

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc Lập – Tự do – Hạnh Phúc**

-----0-----

# **BÁO CÁO ĐỒ ÁN 3**

**Môn: Cơ Sở Trí Tuệ Nhân Tạo**

**Tên đề tài: Máy học**



Thực hiện: 1612286 Nguyễn Hồng Khang

1612778 Nguyễn Anh Tuấn

## 1. Tìm hiểu công cụ Weka:

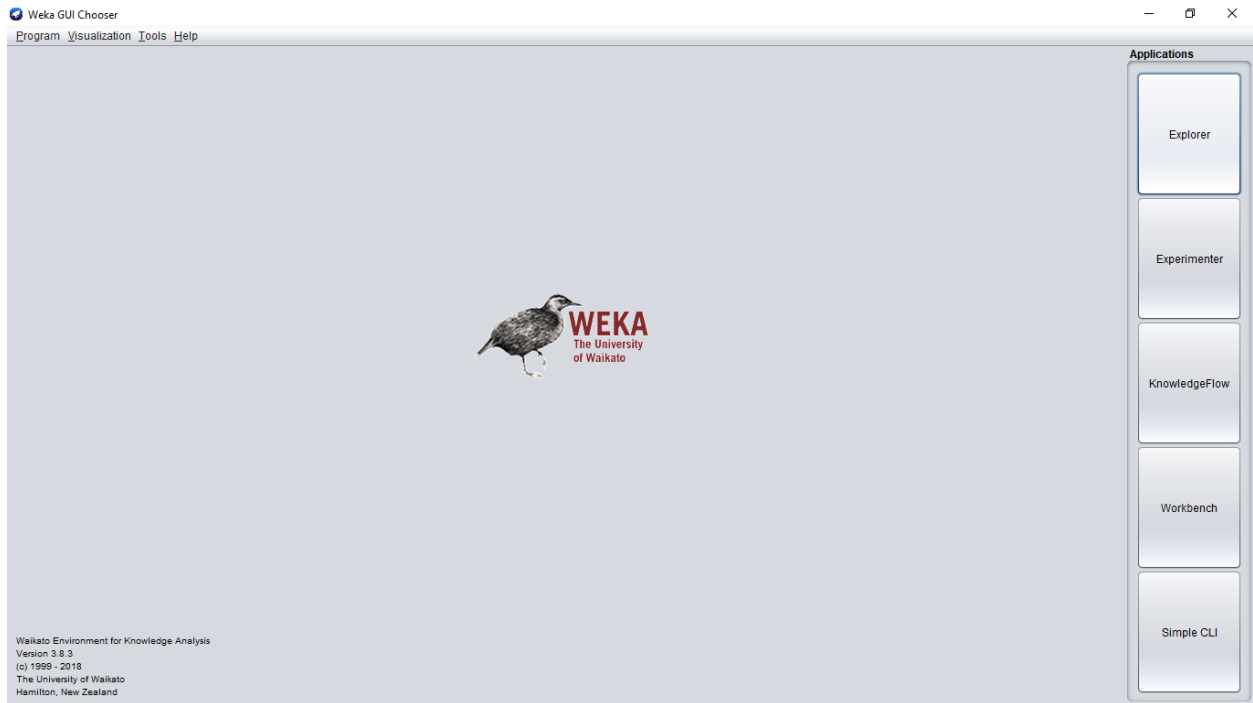
### a. Những điểm ưu việt của Weka:

- Hỗ trợ nhiều thuật toán máy học và khai thác dữ liệu.
- Mã nguồn mở
- Độc lập với môi trường làm việc ( do sử dụng JVM)
- Dễ sử dụng, kiến trúc dạng thư viện dễ dàng cho việc xây dựng ứng dụng thực nghiệm.
- Thư viện của Weka có hơn 500 lớp và tổ chức thành 10 package. Có thể sử dụng các lớp này trực tiếp trên phần mềm hoặc sử dụng các thư viện này để phát triển các ứng dụng.

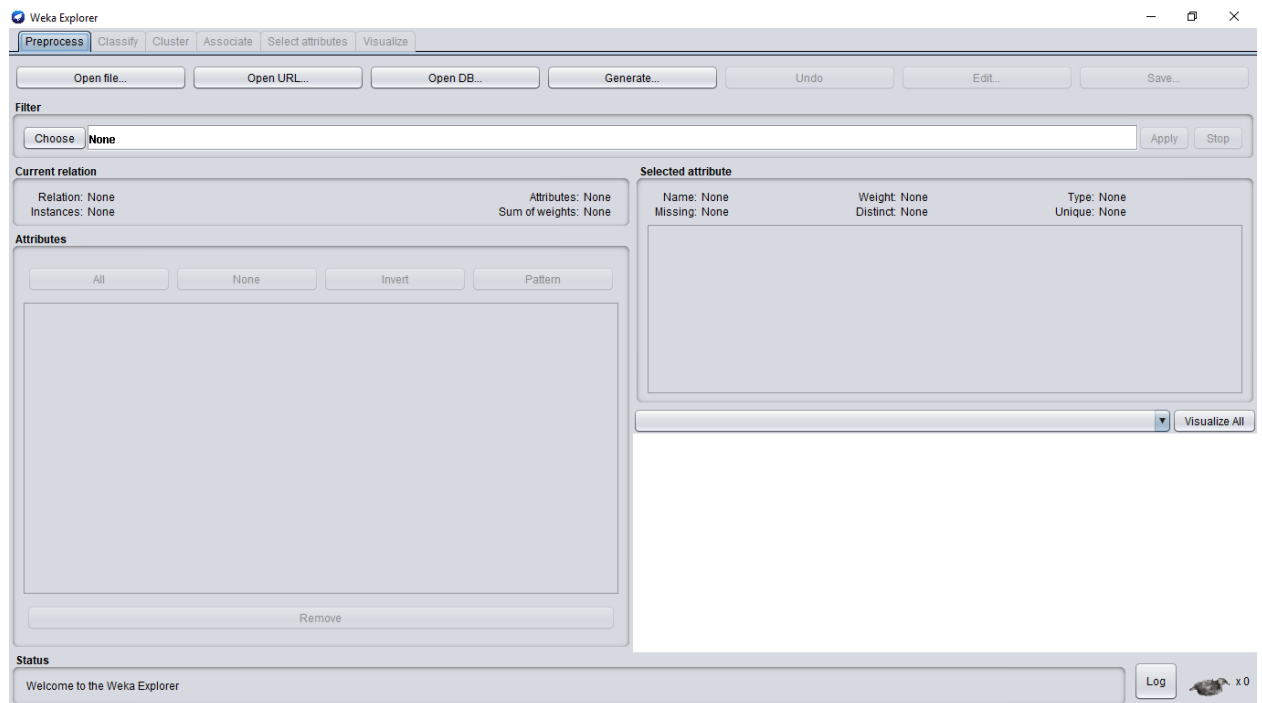
### b. Các môi trường chính của Weka:

- Explorer: cung cấp môi trường để khảo sát dữ liệu
- Experimenter: môi trường để tiến hành các thực nghiệm và kiểm chứng mô hình học
- Knowledflow: tương tự như Explorer nhưng có giao diện kéo thả. Có hỗ trợ incremental learning
- SimpleCLI: cung cấp giao diện dòng lệnh để thực thi Weka

Chúng ta chủ yếu sử dụng trong môi trường Explorer để thao tác với gần như đầy đủ các tính năng của Weka



### c. Các chức năng chính trong môi trường Explorer:



- Tiền xử lý (Preprocessing): Để xử lý dữ liệu trước khi sử dụng các tính năng khác. Tính năng này giúp chúng ta có thể đọc dữ liệu dưới nhiều hình thức, hiệu chỉnh dữ liệu, biểu diễn thông tin về tập dữ liệu, lưu trữ dữ liệu.

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose AddCluster -W "weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 1" Apply Stop

Current relation: Relation: weather.symbolic Instances: 14 Attributes: 5 Sum of weights: 14

Attributes:

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Remove

Selected attribute:

Name: humidity Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	high	7	7.0
2	normal	7	7.0

Class: play (Nom) Visualize All

Status: OK Log x0

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose AddCluster -W "weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 1" Apply Stop

Current relation: Relation: weather.symbolic Instances: 14 Attributes: 5 Sum of weights: 14

Attributes:

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Remove

Viewer

Relation: weather.symbolic

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
1	sunny	hot	high	normal	no
2	sunny	mild	normal	normal	yes
3	overcast	mild	high	normal	yes
4	overcast	cool	normal	normal	yes
5	rainy	cool	high	normal	yes
6	rainy	cool	normal	normal	yes
7	sunny	hot	high	normal	yes
8	sunny	mild	high	normal	yes
9	sunny	cool	normal	normal	yes
10	overcast	hot	high	normal	yes
11	overcast	hot	normal	normal	yes
12	rainy	mild	high	normal	yes
13	rainy	mild	normal	normal	yes
14	rainy	cool	normal	normal	yes

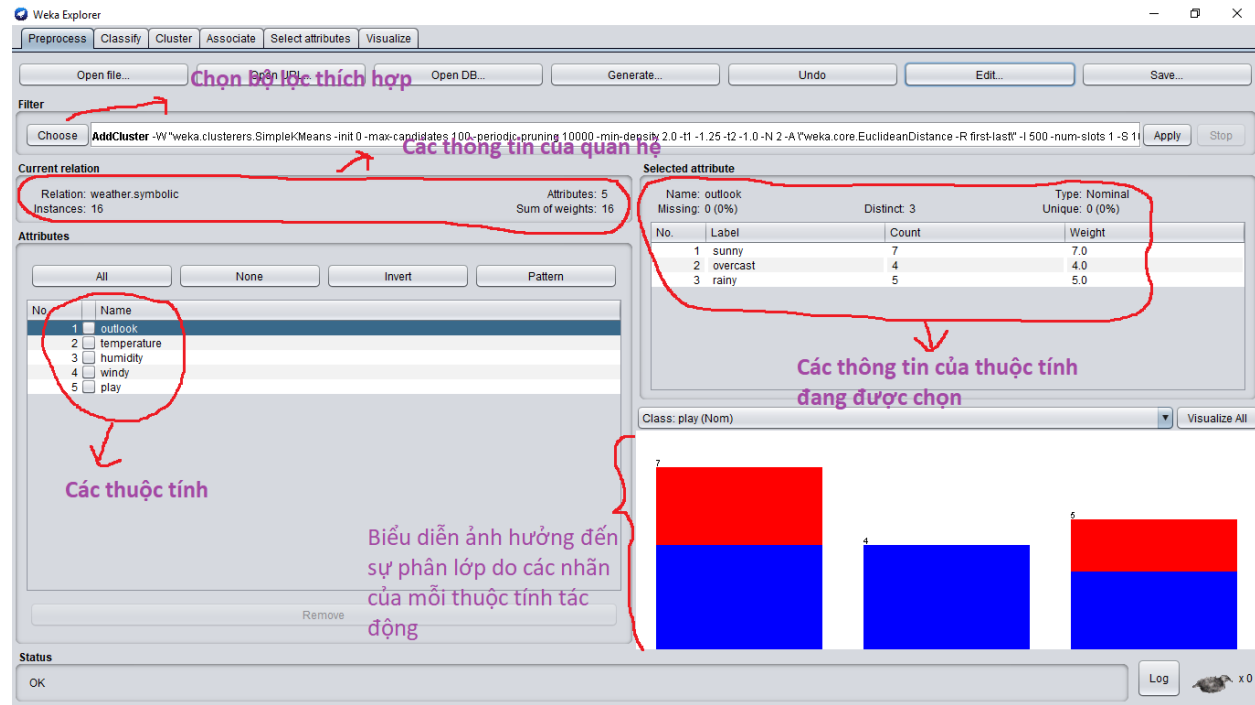
Get mean...  
Set all values to...  
Set missing values to...  
Replace values with...  
Rename attribute...  
Set attribute weight...  
Attribute as class  
Delete attribute  
Delete attributes...  
Sort data (ascending)  
Optimal column width (current)  
Optimal column width (all)

tùy chọn thiết lập giá trị cho thuộc tính

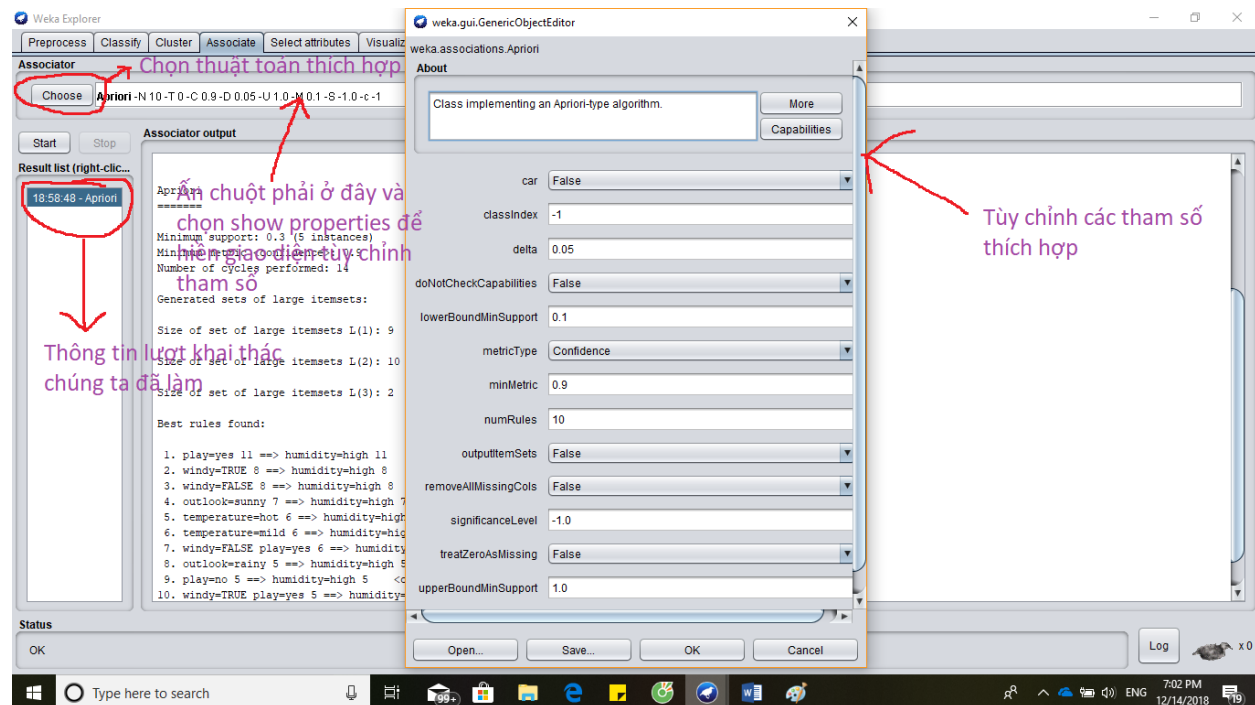
thiết lập các tính chất của thuộc tính

thiết lập độ rộng của cột

Status: OK Log x0



- Khai thác luật kết hợp (Associate): Là chức năng giúp người dùng khai thác luật kết hợp từ một tập dữ liệu đã cho. Chương trình cho phép sử dụng nhiều thuật toán khác nhau để khai báo luật kết hợp như là: Apriori, PredictApriori, Tertius.



- Phân lớp (Classify): là chức năng giúp người dùng huấn luyện và kiểm chứng các mô hình phân lớp và thực hiện hồi quy. Phần mềm cung cấp nhiều thuật toán khác nhau như Cây quyết định, NaiveBayes.

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

**Classifier**

Choose ZeroR

**Test options**

☐ Use training set  
☐ Supplied test set  
☒ Cross-validation Folds 10  
☐ Percentage split % 65  
 More options...

(Nom) play

Start Stop

**Result list (right-click for options)**

0:16:19 - rules.ZeroR

**Classifier output**

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	11	68.75 %
Incorrectly Classified Instances	5	31.25 %
Kappa statistic	0	
Mean absolute error	0.4476	
Root mean squared error	0.4764	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	16	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.688	0.688	?	0.688	?	?	0.227	0.501	

=== Confusion Matrix ===

a b <-- classified as

11	0	a = yes
5	0	b = no

Status

OK

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

**Classifier**

Choose ZeroR

**Test options**

☐ Use training set  
☐ Supplied test set  
☒ Cross-validation Folds 10  
☐ Percentage split % 65  
 More options...

(Nom) play

Start Stop

**Result list (right-click for options)**

20:16:19 - rules.ZeroR

**Classifier output**

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	11	68.75 %
Incorrectly Classified Instances	5	31.25 %
Kappa statistic	0	
Mean absolute error	0.4476	
Root mean squared error	0.4764	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	16	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.688	0.688	?	0.688	?	?	0.227	0.501	

=== Confusion Matrix ===

a b <-- classified as

11	0	a = yes
5	0	b = no

Status

OK

**weka.gui.GenericObjectEditor**

weka.classifiers.rules.ZeroR

About

Class for building and using a 0-R classifier.

batchSize 100

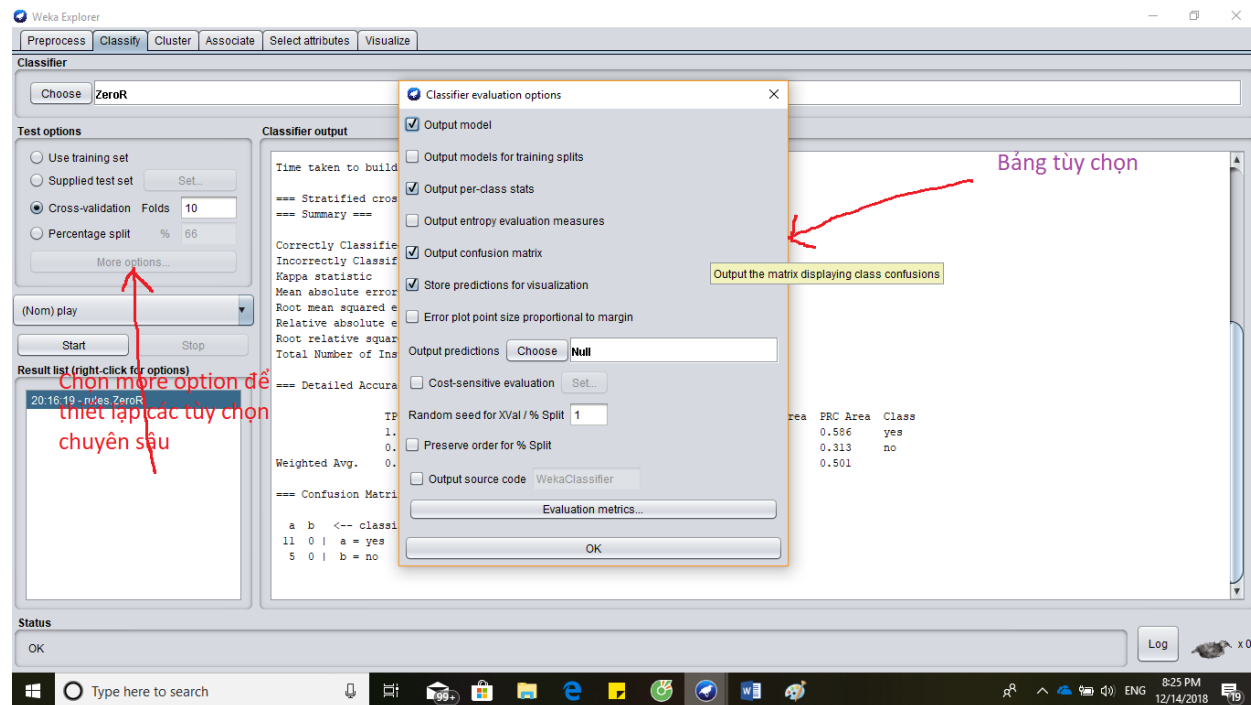
debug False

doNotCheckCapabilities False

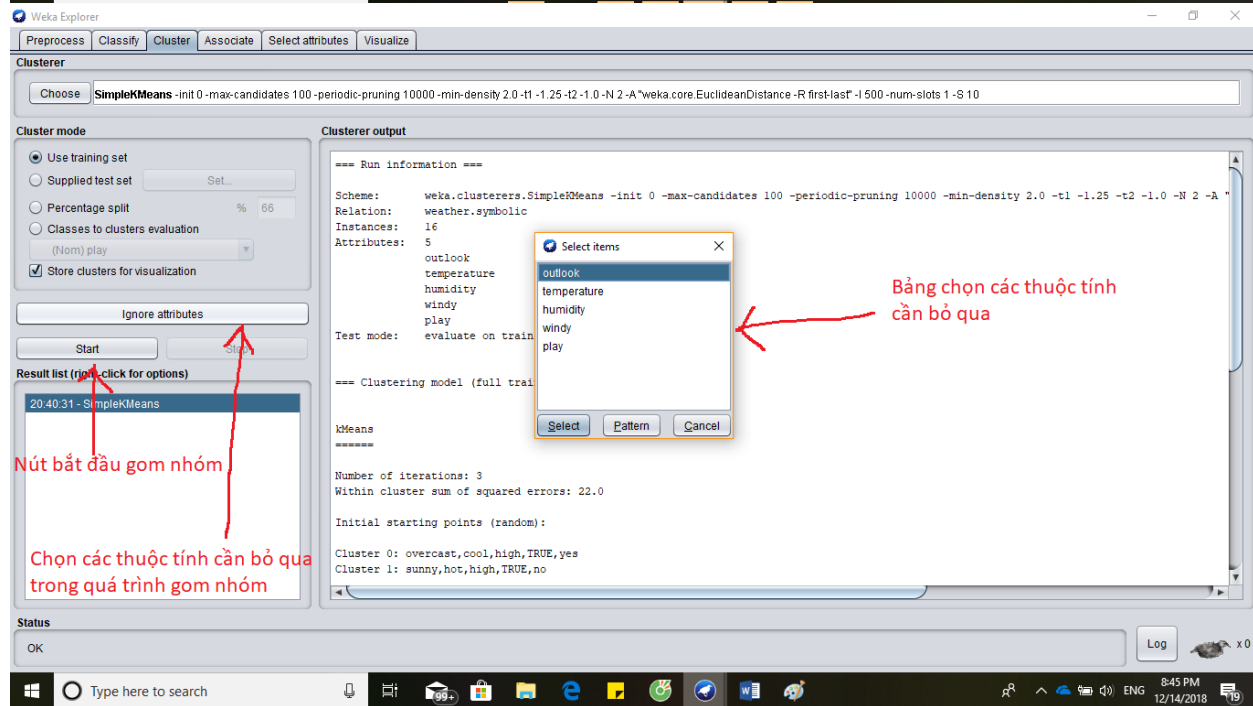
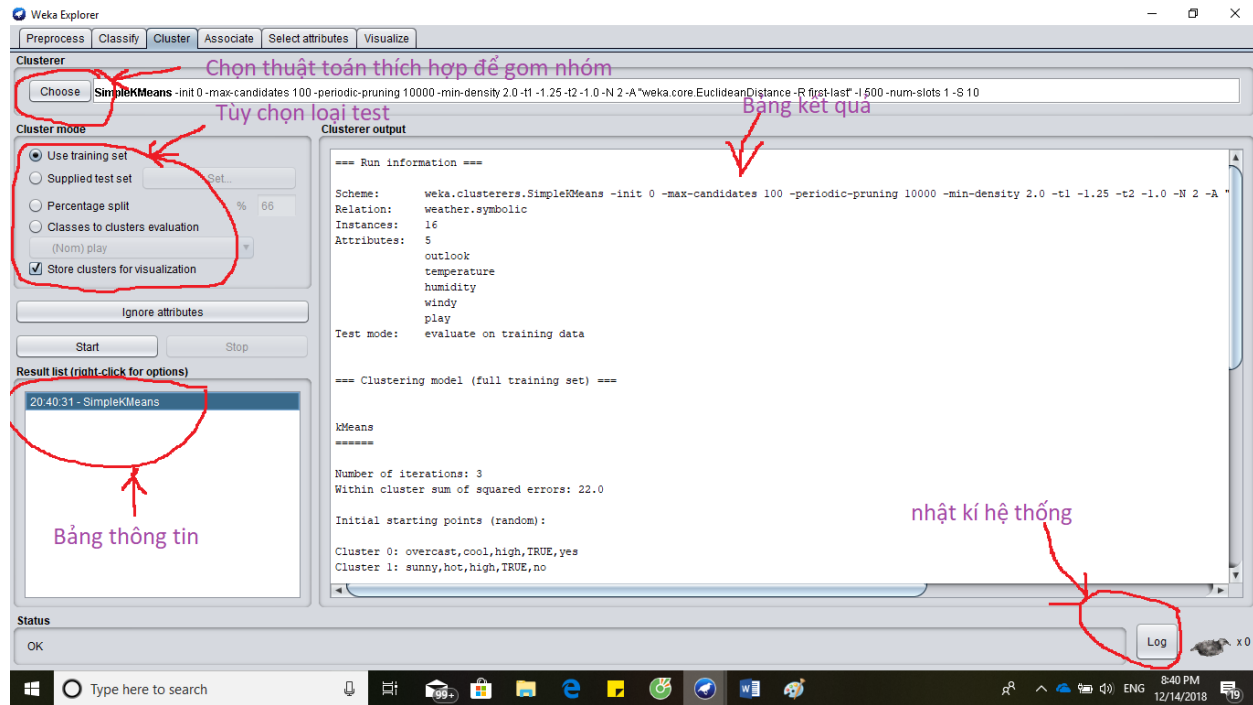
numDecimalPlaces 2

Open... Save... OK Cancel





- Gom nhóm (Cluster): Là chức năng giúp người dùng gom nhóm dữ liệu từ một tập dữ liệu đã cho. Weka sẽ hỗ trợ các thuật toán như DBScan, SimpleKmeans,...



## 2. Sử dụng Weka để chạy thuật toán ID3:

Số mẫu: Total Number of Instances                      101

Tên và ý nghĩa các thuộc tính:

Animal name: Tên động vật (mỗi tên đều độc nhất)

Hair: lông (có hoặc không)

Feathers: lông vũ (Có hoặc không)

Eggs: Đẻ trứng (Có hoặc không)

Milk: Cho sữa (Có hoặc không)

Airborne: Bay được (Có hoặc không)

Aquatic: Sống dưới nước (Có hoặc không)

Predator: Là động vật săn mồi (Có hoặc không)

Toothed: răng (Có hoặc không)

Backbone: xương sống (Có hoặc không)

Breathes: Thở (Có hoặc không)

Venomous: Có độc (Có hoặc không)

Fins: Có vây (Có hoặc không)

Legs: Số chân(0,2,4,5,6,8)

Tail: Đuôi (Có hoặc không)

Domestic: thuần hóa (Có hoặc không)

Catsize: (Có hoặc không)

Type: Phân lớp (1,2,3,4,5,6,7)

Loại phân lớp:

- 1: Mammal(Động vật có vú)
- 2: Fish(Cá)
- 3: Bird(Chim)
- 4: Arthropod(Động vật chân khớp)

- 5: Insect(Côn trùng)
- 6: Amphibian(Động vật lưỡng cư)
- 7: Reptile(Động vật bò sát)

Id3:

```
=== Classifier model (full training set) ===
```

Id3

```
legs = 0
| fins = 0
| | toothed = 0: Arthropod
| | toothed = 1: Reptile
| fins = 1
| | eggs = 0: Mammal
| | eggs = 1: Fish
legs = 2
| hair = 0: Bird
| hair = 1: Mammal
legs = 4
| hair = 0
| | aquatic = 0: Reptile
| | aquatic = 1
| | | toothed = 0: Arthropod
| | | toothed = 1: Amphibian
| hair = 1: Mammal
legs = 5: Arthropod
legs = 6
| aquatic = 0: Insect
| aquatic = 1: Arthropod
legs = 8: Arthropod
```

5 Mẫu:

- o 1. NameIsSecret,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1, ?
- o 2. NameIsSecret,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0, ?
- o 3. NameIsSecret,0,0,1,0,0,0,1,1,1,1,0,0,1,0,0, ?
- o 4. NameIsSecret,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0, ?
- o 5. NameIsSecret,0,0,1,0,0,1,1,1,1,1,0,0,4,1,0,0, ?

Kết quả predict:

=== Predictions on user test set ===

inst#	actual	predicted	error	prediction
1	1:?	1:Mammal		1
2	1:?	3:Bird		1
3	1:?	7:Reptile		1
4	1:?	2:Fish		1
5	1:?	6:Amphibian		1

-

### 3. Chạy thử các thuật toán khác:

1. Naïve Bayes:

Kết quả chạy trong file [NaïveBayes.txt](#)

2. BFTree:

Kết quả chạy trong file [BFTree.txt](#)

Best-First Decision Tree

```
milk=(1): Mammal(41.0/0.0)
milk!=(1)
| feathers=(1): Bird(20.0/0.0)
| feathers!=(1)
| | fins=(1): Fish(13.0/0.0)
| | fins!=(1)
| | | backbone=(1)
| | | | aquatic=(1): Amphibian(4.0/1.0)
| | | | aquatic!=(1): Reptile(4.0/0.0)
| | | backbone!=(1)
| | | | legs=(6)|(2)
| | | | | aquatic=(1): Arthropod(2.0/0.0)
| | | | | aquatic!=(1): Insect(8.0/0.0)
| | | | legs!=(6)|(2): Arthropod(8.0/0.0)
```

Size of the Tree: 15

Number of Leaf Nodes: 8

Time taken to build model: 0.03 seconds

3. SimpleCART:

Kết quả chạy trong file [simpleCART.txt](#)

# CART Decision Tree

```
milk=(1): Mammal(41.0/0.0)
milk!=(1)
| feathers=(1): Bird(20.0/0.0)
| feathers!=(1)
| | fins=(1): Fish(13.0/0.0)
| | fins!=(1)
| | | backbone=(1)
| | | | aquatic=(1): Amphibian(4.0/1.0)
| | | | aquatic!=(1): Reptile(4.0/0.0)
| | | backbone!=(1)
| | | | legs=(6)|(2): Insect(8.0/2.0)
| | | | legs!=(6)|(2): Arthropod(8.0/0.0)
```

Number of Leaf Nodes: 7

Size of the Tree: 13

Time taken to build model: 0.13 seconds