

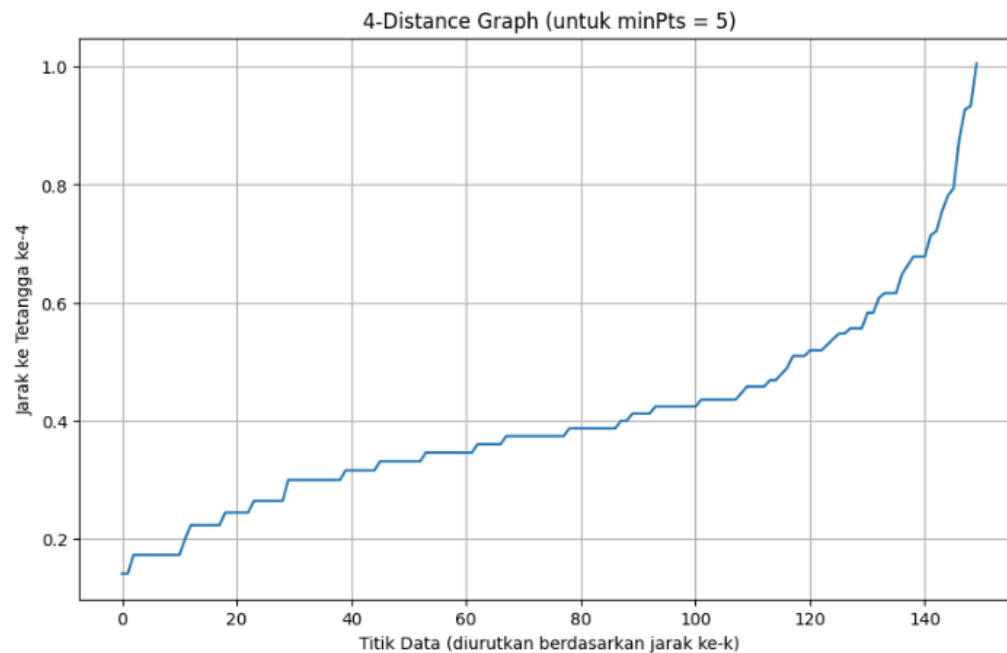
Report ML Assignment

Rayhan Firdaus Ardian - 23/519095/PA/22279 - KOMA

Github : <https://github.com/HappyRehund/DBSCAN-and-KMEANS>

DBSCAN

Parameter Tuning (Pemilihan epsilon dan minPts)



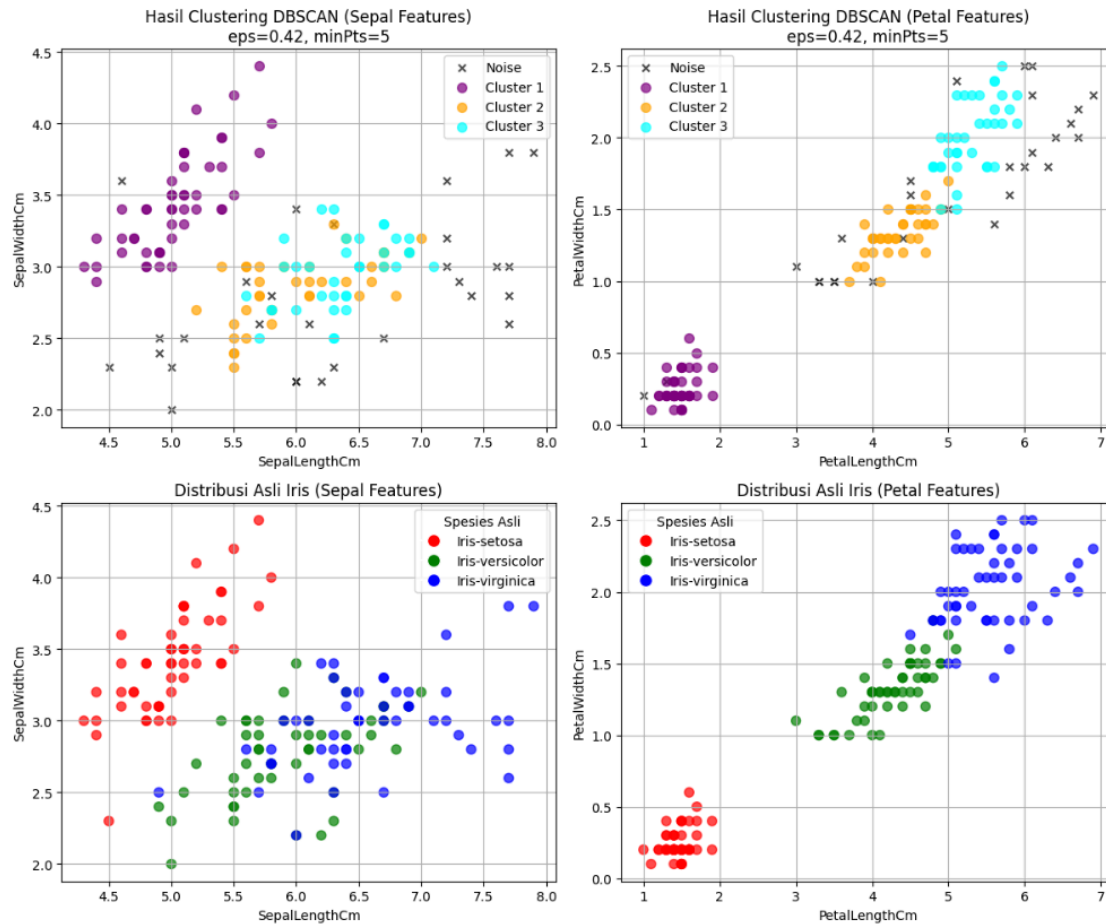
K-Distance Graph:

- Dengan menggunakan minPts = 5 (sehingga $k = 4$), kita memplot jarak ke tetangga ke-4 untuk setiap titik dan mengurutkannya.

Alasan Memilih epsilon = 0.42:

- Dari grafik terlihat adanya “titik siku” di sekitar nilai 0.4 hingga 0.5. Titik ini menunjukkan transisi antara jarak dalam cluster yang relatif kecil dan jarak yang lebih besar (di luar kepadatan cluster).
- Jika epsilon terlalu kecil, banyak titik tidak akan memiliki cukup tetangga sehingga banyak yang dianggap noise.
- Jika eps terlalu besar, cluster bisa menyatu dan struktur asli data hilang.

Hasil Clustering (DBScan)

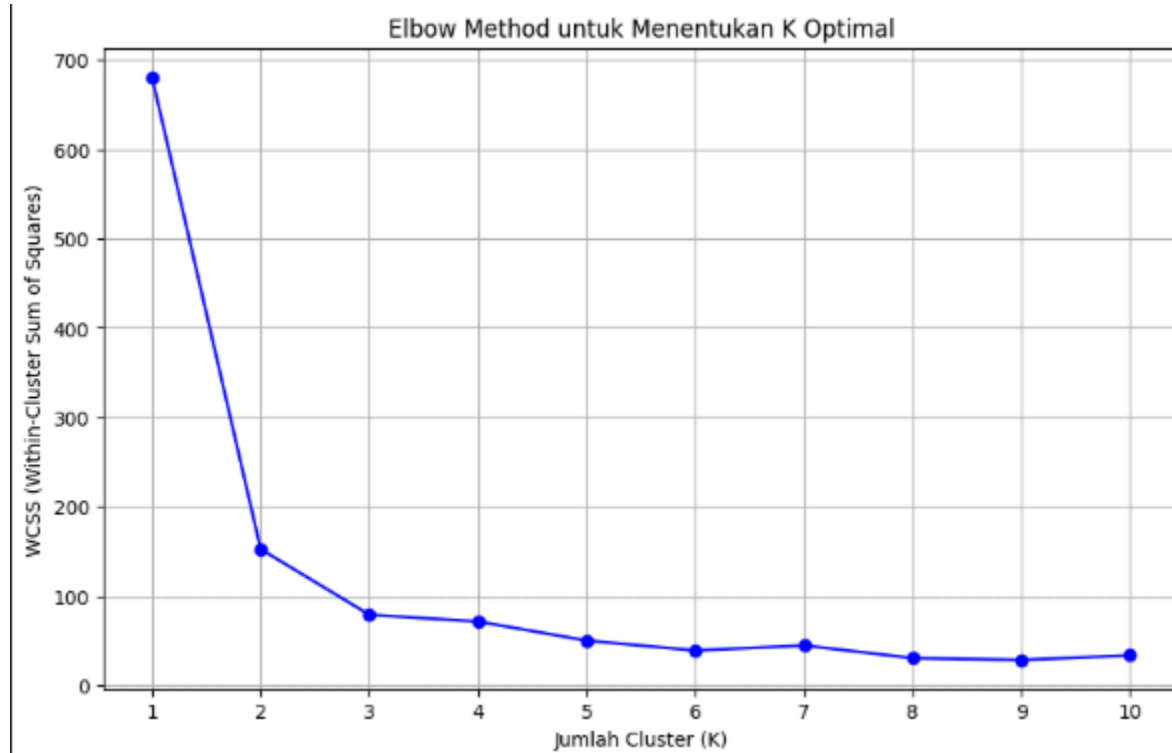


Analisis Cluster

- Setelah menjalankan DBSCAN dengan $\text{eps} = 0.42$ dan $\text{minPts} = 5$, diperoleh 29 cluster dengan beberapa titik yang ditandai sebagai noise (label -1).
- Cluster yang terbentuk cenderung menangkap kepadatan data dengan baik, terutama untuk kelompok data yang memiliki jarak antar titik relatif kecil.
- Perbandingan dengan Label Asli (Spesies Iris):
 - Iris-setosa:
 Biasanya, iris-setosa memiliki karakteristik yang jauh berbeda (dengan sepal dan petal yang lebih kecil) sehingga cenderung terbentuk sebagai cluster terpisah dengan tinggi kepadatan. DBSCAN kemungkinan besar menangkap seluruh iris-setosa sebagai satu cluster inti tanpa banyak titik noise.
 - Iris-versicolor dan Iris-virginica:
 Karena distribusi fitur mereka saling tumpang tindih, DBSCAN menghasilkan satu cluster yang menggabungkan kedua spesies atau membagi menjadi dua cluster dengan batas yang agak samar.

K-MEANS

Elbow Method

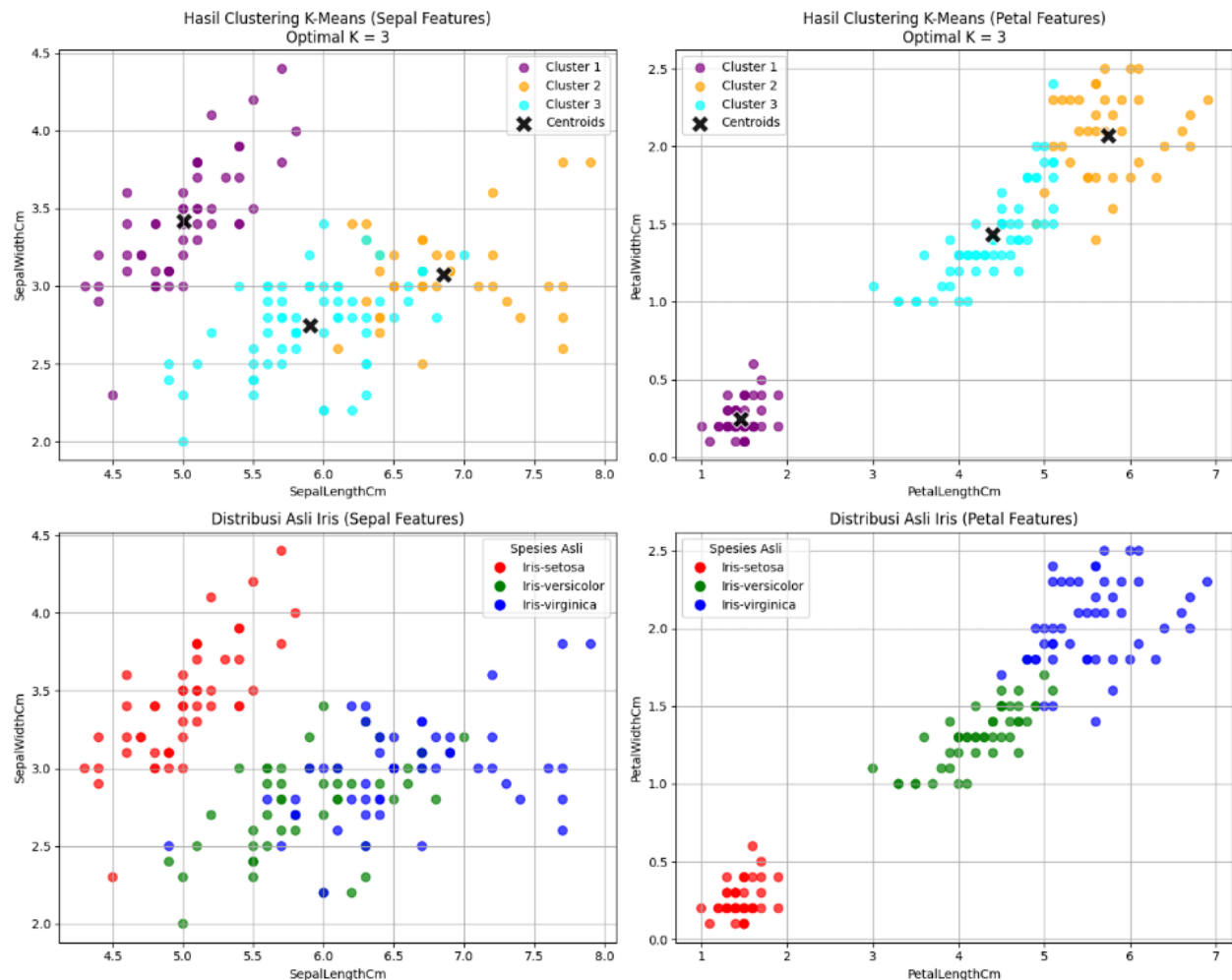


Terlihat bahwa:

- Saat K meningkat dari 1 ke 2, WCSS turun drastis ($\approx 680 \rightarrow 150$).
- Penurunan berikutnya dari K=2 ke K=3 masih signifikan ($\approx 150 \rightarrow 80$).
- Setelah K=3, penurunan WCSS menjadi jauh lebih lambat (K=4...10 bergerak di kisaran 80 \rightarrow 25 dengan kemiringan kecil).
- Titik di mana penurunan WCSS "melandai" = elbow point. Di sini elbow point berada pada K=3

Hasil Clustering (K-Means)

Perbandingan Hasil K-Means (K=3) dengan Data Asli



Analisis Cluster

- Posisi Centroid
 - Centroid cluster 1 (ungu) berada di area Sepal kecil/Petal kecil \Rightarrow mencerminkan Iris-setosa.
 - Centroid cluster 2 (oranye) di area Sepal dan Petal sedang/besar \Rightarrow mayoritas Iris-versicolor.
 - Centroid cluster 3 (cyan) di area Sepal sedang/besar dan Petal sedang \Rightarrow Iris-virginica.
- Visualisasi Sepal vs Petal
 - Pada plot Sepal, cluster 1 terpisah jelas
 - Cluster 2 dan 3 sedikit overlap di Sepal, tetapi pada Petal mereka terpisah lebih baik:
- Perbandingan dengan Label Asli
 - Cluster 1 hampir 100% hanya berisi setosa.
 - Cluster 2 dan 3 mencerminkan versicolor dan virginica, meski ada beberapa poin yang terklasifikasi silang—ini wajar karena kedua species memiliki distribusi fitur yang berdekatan.