



PHÂN LỚP KNN

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

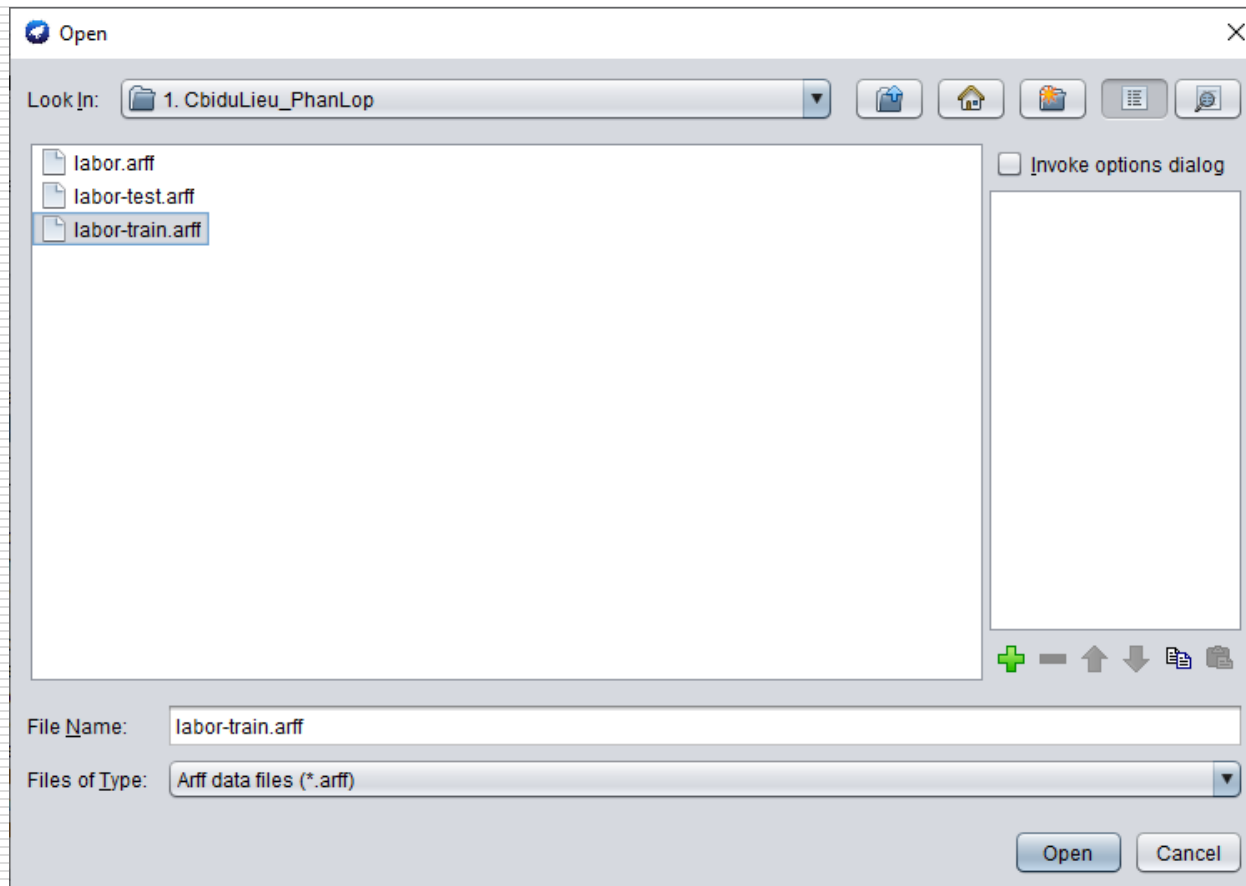
Hà Nội, 2021

Nội dung

- ❖ Mở file labor-train.arff
- ❖ Thiết lập tham số
- ❖ 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

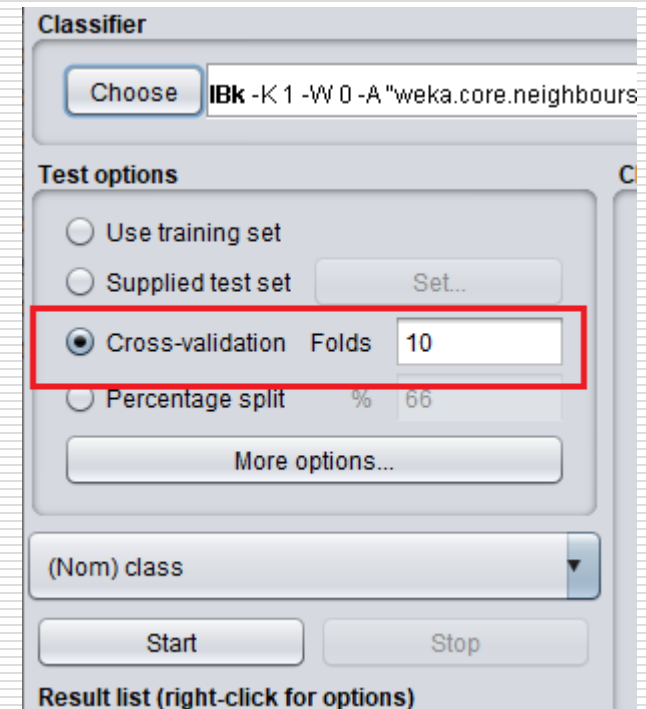
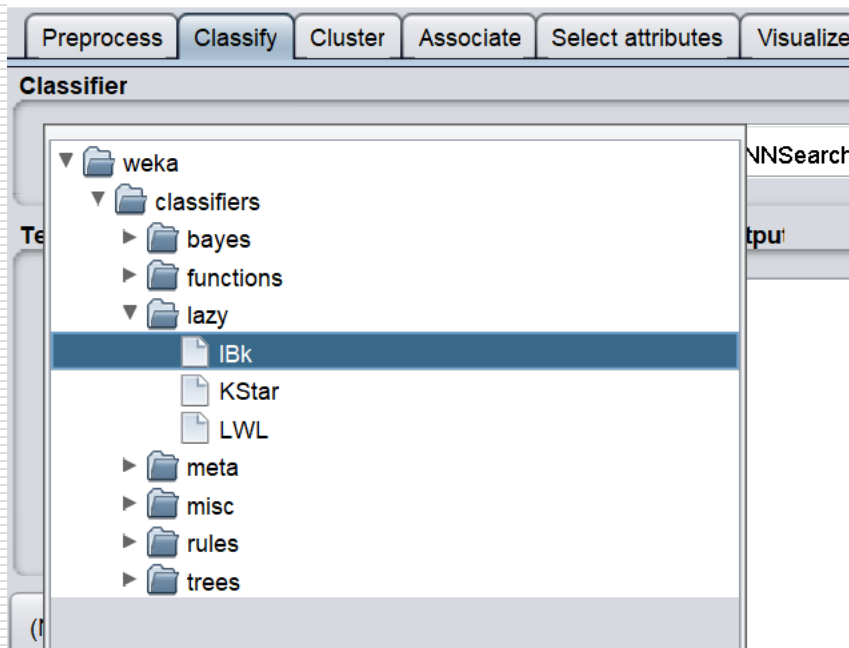
Class: class (Nom) Visualize All

Category	Count
1	8
2	24
3	14

Status: OK Log x 0

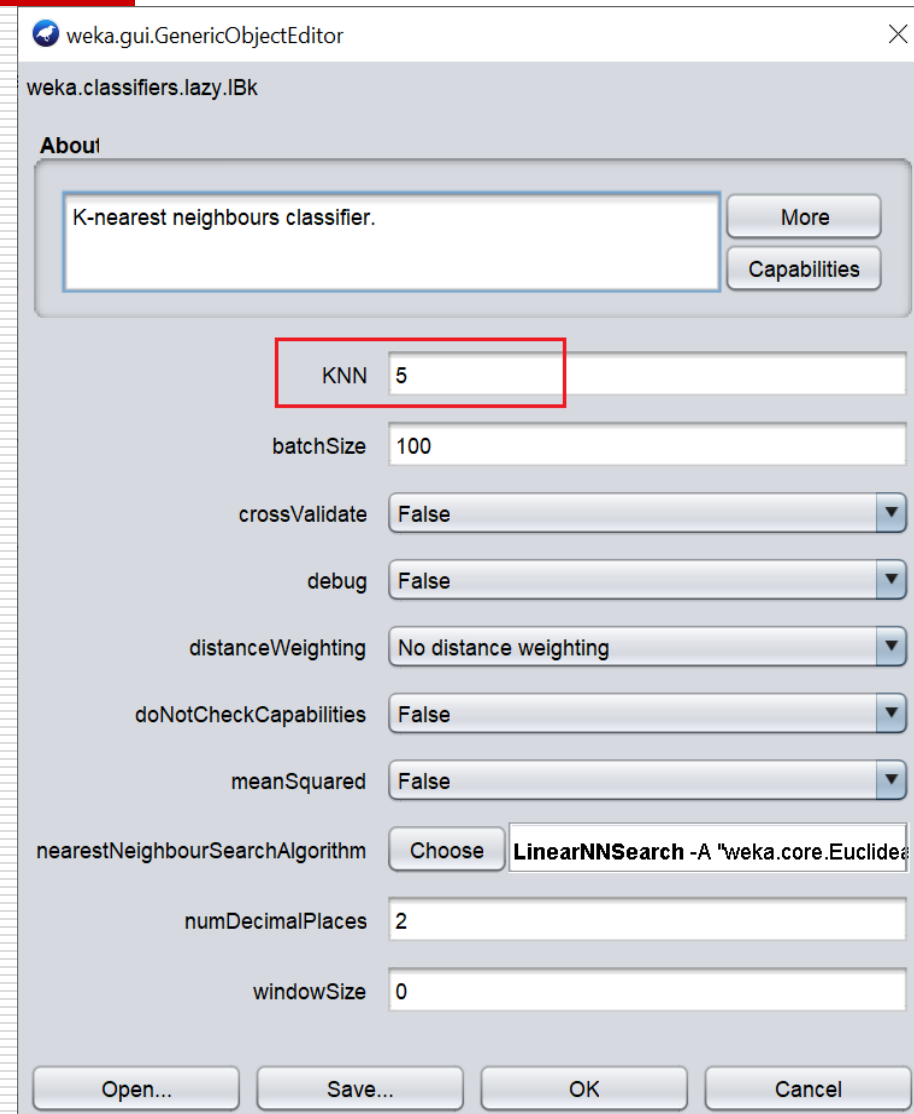
Thiết lập tham số

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



Thiết lập tham số

- ❖ Nhấn chuột vào textbox bên phải nút Choose
- ❖ Nhập số lân cận



Huấn luyện

- ❖ Chọn thuộc tính phân lớp
- ❖ Nhấn Start và xem kết quả:

Choose **IBk -K 5 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

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)

- 19:06:46 - lazy.IBk

Classifier output

```
==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances      40      86.9565 %
Incorrectly Classified Instances    6      13.0435 %
Kappa statistic                    0.7315
Mean absolute error                 0.1986
Root mean squared error             0.2967
Relative absolute error             40.3248 %
Root relative squared error         59.7913 %
Total Number of Instances          46

==== Detailed Accuracy By Class ====
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.077	0.889	0.800	0.842	0.734	0.949	0.931	bad
	0.923	0.200	0.857	0.923	0.889	0.734	0.949	0.957	good
Weighted Avg.	0.870	0.146	0.871	0.870	0.869	0.734	0.949	0.946	

```
==== Confusion Matrix ====

a  b  <-- classified as
```

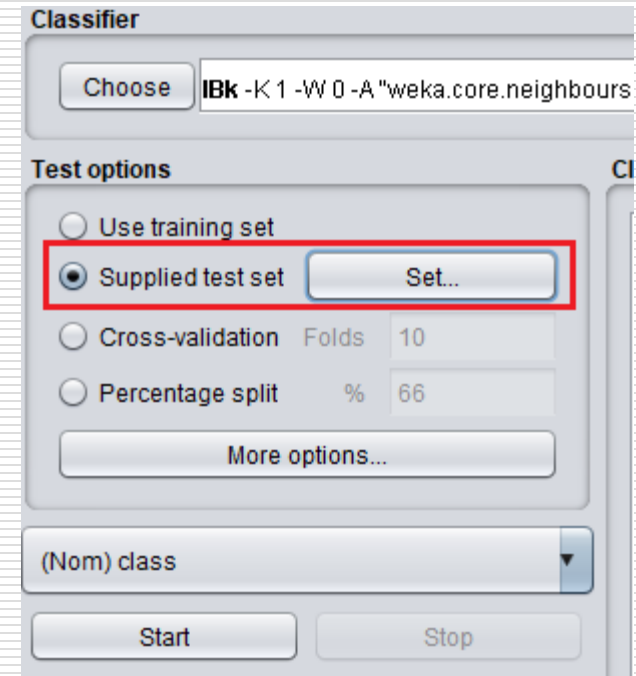
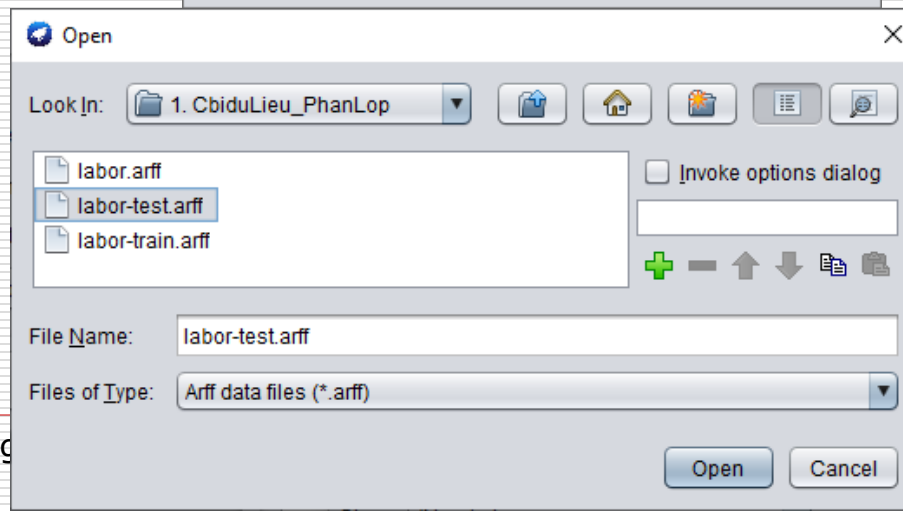
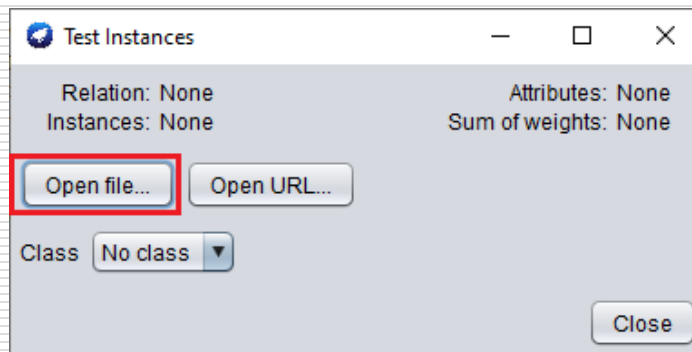
Status

OK

Log x 0

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:

The screenshot shows the Weka GUI. At the top, a text box contains the command: `IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"`. Below this, the "Test options" panel has "Supplied test set" selected. The "Classifier output" panel displays the following text:

```
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      7      63.6364 %
Incorrectly Classified Instances    4      36.3636 %
Kappa statistic                     0
Mean absolute error                 0.3693
Root mean squared error             0.5907
Total Number of Instances          11

=== Detailed Accuracy By Class ===
```

A red box highlights the "Summary" section. Below the "Result list (right-click for options)" panel, a context menu is open for the entry "14:26:44 - lazy.IBk". The menu options are:

- 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

The "Detailed Accuracy By Class" table is partially visible, showing columns for TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, PRC Area, and Class Name. The first row shows values for "bad" and "good" classes.