



Penerapan Algoritma K-Means Untuk Segmentasi Pelanggan Pada Dataset Wholesale Customer

Application of K-Means Algorithm for Customer Segmentation on Wholesale Customer Dataset

Farhan Habli Mariadin¹, Hasbi Firmansyah²

¹Universitas Pancasakti Tegal, Tegal, Indonesia, Email: farhandcd4@gmail.com

²Universitas Pancasakti, Tegal, Tegal, Indonesia, Email: hasbifirmansyah@upstegal.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 03/09/2025

Revisi Akhir: 04/09/2025

Diterbitkan Online: 11/12/2025

KATA KUNCI

Clustering, Data Mining, K-Means, Segmentasi Pelanggan, Wholesale Customer

KORESPONDENSI

Email: farhandcd4@gmail.com

ABSTRAK

Penelitian ini menunjukkan penerapan kalkulasi K-Means dalam pembagian klien dengan memanfaatkan kumpulan data Klien Diskon. Pembagian klien merupakan prosedur penting dalam memahami desain pemanfaatan dan membuat langkah-langkah untuk meningkatkan efektivitas. Dalam pertimbangan ini, penanganan penambangan informasi dilakukan dengan menggunakan program komputer RapidMiner, dengan tahapan pengumpulan informasi, penanganan informasi awal, dan investigasi pengelompokan menggunakan kalkulasi K-Means. Hasil pengelompokan menunjukkan bahwa informasi dikumpulkan menjadi lima kluster dengan kualitas yang dinilai menggunakan File Davies-Bouldin (DBI). Nilai DBI sebesar -0,947 menunjukkan bahwa pengelompokan tersebut memiliki kualitas yang baik, dengan setiap kluster mencerminkan desain pemanfaatan klien yang berbeda. Pertimbangan ini menunjukkan bahwa kalkulasi K-Means dapat digunakan secara efektif untuk pembagian klien, memberikan pengalaman utama bagi perusahaan dalam membuat pilihan bisnis

ABSTRACT

This study demonstrates the application of K-Means calculations to client allocation using the Discount Client dataset. Client allocation is a crucial procedure for understanding utilization design and developing strategies to improve effectiveness. For this purpose, information mining was performed using the RapidMiner computer program, with the stages of information collection, initial data processing, and clustering investigation using K-Means calculations. The clustering results indicate that the information is grouped into five clusters, with quality assessed using the Davies-Bouldin File (DBI). A DBI value of -0.947 indicates good quality clustering, with each cluster reflecting a different client utilization design. This analysis demonstrates that K-Means calculations can be used effectively for client allocation, providing key insights for companies in making business decisions.

1. PENDAHULUAN

Persaingan bisnis yang semakin ketat mendorong perusahaan untuk memahami perilaku pelanggan dengan lebih mendalam. Data pelanggan yang terus bertambah setiap hari memberikan peluang besar untuk menganalisis pola pembelian, melakukan segmentasi pasar, dan membuat keputusan strategis. Namun, tantangan utama adalah bagaimana memanfaatkan data ini secara efektif untuk menghasilkan informasi yang dapat digunakan.

Segmentasi pelanggan menjadi salah satu cara yang efektif untuk mengelompokkan pelanggan berdasarkan kesamaan dalam karakteristik atau pola konsumsi mereka. Dengan segmentasi yang akurat, perusahaan dapat menyusun strategi pemasaran yang lebih spesifik, meningkatkan pengalaman pelanggan, dan memanfaatkan sumber daya secara optimal. Salah satu metode yang sering digunakan dalam segmentasi adalah algoritma K-Means, yang dikenal karena kemampuannya mengelompokkan data dalam jumlah besar secara efisien dan sederhana.

2. TINJAUAN PUSTAKA

A. Data Mining

Data Mining melibatkan penggalian pola dan informasi penting dari kumpulan data yang besar dan kompleks. Proses ini menerapkan berbagai metode seperti statistik, kecerdasan buatan, dan pembelajaran mesin untuk menemukan pola, hubungan, dan tren tersembunyi dalam data yang tidak terlihat secara langsung. Fase penambangan data melibatkan pengumpulan data, pemrosesan awal data untuk menghilangkan anomali dan inkonsistensi, mengubah data ke format yang sesuai, menerapkan teknik analisis seperti pengelompokan dan klasifikasi, dan mendukung pengambilan keputusan strategis. Dalam bisnis, data mining banyak digunakan untuk memahami perilaku pelanggan, menganalisis risiko, dan memprediksi tren masa depan.

B. Clustering

Clustering merupakan strategi penting dalam pengolahan data untuk mengelompokkan informasi ke dalam subset yang sebanding berdasarkan karakteristik atau desain tertentu. Alasan utama pengelompokan adalah untuk mengidentifikasi struktur tersembunyi dalam informasi sehingga kita dapat memperoleh kelompok dan kategori dalam informasi dengan lebih baik. Pedoman pengelompokan yang paling utama adalah untuk memaksimalkan kemiripan dalam setiap klaster dan meminimalkan kemiripan antar klaster.

C. Algoritma K-Means

K-means merupakan teknik pengelompokan data non-hierarki yang berupaya membagi data yang ada menjadi dua kelompok atau lebih. (*clusters*) berdasarkan kemiripan fitur atau karakteristik. Metode ini termasuk ke dalam *unsupervised learning* yang berarti label data dalam proses pembelajaran K-means bertujuan untuk

mempartisi data ke dalam k *cluster*, dimana setiap data akan dimasukkan ke *cluster* dengan jarak terdekat.

$$\begin{aligned} \frac{\partial}{\partial u_k} &= \frac{\partial}{\partial u_k} \sum_{k=1}^k \sum_{i=1}^n (x_i - u_k)^2 \\ &= \sum_{k=1}^k \sum_{i=1}^n \frac{\partial}{\partial u_k} (x_i - u_k)^2 \\ &= \sum_{i=1}^n 2(x_i - u_k) \end{aligned}$$

Gambar 1. Rumus K-Means

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan adalah Wholesale Customer Dataset, yang terdiri dari informasi konsumsi pelanggan negara Portugal pada berbagai kategori produk. Dataset ini memiliki beberapa atribut utama, yaitu: Region: Variabel numerik yang menunjukkan lokasi geografis pelanggan, dengan nilai:

1: Mengacu pada wilayah Lisbon

2: Mengacu pada wilayah lain Oporto

3: Mengacu pada wilayah Other Region

Fresh: Mengindikasikan jumlah konsumsi produk segar, seperti buah dan sayuran, dalam satuan berat atau volume tertentu.

Milk: Jumlah konsumsi produk susu dalam satuan yang sama.

Grocery: Total konsumsi produk kebutuhan sehari-hari, seperti sembako.

Frozen: Konsumsi produk beku, seperti daging atau makanan olahan beku.

Detergents_Paper: Jumlah konsumsi produk pembersih dan kertas.

Delicassen: Mengacu pada konsumsi produk makanan siap saji atau barang khusus lainnya.

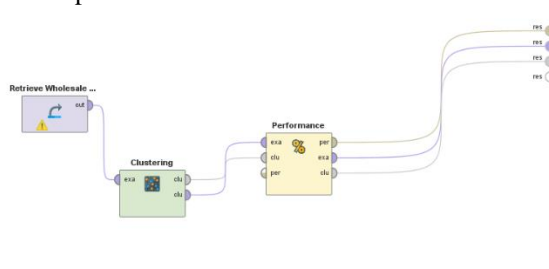
Dataset ini memiliki total 440 baris data, di mana setiap baris mewakili satu pelanggan, dan tidak memiliki atribut target. Oleh karena itu, analisis dilakukan dengan menggunakan metode clustering untuk mengelompokkan pelanggan berdasarkan pola konsumsi mereka

no	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
1	3	12669	9656	7561	214	2674	1338
2	3	7057	9810	9568	1762	3293	1776
3	3	6353	8808	7684	2405	3516	7844
4	3	13265	1196	4221	6404	507	1788
5	3	22615	5410	7198	3915	1777	5185
6	3	9413	8259	5126	666	1795	1451
7	3	12126	3199	6975	480	3140	545
8	3	7579	4956	9426	1669	3321	2566
9	3	5963	3648	6192	425	1716	750
10	3	6006	11093	18881	1159	7425	2098
....
440	3	2787	1698	2510	65	477	52

Gambar 2. Dataset Wholesale Customer

A. Praprocessing Data

Proses penerapan algoritma K-Means ini menggunakan software Rapidminer, dengan cara mengupload dataset lalu melakukan training data dengan menggunakan algoritma clustering K-Means dan menggunakan cluster distance performance.



Gambar 3. Pengolahan data

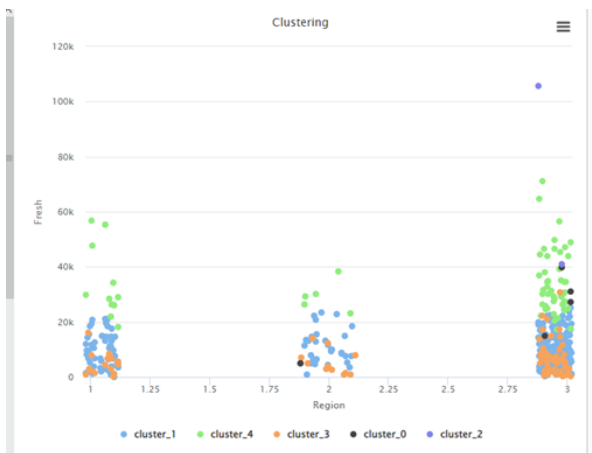
B. Hasil Cluster Model

Setelah data diolah akan menghasilkan cluster model sebagai berikut

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Region	2.800	2.538	3	2.512	2.571
Fresh	25603	8569.241	74499	5178.250	33611.270
Milk	43460.600	3211.381	36788.500	12308.750	4874.397
Grocery	61472.200	4121.105	19159	19113.214	5852.968
Frozan	2636	2745.528	26639.500	1655.060	5729.286
Detergents_Paper	29974.200	1183.227	2593.500	8426.452	1056.730
Delicassen	2708.800	1052.420	28246.500	1980.714	2119.587

Gambar 4. Hasil Cluster Model

Pada Cluster model yang dihasilkan dari penghitungan Rapidminer yaitu Cluster 0: 5 items, Cluster1: 286 items, Cluster2: 2 items, Cluster3: 84 items, Cluster4: 63 items, maka data keseluruhan yang telah dihitung yaitu 440 items



Gambar 5. Visualization Cluster

C. Hasil PerformanceVector

Hasil evaluasi PerformanceVector menunjukkan kualitas cluster yang dihasilkan oleh algoritma K-Means. Berikut penjelasannya:

A. Rata-rata Jarak dalam Klaster (Avg. within centroid

distance)

Nilai ini menunjukkan seberapa dekat data di dalam satu klaster terhadap pusat klasternya (centroid). Setiap klaster memiliki nilai rata-rata yang berbeda. Semakin kecil nilai ini (dalam arti absolut), semakin baik, karena data dalam klaster tersebut lebih terkonsentrasi di sekitar pusatnya.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -130730511.504
Avg. within centroid distance_cluster_0: -11364489819.680
Avg. within centroid distance_cluster_1: -67974049.367
Avg. within centroid distance_cluster_2: -1961379220.000
Avg. within centroid distance_cluster_3: -155767992.849
Avg. within centroid distance_cluster_4: -244303619.133
Davies Bouldin: -0.947
```

Gambar 6. Hasil rata rata dengan performancevector

B. Evaluasi Davies-Bouldin Index

Uji validasi yang diterapkan adalah Davies Bouldin Index (DBI). Dari hasil uji DBI, bisa diketahui seberapa efektif proses pengelompokan yang telah dilakukan, semakin rendah nilai yang diperoleh, semakin baik kualitas pengelompokan tersebut. Pada pengujian kali ini nilai yang didapatkan adalah -0.947, nilai ini menunjukkan kualitas yang baik.

Davies Bouldin

Davies Bouldin: -0.947

Gambar7. Hasil Davies-Bouldin Index

4. KESIMPULAN

Berdasarkan hasil penelitian, algoritma K-Means terbukti mampu mengelompokkan data pelanggan dengan baik menjadi lima klaster berdasarkan pola konsumsi mereka. Proses ini membantu dalam memahami perilaku pelanggan secara lebih mendalam, sehingga memudahkan perusahaan dalam merancang strategi pemasaran yang lebih tepat sasaran. Evaluasi dengan Davies-Bouldin Index menunjukkan nilai sebesar -0.947, yang menandakan bahwa kualitas pengelompokan cukup baik. Dengan demikian, algoritma K-Means dapat menjadi alat yang efektif untuk segmentasi pelanggan, khususnya pada dataset besar seperti Wholesale Customer Dataset.

DAFTAR PUSTAKA

[1] Idham, I., Rosika, H., & Yuliadi, Y. (2024). IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K MEANS. JUTECH: Journal Education and Technology, 5(1), 221-231.

[2] Ubaidillah, M., & Fatah, Z. (2024). IMPLEMENTASI RAPIDMINER PADA KLASERISASI GEMPA BUMI DI INDONESIA BERDASARKAN KEDALAMAN MENGGUNAKAN K-MEANS. Jurnal Ilmiah Multidisiplin Ilmu, 1(6), 84-91.

[3] Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. Jurnal Ilmiah

Informatika dan Ilmu Komputer (JIMA-ILKOM),
3(1), 46-56.

- [4] Nasir, J. (2020). Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means. *Jurnal Simetris*, 11(2), 1-13.
- [5] Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, 2(3).