

Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall

Nur Rohman¹, Arief Wibowo²

¹Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Ciledug Raya, DKI Jakarta, Indonesia

²Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Ciledug Raya, DKI Jakarta, Indonesia

e-mail: 2211601667@student.budiluhur.ac.id¹, arief.wibowo@budiluhur.ac.id²

Received : January, 2024

Accepted : February, 2024

Published : April, 2024

Abstract

Customer segmentation entails grouping consumers based on shared characteristics such as age, gender, annual income, location, and purchasing habits. The results of this segmentation process serve as a guide for developing marketing strategies and introducing new products tailored to each specific customer group. Moreover, segmentation outcomes play a crucial role in creating products specifically targeted at high-value customer segments. In response to this challenge, the researcher utilized clustering techniques, specifically K-Means and K-Medoids, to categorize customers in malls and gain insights into sales strategies. Testing was conducted using customer segmentation data from a mall, comprising 200 data points on Google Colabs. The number of clusters was determined using the Elbow method, resulting in 5 clusters for the K-Means algorithm and 4 clusters for K-Medoids. Evaluation of these clusters involved comparing the Silhouette Coefficient values of both algorithms to determine the optimal number of clusters. A Silhouette Coefficient value approaching 1 indicates more optimal cluster results. The K-Means algorithm demonstrated a higher Silhouette Coefficient at 0.553, while the K-Medoids algorithm produced a Silhouette Coefficient of 0.485. Therefore, based on the obtained Silhouette Coefficient value, the K-Means algorithm is deemed the most suitable for clustering customer segmentation results in shopping malls.

Keywords: K-Means, K-Medoids, Mall, Segmentation, Silhouette Coefficient.

Abstrak

Memahami pelanggan sangat penting untuk mengelola operasi perusahaan. Dengan mengetahui dan memahami setiap pelanggan, dapat meningkatkan komunikasi layanan produk dengan menyesuaikan kebutuhan dan layanan kepada setiap pelanggan. Namun, analisis pelanggan sangat luas sehingga sulit untuk memahami kebutuhan masing-masing pelanggan. Hal ini dapat mencakup berbagai karakteristik dan perilaku pelanggan. Oleh karena itu, diperlukan segmentasi pelanggan untuk mengelompokkan pelanggan berdasarkan perilaku dan karakteristiknya. Tujuan dari penelitian ini adalah membandingkan metode clustering untuk mendapatkan metode yang lebih baik dan optimal dalam mengelompokkan cluster untuk segmentasi pelanggan. Dari permasalahan tersebut, peneliti menerapkan metode CRISP-DM dengan focus pada analisis cluster atau pengelompokkan dengan membandingkan algoritma K-Means dan K-Medoids terhadap Analisa segmentasi pelanggan pada mall. Pada penerapan perbandingan metode K-Means dan K-Medoids, digunakan metode elbow untuk menentukan jumlah cluster yang optimal. Hasil dari metode elbow menunjukkan bahwa penggunaan lima cluster untuk metode K-Means dan empat cluster untuk metode K-Medoids merupakan pilihan yang tepat dalam kasus ini. Langkah selanjutnya adalah mencari nilai *Silhouette Coefficient* setiap metode yang digunakan dalam perbandingan untuk

menentukan metode clustering yang lebih optimal. Hasil nilai yang diperoleh dari metode *Silhouette Coefficient* masing-masing metode adalah k-means adalah 0,553 dan k-medoid adalah 0,485, sehingga algoritma pengelompokan segmentasi pelanggan terbaik pada penelitian ini adalah algoritma K-means karena memiliki nilai koefisien siluet maksimum.

Kata Kunci: K-Means, K-Medoids, Mall, Segmentasi, *Silhouette Coefficient*.

1. PENDAHULUAN

Seiring berjalannya waktu perkembangan teknologi semakin pesat tidak terkecuali dalam beberapa sector, Misalnya saja pada sektor korporasi, pada sektor kesehatan, pada sektor pendidikan, pada sektor pemasaran, bahkan pada sektor pemerintahan[1]. Dalam persaingan bisnis, pusat perbelanjaan seperti mall harus senantiasa melakukan inovasi agar tingkat penjualan setiap produknya dapat mencapai target[2].

Pusat perbelanjaan atau mall modern merupakan tempat yang melayani berbagai kebutuhan sehari-hari masyarakat. Hal ini membuat pusat perbelanjaan menjadi sangat populer [3]. Pusat perbelanjaan seperti mall harus senantiasa berupaya memahami kebutuhan pelanggan dan memastikan setiap penjualan produk memenuhi tujuannya [4]. Pengelola mall juga perlu memahami pendapatan masyarakat sekitar, rentang usia pelanggan yang aktif berbelanja, pengeluaran pelanggan per transaksi di dalam mall, dan jenis kelamin. Mampu mempertimbangkan semua kebutuhan ini sangat membantu ketika merencanakan kebijakan mall [2].

Segmentasi pelanggan dapat diartikan sebagai pengelompokan konsumen ke dalam kategori yang berbeda-beda berdasarkan karakteristik umum yang melibatkan seperti usia, jenis kelamin, pendapatan tahunan, lokasi tinggal, kebiasaan pembelian, dan variable lainnya [2]. Hasil segmentasi pelanggan dapat digunakan untuk mengukur nilai pelanggan, sehingga memungkinkan pihak perusahaan untuk mengidentifikasi pelanggan yang menawarkan manfaat signifikan dan pelanggan yang tidak [5]. Segmentasi pelanggan memungkinkan perusahaan memperoleh keuntungan dengan memasarkan produknya secara lebih spesifik berdasarkan preferensi dan kebutuhan pelanggan individu [6].

Dari konteks tersebut, peneliti memutuskan untuk menggunakan teknik data mining untuk

dilakukan klusterisasi, dengan mencari nilai *Silhouette Coefficient* yang optimal menggunakan algoritma K-Medoids dan K-Means. Data mining merupakan salah satu komponen dalam proses *Knowledge Discovery in Database* (KDD). Penambangan data memiliki peran penting dalam memungkinkan kita mengklasifikasikan, meramal, menyimpulkan, dan mengekstrak informasi berguna dari kumpulan data yang besar[2]. Salah satu teknik data mining untuk menemukan pola atau informasi baru adalah clustering. Clustering adalah suatu proses pengelompokan data (objek) sedemikian rupa sehingga objek yang sejenis (berhubungan) berada dalam satu cluster dan objek yang berbeda (tidak berhubungan) berada dalam cluster yang lain [7]. Tujuan dari metode ini adalah untuk meminimalkan jarak (dissimilarity) seluruh data dari centroidnya masing-masing. Untuk memulai *partitionial clustering*, jumlah kluster yang diinginkan harus ditentukan terlebih dahulu [8]. Teknik *Partitionial Clustering* merupakan proses teknik clustering yang diterapkan untuk mengklasifikasikan pengamatan dalam suatu kumpulan data ke dalam kelompok-kelompok berdasarkan kesamaan data [9]. Algoritma K-Means dan K-Medoids adalah contoh dari *partitionial clustering* [9] [10].

Penelitian yang telah dilakukan sebelumnya meliputi studi perbandingan algoritma K-Means dan K-Medoids untuk memetakan hasil produksi buah [11] dalam klasifikasi produksi buah menggunakan *DBI (Davies Bouldin Index)* dalam rangka mendukung dukungan misi Dinas Ketahanan Pangan dan Pertanian di kabupaten Kotawaringin Timur Untuk menghitung dan meningkatkan produksi pertanian khususnya pada bidang budidaya, pada algoritma K-Means memperoleh nilai DBI yang lebih kecil yaitu 0,296, sedangkan algoritma K-Medoids memperoleh 0,507 artinya hasil tersebut menunjukkan bahwa cluster algoritma K-Means mempunyai nilai DBI yang lebih baik kualitasnya dibandingkan algoritma K-Medoids. Pada penelitian yang dilakukan oleh Mufidah

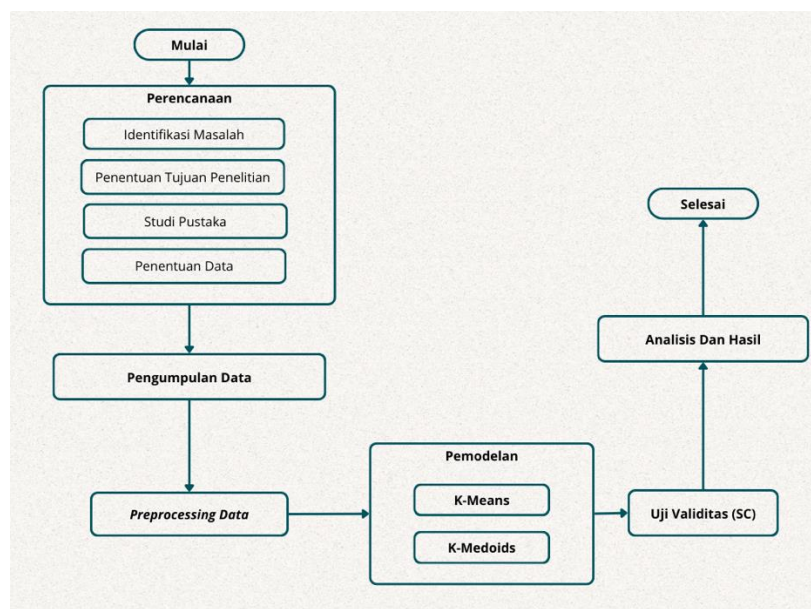
Herviany (dkk) tentang pengelompokan daerah rawan longsor di provinsi Jawa Barat dengan membandingkan algoritma K-Means dan K-Medoids ditemukan bahwa algoritma K-Means lebih optimal dibandingkan algoritma K-Medoids berdasarkan nilai DBI sebesar 0,265 dengan k=6 [12]. Dalam penelitian Nur Awalia Syukri Zainal Abidin (dkk) tentang pengelompokan daerah produksi kakao di provinsi Sulawesi Selatan dengan membandingkan algoritma K-Means dan K-Medoids. Berdasarkan hasil percobaan, algoritma K-Means lebih efektif dibandingkan algoritma K-Medoids dalam mengelompokkan wilayah produksi kakao di provinsi Sulawesi Selatan dengan nilai DBI K-Means sebesar 0,292 dan nilai K-Medoids sebesar 0,365[13].

Dalam era digital saat ini, segmentasi pelanggan menjadi penting dalam meningkatkan efektivitas strategi pemasaran, terutama di pusat perbelanjaan seperti mall. Teknologi data mining telah banyak digunakan dalam analisis segmentasi pelanggan, namun terdapat kesenjangan dalam penelitian yang spesifik mengevaluasi dan membandingkan efektivitas algoritma clustering dalam konteks ini. Studi-studi sebelumnya telah mengeksplorasi penggunaan algoritma tertentu dalam

segmentasi pelanggan, namun masih sedikit penelitian yang secara langsung. Penelitian ini memiliki fokus pada perbandingan algoritma K-Means dan K-Medoids dalam konteks pusat perbelanjaan. Tujuannya adalah untuk mengisi kesenjangan pengetahuan dengan melakukan perbandingan antara kedua algoritma tersebut, dalam hal efisiensi dan efektivitas dalam mengelompokkan pelanggan mall berdasarkan atribut demografis dan perilaku belanja.

2. METODE PENELITIAN

Metode dalam penelitian berperan sebagai sara untuk Menyusun dan mengatur jalannya kegiatan penelitian yang telah ditentukan secara terstruktur. Pada metodologi penelitian melibatkan perancangan alur penelitian yang mencakup seluruh proses dari tahap awal hingga penyelesaian[14]. Fase pertama dari penelitian ini mencakup Analisis Masalah, Pengumpulan Data, Pra-Pemrosesan Data, Proses Pengelompokan dengan menggunakan algoritma K-Means dan K-Medoids, serta Pengujian Akurasi dengan menggunakan metrik kinerja Silhouette Coefficient [15]. Ilustrasi tahapan penelitian ini dapat dilihat dalam Gambar 1 di bawah ini.



Gambar 1. Tahap Penelitian.

2.1. K-Means

Algoritma K-Means dikenal sebagai metode pengelompokan yang populer dan sering digunakan. K-Means juga dianggap sebagai salah satu dari sepuluh algoritma data mining

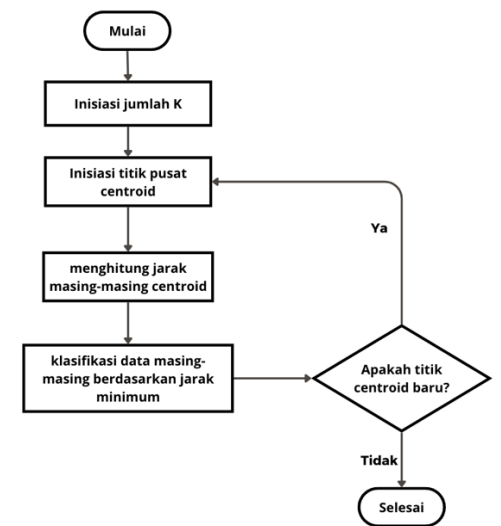
terbaik[16]. K-Means adalah salah satu metode pengelompokan data yang mampu mengelompokkan data ke dalam dua atau lebih kelompok[15]. Dalam metode ini, data yang memiliki karakteristik serupa dimasukkan ke

dalam kelompok yang sama, dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok yang berbeda [17]. Langkah-langkah algoritma K-Means [18] sebagai berikut:

- 1) Tentukan nilai awal kelompok K.
- 2) Tentukan nilai awal K sebagai pusat cluster awal yang dapat dilakukan secara acak.
- 3) Tentukan jarak antara titik-titik setiap objek dan pusat centroid dengan menghitung menggunakan rumus dalam persamaan (1) yang merupakan jarak Euclidean, sampai menemukan jarak terdekat ke nilai centroid untuk kelompok.

$$d = \left(\sum_{i=1}^n (x_i - \mu_i)^2 \right)^{\frac{1}{2}} \quad (1)$$

- 4) Secara acak tentukan pusat kelompok baru dan kemudian lanjutkan untuk menghitung ulang Jarak Euclidean.
- 5) Ulangi langkah b dan langkah c sampai nilai centroid sama dengan rata-rata nilai item kelompok atau sampai nilai kelompok tetap tidak berubah.
- 6) Pembaruan nilai titik centroid dimaksudkan untuk menentukan apakah telah terjadi perubahan pada nilai centroid. Jika tidak ada perubahan, perhitungan centroid dihentikan. Alur algoritma K-Means dapat dilihat pada Gambar 2.



Gambar 2. Proses algoritma K-Means

2.2. K-Medoids

K-Medoids Clustering, juga dikenal sebagai *Partitioning Around Medoids* (PAM), merupakan varian dari metode K-Means. Hal ini didasarkan pada penggunaan k-medoid alih-alih mengamati

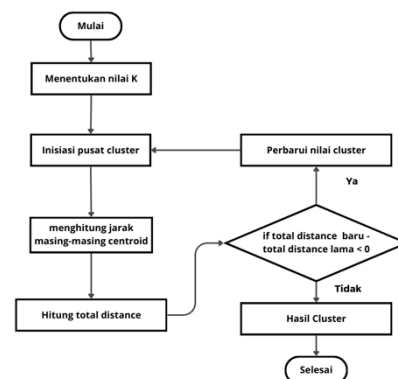
rata-rata yang terkait dengan setiap cluster, dan bertujuan untuk membuat partisi yang dihasilkan kurang sensitif terhadap nilai ekstrem yang ada dalam kumpulan data [19]. Algoritma KMedoids menggunakan mean setiap kelompok sebagai pusat cluster. Algoritma K-Medoids juga menggunakan informasi objek seperti medoids sebagai pusat cluster. Tujuan dari algoritma K-Medoids untuk mengatasi kelemahan perhitungan kmedian, yaitu rentan terhadap outlier karena objek berada jauh dari sebagian besar atribut dalam data. Oleh karena itu, keberadaannya dalam kumpulan data dapat mengganggu rata-rata kluster[20]. Langkah-langkah algoritma K-Medoids[18] sebagai berikut:

- 1) Menentukan nilai jarak K dari setiap pasangan objek berdasarkan pilihan yang berbeda.
- 2) Menghitung persamaan V_j untuk objek j menggunakan rumus berikut: (2):

$$V_j = \sum_{i=1}^n \frac{d_{ij}}{\sum_{i=1}^n d_{ij}} \quad j = 1 \dots n \quad (2)$$

Urutkan V_j dari yang terbesar hingga terkecil, lalu pilih objek k yang memiliki nilai K terkecil sebagai medoid awal.

- 3) Gunakan hasil sebelumnya untuk menginisialisasi kelompok dengan menentukan objek terdekat sebagai medoid.
- 4) Perbarui medoid dari setiap kelompok dengan meminimalkan total jarak dalam kelompok.
- 5) Tetapkan setiap objek ke medoid terdekatnya untuk mendapatkan hasil kelompok dalam tahap kedua. Kemudian hitung jumlah semua jarak antara objek dan medoidnya. Alur algoritma K-Medoids dikenal seperti yang ditunjukkan dalam Gambar 3.



Gambar 3. Proses algoritma K-Medoids

2.3. Silhouette Coefficient

Silhouette Coefficient merupakan salah satu metode analisis untuk mendapatkan validasi nilai pada metode pengelompokan[21]. Evaluasi dan validasi kinerja cluster dengan mengoptimalkan jumlah cluster menggunakan Teknik Silhouette Coefficient. Teknik Silhouette Coefficient berfungsi untuk mengukur akurasi dan kualitas cluster. Untuk menghitung nilai Silhouette Coefficient dapat menggunakan rumus berikut ini [15]:

$$SC = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (3)$$

dimana a(i) adalah jarak rata-rata dari sampel i ke sampel lain dalam suatu cluster, dan b(i) adalah jarak minimum suatu sampel dari sampel i ke cluster lain. Pada table 1 [22] mencerminkan tingkat presisi dalam mengukur Silhouette Coefficient.

Tabel 1. Standar Penilaian *Silhouette Coefficient*

Nilai Silhouette Coefficient	Standar Penilaian
0.7 < Silhouette Coefficient ≤ 1.0	Struktur Kuat
0.5 < Silhouette Coefficient ≤ 0.7 0.25 < Silhouette ≤ 0.5	Struktur Sedang Struktur Lemah
Silhouette ≤ 0.25	Tidak ada Struktur

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil penelitian pengelompokan atau clustering yang membandingkan metode K-Means dan K-Medoids dengan menggunakan bahasa pemrograman Python, penelitian ini diperkuat dengan penggunaan platform Google Colab sebagai alat bantu. Pembahasan penelitian ini meliputi tahap pengumpulan dan preprocessing data, penerapan algoritma DBSCAN dan K-means, serta pengujian dan perbandingan nilai validitas cluster. Isi diskusinya adalah sebagai berikut::

3.1 Pengumpulan Data

Dalam penelitian ini, peneliti menggunakan data yang tersedia untuk umum yang diperoleh dari website kaggle.com. Pada penelitian ini menggunakan 200 data dengan format .csv. Pada data tersebut berisi 5 atribut yang akan digunakan pada penelitian ini yaitu *CustomerID*, *Gender*, *Age*, *AnnualIncome*, *SpendingScore*. Dapat dilihat pada gambar 3.

CustomerID	Genre	Age	AnnualIncome	SpendingScore	
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

Gambar 3. Data segmentation pelanggan

3.2. Preprocessing Data

Pada tahap ini dilakukan *cleaning* data untuk mengurangi noise yang dapat mempengaruhi hasil pengolahan data[12]. Proses pembersihan data dilakukan ketika ditemukan data yang hilang atau nilai yang salah. Hal ini mencakup pengecekan dan koreksi kesalahan serta inkonsistensi data pada saat penulisan data. Jika tidak ada data yang tidak valid, tidak relevan, atau kosong, maka tidak diperlukan langkah pembersihan data. Pada gambar 4 menunjukkan bahwa data yang digunakan tidak mengandung nilai yang kosong atau hilang. Oleh karena itu, tidak ada data yang akan dihapus.

```
[ ] data.isnull().sum()
CustomerID    0
Genre         0
Age           0
AnnualIncome  0
SpendingScore 0
dtype: int64
```

Gambar 4. Mencari *missong value*

Setelah melakukan pengecekan nilai yang *missing value* dari atribut, maka Selain itu, dengan mengevaluasi dimensi data, peneliti dapat menentukan atribut data mana yang harus dan tidak boleh dihapus[23]. Pada gambar 5 menunjukkan atribut *CustomerID* dihapus.

```
data = data.drop('CustomerID', axis=1)
print(data)
```

	Genre	Age	AnnualIncome	SpendingScore
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40
...
195	Female	35	120	79
196	Female	45	126	28
197	Male	32	126	74
198	Male	32	137	18
199	Male	30	137	83

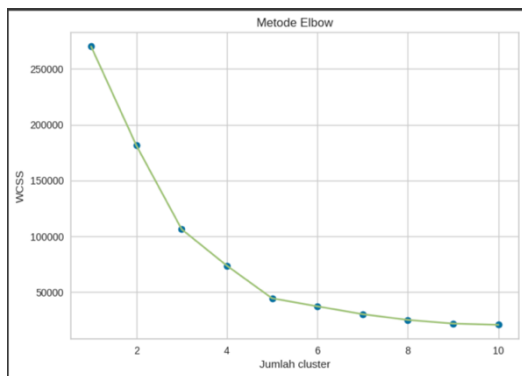
Gambar 5. Menghapus *CustomerID*

3.3. Pemodelan

Selanjutnya adalah fase *modelling*, Pada fase ini, langkah pertama yang dilakukan adalah menentukan jumlah cluster untuk setiap metode. Nantinya akan digunakan untuk proses pengelompokan dan perbandingan antara Metode K-Means dan K-Medoid. Jumlah cluster ditentukan dengan menggunakan metode *elbow* [24].

1) Klasterisasi dengan Algoritma K-Means

Evaluasi metode *elbow* dilakukan dengan menguji 2 hingga 11 cluster. Jumlah cluster dipilih berdasarkan besarnya nilai sudut perbandingan setiap cluster. Pada gambar 6 menunjukkan hasil evaluasi klasterisasi menggunakan metode *elbow* pada algoritma K-Means.



Gambar 6. Grafik cluster K-Means

Lalu nilai mana yang terus berkurang atau bertambah, dan seterusnya. Terlihat pada Gambar 6, nilai antara cluster 4 dan cluster 5 tampak stabil. Data rinci dari Gambar 6 di bawah ini disajikan pada Tabel 2.

Tabel 2. Nilai *elbow* metode K-Means

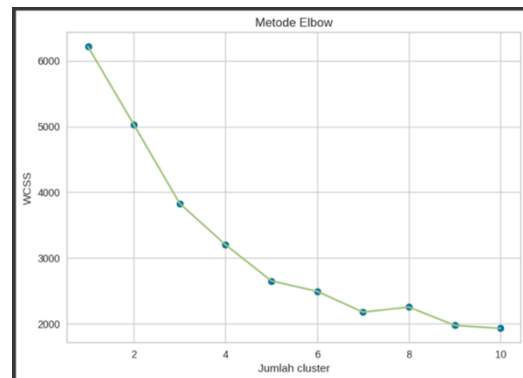
Cluster	Nilai	Selisih
1	269981.280000	88617.684040
2	181363.595960	75015.222897
3	106348.373062	32668.584023
4	73679.789039	29231.333592
5	44448.455448	7182.590243
6	37265.865205	7006.207998
7	30259.657207	5163.953997
8	25095.703210	3265.661232
9	21830.041978	1093.362039

Dari tabel 2 terlihat bahwa cluster 5 merupakan cluster terbaik karena tidak terjadi kenaikan atau penurunan nilai yang signifikan setelah cluster 5. Alternatifnya, perbedaan maksimum

antar cluster diperiksa dan memberikan hasil yang sama. Jadi berdasarkan Tabel 2, cluster 5 merupakan cluster terbaik dengan nilai selisih maksimum sebesar 7182.590243.

2) Klasterisasi dengan algoritma K-Medoids

Tahap selanjutnya klasterisasi algoritma K-Medoids, Dari hasil metode *elbow* yang digunakan untuk *clustering* pada metode K-Means Seperti terlihat pada Gambar 7, dapat disimpulkan bahwa data yang diolah menghasilkan 4 cluster yang optimal.



Gambar 7. Grafik cluster K-Medoids

Dari hasil metode *elbow* diatas, cluster 3 dan 4 tampak stabil. Data rinci dari gambar 7 dibawah ini, disajikan pada tabel 3.

Tabel 3. Nilai *elbow* metode K-Medoids

Cluster	Nilai	Selisih
1	6214.973818	1185.215443
2	5029.758375	1199.528651
3	3830.229725	628.007846
4	3202.221879	550.448215
5	2651.773664	158.243723
6	2493.529941	317.362123
7	2176.167817	78.261387
8	2254.429204	278.800010
9	1975.629194	45.473528

Berdasarkan hasil penggunaan metode *elbow* untuk clustering dengan metode K-Medoids yang ditunjukkan pada Gambar 7 dan perbedaan maksimum antar cluster diperiksa dan memberikan hasil yang sama, dapat disimpulkan bahwa data yang diolah menghasilkan 4 cluster yang optimal.

3.4. Uji Validitas

Teknik silhouette coefficient digunakan dalam langkah evaluasi ini. Tujuannya adalah untuk memeriksa kualitas cluster menggunakan data yang diperoleh sebelumnya pada tahap

pemodelan. Nilai *Silhouette Coefficient* K-means ditunjukkan pada Gambar 8.

```
from sklearn.metrics import silhouette_score
silhouette_score(Seg, y_kmeans)
0.553931997444648
```

Gambar 8. Nilai silhouette coefficient K-Means

Pada pengujian *cluster* menggunakan teknik *silhouette coefficient* pada proses K-Means menghasilkan nilai 0.553. Sedangkan hasil pengujian *cluster* pada proses K-Medoids menghasilkan nilai 0.4843, dapat dilihat pada gambar 9.

```
from sklearn.metrics import silhouette_score
silhouette_score(X, y_kmedoids)
0.4850266531736455
```

Gambar 9. Nilai Silhouette Coefficient K-Medoids

Hasil pada algoritma K-Means yang didapat lebih besar setelah dilakukan pengujian menggunakan teknik *Silhouette Coefficient* yaitu dengan 5 *cluster* senilai 0.553, berdasarkan table 1 mengenai kriteria *Silhouette Coefficient* termasuk dalam Struktur Sedang. Sedangkan untuk pengujian menggunakan K-Medoids termasuk dalam kriteria Struktur Lemah yang menghasilkan 4 *cluster* senilai 0.485. Dengan demikian, hasil perbandingan antara metode K-Means dan K-Medoids menunjukkan bahwa metode K-Means lebih baik, ditandai dengan nilai yang mendekati 1. Hal ini menegaskan bahwa K-Means memiliki keunggulan dibandingkan K-Medoids dalam pengelompokan segmentasi pelanggan di pusat perbelanjaan.

3.5. Analisis dan Hasil

Dari hasil pengolahan data pada tahap sebelumnya, pengetahuan yang diperoleh dari data segmentasi pelanggan di mall akan dikumpulkan dan dikelompokkan menggunakan algoritma K-Means dan K-Medoids. Tujuannya adalah agar informasi ini dapat memberikan manfaat yang signifikan bagi pihak pengelola mall.

Dari hasil analisis menggunakan algoritma K-Means menunjukkan bahwa pelanggan mall dapat dibagi menjadi 5 kelompok atau segmen untuk pemasaran yang ditargetkan, sebagai berikut:

- 1) Cluster 1 merupakan Pelanggan dengan pendapatan rata-rata dan pengeluaran rata-rata. Kelompok ini berhati-hati dalam membelanjakan uangnya saat berbelanja.
- 2) Cluster 2 merupakan Pelanggan berpendapatan tinggi dan belanja tinggi. Kelompok ini berpotensi menghasilkan keuntungan. Diskon dan penawaran lain yang ditargetkan untuk grup ini akan meningkatkan Skor Donor Anda dan memaksimalkan keuntungan Anda.
- 3) Cluster 3 Pelanggan yang mempunyai pendapatan tinggi namun tidak menghabiskan banyak uang di mall. Salah satu asumsinya adalah tidak puas dengan pelayanan mal. Kelompok ini merupakan kelompok sasaran yang ideal bagi tim pemasaran karena dapat mendatangkan keuntungan yang tinggi bagi mal.
- 4) Cluster 4 pelanggan yang berpenghasilan rendah dengan skor pengeluaran rendah. Fenomena ini terjadi karena orang-orang dengan pendapatan rendah akan cenderung membeli lebih sedikit barang di mall.
- 5) Cluster 5 pelanggan berpenghasilan rendah dengan skor pembelanjaan tinggi. Kelompok ini rela menghabiskan pendapatannya untuk berbelanja di mall adalah karena menikmati dan puas dengan layanan yang diberikan selama berbelanja

Sedangkan dari analisis menggunakan algoritma K-Medoids menghasilkan 4 cluster atau segmen, sebagai berikut:

- 1) Cluster 1 merupakan Pelanggan dengan pendapatan rata-rata dan pengeluaran rata-rata. Kelompok ini berhati-hati dalam membelanjakan uangnya saat berbelanja.
- 2) Cluster 2 merupakan Pelanggan berpendapatan tinggi dan belanja tinggi. Kelompok ini berpotensi menghasilkan keuntungan. Diskon dan penawaran lain yang ditargetkan untuk grup ini akan meningkatkan Skor Donor Anda dan memaksimalkan keuntungan Anda.
- 3) Cluster 3 Pelanggan yang mempunyai pendapatan tinggi namun tidak menghabiskan banyak uang di mall. Salah satu asumsinya adalah tidak puas dengan pelayanan mal. Kelompok ini merupakan kelompok sasaran yang ideal bagi tim pemasaran karena dapat mendatangkan keuntungan yang tinggi bagi mal.
- 4) Cluster 4 pelanggan yang berpenghasilan rendah dengan skor pengeluaran rendah.

Fenomena ini terjadi karena orang-orang dengan pendapatan rendah akan cenderung membeli lebih sedikit barang di mall.

Penelitian ini memberikan hasil yang menunjukkan bahwa algoritma K-Means, dengan nilai Silhouette Coefficient yang lebih tinggi (0.553), dibandingkan dengan K-Medoids (0.485), memberikan bukti efektivitasnya dalam pengelompokan segmentasi pelanggan di mall. Namun, Hasil ini membuka pintu untuk diskusi yang lebih mendalam mengenai faktor-faktor yang mempengaruhi efektivitas algoritma dalam berbagai konteks bisnis. Pertama, variabilitas data dan karakteristik demografis pelanggan dapat memengaruhi cara algoritma mengklasifikasikan data. Kedua, aspek seperti interpretasi cluster dan penerapannya dalam strategi pemasaran mall juga penting untuk dipertimbangkan.

Selain itu, penelitian ini membuka diskusi tentang pentingnya memilih algoritma yang tepat sesuai dengan jenis data dan tujuan bisnis. Meskipun K-Means menunjukkan hasil yang lebih baik dalam konteks ini, K-Medoids mungkin lebih sesuai dalam situasi di mana outlier merupakan bagian penting dari analisis. Oleh karena itu, pemilihan algoritma tidak hanya sebatas pada nilai statistik semata, tetapi juga pada pemahaman mendalam tentang data dan kebutuhan bisnis. Dengan memperdalam diskusi ini, penelitian ini tidak hanya menyumbangkan pemahaman tentang efektivitas algoritma tertentu, tetapi juga memberikan insight tentang bagaimana perusahaan bisa menerapkan hasil analisis data dalam strategi pemasaran yang lebih terfokus dan efektif.

Maka hasil dari penelitian ini mendapatkan rekap dari 200 data segmentasi pelanggan pada mall yang sudah melewati tahap CRISP-DM (*Cross Industry Standard Process For Data Mining*) telah menghasilkan *cluster* dan kualitas *cluster* pada algoritma K-Means dan algoritma K-Medoids. Pada uji coba *clustering* dengan algoritma K-Means mendapatkan nilai 0.553 dengan 5 *cluster*, sedangkan untuk algoritma K-Medoids mendapatkan nilai 4 *cluster* dengan nilai 0.483. Hal ini menandakan bahwa penggunaan algoritma K-Means lebih baik untuk segmentasi pada mall karena menghasilkan nilai *Silhouette Coefficient* yang mendekati angka 1.

4. KESIMPULAN

Penelitian dilakukan dengan menggunakan teknik K-Means dan K-Medoids untuk mengelompokkan segmentasi pelanggan pusat perbelanjaan menggunakan tools Google Collabs dan bahasa pemrograman *Python*. Dimana, dalam pengujian tersebut peneliti menggunakan metode *elbow* yang menghasilkan K=5 untuk algoritma K-Means dan K=4 untuk algoritma K-Medoids. Selanjutnya peneliti membandingkan nilai *Silhouette Coefficient* pada kedua algoritma tersebut dengan menggunakan 200 data. Sehingga kesimpulan dalam menentukan nilai terbaik jumlah Klaster (K) bahwa 5 *cluster* dengan menunjukkan nilai sebesar 0.553. Jika nilai *Silhouette Coefficient* semakin besar atau mendekati nilai 1, maka hasil yang didapatkan lebih optimal. Hal yang didapatkan algoritma K-Means mendapatkan hasil nilai *Silhouette Coefficient* yang lebih besar dengan nilai 0.553, sedangkan hasil algoritma K-Medoids dengan nilai *Silhouette Coefficient* sebesar 0.485. Berdasarkan nilai *Silhouette Coefficient* yang diperoleh, algoritma K-means merupakan algoritma yang paling cocok untuk clustering segmentasi pelanggan pusat perbelanjaan. Berdasarkan hasil penelitian tersebut, algoritma K-Means dipilih untuk mengelompokkan pelanggan dengan lebih baik sehingga pengelola mal lebih mudah memperoleh pengetahuan terkait strategi penjualan produk. Penelitian selanjutnya sebaiknya mengembangkan lebih banyak atribut dan sampel data serta membandingkannya dengan algoritma apriori.

DAFTAR PUSTAKA

- [1] R. Wahyusari and S. Wardani, "Comparison of the K-Means Algorithm and the K-Medoid Algorithm for Clustering UMKM in Kebumen," 2023.
- [2] Y. Hapsari *et al.*, "Analisis Segmentasi Pelanggan Mall Menggunakan Algoritma K-Means," 2023. [Online]. Available: <https://www.kaggle.com/code/obrunet/customer-segmentation-k-means/>
- [3] M. Lestari, D. Andarini, A. Camelia, P. Fujianti, D. Arista Putri, and R. Faliria Nandini, "Penerapan Tanggap Darurat Pada Pengunjung Salah Satu Mall di Kota Palembang," 2021. [Online]. Available: <http://ejournal.uika->

- bogor.ac.id/index.php/Hearty/issue/archive
- [4] S. Justine and M. A. Pribadi, "Strategi Komunikasi Pemasaran Pusat Perbelanjaan di Jakarta (Studi Kasus Emporium Pluit Mall)," 2023.
- [5] A. Febriani and S. A. Putri, "Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means," *Jiems (Journal of Industrial Engineering and Management Systems)*, vol. 13, no. 2, Sep. 2020, doi: 10.30813/jiems.v13i2.2274.
- [6] G. Purnama, T. Pudjiantoro, and P. Sabrina, "Segmentasi Pelanggan Menggunakan K-Medoids Berdasarkan Model Length, Recency, Frequency, Monetary (LRFM)," vol. 5, 2021.
- [7] A. S. S. Khan, M. Fatekurohman, and Y. S. Dewi, "Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember," *Jurnal Statistika dan Komputasi*, vol. 2, no. 2, pp. 67–75, Dec. 2023, doi: 10.32665/statkom.v2i2.2301.
- [8] Z. Neng and A. Hadiana, "Kajian Data Mining Menggunakan Algoritma K-Means Dan K-Medoids Dalam Strategi Promosi (Studi Kasus: Universitas Islam Al-Ihya Kuningan)," 2021.
- [9] Adeline Vinda Septiani, R. A. Hasibuan, Anwar Fitrianto, Erfiani, and Alfa Nugraha Pradana, "Penerapan Metode K-Medoids dalam Pengklasteran Kab/Kota di Provinsi Jawa Barat Berdasarkan Intensitas Bencana Alam di Jawa Barat pada Tahun 2020-2021," *Statistika*, vol. 23, no. 2, pp. 147–155, Nov. 2023, doi: 10.29313/statistika.v23i2.3057.
- [10] L. Febby Olivia *et al.*, "Penerapan Metode K-Means Clustering dalam Klasterisasi Pemilihan Pasir Sesuai SNI," *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 3, no. 3, 2023, doi: 10.55606/juisik.v3i3.665.
- [11] E. Prasetyaningrum and P. Susanti, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, pp. 1775–1783, 2023, doi: 10.30865/mib.v7i4.6477.
- [12] M. Herviany, S. Putri Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, pp. 34–40, 2021.
- [13] N. A. S. Z. Abidin, R. D. Avila, A. Hermatyar, and R. Rismayani, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Produksi Kakao," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, Aug. 2022, doi: 10.28932/jutisi.v8i2.4897.
- [14] D. Dwi Aulia and N. Nurahman, "Comparison Performance of K-Medoids and K-Means Algorithms In Clustering Community Education Levels," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 2, pp. 273–282, Jul. 2023, doi: 10.23887/janapati.v12i2.59789.
- [15] R. Adha, N. Nurhaliza, and U. Soleha, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *Jurnal Sains, Teknologi dan Industri*, vol. 18, no. 2, pp. 206–211, 2021, [Online]. Available: <https://covid19.who.int>.
- [16] W. Kwedlo and M. Lubowicz, "Accelerated K-Means Algorithms for Low-Dimensional Data on Parallel Shared-Memory Systems," *IEEE Access*, vol. 9, pp. 74286–74301, 2021, doi: 10.1109/ACCESS.2021.3080821.
- [17] E. Herman, K. E. Zsido, and V. Fenyves, "Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation," *Applied Sciences (Switzerland)*, vol. 12, no. 16, Aug. 2022, doi: 10.3390/app12167985.
- [18] N. Fitrianti Fahrudin and R. Rindiyan, "Comparison of K-Medoids and K-Means Algorithms in Segmenting Customers based on RFM Criteria," *E3S Web of Conferences*, vol. 484, p. 02008, Feb. 2024, doi: 10.1051/e3sconf/202448402008.
- [19] M. Adepeju, S. Langton, and J. Bannister, "Anchored k-medoids: a novel adaptation of k-medoids further refined to measure long-term instability in the

- exposure to crime," *J Comput Soc Sci*, vol. 4, no. 2, pp. 655–680, Nov. 2021, doi: 10.1007/s42001-021-00103-1.
- [20] M. Arifandi, A. Hermawan, and D. Avianto, "Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid19 DKI Jakarta," *Jurnal Teknologi Terapan*], vol. 7, no. 2, 2021.
- [21] H. Tussyakdiah *et al.*, "Implementasi metode K-Means dan K-Medoids Pada Pengelompokan Provinsi Indonesia Berdasarkan Aspek Pendidikan Pemuda," vol. 3, 2023.
- [22] N. Fransiska, D. Anggraeni, and U. Enri, "Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *Jurnal Teknologi Informasi Komunikasi*, vol. 9, pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.921.
- [23] Y. Christian and K. O. Y. R. Qi, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 966, Aug. 2022, doi: 10.30865/jurikom.v9i4.4486.
- [24] H. Alexander, Y. Umaidah, and M. Jajuli, "Implementasi Clustering Untuk Menentukan Efektivitas Nilai Siswa Sesudah pandemi Covid-19 Algoritma K-Means," 2023.