

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1

Cluster mode

☒ Use training set

☐ Supplied test set Set...

☐ Percentage split % 66

☐ Classes to clusters evaluation (Nom) Class

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

Clusterer output

kMeans

Number of iterations: 3

Within cluster sum of squared errors: 18.698918560481786

Initial starting points (random):

Cluster 0: 4,Dang,M,25,82,A

Cluster 1: 9,Gaga,F,14,70,C

Cluster 2: 12,Rika,F,19,77,B

Cluster 3: 7,Sompong,M,18,78,B

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	0	1	2	3
	(15.0)	(1.0)	(5.0)	(4.0)	(5.0)	
ID	8	4	10.2	11	4.2	
Name	John	Dang	David	Sunee	John	
SEX	F	M	F	F	M	
Age	16.8667	25	12	24	14.4	
Score	70.4	82	61.6	83	66.8	
Class	B	A	C	A	B	

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	1 (7%)
1	5 (33%)
2	4 (27%)
3	5 (33%)

Status OK

Log x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set Set...

☐ Percentage split % 66

☐ Classes to clusters evaluation (Nom) Class

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

17:40:43 - SimpleKMeans

Clusterer output

kMeans

Number of iterations: 3

Within cluster sum of squared errors: 18.698918560481786

Initial starting points (random):

Cluster 0: 4,Dang,M,25,82,A

Cluster 1: 9,Gaga,F,14,70,C

Cluster 2: 12,Rika,F,19,77,B

Cluster 3: 7,Sompong,M,18,78,B

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	0	1	2	3
	(15.0)	(1.0)	(5.0)	(4.0)	(5.0)	
ID	8	4	10.2	11	4.2	
Name	John	Dang	David	Sunee	John	
SEX	F	M	F	F	M	
Age	16.8667	25	12	24	14.4	
Score	70.4	82	61.6	83	66.8	
Class	B	A	C	A	B	

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	1 (7%)
1	5 (33%)
2	4 (27%)
3	5 (33%)

Status OK

Log x 0

5

View in main window

View in separate window

Save result buffer

Delete result buffer(s)

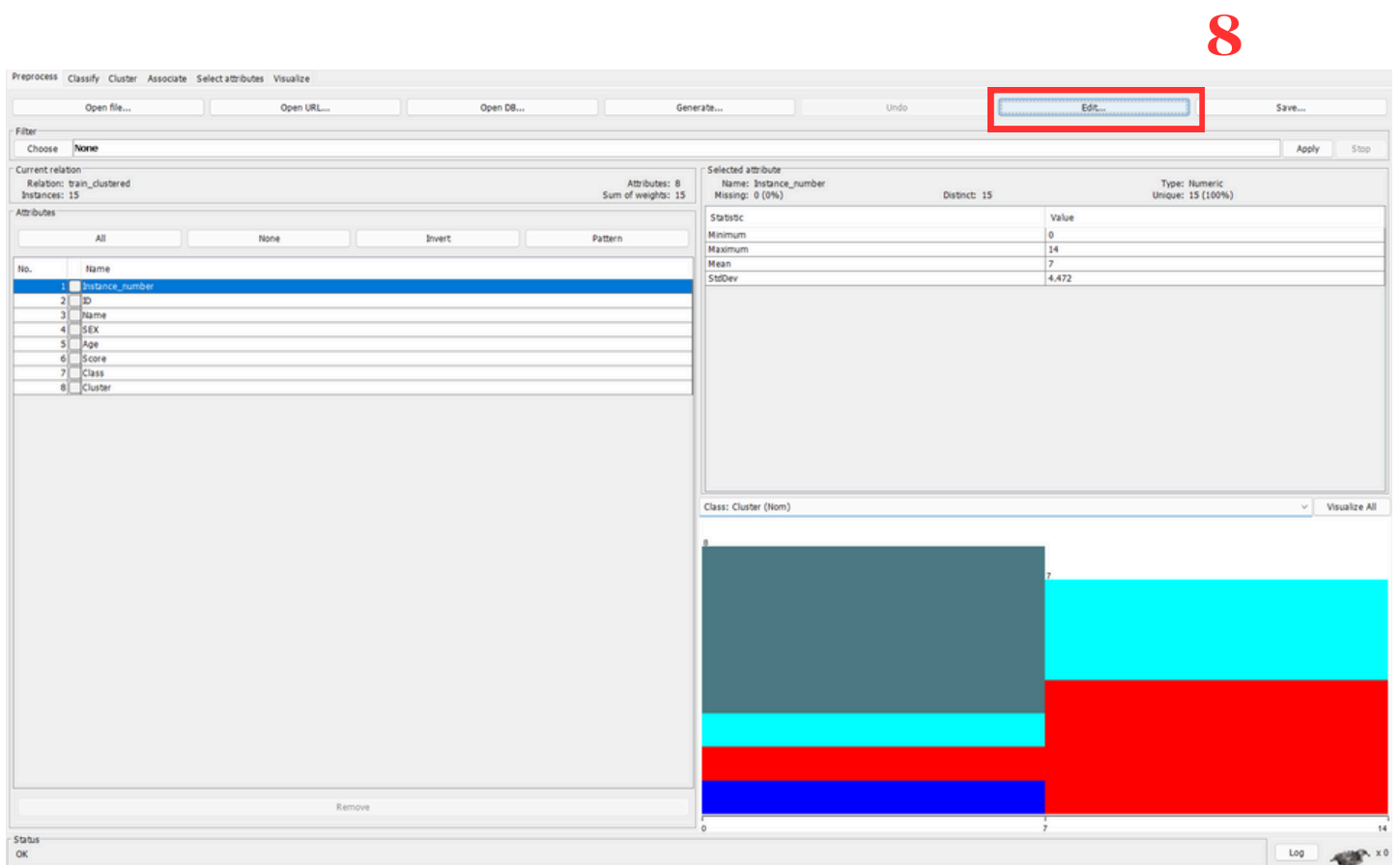
Load model

Save model

Re-evaluate model on current test set

Re-apply this model's configuration

Visualize cluster assignments



Relation: train_clustered

No.	1: Instance_number Numeric	2: ID Numeric	3: Name Nominal	4: SEX Nominal	5: Age Numeric	6: Score Numeric	7: Class Nominal	8: Cluster Nominal
1	0.0	1.0	John	M	17.0	71.0	B	cluster3
2	1.0	2.0	Somsak	M	18.0	75.0	B	cluster3
3	2.0	3.0	David	M	13.0	69.0	C	cluster1
4	3.0	4.0	Dang	M	25.0	82.0	A	cluster0
5	4.0	5.0	Mana	M	10.0	50.0	D	cluster3
6	5.0	6.0	Dum	M	9.0	60.0	D	cluster3
7	6.0	7.0	Sompong	M	18.0	78.0	B	cluster3
8	7.0	8.0	Sunee	F	26.0	85.0	A	cluster2
9	8.0	9.0	Gaga	F	14.0	70.0	C	cluster1
10	9.0	10.0	Lumyai	F	11.0	52.0	D	cluster1
11	10.0	11.0	Momoko	F	24.0	87.0	A	cluster2
12	11.0	12.0	Rika	F	19.0	77.0	B	cluster2
13	12.0	13.0	Manee	F	27.0	83.0	A	cluster2
14	13.0	14.0	Risa	F	14.0	62.0	C	cluster1
15	14.0	15.0	Jessica	F	8.0	55.0	D	cluster1

Clusterer output

```

=== Run information ===
Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning
Relation:    train
Instances:   15
Attributes:  6
              ID
              Name
              SEX
              Age
              Score
              Class
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 18.698918560481786
Initial starting points (random):
Cluster 0: 4,Dang,M,25,82,A
Cluster 1: 9,Gaga,F,14,70,C
Cluster 2: 12,Rika,F,19,77,B
Cluster 3: 7,Sompong,M,18,78,B
Missing values globally replaced with mean/mode
  
```

ส่วนนี้คือ ข้อมูลทั่วไปของการรัน เช่นใช้วิธีอะไร ใช้ไฟล์อะไร

คุณใช้เทคนิค SimpleKMeans (K-Means Clustering)

ชื่อไฟล์ที่ใช้ คือ train.arff

มีข้อมูลทั้งหมด 15 แถว (15 คน)

มี 6 ตัวแปร: ID, Name, SEX, Age, Score, Class

พีเจอร์ทั้งหมด 6 ตัว ได้แก่ ID, Name, SEX, Age, Score, Class

ส่วนนี้คือผลลัพธ์จริง

K-Means ต้องวนซ้ำกี่รอบกว่าจะ “จัดกลุ่มลงตัว”

ค่าความคลาดเคลื่อนภายในกลุ่ม ยิ่ง น้อย แปลว่าแบ่งกลุ่มได้ดีขึ้น

จุดเริ่มต้น (ตัวแทนเก่า) ของแต่ละคลัสเตอร์

K-Means จะเลือก “ข้อมูลจริง” มาเป็น centroid เริ่มต้น

- Cluster 0 เริ่มต้นจากคนที่ ID = 4 → Dang
- Cluster 1 เริ่มต้นจาก Gaga
- Cluster 2 เริ่มต้นจาก Rika
- Cluster 3 เริ่มต้นจาก Sompong

Final cluster centroids: รูปลักษณะเฉลี่ยของแต่ละคลัสเตอร์

Attribute	Cluster#				
	Full Data (15.0)	0 (1.0)	1 (5.0)	2 (4.0)	3 (5.0)
ID	8	4	10.2	11	4.2
Name	John	Dang	David	Sunee	John
SEX	F	M	F	F	M
Age	16.8667	25	12	24	14.4
Score	70.4	82	61.6	83	66.8
Class	B	A	C	A	B

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

```
0      1 ( 7%)
1      5 ( 33%)
2      4 ( 27%)
3      5 ( 33%)
```

```
Full Data = 8
Cluster 0 = 4
Cluster 1 = 10.2
Cluster 2 = 11
Cluster 3 = 4.2
```

```
Full Data = John
Cluster 0 = Dang
Cluster 1 = David
Cluster 2 = Sunee
Cluster 3 = John
```

```
Full Data = F
Cluster 0 = M
Cluster 1 = F
Cluster 2 = F
Cluster 3 = M
```

```
Full Data = 16.8667
Cluster 0 = 25
Cluster 1 = 12
Cluster 2 = 24
Cluster 3 = 14.4
```

ค่าเฉลี่ยของ ID ในแต่ละคลัสเตอร์

ตัวอย่างชื่อจากแต่ละกลุ่ม
Weka เลือกชื่อที่ “มักจะถูกพบมากที่สุด” หรือ “ค่าที่เป็น mode”

- Cluster 0 = ส่วนใหญ่เป็นผู้ชาย
- Cluster 1 = ผู้หญิง
- Cluster 2 = ผู้หญิง
- Cluster 3 = ผู้ชาย

- Cluster 0 = อายุเฉลี่ย 25 (Dang คนเดียว)
- Cluster 1 = อายุเฉลี่ย 12 → เด็กวัยรุ่น
- Cluster 2 = อายุเฉลี่ย 24 → ผู้ใหญ่
- Cluster 3 = อายุเฉลี่ย 14.4 → วัยรุ่นผู้ชาย

```
Full Data = 70.4
Cluster 0 = 82
Cluster 1 = 61.6
Cluster 2 = 83
Cluster 3 = 66.8
```

```
0      1 ( 7%)
1      5 ( 33%)
2      4 ( 27%)
3      5 ( 33%)
```

- Cluster 0 = คะแนนสูง (82)
- Cluster 1 = คะแนนต่ำ (61)
- Cluster 2 = คะแนนสูงมาก (83)
- Cluster 3 = คะแนนกลาง (66)

- Cluster 0 มี 1 คน (7%)
- Cluster 1 มี 5 คน (33%)
- Cluster 2 มี 4 คน (27%)
- Cluster 3 มี 5 คน (33%)