

Analisis Klustering Dataset Kampanye Pemasaran Menggunakan Algoritma K-Means Pada Rapidminer

Rifqi Aditya Wilalung

¹Program Studi S1 Informatika, Universitas Pancasakti, Kota Tegal, Indonesia

Email : Rifqiaditya1809@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis data kampanye pemasaran dengan menerapkan algoritma K-Means melalui platform RapidMiner. K-Means dipilih karena efisiensinya dalam mengelola dataset besar dan kemampuannya menghasilkan kluster yang bermakna. Dataset yang dianalisis mencakup informasi mengenai pelanggan, status perkawinan, dan tingkat pendidikan. Sebelum melakukan analisis, langkah preprocessing dilakukan untuk memastikan kualitas data yang optimal. Dalam studi ini, evaluasi Davies-Bouldin digunakan untuk menentukan jumlah kluster yang paling tepat, dengan eksplorasi varians jumlah kluster antara (2,3,4,5). Hasil analisis menunjukkan bahwa jumlah kluster yang optimal adalah 2, di mana masing-masing kluster memperlihatkan karakteristik unik hasil klusterisasi. Proses ini memberikan wawasan berharga tentang perilaku pelanggan yang dapat digunakan untuk merumuskan strategi pemasaran yang lebih efektif dan efisien. Temuan penelitian ini menegaskan bahwa algoritma K-Means dapat mengelompokkan data secara akurat, memberikan fondasi yang kuat untuk pengambilan keputusan strategis di bidang pemasaran.

Kata Kunci : klusterisasi, algoritma K-Means, RapidMiner, Davies-Bouldin Index.

Abstract

This study aims to analyze marketing campaign data by applying the K-Means algorithm through the RapidMiner platform. K-Means was chosen due to its efficiency in managing large datasets and its ability to generate meaningful clusters. The dataset includes information on customers, marital status, and education level. Prior to analysis, preprocessing was performed to ensure optimal data quality. In this study, the Davies-Bouldin evaluation was used to determine the most appropriate number of clusters, exploring the variance of cluster numbers between (2, 3, 4, or 5). The analysis results indicated that the optimal number of clusters was 2, with each cluster displaying unique characteristics resulting from the clustering process. This process provides valuable insights into customer behavior that can be used to design more effective and efficient marketing strategies. The findings of this study confirm that the K-Means algorithm can accurately cluster data, providing a strong foundation for strategic decision-making in the marketing field.

Keywords: clustering, K-Means algorithm, RapidMiner, Davies-Bouldin Index.

PENDAHULUAN

1. Latar Belakang

Di era digital saat ini, pemasaran yang efektif memerlukan pemahaman mendalam mengenai perilaku dan preferensi pelanggan. Dengan meningkatnya jumlah data yang dihasilkan dari interaksi pelanggan melalui berbagai saluran digital,

perusahaan dituntut untuk mampu menganalisis dan mengelompokkan informasi tersebut guna merumuskan strategi pemasaran yang lebih efisien. Salah satu metode yang banyak digunakan untuk mencapai tujuan ini adalah klustering, terutama dengan memanfaatkan algoritma K-Means.

Algoritma K-Means telah menjadi pilihan utama dalam analisis data karena efisiensinya dalam menangani volume besar serta kemampuannya untuk menghasilkan kluster yang bermakna. Selain itu, K-Means merupakan algoritma yang paling mudah digunakan, terutama bagi pemula dalam bidang klusterisasi. Namun, satu tantangan utama dalam implementasinya adalah menentukan jumlah kluster yang optimal. Jumlah kluster yang tidak tepat dapat berakibat pada hasil klusterisasi yang tidak representatif atau sulit untuk diinterpretasikan.

Melalui penelitian ini, diharapkan dapat tercapai pemahaman yang lebih mendalam mengenai karakteristik segmen pelanggan dalam dataset kampanye pemasaran. Hasil dari klusterisasi ini tidak hanya akan memberikan wawasan tentang perilaku pelanggan, tetapi juga akan menjadi dasar yang kuat untuk pengambilan keputusan strategis dalam merancang kampanye pemasaran yang lebih efektif dan efisien.

2. Identifikasi Masalah

- a. Kesulitan mengelola data pelanggan yang besar dan kompleks.
- b. Minimnya pemahaman terhadap segmentasi pasar.
- c. Kurangnya pemanfaatan teknologi modern.
- d. Penentuan jumlah kluster yang tepat.

3. Tujuan Penelitian

- a. Mengimplementasikan algoritma K-Means untuk melakukan klustering pada dataset kampanye pemasaran guna mengidentifikasi pola segmentasi pelanggan.
- b. Menentukan jumlah cluster yang optimal dalam dataset menggunakan indeks Davies-Bouldin sebagai metode evaluasi.
- c. Mengevaluasi efektivitas algoritma K-Means dalam mengelompokkan pelanggan berdasarkan karakteristik serupa.

METODE

1. Data Mining

Data mining adalah proses untuk mengekstraksi informasi, pengetahuan, dan pola dari kumpulan data yang besar. Tujuannya adalah untuk memperoleh wawasan dan memperluas data serta informasi yang ada.

Data mining, sebagai salah satu metode analisis data, telah menjadi alat bagi perusahaan dalam menangani volume data yang sangat besar. Dengan teknik ini, Perusahaan dapat mengidentifikasi pola-pola tersembunyi yang dapat dimanfaatkan untuk mendukung pengambilan keputusan strategis[1].

2. Algoritma K-means

Salah satu algoritma data mining yang paling umum digunakan adalah K-means Clustering. Algoritma ini berfungsi untuk mengelompokkan data menjadi beberapa kategori berdasarkan kesamaan karakteristik tertentu. K-means Clustering diakui sebagai salah satu metode yang paling sederhana dan umum, berkat kemampuannya untuk mengelompokkan data dalam jumlah besar dengan cara yang efisien.[2].

Penelitian ini memanfaatkan metode K-Means Clustering untuk melakukan segmentasi pasar, yang bertujuan untuk membantu merancang strategi pemasaran yang lebih terarah. Berikut adalah langkah-langkah dalam mengimplementasikan algoritma K-Means untuk pengelompokan:

- a. Menetapkan Jumlah Cluster (K): Langkah Pertama adalah menentukan jumlah cluster yang diinginkan sesuai dengan kebutuhan analisis yang akan dilakukan.
- b. Inisialisasi Centroid: Secara acak pilih centroid awal dari dataset.
- c. Menetapkan Setiap Data ke Cluster Terdekat: Hitunglah jarak setiap titik data ke centroid yang terdekat, kemudian alokasikan setiap titik tersebut ke centroid yang memiliki jarak terpendek.
- d. Memperbarui Centroid: Hitung rata-rata dari data yang telah dikelompokkan ke dalam setiap cluster, lalu perbarui posisi centroid sesuai dengan hasil perhitungan tersebut.
- e. Mengulang Proses: Lakukan langkah-langkah penetapan data dan pembaruan centroid secara berulang hingga tidak ada perubahan signifikan pada posisi centroid atau hingga mencapai jumlah iterasi yang ditentukan.

- f. **Evaluasi Hasil Clustering:** Sebagai Langkah terakhir, lakukan evaluasi terhadap hasil pengelompokan untuk memastikan bahwa data telah dikelompokkan dengan baik dan tujuan penelitian tercapai.

3. Metode Davies Bouldin

Davies Bouldin Index merupakan ukuran untuk menilai efektivitas pengelompokan. Terapkan DBI sebagai metrik pengelompokan karena metode umum dalam validasi pengelompokan meliputi dua kategori utama validasi eksternal dan validasi internal yang digunakan untuk mengevaluasi kinerja hasil clustering[3].

Menggunakan Davies Bouldin Index untuk mengevaluasi performance clustering K-Means karena di dalam Davies Bouldin ini sendiri terdapat dua indikator penting yang digunakan yaitu :

1. **Indikator Homogenitas**

Dalam indikator ini akan dilihat seberapa mirip/kesamaan antara satu anggota dengan anggota yg lain dalam satu cluster. Makin kecil nilai Davies Bouldinnya berarti makin mirip antar anggotanya yang artinya semakin baik performance-nya, sebaliknya makin besar nilai Davies Bouldinnya maka makin tidak mirip antar anggota yang berarti kurang baik performance-nya.

2. **Indikator Heterogenitas**

Dalam Indikator ini akan dilihat jarak antar cluster satu dengan lainnya benar-benar berjarak atau tidak. Makin besar jaraknya maka makin bagus, begitupun sebaliknya.

4. Rapidminer

RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platform dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. RapidMiner memiliki sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi.

RapidMiner memiliki sekitar 500 administrator untuk penambahan informasi, yang mencakup berbagai peran seperti administrator input, hasil, prapemrosesan data, dan representasi. Platform ini dibangun dengan menggunakan bahasa Java, yang memungkinkannya untuk beradaptasi dengan berbagai kerangka kerja[4].

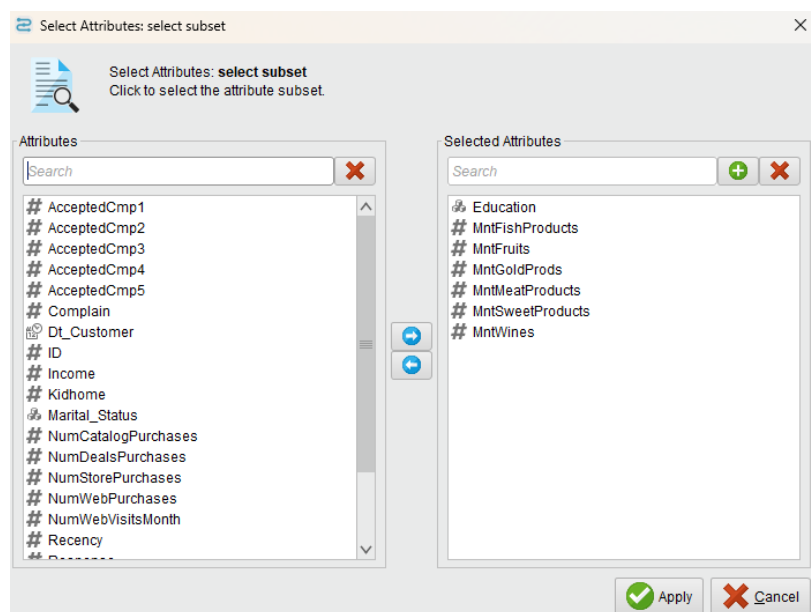
5. Persiapan Dataset

Dataset diambil dari sumber terpercaya yang mencakup informasi seperti data pelanggan, status perkawinan, dan pendidikan dengan pengambilan sampel data dari kaggle. Data tersebut akan di proses klasterisasi setelah melewati beberapa Langkah selanjutnya.

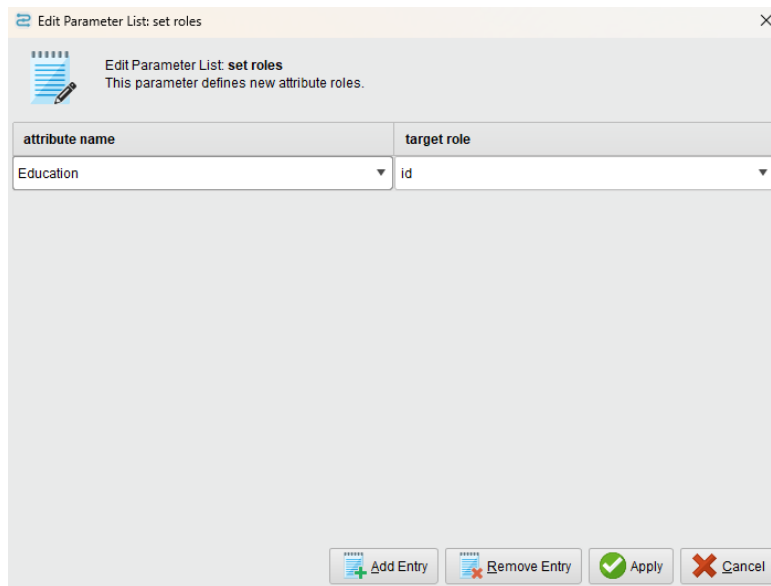
6. Preprocessing Data

Proses preprocessing dilakukan untuk memastikan kualitas data yang optimal. Tahap ini mencakup seleksi data dengan memilih dua atribut utama, yaitu positive dan negative review dan multiply data. Tahapan ini dilakukan menggunakan operator Select Attributes, Set Role dan multiply untuk menguji berbagai jumlah cluster di RapidMiner.

Gambar 1. Select Attributes



Gambar 2. Set Role



Sumber : Dokumentasi Pribadi

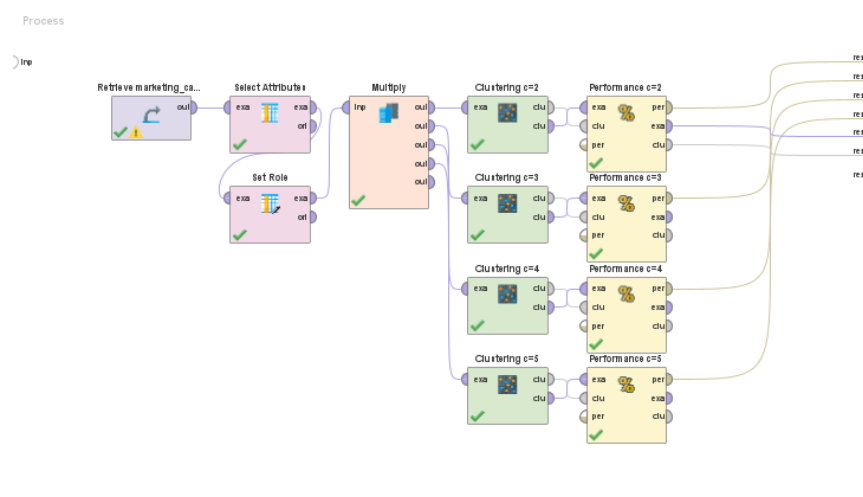
HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menentukan jumlah cluster optimal pada dataset Kampanye Pemasaran menggunakan algoritma K-Means pada RapidMiner. Hasil penelitian diperoleh dari serangkaian pengumpulan dataset hingga analisis klasterisasi dengan variasi jumlah cluster (2, 3, 4, dan 5). Dimana nanti dari keempat operator K-Means ini akan dipilih salah satu yang paling optimal melalui evaluasi Davies Bouldin.

1. Hasil Penerapan Algoritma K-Means

Algoritma K-Means diterapkan pada masing-masing dataset yang digandakan. Setelah itu, Desain proses klasterisasi dibuat dalam bentuk workflow di RapidMiner untuk memastikan langkah-langkah berjalan sistematis dan terstruktur. Proses ini ditampilkan pada Gambar 3, yang menggambarkan alur lengkap dari preprocessing data hingga evaluasi cluster.

Gambar 3. Proses klustering



Sumber : Dokumentasi Pribadi

Keterangan :

- 1) Read CSV : Digunakan untuk mengimport dataset kampanye pemasaran.
- 2) Select Attributes : Memilih atribut apa saja yang akan digunakan.
- 3) Set Role : menentukan fungsi atau peran atribut dalam dataset.
- 4) Multiply : Untuk share dataset agar bisa digunakan oleh 4 K-means.
- 5) K-means : Untuk clustering data.
- 6) Performance : Untuk mengevaluasi kinerja clustering K-means.

Proses klustering pada dataset Kampanye Pemasaran menghasilkan performa untuk masing-masing cluster berdasarkan analisis rata-rata jarak tiap cluster. Evaluasi ini dilakukan menggunakan *operator Cluster Distance Performance* di RapidMiner.

Hasil performa untuk masing-masing jumlah cluster yang berdasarkan avg. within centroid distance ditunjukkan pada Tabel 1 berikut:

Tabel 1. Hasil pengujian performance

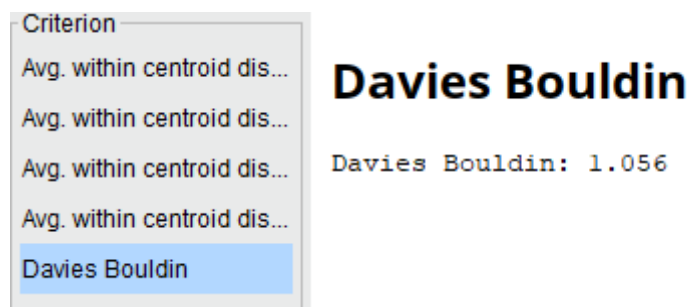
#	Cluster	Avg.centroid
1	2	52068.549
2	3	36254.943
3	4	33384.412
4	5	69841.010

2. Hasil Evaluasi dengan metode Davies Bouldin

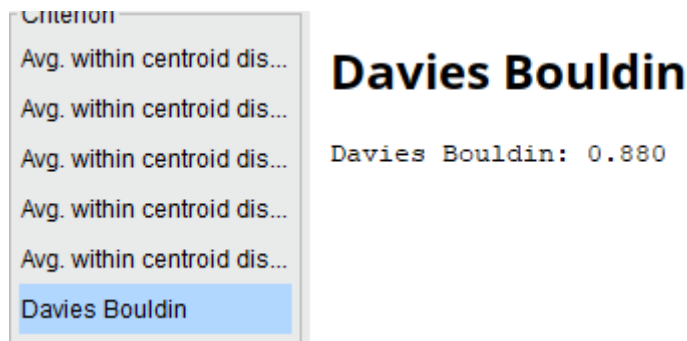
Berikut adalah hasil evaluasi untuk k-means dengan jumlah kluster 2, 3, 4, dan 5:



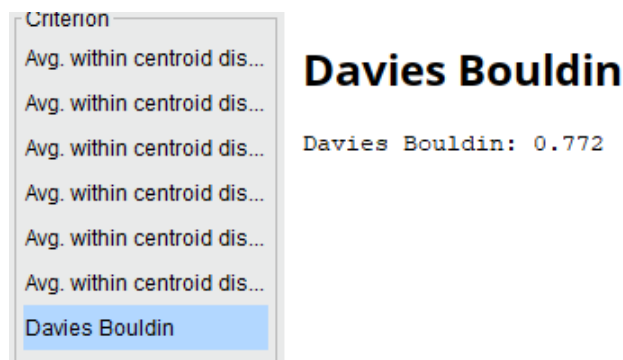
Gambar 4. Evaluasi Davies Bouldin Index K-Means 2



Gambar 5. Evaluasi Davies Bouldin Index K-Means 3



Gambar 6. Evaluasi Davies Bouldin Index K-Means 4

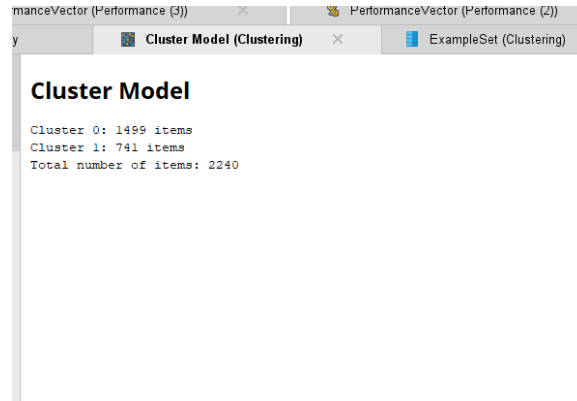


Gambar 7. Evaluasi Davies Bouldin Index K-Means 5

Sumber : Dokumentasi Pribadi

3. Hasil Cluster Optimal

Setelah melalui proses evaluasi menggunakan metode davies bouldin, jumlah cluster optimal yang dipilih adalah 2. Setiap cluster yang terbentuk memiliki karakteristik unik berdasarkan hasil klasterisasi terhadap dataset Kampanye Pemasaran.



Gambar 8. Cluster Model

Sumber : Dokumentasi Pribadi

Gambar 8 menunjukkan distribusi data hasil clustering menunjukkan total 2240 item terbagi menjadi dua cluster : Cluster 0 berisi 1499 item, dan cluster 1 berisi 741 item. Evaluasi clustering menggunakan metode Davies Bouldin dapat digunakan untuk memastikan pembagian ini menghasilkan cluster yang optimal.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means untuk melakukan klasterisasi pada dataset kampanye pemasaran, dan selanjutnya mengevaluasi hasilnya dengan menggunakan metode Davies Bouldin. Temuan menunjukkan bahwa konfigurasi optimal terdiri dari dua klaster, di mana masing-masing klaster memiliki karakteristik unik yang dapat dimanfaatkan untuk mendukung strategi pemasaran. Evaluasi melalui indeks Davies-Bouldin memberikan validasi yang kuat mengenai kualitas klasterisasi yang dilakukan. Oleh karena itu, analisis ini berpotensi menjadi landasan yang kokoh bagi pengambilan keputusan yang lebih terarah dalam pengelolaan kampanye pemasaran.

DAFTAR PUSTAKA

- [1] A. A. Alya Putri and S. A. Rahmah, "Implementasi Data Mining Dengan Algoritma K-Means Clustering Untuk Analisis Bisnis Pada Perusahaan Asuransi," *Djtechno J.*

- Teknol. Inf.*, vol. 5, no. 1, pp. 139–152, 2024, doi: 10.46576/djtechno.v5i1.4537.
- [2] N. Aini, N. Suarna, and W. Prihartono, “Strategi Promosi Online Shop Di Media Sosial Melalui Penerapan Data Mining Dengan Metode K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3628–3632, 2024, doi: 10.36040/jati.v7i6.8301.
- [3] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [4] P. S. Ananda, E. Sediono, and I. Sembiring, “KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal,” *J. BATIRSI*, vol. 6, no. 2, p. 10, 2023.