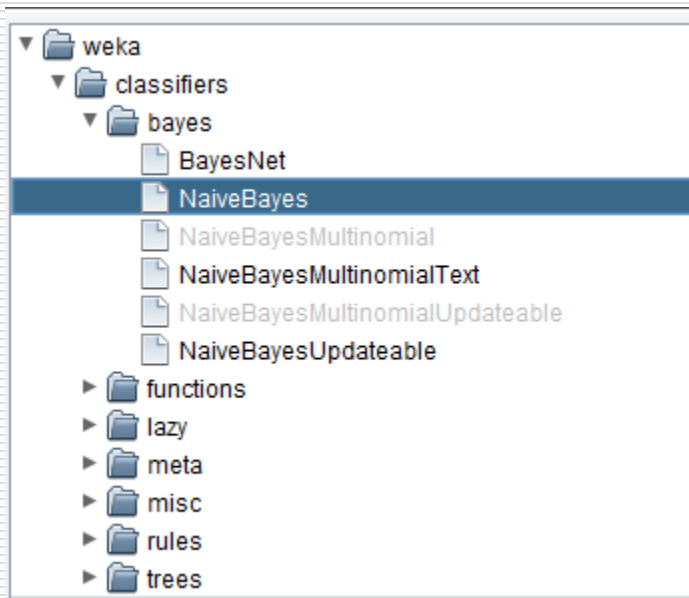
Class: class (Nom) [Visualize All]

1 2 3

Status: OK [Log] x 0

Thiết lập tham số

- ❖ Chuyển Tab Classify, chọn Choose > NaviveBayes
- ❖ Chọn Cross-validation



Huấn luyện

❖ Nhấn Start và xem kết quả:

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)

17:29:39 - trees.J48

Classifier output

Correctly Classified Instances	36	78.2609 %
Incorrectly Classified Instances	10	21.7391 %
Kappa statistic	0.5525	
Mean absolute error	0.2646	
Root mean squared error	0.4466	
Relative absolute error	53.7448 %	
Root relative squared error	89.9942 %	
Total Number of Instances	46	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
bad	0.700	0.154	0.778	0.700	0.737	0.555	0.750	0.657	bad
good	0.846	0.300	0.786	0.846	0.815	0.555	0.750	0.766	good
Weighted Avg.	0.783	0.236	0.782	0.783	0.781	0.555	0.750	0.718	

=== Confusion Matrix ===

```
a b <-- classified as
14 6 | a = bad
 4 22 | b = good
```

Status

OK Log x0

View cây quyết định

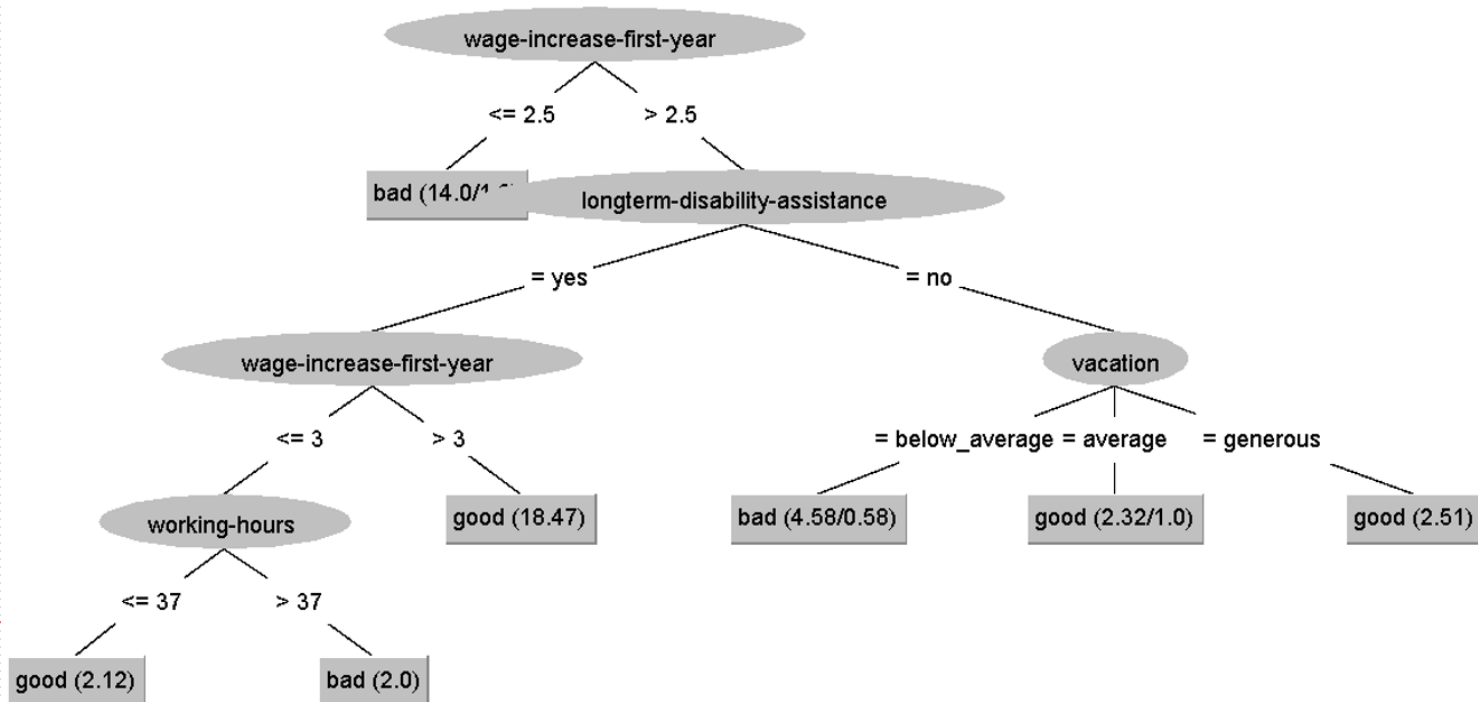
Result list (right-click for options) 0.700
0.846

17:29:39 - trees.J48

- 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

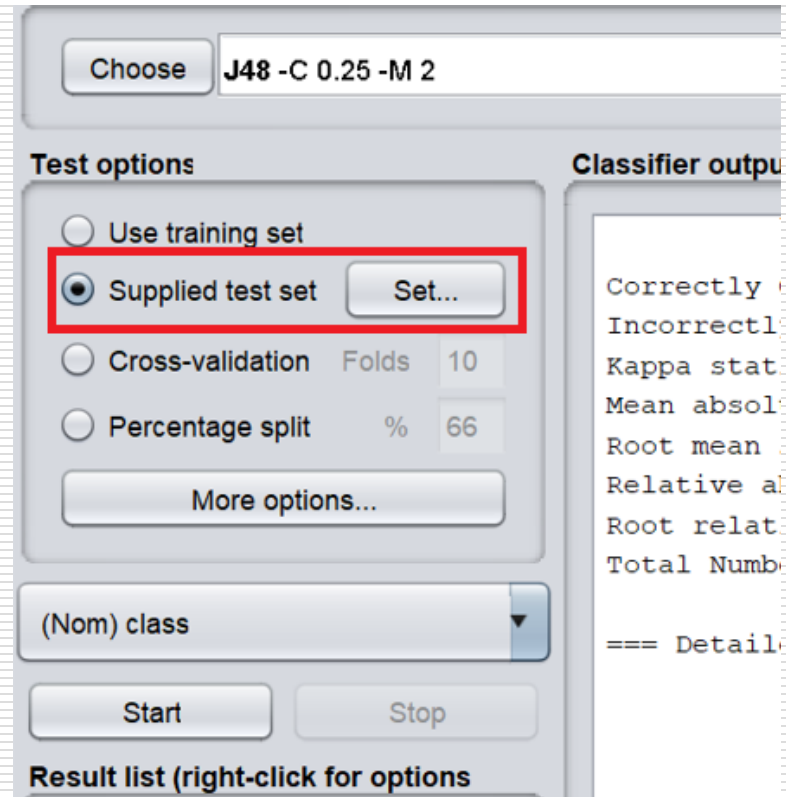
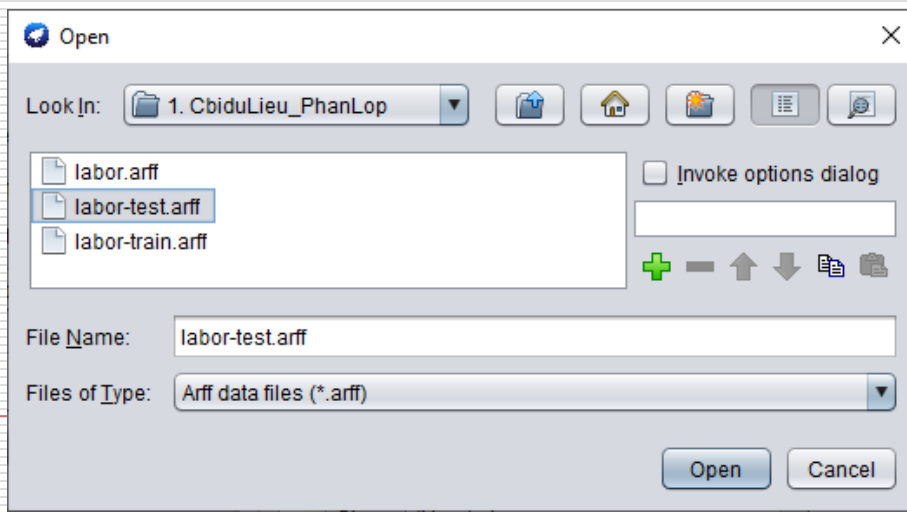
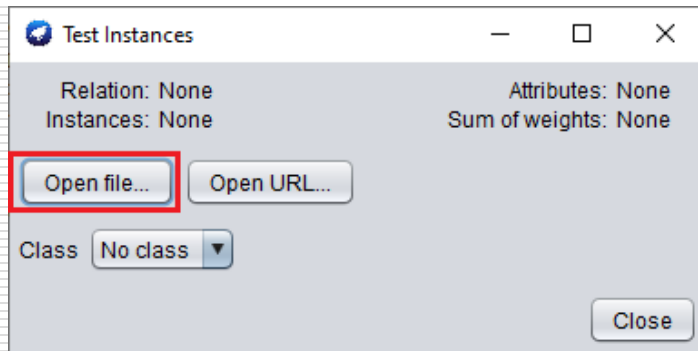
Status

OK



Kiểm thử

- ❖ Chọn Supplied test set
- ❖ Nhấn nút Set
- ❖ Nhấn nút Open file để chọn file test và nhấn Close



Kiểm thử

- ❖ Nhấn chuột phải vào dòng model muốn sử dụng trong Result list và xem kết quả:

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)

17:29:39 - trees.J48

Classifier output

Correctly Classified Instances	9	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0899		
Root mean squared error	0.1214		
Total Number of Instances	9		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MC
	1.000	0.000	1.000	1.000	1.000	1.
	1.000	0.000	1.000	1.000	1.000	1.
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.

=== Confusion Matrix ===

```
a b <-- classified as
3 0 | a = bad
0 6 | b = good
```

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)

17:29:39 - trees.J48

- 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

Status

Save và load model

The screenshot shows the Weka GUI with the 'Classifier output' tab selected. The 'Test options' section on the left shows 'Supplied test set' selected. The 'Classifier output' section displays the following results:

Metric	Value	Percentage
Correctly Classified Instances	9	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0.0899	
Root mean squared error	0.1214	
Total Number of Instances	9	

Below the main output, there is a section titled '=== Detailed Accuracy By Class ===' with a table showing performance metrics for 'bad' and 'good' classes, and a 'Weighted Avg.' row. All metrics are 1.000.

A context menu is open over the 'Result list' on the left, showing options for the file '17:29:39 - trees.J48'. The 'Save model' option is highlighted.

The 'Status' bar at the bottom left shows 'OK'. The 'Log' button and a small icon are visible at the bottom right.

❖ Nếu cần sử dụng lại, ta Load model