

SEGMENTASI PELANGGAN TOKO MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Mulyati Eka Saputri

Program Studi Sains Data
Fakultas Sains Dan Teknologi
Universitas Koperasi Indonesia
Email : mulyatiekasaputri@gmail.com

ABSTRAK

Penelitian ini menerapkan algoritma K-Means Clustering untuk segmentasi pelanggan pada sebuah toko, dengan fokus pada tiga fitur utama: usia, pendapatan tahunan, dan skor pengeluaran. Menggunakan Elbow Method, jumlah cluster optimal ditentukan sebanyak empat cluster. Hasil clustering dalam dua dimensi (2D) mengidentifikasi empat segmen pelanggan yang berbeda berdasarkan usia dan skor pengeluaran, sedangkan analisis tiga dimensi (3D) yang menambahkan pendapatan tahunan sebagai fitur tambahan memberikan segmentasi yang lebih rinci dan komprehensif. Segmentasi ini memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang lebih tepat sasaran, seperti memberikan promosi eksklusif untuk pelanggan dengan pendapatan dan pengeluaran tinggi atau insentif untuk pelanggan dengan skor pengeluaran rendah. Standarisasi data menggunakan StandardScaler memastikan setiap fitur berkontribusi secara proporsional dalam analisis, menghasilkan segmentasi yang lebih akurat dan bermakna. Secara keseluruhan, penerapan K-Means Clustering dalam penelitian ini memberikan wawasan berharga yang dapat meningkatkan efektivitas strategi pemasaran, kepuasan pelanggan, dan potensi keuntungan bisnis.

Kata Kunci : *K-Means Clustering, segmentasi pelanggan, Elbow Method, standarisasi data, strategi pemasaran, pengeluaran pelanggan, pendapatan tahunan.*

1. PENDAHULUAN

Dalam skenario kompetitif zaman sekarang, perusahaan mulai mengevaluasi dan mengelola pengalaman pelanggan melalui teknik-teknik pemasaran relasional dalam rangka menjalin dan meningkatkan hubungan dengan pelanggan. Perusahaan atau pelaku bisnis perlu memahami data pelanggan dengan lebih baik dalam semua hal aspek seperti mendeteksi persamaan dan perbedaan di antara para pelanggan, memprediksi perilaku mereka, mengusulkan pilihan dan kesempatan yang lebih baik untuk pelanggan menjadi sangat penting untuk keterlibatan hubungan pelanggan dengan perusahaan. Segmentasi pelanggan dari data mereka menjadi penting dalam konteks hubungan perusahaan usaha dengan pelanggan.

Dalam melakukan segmentasi pelanggan, umumnya dapat menggunakan metode clustering dengan tujuan dapat memberikan informasi dengan jelas dan kredibel. Teknik data mining, khususnya teknik clustering, memungkinkan semua pelanggan untuk dibagi menjadi beberapa kelompok (cluster) berdasarkan beberapa kesamaan di antara mereka. Algoritma K-Means adalah algoritma klasik untuk menyelesaikan masalah clustering. K-Means memiliki metode iteratif yang sederhana, seperti membagi dataset tertentu menjadi sejumlah cluster tertentu, atau menggunakan metode Elbow untuk menentukan jumlah cluster yang optimal.

Berdasarkan latar belakang tersebut maka tujuan melakukan penelitian ini yaitu mengetahui hasil segmentasi pelanggan menggunakan clustering dengan algoritma K-Means.

2. TINJAUAN PUSTAKA

2.1 Segmentasi Pelanggan

Proses mengkategorikan pelanggan dengan karakteristik heterogen ke dalam kelompok yang jelas dan nyata, sedangkan karakteristik homogen dimasukkan ke kelompok yang berdasarkan atribut umum merupakan definisi dari segmentasi pelanggan. Segmentasi pelanggan dianggap sebagai metode yang efektif untuk mengembangkan strategi pemasaran yang berbeda berdasarkan karakteristik pelanggan. Segmentasi pelanggan yang efektif berkontribusi dalam meningkatkan tidak hanya kepuasan pelanggan, tetapi juga keuntungan yang diharapkan dari organisasi atau pelaku bisnis.

2.2 Standardization Scale

Atribut RFM memiliki satuan yang berbeda, sehingga dilakukan standarisasi data. Sesuai dengan tujuan standarisasi yaitu untuk menyesuaikan ukuran (magnitude) dan bobot relatif dari variabel input. Perhitungan standarisasi menggunakan persamaan skor-z.

$$x' = \frac{x - x_{mean}}{\sigma}$$

Dimana :

x' : hasil nilai standarisasi

x : nilai yang akan ditransformasi dalam atribut

x_{mean} : rata-rata nilai atribut yang akan ditransformasi

σ : standar deviasi atribut yang akan ditransformasi

2.3 K-Means Clustering

Algoritma K-Means adalah algoritma klasik untuk memecahkan masalah clustering. Pada K-Means terdapat metode iterasi sederhana untuk mempartisi dataset yang diberikan ke dalam sejumlah cluster yang ditentukan oleh pengguna. Prosedur algoritma K-Means sebagai berikut.

1. Tentukan banyaknya cluster.

Untuk menentukan jumlah cluster atau K dapat dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin dicetuskan menjadi penentuan berapa jumlah cluster yang akan digunakan.

2. Tentukan titik centroid k (pusat cluster) secara acak.
3. Hitung jarak setiap titik ke pusat cluster, jarak antara satu data dengan satu cluster lainnya akan menentukan data mana yang masuk ke cluster yang mana. Perhitungan jarak antara data dan pusat cluster menggunakan Euclidian Distance dengan rumus:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Dimana :

$D(i, j)$: jarak dari data ke-i ke pusat cluster j

X_{ki} : koordinat data

X_{kj} : koordinat *centroid*

4. Kelompokkan data berdasarkan kedekatan dengan centroid kemudian perbaharui nilai centroid baru dengan lokasi dari pusat cluster menggunakan persamaan:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} X_j$$

Dimana :

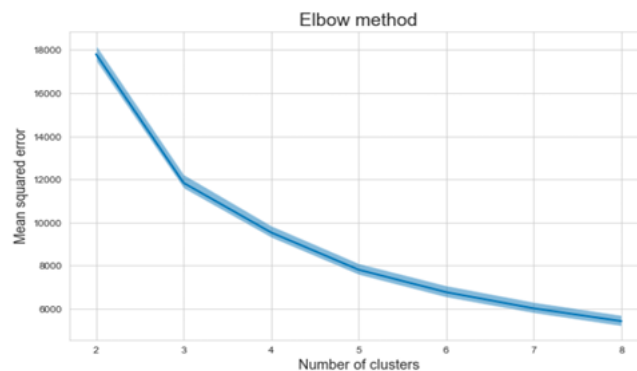
$\mu_j(t+1)$: *centroid* baru pada iterasi ke-(t+1)

N_{sj} : banyaknya data pada cluster s_j

5. Lakukan langkah 2 sampai 4 sampai anggota tiap cluster tidak ada yang berubah.

2.3 Elbow Method

Metode Elbow merupakan salah satu cara untuk menentukan nilai K atau jumlah cluster yang optimal. Metode elbow menggunakan sum of squared error (SSE) untuk memilih nilai K atau jumlah cluster yang ideal berdasarkan jarak antara titik data dan cluster yang ditetapkan. Nilai K atau jumlah cluster akan dipilih Ketika SSE mulai mendatar dan terlihat titik belok. Ketika divisualisasikan grafik ini akan terlihat seperti siku, seperti nama metodenya.



Gambar 2.3. Grafik Elbow Method

3. METODE PENELITIAN

Penelitian ini menggunakan Google Colab yang bertujuan untuk mengimplementasikan algoritma K-Means Clustering dalam mengelompokkan pelanggan. Data ini diperoleh dari situs Kaggle (<https://www.kaggle.com/datasets/datascientistanna/customers-dataset>). Data tersebut memiliki 2000 data dengan 8 variabel yaitu CustomerID (ID Pelanggan),

Gender (Jenis Kelamin), Age (Usia), Annual Income (Pendapatan Tahunan), Spending Score (Skor Pengeluaran), Profession (Profesi), Work Experience (Pengalaman Kerja), Family Size (Ukuran Keluarga). Adapun variabel yang digunakan adalah variabel Usia, Pendapatan Tahunan dan Skor Pengeluaran. Langkah-langkah metodologi yang dilakukan adalah sebagai berikut:

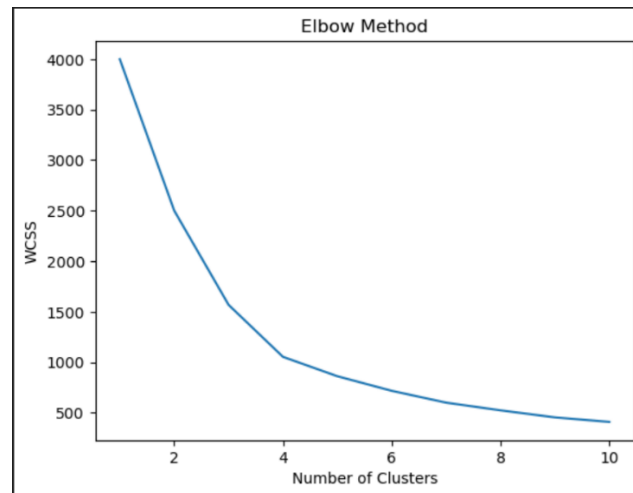
- 1) Data pelanggan dimuat menggunakan library Pandas.
- 2) Data dipecah menjadi dua dataframe: satu untuk analisis dua fitur (usia dan skor pengeluaran) dan satu lagi untuk analisis tiga fitur (usia, pendapatan tahunan, dan skor pengeluaran). Kedua dataframe ini kemudian distandarisasi menggunakan StandardScaler untuk memastikan setiap fitur memiliki skala yang sama.
- 3) Penentuan jumlah cluster optimal menggunakan Metode Elbow
- 4) Clustering Menggunakan K-Means
- 5) Hasil clustering divisualisasikan dalam bentuk 2D (usia dan skor pengeluaran) dan 3D (usia, pendapatan tahunan, dan skor pengeluaran)

4. HASIL PENELITIAN DAN PEMBAHASAN

Dalam penelitian ini, algoritma K-Means Clustering diterapkan pada data pelanggan untuk mengidentifikasi segmentasi pelanggan berdasarkan beberapa fitur utama: usia, pendapatan tahunan, dan skor pengeluaran. Berikut adalah hasil dan pembahasan dari proses clustering yang telah dilakukan.

4.1 Metode *Elbow*

Elbow Method digunakan untuk menentukan jumlah cluster yang optimal dengan melihat penurunan Within-Cluster Sum of Squares (WCSS). Grafik hasil Elbow Method ditampilkan pada Gambar 4.1.

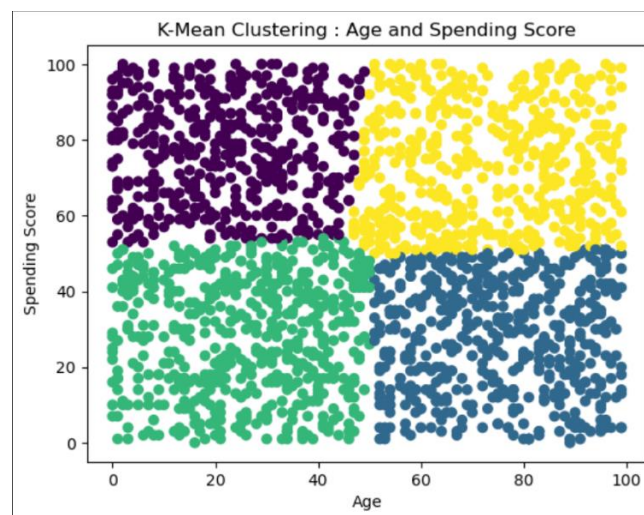


Gambar 4.1 Grafik hasil Elbow Method

Berdasarkan hasil grafik Elbow, jumlah cluster yang optimal untuk data pelanggan ini adalah 4. Setelah jumlah cluster 4, penurunan WCSS tidak signifikan, menunjukkan bahwa menambah lebih banyak cluster tidak memberikan keuntungan berarti dalam hal pengurangan variasi dalam cluster.

4.2 K-Means Clustering Berdasarkan Usia dan Skor Pengeluaran (2 Dimensi)

Cluster pertama dibuat menggunakan dua fitur: usia dan skor pengeluaran pelanggan. Hasil clustering ditampilkan pada Gambar 4.2.



Gambar 4.2 Hasil clustering Berdasarkan Usia dan Skor Pengeluaran

Sumbu X mewakili usia pelanggan (Age), sedangkan sumbu Y mewakili skor pengeluaran pelanggan (Spending Score). Warna titik menunjukkan cluster mana pelanggan tersebut termasuk. Dalam visualisasi ini, ada empat warna berbeda yang masing-masing mewakili satu cluster.

- Cluster 1 (Ungu) : Pelanggan dengan skor pengeluaran tinggi dan usia yang bervariasi, ini mungkin mencakup pelanggan yang sering berbelanja atau melakukan pembelian dalam jumlah besar.
- Cluster 2 (Kuning): Pelanggan dengan skor pengeluaran tinggi dan usia yang lebih tua, ini bisa mencakup pelanggan setia yang memiliki daya beli yang kuat.
- Cluster 3 (Hijau) : Pelanggan dengan skor pengeluaran rendah dan usia yang bervariasi, menunjukkan bahwa mereka mungkin berbelanja lebih jarang atau membeli barang dalam jumlah yang lebih kecil.
- Cluster 4 (Biru) : Pelanggan dengan skor pengeluaran rendah dan usia yang lebih tua, menunjukkan bahwa mereka mungkin berbelanja lebih sedikit baik dalam frekuensi maupun jumlah.

4.3 K-Means Clustering Berdasarkan Usia, Pendapatan Tahunan, dan Skor Pengeluaran (3 Dimensi)

Cluster kedua dibuat menggunakan tiga fitur: usia, pendapatan tahunan, dan skor pengeluaran. Hasil clustering dalam bentuk 3D ditampilkan pada Gambar 4.3.



Gambar 4.3 Hasil clustering Berdasarkan Usia, Pendapatan Tahunan, dan Skor Pengeluaran

Visualisasi 3D ini memberikan wawasan lebih dalam mengenai struktur data dan hubungan antar fitur. Sumbu X mewakili usia pelanggan (Age), sumbu Y mewakili pendapatan tahunan pelanggan (Annual Income), dan sumbu Z mewakili skor pengeluaran pelanggan (Spending Score).

- Cluster 1 (Ungu) : Pelanggan dengan pendapatan tahunan rendah dan skor pengeluaran rendah, ini mungkin mencakup pelanggan yang berbelanja lebih sedikit baik dalam frekuensi maupun jumlah.
- Cluster 2 (Kuning): Pelanggan dengan pendapatan tahunan tinggi dan skor pengeluaran tinggi, menunjukkan bahwa mereka mungkin adalah pelanggan setia yang memiliki daya beli yang kuat dan sering berbelanja dalam jumlah besar.
- Cluster 3 (Hijau) : Pelanggan dengan pendapatan tahunan menengah dan skor pengeluaran bervariasi, pelanggan dalam cluster ini mungkin memiliki pola belanja yang tidak konsisten.
- Cluster 4 (Biru) : Pelanggan dengan pendapatan tahunan rendah hingga menengah dan skor pengeluaran bervariasi, menunjukkan beragam perilaku belanja.

5. KESIMPULAN

Penelitian ini menggunakan algoritma K-Means Clustering untuk segmentasi pelanggan toko berdasarkan usia, pendapatan tahunan, dan skor pengeluaran. Elbow Method menunjukkan bahwa empat cluster adalah jumlah optimal. Hasil clustering dua dimensi menunjukkan empat segmen pelanggan berbeda berdasarkan usia dan skor pengeluaran, sedangkan analisis tiga dimensi menambahkan pendapatan tahunan sebagai fitur tambahan, menghasilkan segmentasi yang lebih rinci. Segmentasi ini memungkinkan perusahaan mengembangkan strategi pemasaran yang lebih tepat sasaran, seperti promosi eksklusif untuk pelanggan dengan pendapatan dan pengeluaran tinggi atau insentif untuk pelanggan dengan skor pengeluaran rendah. Standarisasi data memastikan setiap fitur berkontribusi secara proporsional dalam analisis, menghasilkan segmentasi yang lebih akurat dan bermakna. Secara keseluruhan, penerapan K-Means Clustering memberikan wawasan berharga yang dapat meningkatkan efektivitas strategi pemasaran dan kepuasan pelanggan.

6. DAFTAR PUSTAKA

- Granata, G. 2020. The Digital Evolution of Consumer Purchasing Methods and the Impact on Retail. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 10(4).
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. 2008. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.
- Hong, T., & Kim, E. 2012. Segmenting customers in online stores based on factors that affect the customer's intention to purchase. *Expert Systems with Applications*, 39(2), 2127–2131.
- Chen, D., Sain, S. L., & Guo, K. 2012. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing and Customer Strategy Management*, 19(3), 197–208.
- Miuiigan, G. W., Cooper, M. C., Milligan, G. W., Milligan, G. W., & Cooper, M. C. 1988. A Study of Standardization of Variables in Cluster Analysis. In *Journal of Classification* (Vol. 5).
- Fithri, F. A., & Wardhana, S. 2019. Cluster Analysis of Sales Transaction Data Using K-Means Clustering at Toko Usaha Mandiri
- Cui, M. 2020. Introduction to the K-Means Clustering Algorithm Based on the Elbow Method.