

## PERTEMUAN 5

### TEKNIK *CLUSTERING* DALAM WEKA

#### TUJUAN PRAKTIKUM

Mahasiswa akan dapat menggunakan teknik *clustering* dengan *tools* Weka.

#### TEORI PENUNJANG

##### 1. Simple KMeans pada Weka

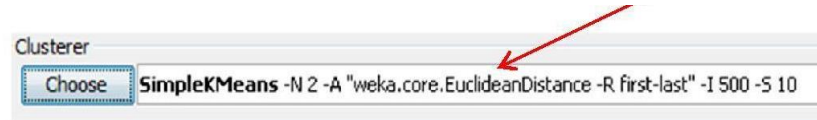
- Buka fail iris.arff
- Hilangkan atribut *Class* pada data set tersebut
- *Save as* menjadi databaru.arff
- Pilih tab *Cluster*
- Pada *Clusterer* klik *Choose*



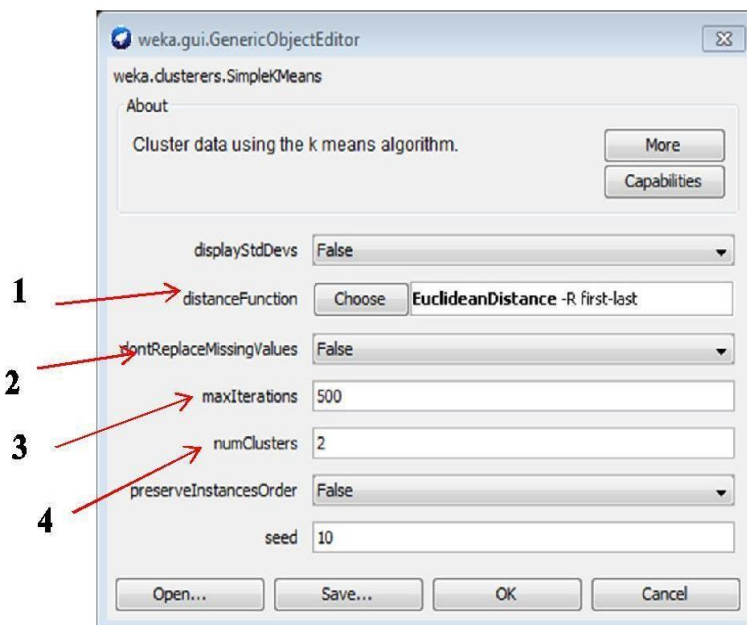
- Pilih algoritme “SimpleKMeans”



- Setelah itu, pilih *option* dari SimpleKMeans



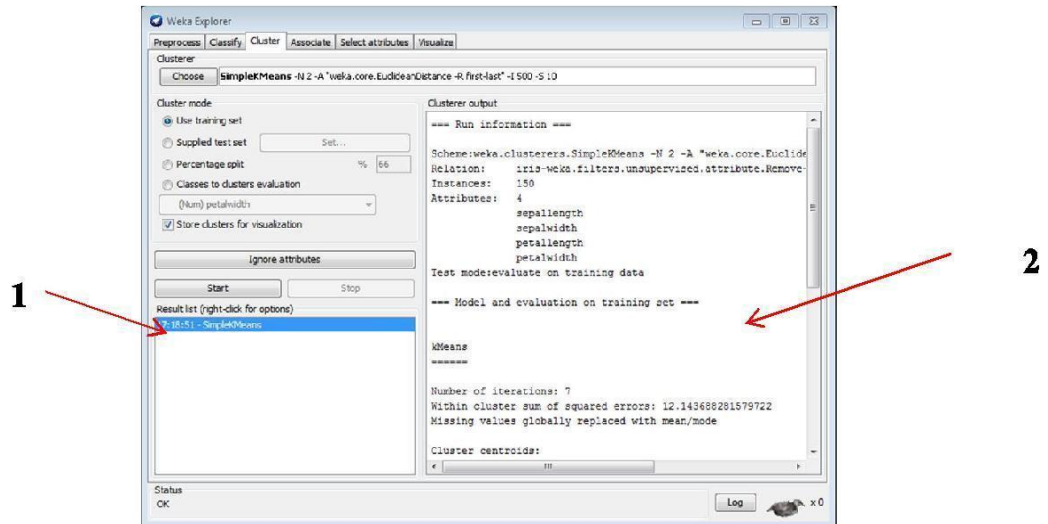
- Maka akan muncul tampilan seperti berikut:



Keterangan:

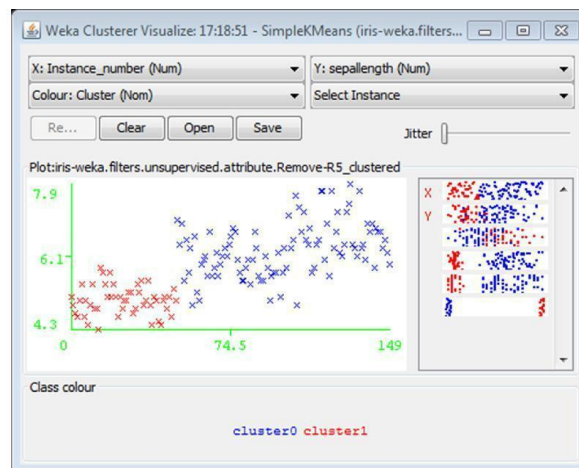
- 1 : untuk menentukan fungsi jarak yang akan dipakai
- 2 : untuk menentukan *missing value* akan diisi nilai atau tidak
- 3 : untuk menentukan maksimum jumlah iterasi
- 4 : untuk menentukan jumlah *cluster*

- Untuk awal, biarkan seperti nilai bawaan, klik OK >> klik Start >> maka akan muncul tampilan seperti berikut:



- 1 : daftar hasil *clustering* yang telah dilakukan, yang akan muncul *option* ketika diklik kanan
- 2 : *clusterer output*, menunjukkan hasil proses *clustering*

- Klik kanan pada nomor 1 >> *visualize cluster assignment*, maka akan muncul tampilan seperti berikut:



- Perhatikan nomor 2, scroll layar ke bawah sampai terlihat hasil pengklusteran seperti berikut:

Cluster output

Cluster centroids:

Attribute	Full Data (150)	Cluster#	
		0 (100)	1 (50)
sepal.length	5.8433	6.262	5.006
sepal.width	3.054	2.872	3.418
petal.length	3.7587	4.906	1.464
petal.width	1.1987	1.676	0.244

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0	100 ( 67%)
1	50 ( 33%)

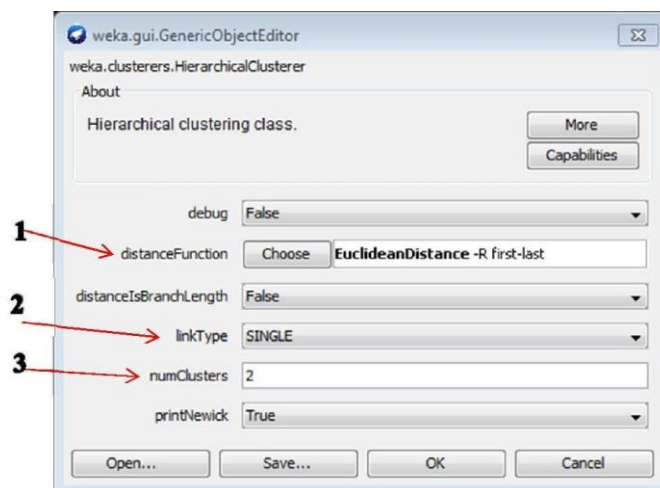
- Gambar di atas menunjukkan hasil pengklusteran dari data set iris beserta *centroid* dari masing-masing *cluster*.

## 2. Hierarchical Clustering pada Weka

- Buka kembali databaru.arff
- Pilih *tab Cluster*
- Pilih *Choose*, dan pada *clusters* pilih *Hierarchical Clusterer*

..... ☐ FilteredClusterer  
 ..... ☒ HierarchicalClusterer  
 ..... ☐ MakeDensityBasedClusterer

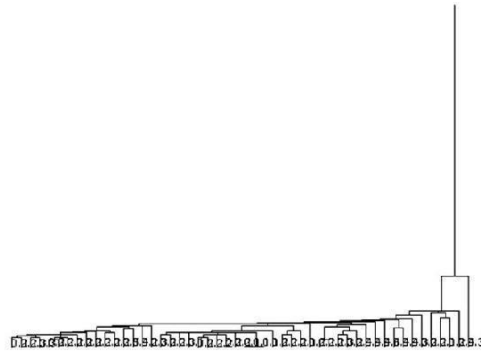
- Klik *option* dari *Hierarchical Clusterer*, maka akan muncul tampilan seperti berikut:



Keterangan:

- 1 : untuk menentukan fungsi jarak yang akan dipakai
- 2 : untuk menentukan tipe penentuan pemilihan jarak yang dipakai
- 3 : untuk menentukan jumlah *cluster*

- Biarkan saja nilai dari nilai bawaan, klik OK >> klik *Start*
- Hasil *clustering* dapat dilihat pada panel *clusterer output*
- Untuk visualisasi hasil dendrogram yang dibentuk dapat dilakukan dengan klik kanan ***Hierarchical Clusterer*** >> klik *visualize cluster assignment* atau klik *visualize tree*



#### LAPORAN PENDAHULUAN

1. Apakah setiap akan dilakukan proses *clustering* harus dilakukan praproses?
2. Jenis *clustering* apa saja yang Anda ketahui?
3. Algoritme *clustering* apa saja yang Anda ketahui?
4. Tahap apa saja yang harus dilakukan pada proses *clustering*?

#### MATERI PRAKTIKUM

1. Simple K-Means pada Weka
2. *Hierarchical Clustering* pada Weka

#### TUGAS PRAKTIKUM

#### DAFTAR PUSTAKA

- Bouckaert, R. R.. et al. 2013. *WEKA Manual for Version 3-6-9*. Edition of January 21, 2013.  
<http://jaist.dl.sourceforge.net/project/weka/documentation/3.6.x/WekaManual-3-6-9.pdf> . Accessed on 27 January 2013.
- Han, J. (2006) *Data Mining: Concepts and Technique*. [Internet]. [diunduh 2014 Mar8]. Tersedia pada: <http://www.cs.uiuc.edu/homes/hanj/bk2/slidesindex.htm>

#### SENARAI

1. *Clustering* : proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan (*similarity*) antar anggota satu kelas dan meminimumkan kesamaan antar kelas/*cluster*

2. *Flat / partitioning clustering* : mengelompokkan data dengan memilah-milah data yang dianalisis ke dalam *cluster-cluster* yang ada
3. *Hierarchical clustering* : mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan