



PHÂN LỚP BAYES

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

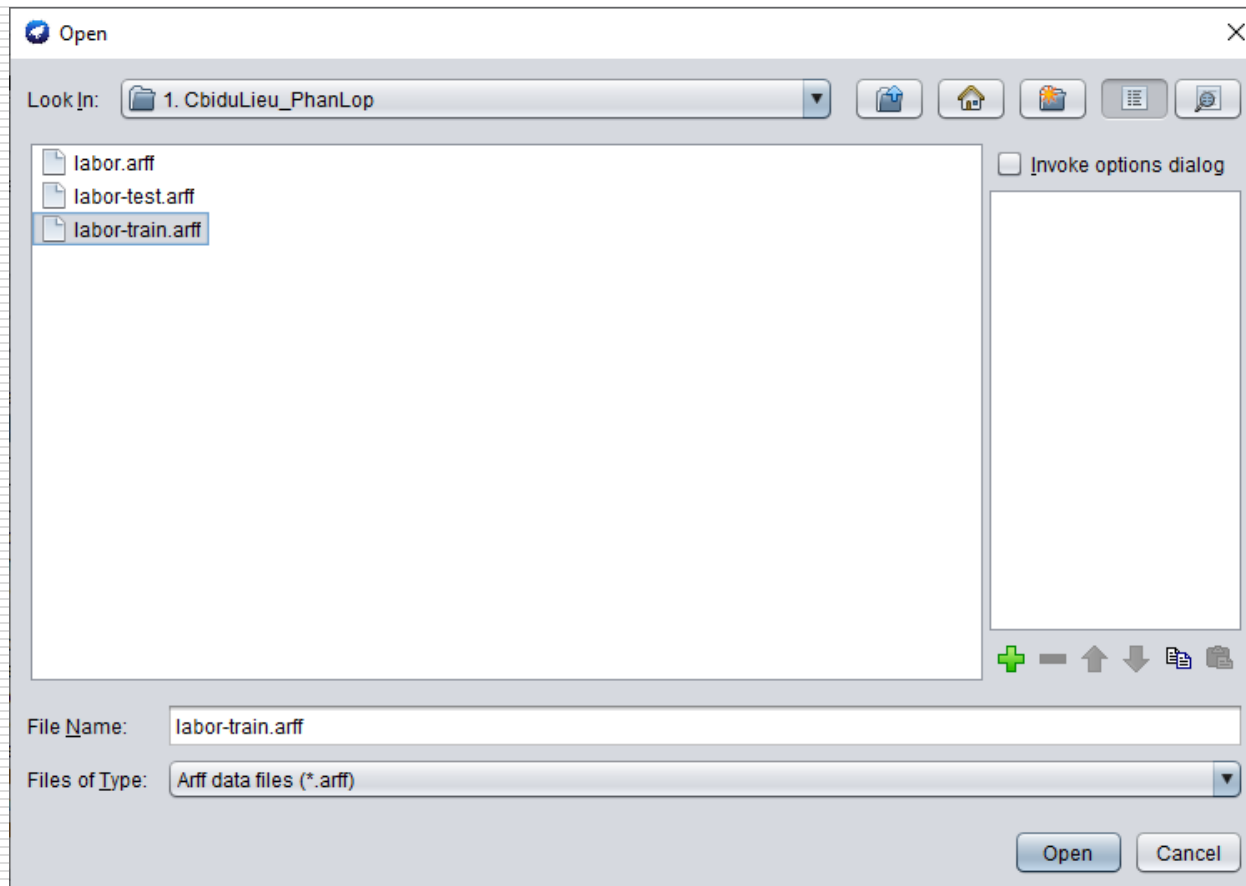
Hà Nội, 2021

Nội dung

- ❖ Mở file labor-train.arff
- ❖ Thiết lập tham số
- ❖ Huấn luyện
- ❖ Kiểm thử

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

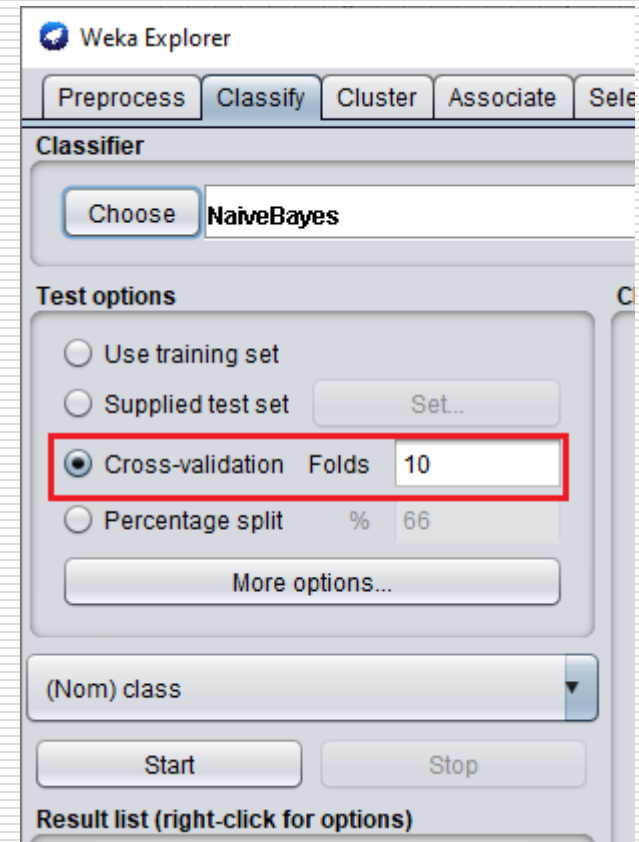
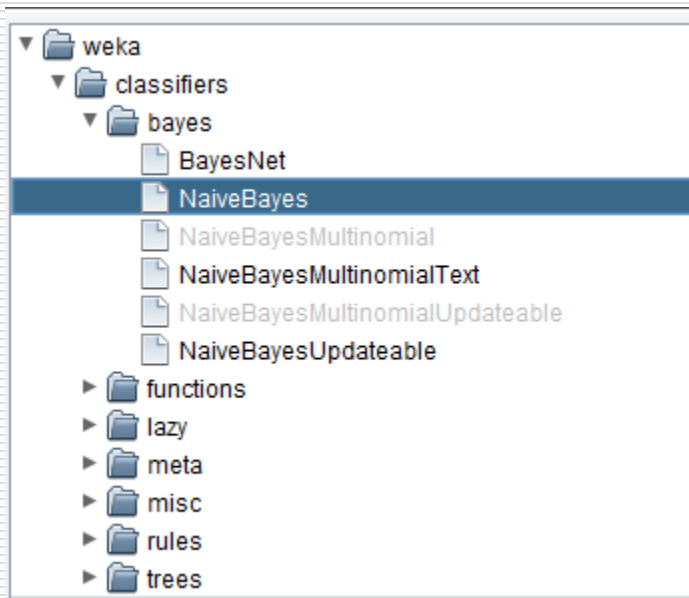
Class: class (Nom) Visualize All

Category	Count
1	8
2	24
3	14

Status: OK Log x 0

Thiết lập tham số

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



Huấn luyện

❖ Nhấn Start và xem kết quả:

The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected. The classifier chosen is 'NaiveBayes'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Start' button is highlighted with a red box. The 'Classifier output' pane shows the following results:

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances	42	91.3043 %
Incorrectly Classified Instances	4	8.6957 %

Kappa statistic 0.8231
Mean absolute error 0.0815
Root mean squared error 0.2536
Relative absolute error 16.5411 %
Root relative squared error 51.1015 %
Total Number of Instances 46

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.900	0.077	0.900	0.900	0.900	0.823	0.971	0.976	bad
	0.923	0.100	0.923	0.923	0.923	0.823	0.971	0.973	good
Weighted Avg.	0.913	0.090	0.913	0.913	0.913	0.823	0.971	0.974	

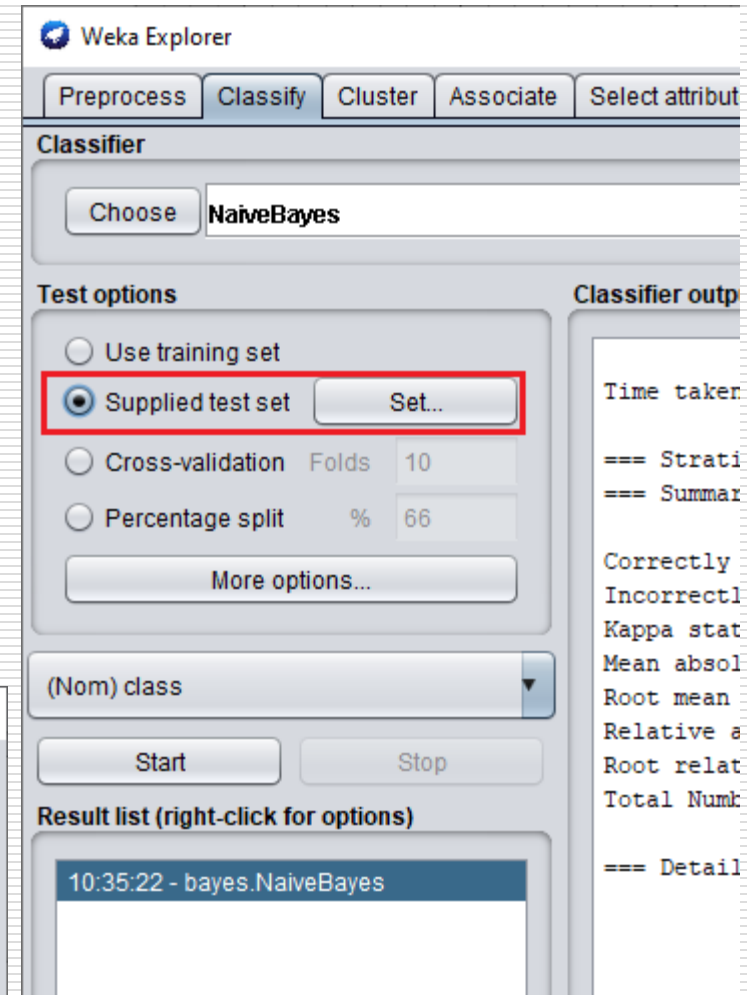
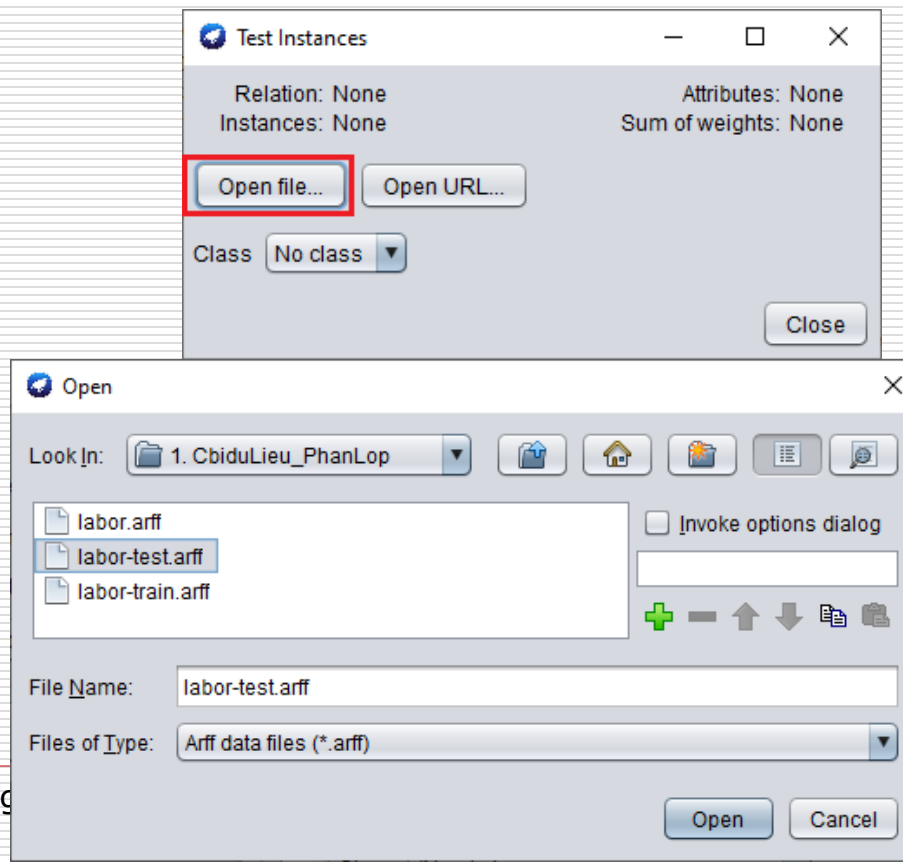
=== Confusion Matrix ===

```
a b <-- classified as
18 2 | a = bad
 2 24 | b = good
```

The 'Result list' on the left shows a single entry: '10:35:22 - bayes.NaiveBayes'.

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:

Classifier

Choose **NaiveBayes**

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)

10:35:22 - bayes.NaiveBayes

Classifier output

User supplied test set

Relation: labor-neg-data-weka.filters.unsupervised.instance.RemovePercentage-P20.0-V

Instances: unknown (yet). Reading incrementally

Attributes: 17

=== Summary ===

Correctly Classified Instances	9	81.8182 %
Incorrectly Classified Instances	2	18.1818 %
Kappa statistic	0	
Mean absolute error	0.1692	
Root mean squared error	0.356	
Total Number of Instances	11	

=== Detailed Accuracy By Class ===

	ate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
8	0.182	0.000	?	?	?	?	?	?	bad
8	?	1.000	0.818	0.900	?	?	?	1.000	good
8	?	1.000	0.818	0.900	?	?	?	1.000	

Context Menu:

- 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

Status

OK

Log