

Modul 6
Clustering
Muhammad Fatih 'Ad-Li Pmm3

- A. Mata Kuliah
Data Mining
- B. Kompetensi
Clustering (DBSCAN)
- C. Teori dan Praktikum

Konsep Dasar	
1	<p>Jelaskan Konsep Dasar Clustering</p> <p>a. DBSCAN (Konsep dasar, DBSCAN adalah teknik pengelompokan berbasis kepadatan yang dimulai dari objek mana saja dan jikatetanggadi sekitarnya dalam radius tertentu (ϵ) memenuhi setidaknya jumlah minimum objek ($minPts$). Algoritma ini mencari objek inti untuk mengekstrak daerah padat dan akhirnya merupakan bagian dari kluster. Memeriksa setiap objek untuk kondisi inti, artinya mencari seluruh dataset untuk mendapatkan tetangga dalam ambang ϵ yang ditentukan.[1]</p> <p>b. Jelaskan perbaris Pseudo code</p> <hr/> <p>Algoritma: : DBSCAN: algoritma kluster berbasis Densitas. Input: D: Kumpulan data da yang berisi n objek; ϵ :parameter radius, $MinPts$: ambang Densitas lingkungan; Output: Satu set kluster berbasis Densitas ; Metode:</p> <ol style="list-style-type: none"> 1. tandai semua objek sebagai belum dikunjungi; 2. do secara acak memilih objek yang belum dikunjungi p beri tanda begitu dikunjungi; 3. if ϵ-neighborhood p memiliki setidaknya objek $MinPts$, buat kluster baru C, dan tambahkan p ke C tetapkan N menjadi himpunan objek di ϵ-neighborhood p 4. for each titik p' di N; 5. if p tidak dikunjungi tandai p' sebagai dikunjungi; 6. if di neighborhood dari p' memiliki setidaknya poin $MinPts$ tambahkan titik-titik itu ke N; 7. if p' belum menjadi anggota kluster mana pun, tambahkan p' ke C 8. End for output C 9. else tandai p sebagai kebisingan; 10. until tidak ada objek yang belum dikunjungi; <hr/> <p>c. Mekanisme kerja, Keterjangkauan dalam hal kepadatan menetapkan titik yang dapat dijangkau dari yang lain jika terletak dalam jarak tertentu (ϵ)s) darinya. Konektivitas, di sisi lain, melibatkan pendekatan rantai berbasis transitivitas untuk menentukan apakah titik-titik terletak di cluster tertentu. Sebagai contoh, titik p dan q dapat dihubungkan jika $p \rightarrow r \rightarrow s \rightarrow t \rightarrow q$, di mana $a \rightarrow b$ berarti b berada di lingkungan a.</p> <p>a. Ada tiga jenis titik setelah pengelompokan DBSCAN selesai:</p>

	<div data-bbox="606 190 1236 593" data-label="Diagram"> <p>The diagram shows three clusters of points. Each cluster is represented by a dashed circle. A 'Core point' is marked with a blue square at the center of each cluster. 'Border points' are marked with black dots on the perimeter of the clusters. 'Noise points' are marked with open circles and are not part of any cluster. A label 'MinPts = 3' indicates the minimum number of points required to form a cluster. An arrow labeled 'ε' points from a core point to its cluster boundary.</p> </div> <p>Tiga jenis titik DBSCAN</p> <p>Core - Ini adalah titik yang memiliki setidaknya m titik dalam jarak n dari dirinya.</p> <p>Border - Ini adalah titik yang memiliki setidaknya satu titik inti pada jarak n</p> <p>Noise - Ini adalah titik yang bukan inti maupun Perbatasan. Dan itu memiliki kurang dari m titik dalam jarak n dari dirinya sendiri.</p> <p>Langkah-langkah algoritmik untuk pengelompokan DBSCAN</p> <ol style="list-style-type: none"> 1) Algoritma melanjutkan dengan mengambil titik dalam kumpulan data secara sewenang - wenang (hingga semua titik telah dikunjungi). 2) Jika ada setidaknya titik 'titik kecil' dalam radius 'ε' ke titik tersebut, maka dianggap semua titik ini sebagai bagian dari kelompok yang sama. 3) Cluster tersebut kemudian diperluas dengan mengulangi penghitungan lingkungan secara rekursif untuk setiap titik tetangga [2] <p>d. Contoh Penerapan</p> <p>melakukan analisis daerah rentan penyakit menggunakan algoritma Density Based Spatial Clustering of Applications With Noise (DBSCAN). Tujuan dari penelitian ini adalah untuk menganalisis daerah rentan penyakit menggunakan algoritma Density Based Spatial Clustering of Applications With Noise (DBSCAN) sebagai alat bantu untuk memberikan kebijakan dan memonitoring penyuluhan.</p> <p>Sumber data utama yang digunakan yaitu laporan data bulan Oktober 2019 sampai dengan Desember 2019 dan hanya diambil berdasarkan diagnosis Asma, Aspa, Pnumonia, dan TBC dengan 736 record. Data tersebut nantinya akan diproses untuk dapat menentukan daerah rentan penyakit sesuai dengan sistem data mining yang digunakan</p>
2	<p>Buatlah Contoh Perhitungan manual menggunakan Excel mengenai DBSCAN</p> <p>a. Kerjakan d Excel dengan data</p>

Nama	Parameter	
	X1	X2
A	2	1
B	3	2
C	2	4
D	4	3
E	3	5
F	4	6
G	3	8
H	4	9
I	9	1
J	9	3
K	11	2
L	10	4
M	8	9
N	9	10
O	10	11
P	8	11

Eps	3
MinPts	4

Iterasi 1			
		Xp	Yp
Titik Pusat awal	B	3	2

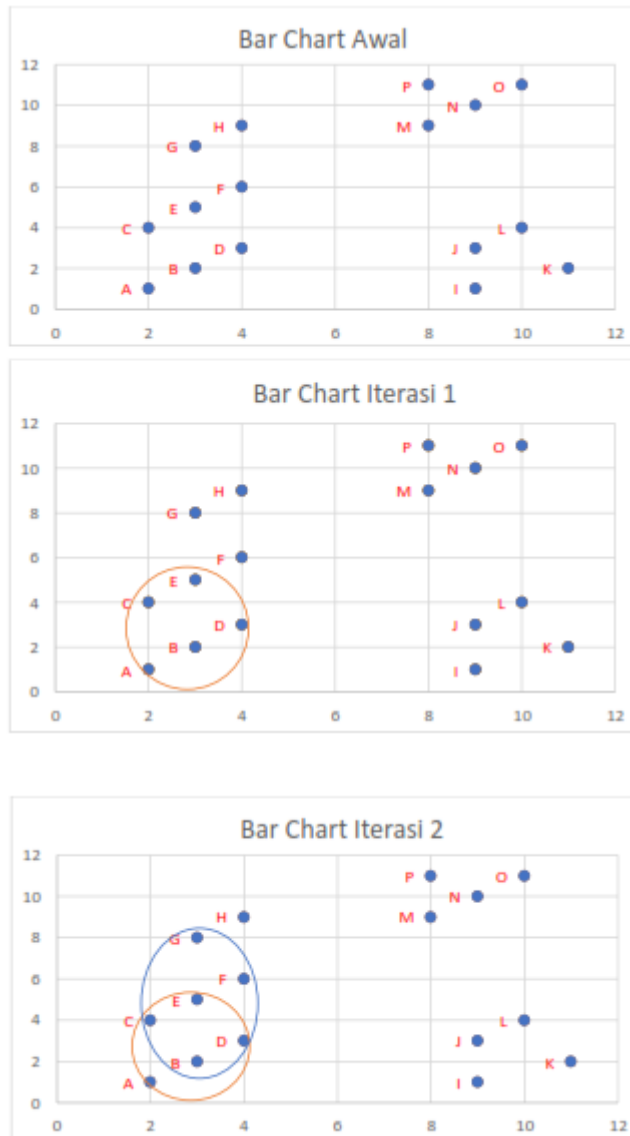
Jarak ke titik B		Dalam Radius		Nilai
A-B	1,414	A	ya	1,414
B-B	0,000	B	ya	0,000
C-B	2,236	C	ya	2,236
D-B	1,414	D	ya	1,414
E-B	3,000	E	ya	3,000

F-B	4,123	F	tidak	
G-B	6,000	G	tidak	
H-B	7,071	H	tidak	
I-B	6,083	I	tidak	
J-B	6,083	J	tidak	
K-B	8,000	K	tidak	
L-B	7,280	L	tidak	
M-B	8,602	M	tidak	
N-B	10,000	N	tidak	
O-B	11,402	O	tidak	
P-B	10,296	P	tidak	

Iterasi 2			
		Xp	Yp
Titik Pusat awal	E	3	5

Jarak ke titik E		Dalam Radius		Nilai
A-E	4,123	A	tidak	
B-E	3,000	B	ya	3,000
C-E	1,414	C	ya	1,414
D-E	2,236	D	ya	2,236
E-E	0,000	E	ya	0,000
F-E	1,414	F	ya	1,414
G-E	3,000	G	ya	3,000
H-E	4,123	H	tidak	
I-E	7,211	I	tidak	
J-E	6,325	J	tidak	
K-E	8,544	K	tidak	
L-E	7,071	L	tidak	
M-E	6,403	M	tidak	
N-E	7,810	N	tidak	
O-E	9,220	O	tidak	
P-E	7,810	P	tidak	

- Lakukan iterasi hingga tidak ada titik diluar cluster
 - Pada contoh data diatas dilakukan iterasi sebanyak 6 kali dan yang menjadi titik pusat yaitu titik : B – E – G – J – K – N
- b. Visualisasikan menggunakan Bar Chart



3. Buatlah Contoh Perhitungan mengenai K-mens pada python
- a. Script *jangan Prins Screen

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.cluster import DBSCAN
import pandas as pd
```

```
data = pd.read_csv('mall_cust.csv')
```

```
data
```

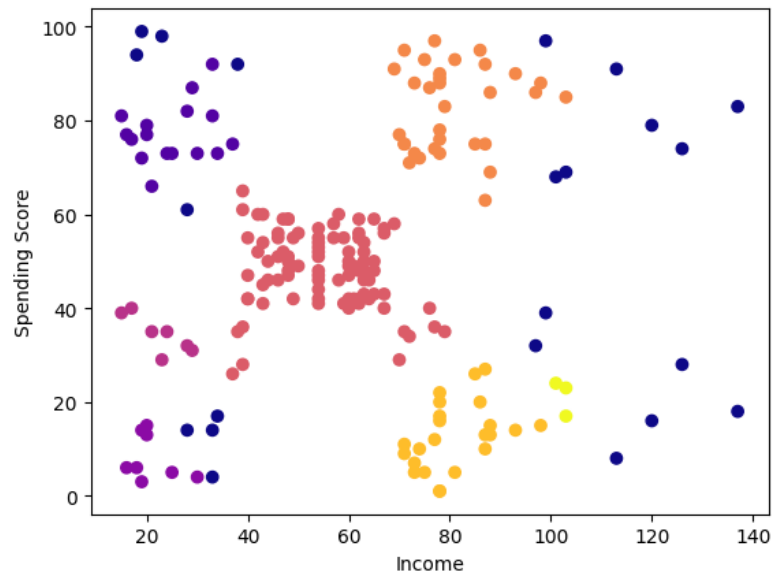
```
x = data.loc[:, ['Annual Income (k$)',
                 'Spending Score (1-100)']].values
```

```
dbscan = DBSCAN(eps = 8, min_samples = 4).fit(x)
labels = dbscan.labels_
```

```
plt.scatter(x[:, 0], x[:, 1], c = labels, cmap= "plasma")
plt.xlabel("Income")
```

```
plt.ylabel("Spending Score")  
plt.show()
```

b. Visualisasi menggunakan Scatter Plot



D. Daftar Pustaka

- [1] N. Salman, "Density-Based Clustering Analysis," vol. 8, pp. 1–8, 2023.
- [2] T. A. Cinderatama, R. Z. Alhamri, and Y. Yunhasnawa, "Implementasi Metode K-Means, DbSCAN, Dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System," *INOVTEK Polbeng - Seri Inform.*, vol. 7, no. 1, p. 169, 2022, doi: 10.35314/isi.v7i1.2336.