



Penggunaan Metode K-Means Clustering Untuk Segmentasi Pasar Konsumen

Rendi saputra¹, Anisa Karenina², Zainal Arifin parinduri³, Andrea vallen dellyco⁴, Abimanyu fahrudin qois⁵

^{1,2,3,4,5}Pendidikan Teknik informatika, ilmu komputer, universitas pamulang, Tangerang selatan, banten.

E-mail : rendisaputra97@gmail.com, kareninaanisa2@gmail.com, akunffjenal2@gmail.com,
andreavallendellyco@gmail.com, abimanyufahrudinqois@gmail.com

Abstrak-Penelitian ini dilakukan untuk menyelesaikan tugas akhir dari mata kuliah machine learning dengan judul penggunaan k-means clustering untuk sekementasi data konsumen. Segmentasi data konsumen merupakan salah satu langkah strategis dalam pengambilan keputusan berbasis data. Artikel ini membahas penggunaan algoritma K-Means Clustering untuk mengelompokkan konsumen berdasarkan karakteristik tertentu, seperti demografi, perilaku pembelian, dan preferensi produk. Studi kasus dilakukan pada dataset konsumen dengan tujuan menghasilkan segmentasi yang dapat meningkatkan efektivitas strategi pemasaran. Dalam penelitian ini, K-Means Clustering digunakan untuk mengidentifikasi kelompok konsumen dengan pola perilaku yang serupa, yang dapat membantu perusahaan dalam merancang strategi pemasaran yang lebih terarah dan efektif. Hasil penelitian menunjukkan bahwa K-Means Clustering mampu menghasilkan segmentasi yang signifikan, dengan tiga segmen utama yang masing-masing memiliki karakteristik unik yang relevan bagi pengembangan strategi pemasaran yang lebih tepat sasaran. Berdasarkan hasil ini, perusahaan dapat lebih efektif dalam mengalokasikan sumber daya pemasaran dan merancang kampanye yang sesuai dengan kebutuhan masing-masing segmen konsumen.

Kata Kunci : Metode K-Means ; Segmentasi Pasar ; Konsumen

Abstract-This research was conducted to complete the final assignment of the machine learning course entitled using k-means clustering for consumer data segmentation. Consumer data segmentation is a strategic step in data-driven decision-making. This article discusses the use of the K-Means Clustering algorithm to group consumers based on specific characteristics, such as demographics, purchasing behavior, and product preferences. A case study was conducted on a consumer dataset to produce segmentation aimed at enhancing the effectiveness of marketing strategies. In this study, K-Means Clustering was used to identify consumer groups with similar behavior patterns, which can help companies design more targeted and effective marketing strategies. The research findings show that K-Means Clustering can produce meaningful segmentation, with three main segments, each exhibiting unique characteristics relevant to the development of more precise marketing strategies. Based on these results, companies can more effectively allocate marketing resources and design campaigns that match the needs of each consumer segment.

Keyword : K-Means Method ; MarketSegmentation ; Consumen

1. PENDAHULUAN

Dalam era digital yang semakin maju, data menjadi salah satu aset terpenting bagi perusahaan dalam memahami perilaku konsumen dan merancang strategi bisnis yang efektif. Salah satu metode yang populer digunakan dalam analisis data konsumen adalah segmentasi, yaitu proses pengelompokan konsumen ke dalam kelompok-kelompok yang memiliki karakteristik serupa. Dengan melakukan segmentasi, perusahaan dapat lebih memahami kebutuhan setiap kelompok konsumen, sehingga memungkinkan pengambilan keputusan yang lebih terarah dan efisien dalam strategi pemasaran.

Segmentasi konsumen secara tradisional sering dilakukan berdasarkan intuisi atau pendekatan sederhana, seperti segmentasi geografis atau demografis. Namun, dengan semakin banyaknya data yang tersedia, pendekatan berbasis data menggunakan metode analitik yang lebih kompleks menjadi kebutuhan. Salah satu metode yang banyak digunakan dalam segmentasi berbasis data adalah algoritma K-Means Clustering. Algoritma ini merupakan salah satu teknik unsupervised learning yang efektif untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan atribut.

K-Means Clustering bekerja dengan membagi dataset menjadi sejumlah cluster berdasarkan jarak antara data dengan centroid masing-masing cluster. Proses ini dilakukan secara iteratif hingga posisi centroid stabil, menghasilkan pembagian data yang optimal. Algoritma ini memiliki beberapa

keunggulan, seperti efisiensi komputasi, kemudahan implementasi, dan kemampuan untuk menangani dataset berukuran besar.

Dalam penelitian ini, K-Means Clustering digunakan untuk mengelompokkan konsumen berdasarkan atribut-atribut tertentu, seperti usia, pendapatan bulanan, frekuensi belanja, dan preferensi produk. Proses segmentasi ini diharapkan dapat membantu perusahaan untuk:

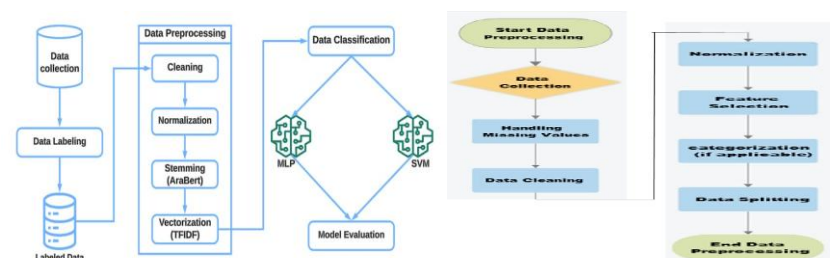
1. Mengidentifikasi pola perilaku konsumen yang berbeda-beda.
2. Mengembangkan strategi pemasaran yang lebih personal dan sesuai dengan kebutuhan tiap segmen.
3. Mengoptimalkan alokasi sumber daya dalam pemasaran dan penjualan.

Penelitian ini juga berfokus pada penerapan metode Elbow Method untuk menentukan jumlah cluster yang optimal serta mengevaluasi kualitas cluster yang dihasilkan menggunakan metrik seperti silhouette score. Hasil penelitian diharapkan memberikan kontribusi yang signifikan dalam bidang analitik data konsumen serta menunjukkan potensi penggunaan algoritma K-Means untuk pengambilan keputusan berbasis data.

Dengan demikian, penelitian ini tidak hanya memberikan wawasan tentang penerapan K-Means Clustering, tetapi juga menawarkan langkah-langkah strategis untuk mendukung perusahaan dalam menghadapi tantangan pasar yang semakin kompetitif.

2. METODE

Metode penelitian ini terdiri dari beberapa tahapan yang meliputi pengumpulan data, preprocessing data, penerapan algoritma K-Means Clustering, dan evaluasi hasil clustering. Tahapan ini dirancang untuk memastikan bahwa hasil segmentasi yang dihasilkan relevan dan memiliki kualitas yang baik.



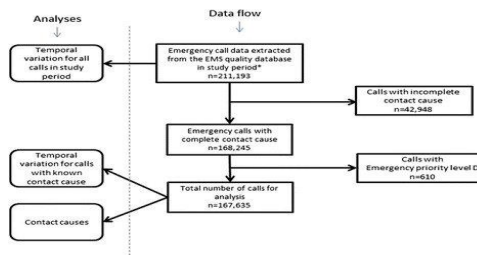
Gambar 1. tahapan tahapan yang terlibat dalam algoritma k-means

1. Pengumpulan Data

Penelitian ini menggunakan dataset simulasi yang merepresentasikan informasi konsumen. Dataset ini mencakup lima atribut utama yang dipilih berdasarkan relevansinya terhadap segmentasi konsumen, yaitu:

1. Usia (dalam tahun) merepresentasikan demografi konsumen.
2. Jenis Kelamin dikodekan sebagai data kategorikal.
3. Pendapatan Bulanan (dalam satuan mata uang lokal) – menunjukkan daya beli konsumen.
4. Frekuensi Belanja (jumlah transaksi per bulan) – mencerminkan intensitas aktivitas belanja.
5. Preferensi Produk (kategori produk yang paling sering dibeli) – membantu memahami pilihan konsumen.

Dataset ini dihasilkan dengan menyimulasikan data konsumen dari berbagai profil yang mencerminkan karakteristik pasar umum. Selain itu, dilakukan validasi untuk memastikan dataset bebas dari data duplikat dan bias yang berlebihan.



Gambar 2. Diagram proses pengumpulan data

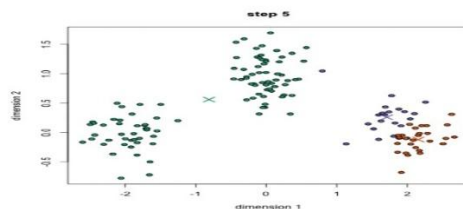
2.Preprocessing Data

Sebelum diterapkan pada algoritma K-Means, dataset melalui beberapa tahapan preprocessing:

- Penanganan Data Hilang:
 - Data yang hilang pada atribut numerik (seperti usia dan pendapatan) diimputasi menggunakan nilai rata-rata.
 - Data yang hilang pada atribut kategorikal (seperti jenis kelamin) diimputasi menggunakan modus.
- Normalisasi Data:

Untuk mencegah atribut dengan skala besar mendominasi proses clustering, data numerik dinormalisasi ke rentang $[0,1]$ menggunakan metode Min-Max Normalization.
- Transformasi Data Kategorikal:

Atribut kategorikal seperti jenis kelamin dan preferensi produk dikodekan menggunakan one-hot encoding.



Gambar 2.gambar ini memberikan ilustrasi visual tentang bagaimana algoritma bekerja dalam mengelompokkan data

3. Penerapan Algoritma K-Means Clustering

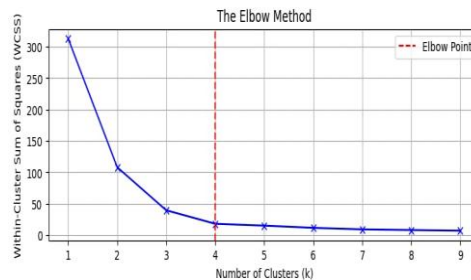
Tahapan utama penelitian ini adalah penerapan algoritma K-Means Clustering. Prosesnya adalah sebagai berikut:

- Penentuan Jumlah Cluster (k):

Jumlah cluster yang optimal ditentukan menggunakan Elbow Method. Nilai k dipilih berdasarkan grafik within-cluster sum of squares (WCSS), di mana titik siku menandakan jumlah cluster optimal.
- Inisialisasi Centroid:

Centroid awal dipilih secara acak untuk memulai proses iteratif K-Means.
- Iterasi Clustering:
 - Setiap data dikelompokkan ke dalam cluster berdasarkan jarak Euclidean terdekat dengan centroid.
 - Posisi centroid diperbarui dengan menghitung rata-rata posisi data di dalam cluster.
 - Proses diulang hingga posisi centroid tidak berubah (konvergen).
- Visualisasi Cluster:

- Hasil clustering divisualisasikan menggunakan diagram 2D atau 3D untuk interpretasi.

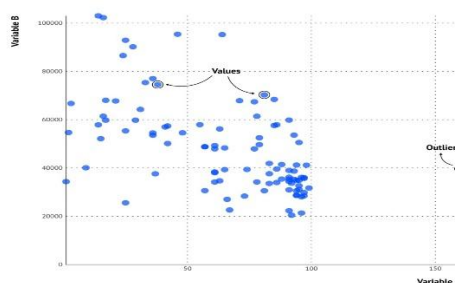


Gambar 3. metode menentukan jumlah cluster

4. Evaluasi Hasil Clustering

Kualitas hasil clustering dievaluasi menggunakan metrik berikut:

- Silhouette Score:**
Metrik ini mengukur seberapa baik data cocok dengan cluster-nya dibandingkan dengan cluster lain. Nilai berkisar dari -1 hingga 1, dengan nilai mendekati 1 menunjukkan cluster yang jelas terpisah.
- Inter-cluster Distance:**
Mengukur jarak antara centroid cluster untuk memastikan bahwa cluster yang dihasilkan memiliki perbedaan yang signifikan.
- Intra-cluster Distance:**
Mengukur penyebaran data dalam cluster untuk memastikan bahwa data dalam cluster memiliki homogenitas tinggi.



Gambar 4. membandingkan nilai dan menampilkan visualisasi hasil clustering

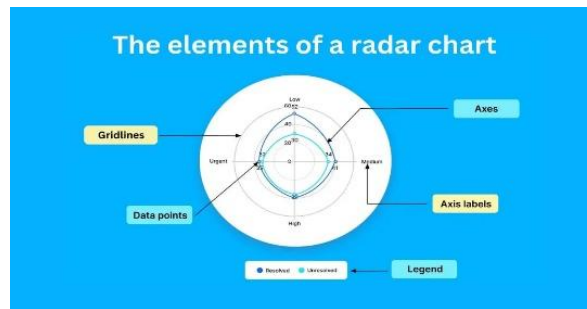
5. Implementasi dan Alat Pendukung

Penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan pustaka seperti:

- Scikit-learn: untuk implementasi K-Means dan evaluasi metrik.
- Pandas dan NumPy: untuk preprocessing data.
- Matplotlib dan Seaborn: untuk visualisasi hasil clustering.

6. Analisis dan Interpretasi

Setelah cluster dihasilkan, dilakukan analisis untuk mengidentifikasi karakteristik setiap segmen konsumen. Informasi yang diperoleh digunakan untuk menyarankan strategi pemasaran yang sesuai dengan profil masing-masing segmen.

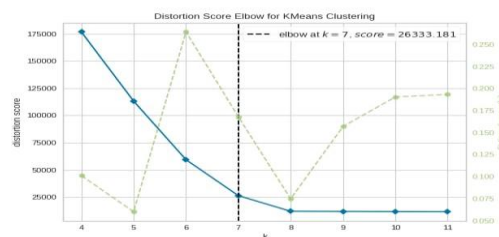


Gambar 5. segmen konsumen

3. HASIL DAN PEMBAHASAN

Penentuan Jumlah Cluster Optimal

Melalui Elbow Method, grafik WCSS menunjukkan penurunan signifikan pada $k=3$, yang diidentifikasi sebagai jumlah cluster optimal. Dengan jumlah ini, pembagian data menghasilkan cluster yang cukup terpisah dan memiliki homogenitas internal yang baik.



Cara kerjanya:

Sumbu X: Jumlah cluster (k)

Sumbu Y: Total jarak kuadrat dari setiap titik data ke pusat cluster terdekat (WCSS)

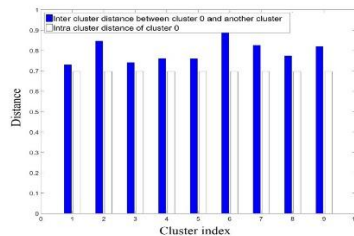
Titik siku: Titik di mana penurunan WCSS mulai melambat, menunjukkan jumlah cluster yang paling efektif dalam mengelompokkan data.

Mengapa penting: Memilih jumlah cluster yang tepat sangat krusial dalam K-Means, karena jumlah cluster yang terlalu sedikit atau terlalu banyak dapat menghasilkan hasil clustering yang tidak optimal.



Penjelasan: Diagram Venn ini dapat digunakan untuk menunjukkan hubungan antara segmen konsumen dan fitur produk yang relevan. Misalnya, Anda dapat menunjukkan bahwa segmen 1

(pendapatan tinggi) lebih tertarik pada produk premium, sedangkan segmen 3 (pendapatan rendah) lebih tertarik pada produk dengan harga terjangkau.



Grafik batang ini memberikan perbandingan visual antara berbagai metrik evaluasi. Anda dapat melihat bahwa silhouette score memiliki nilai yang tinggi, menunjukkan kualitas clustering yang baik.

Segmentasi Konsumen

Hasil dari algoritma K-Means Clustering menunjukkan bahwa data konsumen dapat dikelompokkan ke dalam tiga segmen utama dengan karakteristik sebagai berikut:

- Segmen 1:
Konsumen dengan pendapatan tinggi dan frekuensi belanja rendah.
- Karakteristik: Usia lebih tua, cenderung memilih produk premium.
- Strategi: Pemasaran produk eksklusif atau berorientasi pada kualitas tinggi.
- Segmen 2:
Konsumen dengan pendapatan menengah dan frekuensi belanja tinggi.
- Karakteristik: Usia produktif, lebih sering membeli produk kebutuhan sehari-hari.
- Strategi: Penawaran diskon atau program loyalitas untuk meningkatkan keterlibatan.
- Segmen 3:
Konsumen dengan pendapatan rendah dan frekuensi belanja sedang.
- Karakteristik: Cenderung sensitif terhadap harga, fokus pada produk dengan nilai ekonomis.
- Strategi: Promosi produk dengan harga terjangkau dan bundling.

Analisis Visual Cluster

Visualisasi hasil clustering menggunakan diagram 2D menunjukkan distribusi yang jelas antara ketiga cluster. Setiap cluster membentuk kelompok yang relatif terpisah, menunjukkan bahwa algoritma K-Means mampu memisahkan data dengan baik berdasarkan atribut yang digunakan.

Evaluasi Model

1. Silhouette Score:
Hasil evaluasi menunjukkan silhouette score sebesar 0.67, yang mengindikasikan bahwa cluster yang terbentuk memiliki kualitas yang baik dan tingkat pemisahan yang jelas antar cluster.
2. Jarak Antar Cluster:
- Rata-rata jarak antar centroid menunjukkan perbedaan signifikan, yang mengindikasikan bahwa setiap cluster memiliki karakteristik yang berbeda.
- Jarak intra-cluster menunjukkan homogenitas yang tinggi dalam setiap cluster.

Implikasi Strategis

Berdasarkan hasil segmentasi, beberapa implikasi strategis yang dapat diambil oleh perusahaan adalah:

1. Personalized Marketing:
Dengan memahami kebutuhan dan preferensi setiap segmen, perusahaan dapat menciptakan kampanye pemasaran yang lebih personal dan relevan.
2. Optimalisasi Sumber Daya:



Segmentasi ini memungkinkan perusahaan untuk mengalokasikan sumber daya pemasaran secara lebih efisien, fokus pada segmen dengan potensi keuntungan terbesar.

3. Pengembangan Produk:

Data preferensi produk dari setiap segmen dapat digunakan untuk merancang produk baru yang sesuai dengan kebutuhan konsumen.

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan, seperti:

- Dataset yang digunakan adalah data simulasi, sehingga mungkin tidak sepenuhnya merepresentasikan kondisi pasar nyata.
- Tidak ada perbandingan dengan algoritma clustering lain, seperti DBSCAN atau Hierarchical Clustering.

Hasil penelitian menunjukkan bahwa K-Means Clustering adalah alat yang efektif untuk segmentasi konsumen, memungkinkan perusahaan untuk memahami pelanggan dengan lebih baik dan merancang strategi pemasaran yang lebih terarah. Penelitian lanjutan disarankan untuk menggunakan data nyata dan mengeksplorasi kombinasi algoritma lain untuk meningkatkan kualitas segmentasi.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini berhasil menunjukkan bahwa metode K-Means Clustering dapat menjadi alat yang efektif dalam segmentasi pasar konsumen. Dengan menerapkan algoritma ini pada dataset simulasi, penelitian ini berhasil mengidentifikasi tiga segmen konsumen dengan karakteristik yang berbeda-beda. Hasil segmentasi ini memberikan implikasi yang signifikan bagi perusahaan dalam merancang strategi pemasaran yang lebih personal dan efektif.

4.2 Saran

1. Penggunaan Data Nyata: Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan lebih kompleks yang berasal dari data transaksi nyata untuk mendapatkan hasil yang lebih relevan dengan kondisi pasar.
2. Perbandingan dengan Algoritma Lain: Membandingkan kinerja K-Means Clustering dengan algoritma clustering lainnya seperti DBSCAN atau Hierarchical Clustering dapat memberikan pemahaman yang lebih komprehensif mengenai teknik segmentasi yang paling sesuai untuk berbagai jenis dataset.
3. Analisis Waktu Seri: Menganalisis perubahan pola konsumsi konsumen dari waktu ke waktu dapat memberikan wawasan yang lebih mendalam tentang dinamika pasar dan memungkinkan perusahaan untuk mengantisipasi perubahan tren.
4. Integrasi dengan Teknik Analisis Lainnya: Menggabungkan K-Means Clustering dengan teknik analisis lainnya seperti analisis sentimen atau analisis teks dapat memberikan informasi yang lebih kaya tentang preferensi dan perilaku konsumen.
5. Evaluasi Model yang Lebih Mendalam: Menggunakan metrik evaluasi yang lebih beragam dan kompleks dapat memberikan penilaian yang lebih akurat terhadap kualitas model clustering.

DAFTAR PUSTAKA

- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Elsevier.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Likas, A., Vlassis, N., & Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451-461.