

IF4071 Pembelajaran Mesin

Tugas Kecil 1

Eksplorasi scikit-learn untuk Clustering pada Jupyter Notebook

Rizki Ihza Parama / 13515104

I. Algoritma dan *library* yang digunakan

- A. K-means: menggunakan *library* sklearn
- B. Agglomerative Clustering: menggunakan *library* sklearn
- C. DBSCAN: menggunakan *library* sklearn
- D. Gaussian Mixture: menggunakan *library* sklearn
- E. K-medoids: menggunakan *library* pyclustering
- F. MST Clustering: menggunakan *library* mst_clustering
- G. BANG (Grid Clustering): menggunakan *library* pyclustering

II. Data yang akan dianalisa

A. Data iris

Data iris merupakan data ukuran bagian-bagian bunga iris dengan bentuk sebagai berikut

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

B. Data tenis

merupakan data keadaan cuaca lapangan tenis dengan bentuk sebagai berikut

	outlook	temp	humidity	windy
0	2	2	1	0
1	2	2	1	1
2	1	2	1	0
3	0	1	1	0
4	0	0	0	0

III. Parameter yang digunakan dalam tiap algoritma

A. K-means

Parameter yang digunakan untuk tiap data adalah:

- Data iris:
 - n_clusters = 3
 - random_state = 0
- Data tenis:
 - n_clusters = 2
 - random_state = 0

B. Agglomerative Clustering

Parameter yang digunakan untuk tiap data adalah:

- Data iris:
 - n_clusters = 3
 - memory = None
 - affinity = "euclidean"
 - linkage = "average"
- Data tenis:
 - n_clusters = 3
 - memory = None
 - affinity = "manhattan"
 - linkage = "complete"

C. DBSCAN

- Data iris:
 - eps = 0.41
 - min_samples = 3
 - n_jobs = -1
 - metric = "euclidean"
- Data tenis:
 - eps = 1
 - min_samples = 2
 - n_jobs = -1
 - metric = "manhattan"

D. Gaussian Mixture

- Data iris:
 - max_iter = 200
 - random_state = 0
 - n_components = 3

- Data tenis:
 - max_iter = 200
 - random_state = 0
 - n_components = 2

E. K-medoids

- Data iris:
 - initial_index_medoids = [1, 90, 120]
 - tolerance = 0.02
- Data tenis:
 - initial_index_medoids = [1, 13]
 - tolerance = 0.02

F. MST Clustering

- Data iris:
 - cutoff = 3
- Data tenis:
 - cutoff = 2

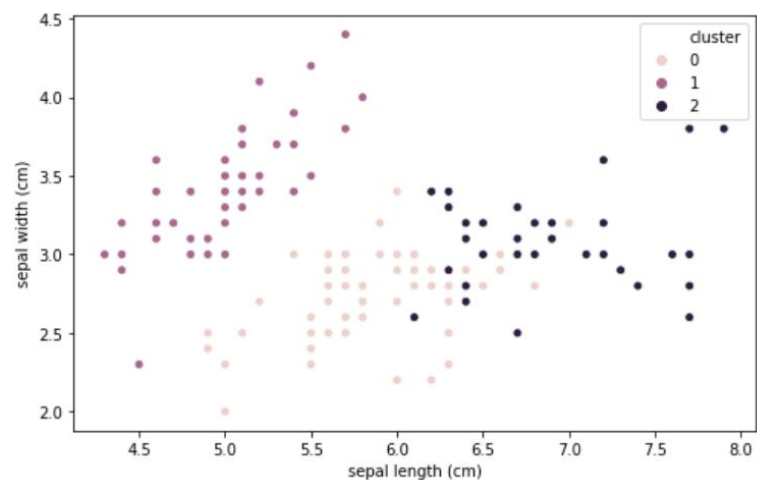
G. BANG (Grid Clustering)

- Data iris:
 - levels = 3
- Data tenis:
 - levels = 2

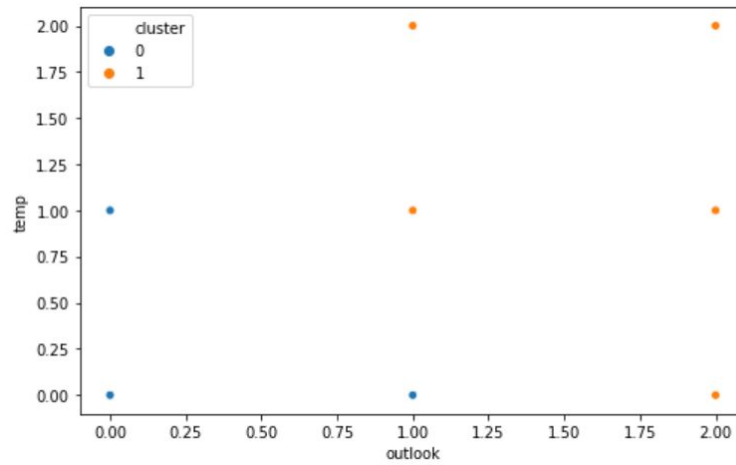
IV. Hasil *clustering*

A. K-means

- Data iris

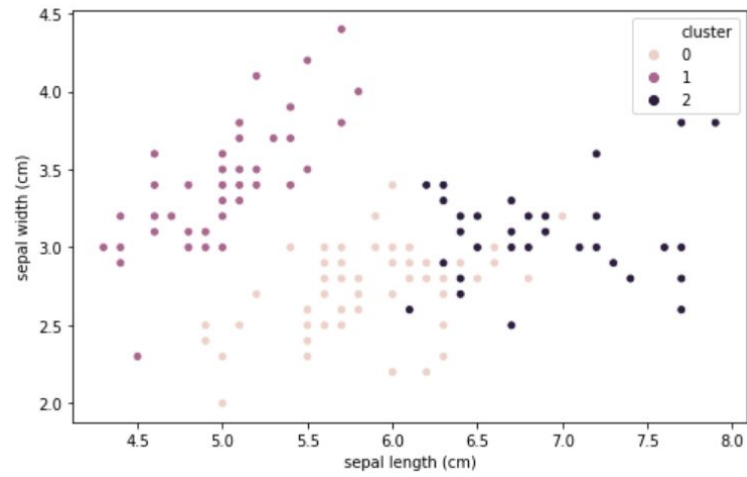


- Data tenis

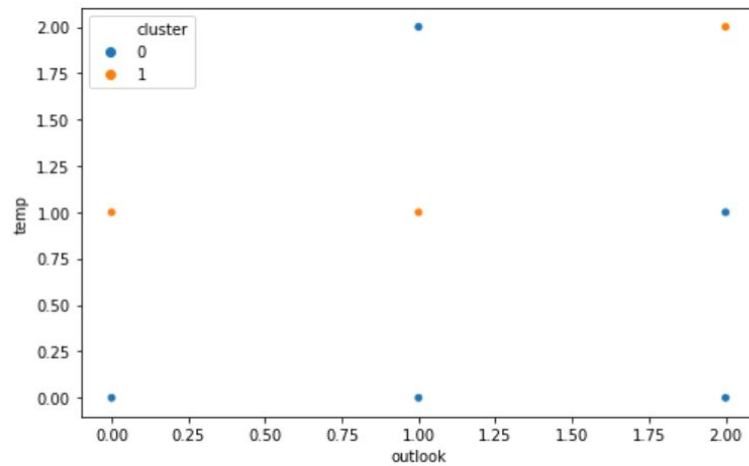


B. Agglomerative Clustering

- Data iris

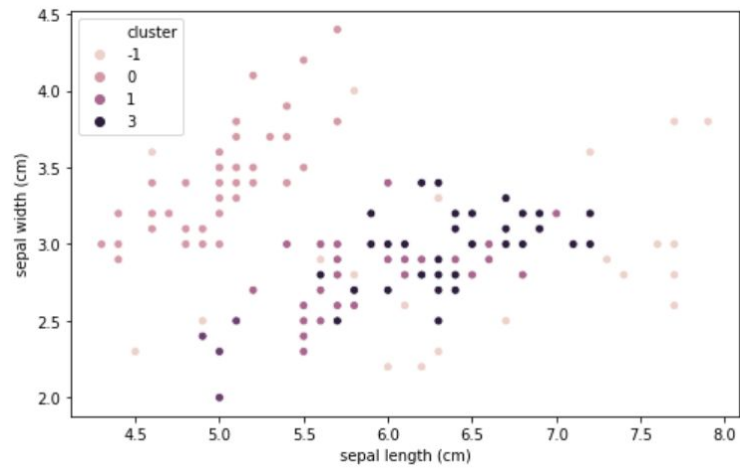


- Data tennis

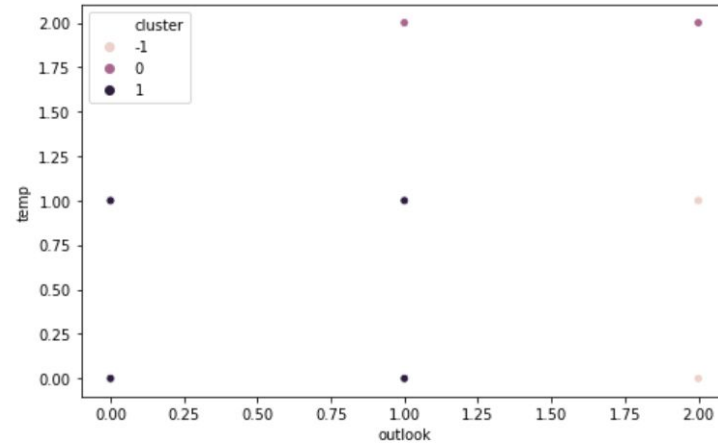


C. DBSCAN

- Data iris

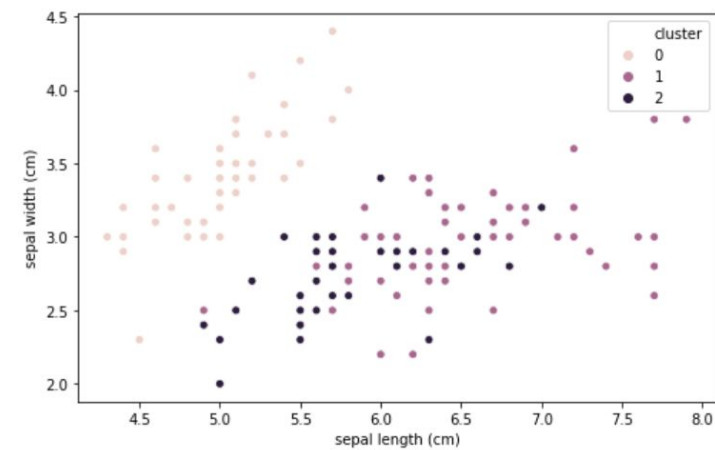


- Data tennis

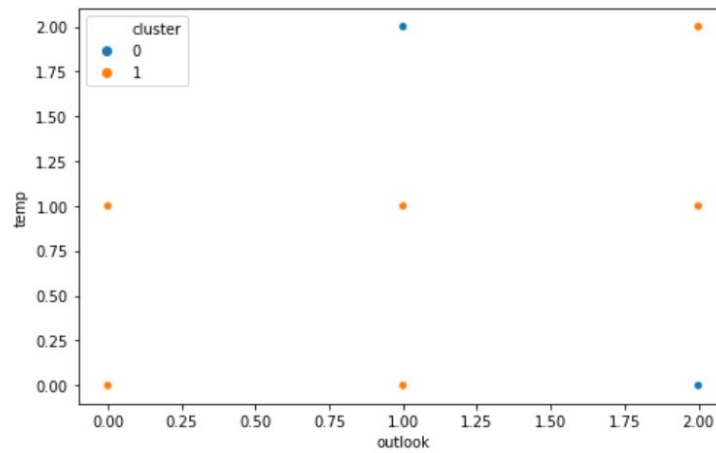


D. Gaussian Mixture

- Data iris

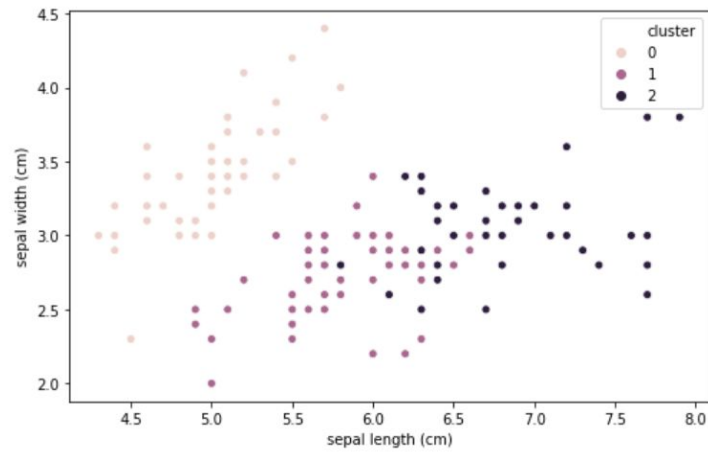


- Data tenis

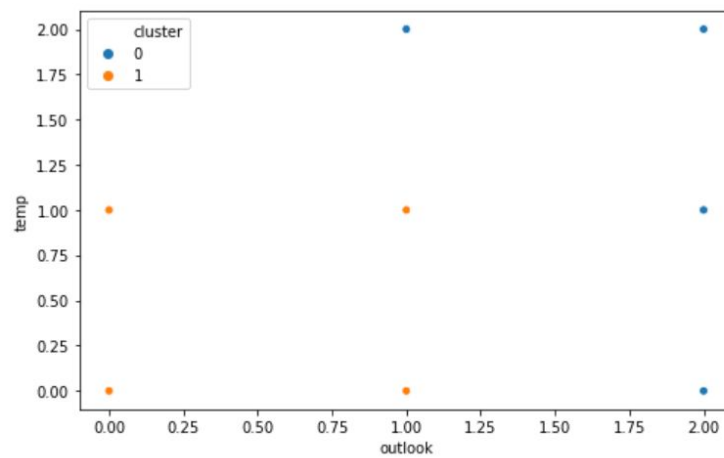


E. K-medoids

- Data iris

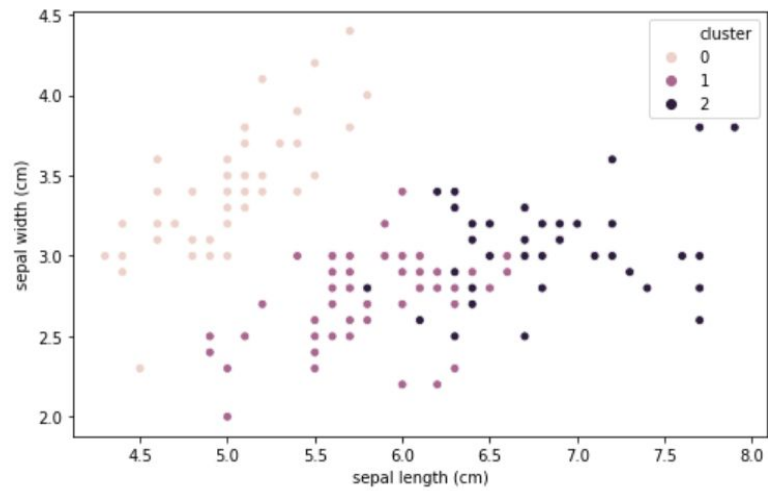


- Data tenis

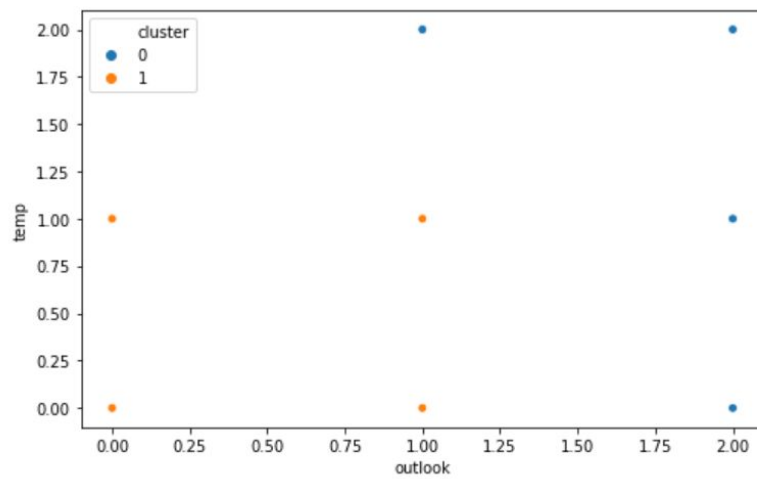


F. MST Clustering

- Data iris

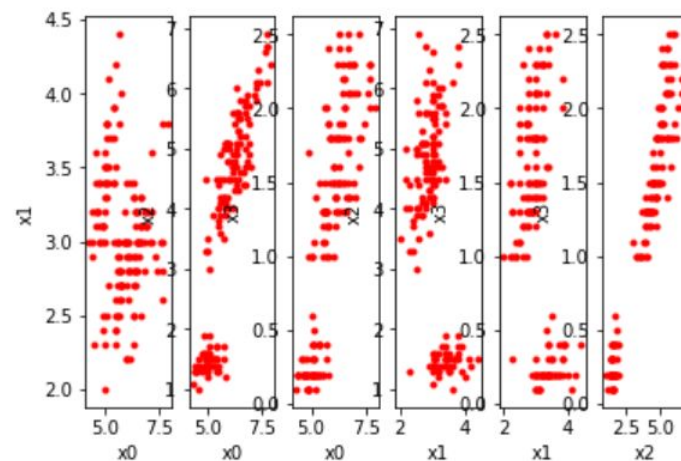


- Data tenis

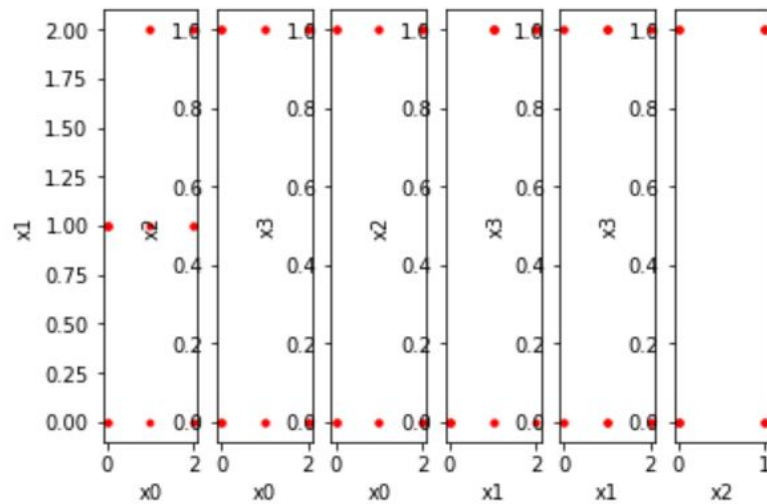


G. BANG (Grid Clustering)

- Data iris



- Data tenis



V. Menyimpan model hasil *clustering*

Model yang telah dilatih dengan data iris dan data tenis akan disimpan agar dapat digunakan untuk keperluan selanjutnya. Penyimpanan model dilakukan dengan bantuan *library* pickle pada bahasa pemrograman python. Direktori penyimpanan model adalah sebagai berikut

- Model data iris

agglo_iris.sav	9/26/2018 2:10 AM	SAV File	3 KB
bang_iris.sav	9/26/2018 2:10 AM	SAV File	7 KB
dbscan_iris.sav	9/26/2018 2:10 AM	SAV File	7 KB
em_iris.sav	9/26/2018 2:10 AM	SAV File	3 KB
kmeans_iris.sav	9/26/2018 2:10 AM	SAV File	2 KB
kmedoids_iris.sav	9/25/2018 12:43 PM	SAV File	0 KB
mst_iris.sav	9/26/2018 2:10 AM	SAV File	11 KB

- Model data tenis

Name	Date modified	Type	Size
agglo_tennis.sav	9/26/2018 2:10 AM	SAV File	1 KB
bang_tennis.sav	9/26/2018 2:10 AM	SAV File	3 KB
dbscan_tennis.sav	9/26/2018 2:10 AM	SAV File	1 KB
em_tennis.sav	9/26/2018 2:10 AM	SAV File	2 KB
kmeans_tennis.sav	9/26/2018 2:10 AM	SAV File	1 KB
mst_tennis.sav	9/26/2018 2:10 AM	SAV File	2 KB