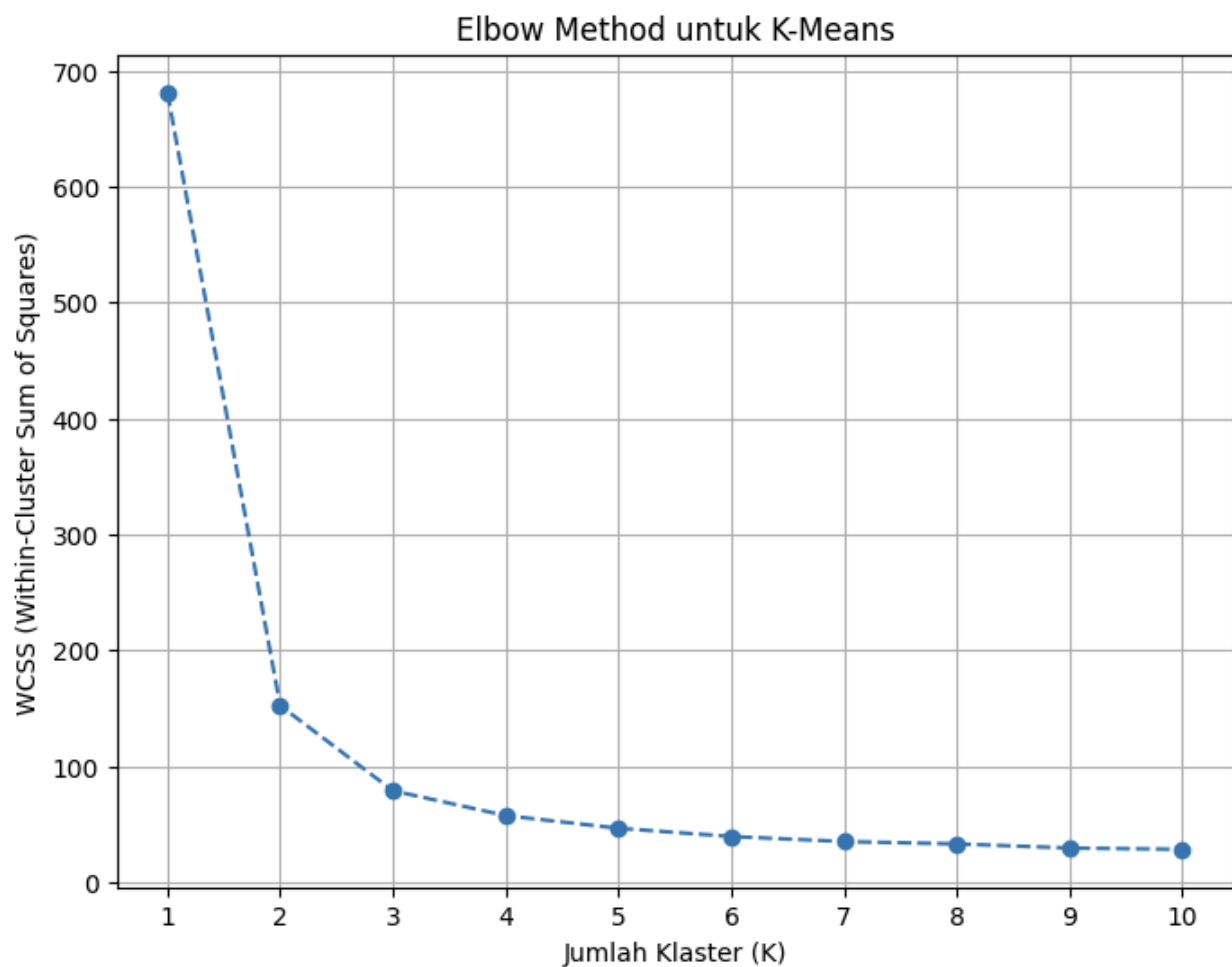


Laporan Assignment ML: K-means & DBScan

Daffa Indra Wibowo (23/518514/PA/22253) - KOM A

K-Means

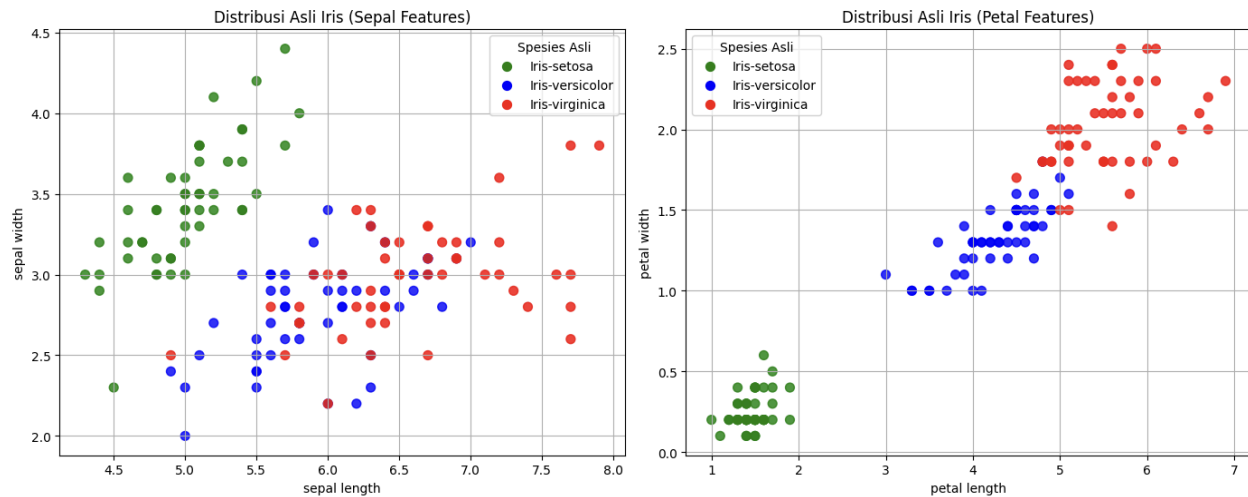
Elbow Method



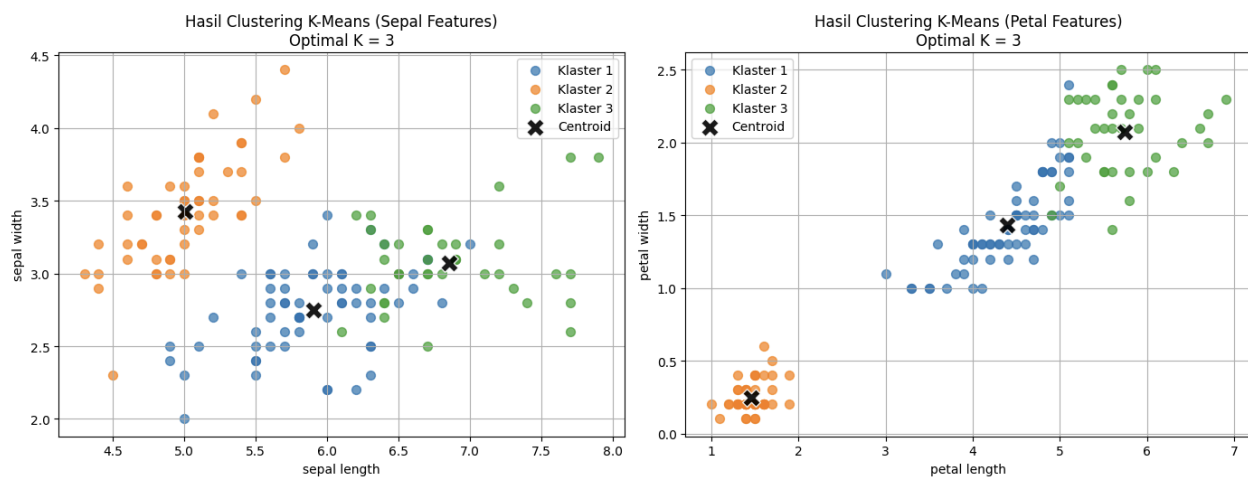
Di sini dapat kita perhatikan bahwa saat k berubah dari 1 ke 2, nilai WCSS turun drastis dari sekitar 690 menjadi sekitar 160. Begitu juga dari $k = 2$ ke $k = 3$, nilai WCSS masih turun cukup drastis yakni dari sekitar 160 menjadi sekitar 80. Akan tetapi setelah dari $k = 3$, perubahannya tidak terlalu signifikan/mulai melandai yang menandakan $k = 3$ ini adalah titik siku/elbow point.

Hasil Clustering K-means dengan Elbow Method

Visualisasi Data Asli Iris Sebelum Clustering



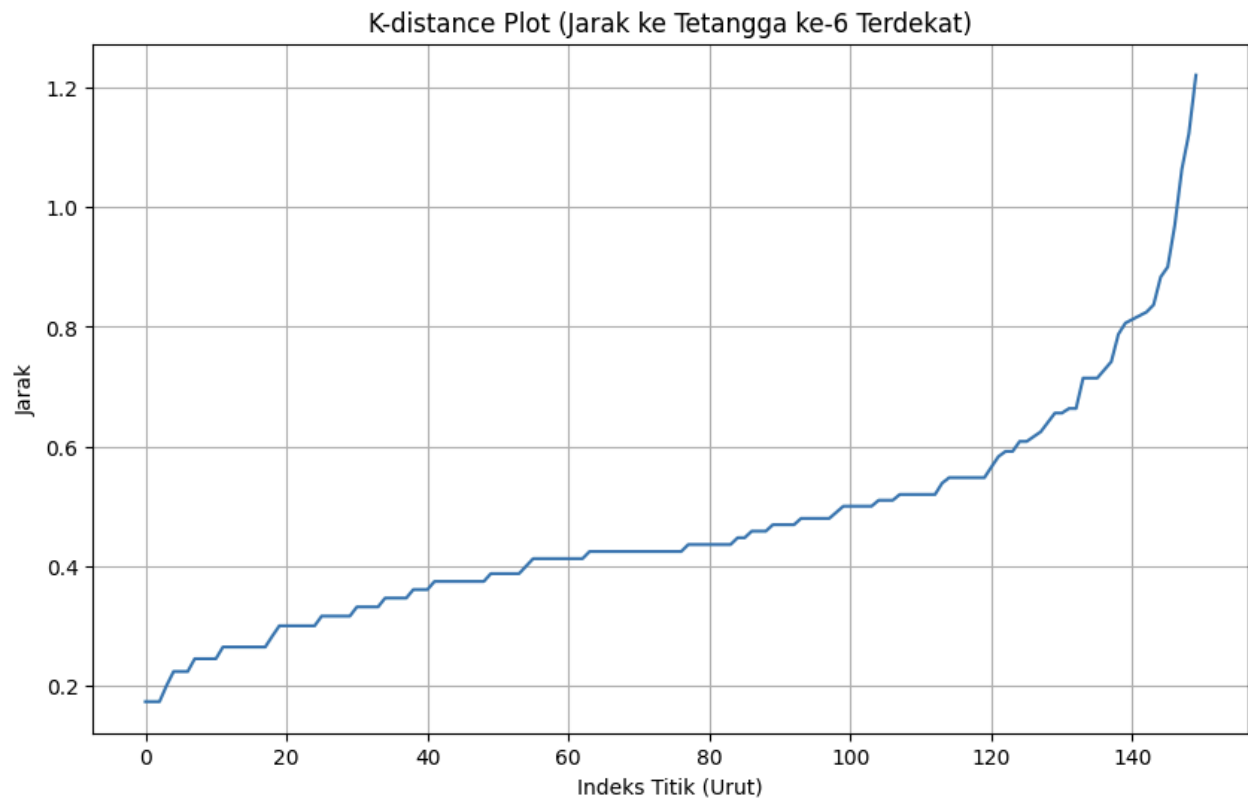
Hasil Clustering K-Means dengan K=3



Di sini dapat kita amati bahwa:

- Centroid untuk klaster 2 berada pada sepal kecil/sedang dan petal kecil. Centroid ini jelas mewakili iris-setosa
- Centroid untuk klaster 1 berada pada sepal sedang dan petal sedang/besar. Centroid ini mewakili iris-versicolor meskipun sedikit tumpang tindih dengan beberapa iris-virginica.
- Centroid untuk klaster 3 berada pada sepal sedang/besar dan petal sedang/besar. Centroid ini mewakili iris-virginica meskipun sedikit tumpang tindih dengan beberapa iris-versicolor.
- Untuk klaster 2 bisa cukup bersih/hanya terdiri dari iris-setosa karena memang cirinya cukup berbeda dengan 2 spesies lainnya terutama di bagian petal. Sedangkan untuk klaster 1 & 3 wajar terjadi tumpang tindih karena iris-virginica dengan iris-versicolor memiliki ciri yang hampir sama.

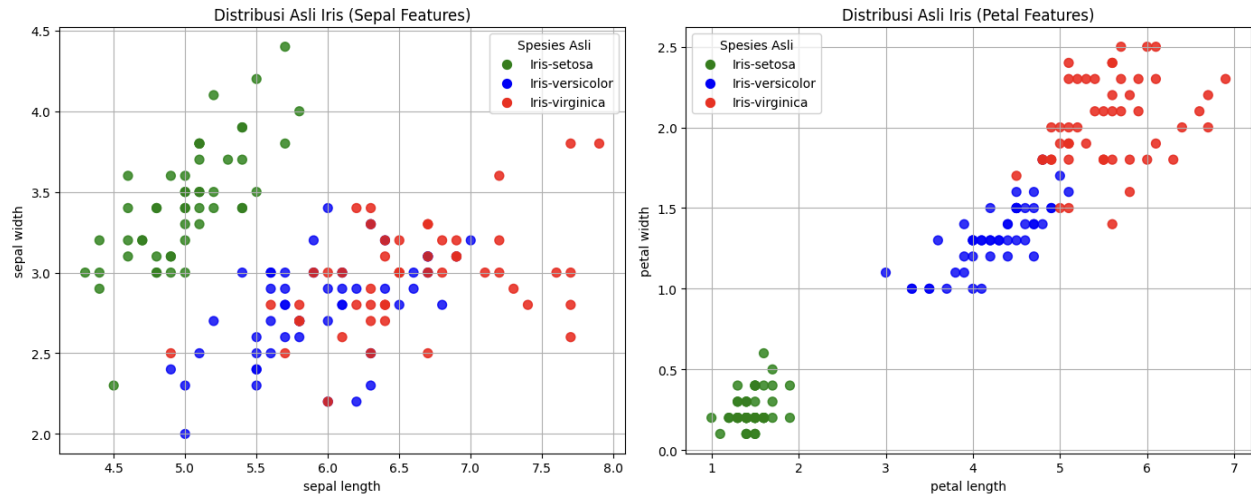
Tuning Parameter



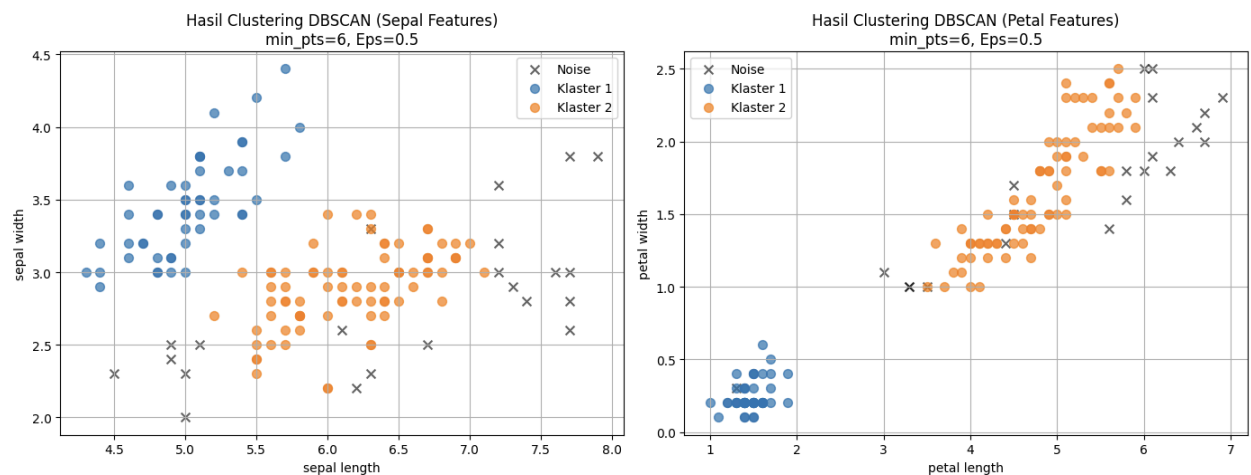
Tuning parameter di sini saya menggunakan $\text{minPts} = 6$ (yang berarti $k = 5$), di mana berarti plotting jarak ke tetangga ke-5 per-titik untuk kemudian mengurutkannya. Untuk epsilon-nya sendiri saya memilih 0.5 karena dari grafik, titik siku dapat terlihat di sekitar 0.5. Pemilihan ini juga berdasarkan jika epsilon-nya terlalu kecil, maka banyak titik yang tidak cukup meraih tetangga sehingga akan dianggap sebagai noise, tapi jika epsilon terlalu besar, maka clusternya bisa jadi terlalu besar dan malah menyatu.

Hasil Clustering DBScan

Visualisasi Data Asli Iris Sebelum Clustering



Hasil Clustering DBSCAN (min_pts=6, Eps=0.5)



Di sini hasil dari clustering mengelompokkan/membentuk klaster menjadi 2 dan titik noise sebanyak 22, di mana klaster 1 terdiri dari iris-setosa dan klaster 2 terdiri dari iris-versicolor dan iris-virginica. Di sini saya belum menemukan kombinasi yang cocok untuk menjadikannya lebih akurat dan membentuk 3 klaster yang dapat memisahkan antara iris-versicolor dan iris-virginica dengan ciri mereka yang hampir sama.