



Tugas *Task*

Mata Kuliah / Course Name	: Data Science
Sesi Pertemuan/ Session Number	: Sesi 9
Judul Materi / Title	:
Nama Dosen / Lecturer	: Alun Sujjada, S.Kom, M.T
Jenis Tugas / Type of task	: Case Based / Project Based
Nama Mahasiswa	: Edli Perdiansah
NIM	: 20230040033
Kelas	: TI 23C

Analisis Unsupervised Learning Menggunakan Agglomerative Clustering pada Dataset “Customer Segmentation Dataset”

1. Pendahuluan

Perkembangan teknologi informasi menyebabkan jumlah data yang dihasilkan semakin meningkat. Data tersebut sering kali tidak memiliki label, sehingga diperlukan metode khusus untuk mengelompokkannya. Salah satu metode yang dapat digunakan adalah unsupervised learning, yaitu metode pembelajaran mesin yang bertujuan untuk menemukan pola tersembunyi dalam data.

Agglomerative Clustering merupakan salah satu algoritma clustering yang bekerja dengan cara menggabungkan data yang memiliki kemiripan secara bertahap hingga terbentuk beberapa kelompok atau cluster. Metode ini banyak digunakan dalam berbagai bidang, salah satunya adalah segmentasi pelanggan. Oleh karena itu, pada tugas ini dilakukan analisis segmentasi pelanggan menggunakan Agglomerative Clustering untuk mengetahui pola dan karakteristik setiap kelompok pelanggan.

2. Analisis Dataset

Dataset yang di gunakan adalah Customer Segmentation, data ini terdiri dari beberapa atribut utama yaitu jenis kelamin, usia, pendapatan tahunan, dan skor pengeluaran pelanggan. Hasil pengecekan missing value menunjukkan bahwa seluruh data lengkap sehingga tidak diperlukan proses imputasi atau penghapusan data.

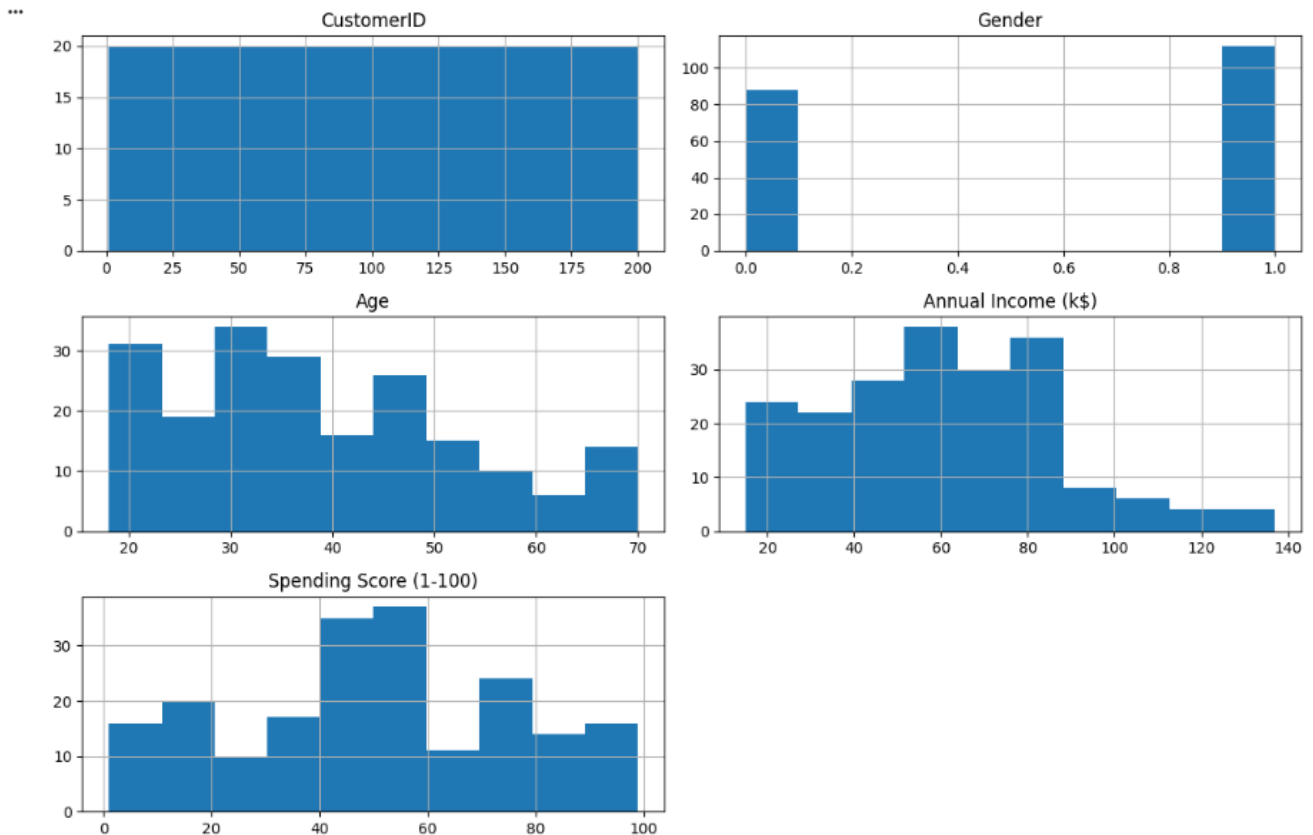
Sebelum proses clustering dilakukan, atribut kategorikal yaitu *Gender* diubah menjadi bentuk numerik agar dapat diproses oleh algoritma Agglomerative Clustering. Selanjutnya, seluruh atribut dilakukan standarisasi menggunakan metode *StandardScaler*. Langkah ini bertujuan untuk menyamakan skala antar fitur sehingga tidak ada atribut yang memiliki pengaruh lebih dominan dalam proses pembentukan cluster.

3. Analisis Deskriptif dan Visualisasi

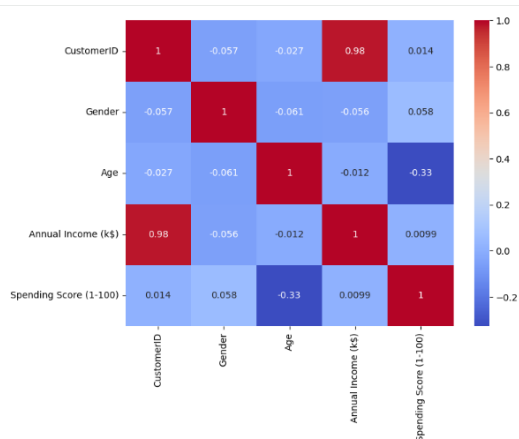
Hasil analisis deskriptif menunjukkan bahwa setiap atribut memiliki rentang nilai yang berbeda-beda. Atribut *Annual Income* dan *Spending Score* memiliki variasi yang cukup besar dibandingkan atribut lainnya. Hal ini mengindikasikan bahwa kedua atribut tersebut berpotensi menjadi faktor utama dalam membedakan karakteristik pelanggan.

Sebelum proses clustering dilakukan, atribut kategorikal yaitu *Gender* diubah menjadi bentuk numerik agar dapat diproses oleh algoritma Agglomerative Clustering. Selanjutnya, seluruh

atribut dilakukan standarisasi menggunakan metode **StandardScaler**. Langkah ini bertujuan untuk menyamakan skala antar fitur sehingga tidak ada atribut yang memiliki pengaruh lebih dominan dalam proses pembentukan cluster



Berdasarkan hasil visualisasi histogram, dapat diketahui bahwa setiap variabel memiliki distribusi nilai yang berbeda. Variabel *Age* tersebar pada rentang usia muda hingga dewasa, sedangkan *Annual Income* memiliki variasi yang cukup lebar dari pendapatan rendah hingga tinggi. Sementara itu, *Spending Score* menunjukkan perbedaan perilaku belanja pelanggan, mulai dari yang rendah hingga sangat tinggi. Hasil ini menunjukkan bahwa dataset memiliki keragaman karakteristik yang cukup baik untuk dilakukan proses clustering.

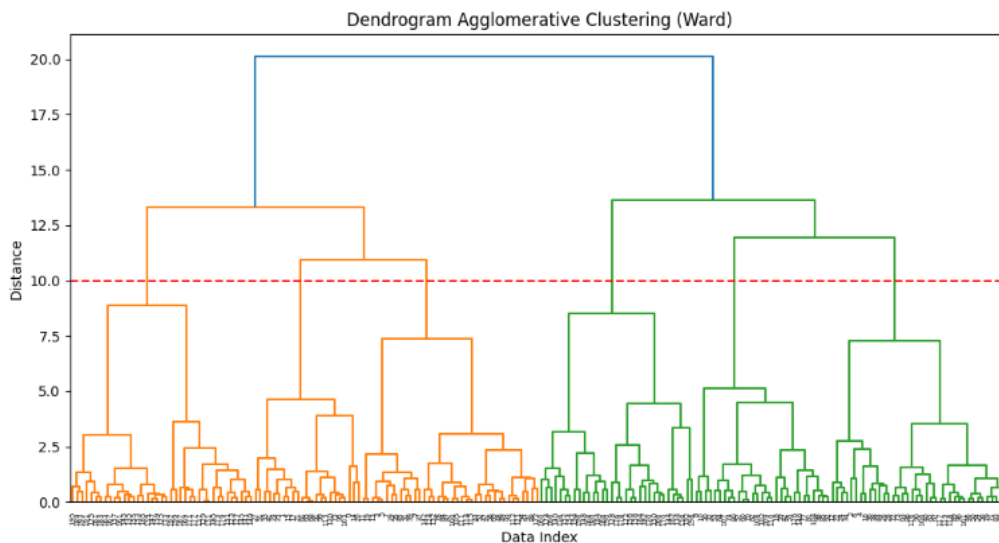


Hasil heatmap korelasi menunjukkan bahwa sebagian besar variabel tidak memiliki korelasi yang kuat satu sama lain. Korelasi yang cukup terlihat hanya pada *CustomerID* dan *Annual Income*, yang tidak mempengaruhi proses clustering karena *CustomerID* tidak digunakan sebagai fitur utama. Selain itu, korelasi antara *Age* dan *Spending Score* cenderung negatif, yang mengindikasikan bahwa pelanggan yang lebih muda cenderung memiliki skor pengeluaran yang lebih tinggi. Secara umum, seluruh fitur dianggap relevan untuk proses clustering.

4. Agglomerative Clustering

Pada tahap ini dilakukan pengelompokan data menggunakan metode Agglomerative Clustering dengan empat variasi linkage, yaitu single, complete, average, dan ward. Jumlah cluster ditentukan sebanyak tiga cluster berdasarkan hasil analisis dendrogram.

Dendrogram menunjukkan adanya jarak penggabungan yang cukup besar sebelum terbentuk cluster berikutnya. Hal ini menandakan bahwa pemotongan dendrogram pada jumlah tiga cluster dapat menghasilkan pemisahan kelompok data yang cukup jelas dan optimal.

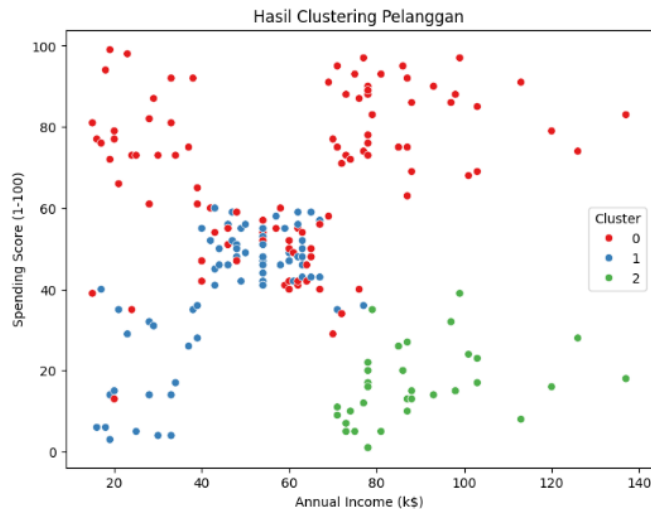


Dendrogram yang dihasilkan menggunakan metode Ward menunjukkan struktur pengelompokan data yang cukup jelas. Terlihat adanya lonjakan jarak penggabungan pada jarak tertentu, sehingga pemotongan dendrogram dilakukan pada jarak tersebut untuk menghasilkan tiga cluster. Pemilihan jumlah cluster ini dianggap optimal karena mampu memisahkan data pelanggan ke dalam kelompok yang berbeda dengan jarak antar cluster yang cukup signifikan.

5. Hasil Cluster

Hasil clustering menggunakan Ward linkage menghasilkan tiga kelompok pelanggan dengan karakteristik yang berbeda, yaitu:

- Cluster 1
Kelompok ini berisi pelanggan dengan pendapatan dan skor pengeluaran yang relatif rendah. Pelanggan dalam cluster ini cenderung memiliki aktivitas belanja yang rendah.
- Cluster 2
Cluster ini terdiri dari pelanggan dengan pendapatan menengah hingga tinggi dan skor pengeluaran yang tinggi. Kelompok ini memiliki potensi besar sebagai target pelanggan utama atau premium.
- Cluster 3
Cluster ini berisi pelanggan dengan pendapatan tinggi namun skor pengeluaran yang sedang. Pola ini menunjukkan adanya peluang untuk meningkatkan loyalitas dan frekuensi pembelian melalui strategi pemasaran yang lebih tepat.



Berdasarkan visualisasi scatter plot antara *Annual Income* dan *Spending Score*, terlihat bahwa pelanggan terbagi ke dalam tiga cluster yang memiliki karakteristik berbeda. Setiap cluster menunjukkan pola pengeluaran dan tingkat pendapatan yang berbeda, sehingga hasil clustering dapat digunakan untuk mengidentifikasi kelompok pelanggan potensial, pelanggan reguler, dan pelanggan dengan aktivitas belanja rendah. Hal ini menunjukkan bahwa metode Agglomerative Clustering dengan Ward linkage mampu menghasilkan segmentasi pelanggan yang cukup jelas dan mudah diinterpretasikan

6. Insight

Berdasarkan hasil analisis clustering, dapat diperoleh beberapa insight penting, antara lain:

- Agglomerative Clustering mampu mengelompokkan pelanggan berdasarkan pola perilaku belanja dengan cukup baik.
- Atribut pendapatan dan skor pengeluaran merupakan faktor yang paling berpengaruh dalam pembentukan cluster.
- Hasil clustering dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan strategi pemasaran, seperti penentuan target promosi dan program loyalitas pelanggan.

7. Kesimpulan

Hasil pengolahan data dan analisis menunjukkan bahwa metode Agglomerative Clustering, khususnya dengan Ward linkage, memberikan hasil pengelompokan yang paling optimal pada dataset segmentasi pelanggan. Metode ini terbukti efektif dalam menemukan pola tersembunyi pada data tanpa label dan relevan untuk diterapkan pada kasus nyata.