JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Penerapan Data Mining Dalam Pengelompokan Wilayah Toko Penjualan Air Mineral Kemasan Menggunakan Algoritma K-Means Clustering

Feby Desryani Br Ginting [1], Mukhlis Ramadhan [2], Rina mahyuni[3]
[1,2,3] Sistem Informasi , Stmik Triguna Dharma
Email: [1] febyginting67@gmail.com, [2]mukhlisramadhan.tgd@gmail.com, [3]rinamahyuni14@gmail.com
Email Penulis Korespondensi: febyginting67@gmail.com

Abstrak
Distribusi yang tepat menjadi salah satu hal penting dalam mendukung keberhasilan penjualan produk, termasuk dalam industri air mineral kemasan. PT. Tirta Alpin Makmur (Amoz) sebagai produsen air mineral perlu mengetahui wilayah toko mana saja yang memiliki potensi tinggi agar strategi distribusi dan pemasaran bisa berjalan lebih efektif dan efisien. Untuk itu, dibutuhkan pengelompokan wilayah berdasarkan data penjualan yang tersedia. Penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan wilayah toko berdasarkan tiga atribut, yaitu jumlah toko distribusi, jumlah penjualan per bulan, dan frekuensi pemesanan. Data diambil dari 30 desa yang berada di Kecamatan Namorambe dan kemudian diolah secara iteratif untuk mendapatkan hasil pengelompokan yang optimal. Dari proses tersebut, terbentuk tiga kelompok wilayah yaitu: wilayah dengan permintaan rendah, permintaan sedang, dan permintaan tinggi. Hasil ini memberikan gambaran yang lebih jelas mengenai karakteristik masing-masing wilayah toko yang dapat menjadi pertimbangan dalam pengambilan keputusan perusahaan. Dengan adanya pengelompokan ini, diharapkan PT. Tirta Alpin Makmur (Amoz) dapat merancang strategi distribusi dan promosi yang lebih tepat sasaran sesuai dengan kondisi wilayah toko. Selain itu, metode ini juga bisa dijadikan sebagai acuan dalam penerapan analisis data untuk kebutuhan distribusi di masa mendatang.

Kata Kunci: Data mining, K-Means Clustering , Pengelompokan Wilayah Toko, Air Mineral Kemasan, Distribusi Penjualan.

Abstract
An appropriate distribution strategy is one of the key factors in supporting the success of product sales, including in the bottled mineral water industry. PT. Tirta Alpin Makmur (Amoz), as a producer of bottled water, needs to identify which store areas have high potential so that distribution and marketing strategies can be carried out more effectively and efficiently. Therefore, it is necessary to group regions based on the available sales data. This study uses the K-Means Clustering method to group store areas based on three attributes: the number of distribution stores, monthly sales volume, and order frequency. The data were collected from 30 villages located in the Namorambe District and then processed iteratively to obtain optimal clustering results. From this process, three regional groups were formed: areas with low demand, medium demand, and high demand. These results provide a clearer picture of the characteristics of each store area, which can be considered in the company's decision-making process. With this clustering, it is expected that PT. Tirta Alpin Makmur (Amoz) will be able to design more targeted distribution and promotion strategies according to the conditions of each store area. In addition, this method can also serve as a reference for future data analysis in distribution planning.

Keywords: *Data mining, K-Means Clustering, Store Area Clustering, Bottled Mineral Water, Sales Distribution.*

1. PENDAHULUAN

Data mining adalah proses mendapatkan informasi baru dari kumpulan data dengan menggunakan algoritma dan teknik dari ilmu statistik, mesin pembelajaran, dan sistem manajemen database. Ini digunakan untuk mengekstraksi informasi penting yang tersembunyi dari dataset yang sangat besar. Proses ini menghasilkan permata pengetahuan dari kumpulan data yang sangat besar[1].

Algoritma K-Means adalah salah satu teknik clustering yang digunakan dalam data mining, yang berfungsi untuk mempartisi data menjadi kelompok-kelompok berdasarkan kesamaan. Prosesnya melibatkan pembagian data ke dalam cluster yang memiliki kesamaan dan memisahkan data yang tidak sama ke dalam kluster yang berbeda. K-Means populer karena mudah dipahami dan diterapkan, serta memiliki efek pengelompokan yang baik. Namun, algoritma ini juga memiliki kekurangan, yaitu ketergantungan yang kuat pada pemilihan pusat cluster awal[2].

Air minum dalam kemasan adalah air minum yang telah diproses dan dikemas dalam berbagai bentuk, seperti botol, galon, atau kemasan lainnya, sehingga aman untuk dikonsumsi langsung. Menurut Standar Industri Indonesia (SII) 2040-90, didefinisikan sebagai air yang telah diproses, dikemas, dan aman untuk diminum langsung[3].

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

2. METODOLOGI PENELITIAN

2.1 Teknik Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang telah ditentukan sebelumnya pada proses pendahuluan. Proses pengumpulan data terbagi menjadi 2 tahap yaitu dijabarkan sebagai berikut:

a. Studi literatur
Studi literatur yaitu dasar teori yang digunakan dalam penelitian untuk menyelesaikan permasalahan dan merupakan referensi yang kuat dalam melakukan analisis. Studi literatur yang dilakukan adalah memilih, menetapkan dan mempersiapkan daftar pustaka berupa buku, jurnal sebagai landasan teori.

b. Observasi dan Wawancara
Kegiatan wawancara dilakukan secara langsung dengan pihak terkait di PT. Tirta Alpin Makmur yang berlokasi di Ujung Labuhan, Kec. Namorambe, Kab. Deli Serdang , Sumatera Utara.

Tabel 2 Data Wilayah Toko Penjualan Air Mineral Kemasan

No	Nama Wilayah	Jumlah Toko Distribusi	Jumlah Penjualan (pallet/bulan)	Frekuensi Pemesanan (kali/bulan)
1	Desa Batu Gemuk	5	12	7
2	Desa Batu Mbelin	6	14	8
3	Desa Batu Penjemuran	8	16	9
4	Desa Batu Rejo	6	13	8
5	Desa Bekukul	5	9	6
6	Desa Cinta Rakyat	7	18	10
7	Desa Gunung Berita	6	12	7
8	Desa Gunung Kelawas	5	9	6
9	Desa Jaba	6	12	7
10	Desa Jati Kesuma	8	17	9
11	Desa Kuta Tualah	7	15	8
12	Desa Kwala Simeme	6	13	8
13	Desa Lubang Ido	5	11	7
14	Desa Namo Batang	6	14	8
15	Desa Namo Landur	5	9	6
16	Desa Namo Mbaru	5	11	7
17	Desa Tangkahan Baru	6	12	7
18	Desa Namo Pakam	6	14	8
19	Desa Namo Pinang	5	11	7
20	Desa Rimo Mungkur	5	9	6
21	Desa Namorambe	9	21	11
22	Desa Rumah Mbacang	5	9	6
23	Desa Salang Tungir	5	10	7
24	Desa Sudirejo	6	12	7
25	Desa Suka Mulia Hilir	5	13	8
26	Desa Suka Mulia Hulu	6	13	8
27	Desa Tanjung Selamat	5	11	7
28	Desa Ujung Labuhan	8	19	10
29	Desa Uruk Gedang	5	11	7
30	Desa Kuta Tengah	7	14	8

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

2.2 Data Mining

Data mining adalah proses mendapatkan informasi baru dari kumpulan data dengan menggunakan algoritma dan teknik dari ilmu statistik, mesin pembelajaran, dan sistem manajemen *database*. Ini digunakan untuk mengekstraksi informasi penting yang tersembunyi dari dataset yang besar. Proses ini menghasilkan permata pengetahuan dari kumpulan data yang sangat besar. Data mining adalah analisis data untuk menemukan hubungan yang jelas dan menghasilkan kesimpulan yang belum diketahui sebelumnya sehingga pemilik data dapat memahaminya dengan mudah[4].

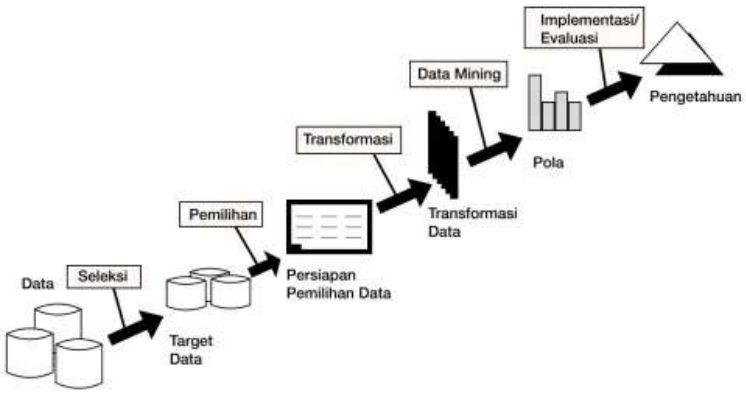
Data mining merupakan proses yang secara otomatis menggunakan satu atau lebih metode pembelajaran komputer untuk menganalisis dan menghasilkan pengetahuan. Knowledge Discovery in Databases (KDD) adalah pendekatan saintifik untuk data mining. Dalam kasus ini, data mining adalah merupakan satu langkah dari proses KDD, yang terdiri dari proses tahap berikut[5]:

1. Pembersihan data adalah proses menghilangkan data yang tidak konsisten dan *noise*.
2. Integrasi data adalah proses menggabungkan data dari berbagai sumber.
3. Transformasi data adalah proses mengubah data menjadi bentuk yang sesuai untuk mining.
4. Penggunaan teknik data mining, yang merupakan proses mengekstrak pola dari data yang ada.
5. Evaluasi pola yang ditemukan, yaitu proses menginterpretasikan pola menjadi pengetahuan yang dapat digunakan untuk membantu pengambilan keputusan.
6. Presentasi pengetahuan, yang merupakan proses menyampaikan pengetahuan melalui teknik visualisasi.

Tahap ini merupakan bagian dari proses pencarian pengetahuan, yang mencakup menentukan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Presentasi informasi dalam format yang mudah dipahami pengguna adalah langkah terakhir dalam KDD.

2.2.1 Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) merupakan proses Analisa terstruktur untuk memperoleh informasi yang benar, baru, bermanfaat dan menemukan pola dari data yang besar dan komplek. Data Mining menjadi inti dari proses KDD, yakni dengan menggunakan algoritma tertentu untuk mengeksplorasi data, membangun model dan menemukan pola yang belum diketahui. Model digunakan untuk memahami fenomena data, Analisa maupun prediksi[6].



Gambar 1 Proses *Knowleadge Discovery Database*

Pada proses tahapan perancangan data ini berdasarkan *Knowledge Discovery in Database* (KDD) adalah proses analisa terstruktur untuk memperoleh informasi yang benar, baru, bermanfaat dan menemukan pola dari data yang besar dan komplek. Berikut ini ada beberapa tahapan *Knowledge Discovery in Database* antara lain[7]:

1. Seleksi Data (*Selection*)
   Proses pembacaan dataset dalam format *file Excel* dilakukan dengan memanfaatkan operator *Read Excel* dari lingkungan kerja yang digunakan.
2. Pemilihan Data (*Prepocessing/Cleaning*)
   Proses *cleansing* merupakan tahapan pembersihan data dari nilai yang hilang *(missing)* atau tidak konsisten dalam tahap *preprocessing*. Sebelum tahap ini dilakukan, terlebih dahulu dilakukan analisis untuk memastikan apakah *dataset* mengandung data yang hilang atau inkonsistensi nilai.
3. Data transformasi
   Transformasi data merupakan tahapan di mana data yang telah dipilih diubah ke dalam format yang sesuai untuk proses data mining. Pada tahap ini, data dengan tipe *polinomial* dikonversi menjadi bentuk numerik agar dapat dianalisis berdasarkan perhitungan jarak.
4. Data mining.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Data mining merupakan proses eksplorasi dan analisis terhadap sejumlah besar data dengan tujuan untuk menemukan pola tertentu. Pemilihan teknik, metode, atau algoritma yang digunakan sangat bergantung pada tujuan serta keseluruhan *proses knowledge discovery in databases* (KDD). Tahapan ini menjadi inti dari proses KDD, di mana analisis dilakukan terhadap data yang telah melalui proses pembersihan sebelumnya.

5. *Evaluation*
   Tahap interpretasi bertujuan untuk menerjemahkan pola atau alur informasi yang dihasilkan dari proses data mining ke dalam bentuk yang mudah dipahami oleh pihak yang berkepentingan. Langkah ini merupakan bagian dari proses KDD yang juga mencakup verifikasi apakah informasi yang ditemukan sesuai atau bertentangan dengan fakta maupun hipotesis sebelumnya.

2.2.1 Metode Data Mining

Pada umumnya metode data mining dapat dikelompokkan ke dalam dua kategori yaitu: *deskriptif* dan *prediktif*. Metode deskriptif berfokus pada pemahaman karakteristik dan pola yang terdapat dalam data yang diamati
Berikut merupakan jenis metode - metode yang ada dalam data mining adalah sebagai berikut.

1. *Classification*
   Klasifikasi adalah proses mengelompokkan setiap entitas data berdasarkan sekumpulan atribut, termasuk atribut kelas *(class attribute).* Tujuan dari metode ini adalah membangun sebuah model atau fungsi yang mampu menggambarkan serta membedakan masing-masing kelas data atau konsep, sehingga model tersebut dapat digunakan untuk memprediksi kelas dari data yang belum diketahui labelnya [8].
2. *Clustering*
   Pengelompokan (*Clustering*) merupakan teknik dalam data mining yang berfungsi untuk mengelompokkan sekumpulan objek atau data ke dalam grup (*cluster*) berdasarkan kemiripan karakteristik antar data tersebut.
3. *Assosiation*
   Metode ini bertujuan untuk menghasilkan sejumlah aturan *(rule)* yang mampu menggambarkan hubungan kuat antar data. *Association* digunakan, salah satunya, untuk mengidentifikasi pola pembelian barang yang dilakukan secara bersamaan oleh pelanggan.
4. Regression
   Regression adalah metode untuk memprediksi nilai variabel kontinyu berdasarkan variabel lain, dengan asumsi hubungan *linear* atau *nonlinear*.[9].
5. *Forecasting*
   Prediksi (*Forecasting*) berfungsi untuk melakukan prediksi kejadian yang akan datang berdasarkan data sejarah yang ada. Metode ini sering digunakan dalam analisis tren dan peramalan permintaan produk.
6. *Sequence Analysis*
   Metode ini berfungsi untuk menganalisis urutan peristiwa atau kejadian dalam dataset, membantu dalam memahami pola temporal dan hubungan antar kejadian.
7. *Deviation Analysis*
   Metode ini bertujuan untuk mengidentifikasi penyebab adanya perbedaan atau penyimpangan antar data, dan sering kali disebut juga sebagai deteksi *outlier*. Contohnya adalah mendeteksi kemungkinan terjadinya penipuan pada pengguna kartu kredit dengan meninjau riwayat transaksi yang tercatat dalam basis data perusahaan.

2.2.2 Clustering

Clustering adalah proses pengelompokan data yang serupa kedalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam sub set, sehingga data dalam setiap sub set mempunyai arti yang bermanfaat, yang mana dalam *cluster* terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan benda yang terdapat pada *cluster* lainnya[10].

2.3 Algoritma K-Means Clustering

K-Means adalah algoritma yang populer dalam dunia riset dan masuk kedalam kelompok *unsupervised learning* yang dilihat dari kemiripan data yang dimiliki. Metode *K-means clustering* merupakan metode pengelompokan data iteratif yang melakukan partisi set data ke dalam sejumlah k *cluster* yang sudah ditentukan. *K-means clustering* bersifat sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi dan umum digunakan. Berikut ini persamaan yang dapat digunakan untuk pengukuran jarak[11]:

$$d(x,y) = \left|\left|x-y\right|\right|^2 = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Dimana:
d= *Euclidian Distance*
x= data
y= Pusat *cluster*

Algoritma ini bertujuan untuk memaksimalkan perbedaan antar kluster serta meminimalkan nilai fungsi objektif di dalam masing-masing kluster. Salah satu keunggulan *dari K-Means* adalah kemudahannya dalam proses implementasi. Secara matematis, rumus *K-Means* adalah sebagai berikut[12]:

$$Ci = \frac{1}{M}\sum_{j=1}^{M} x_j$$

Berdasarkan rumus yang ditampilkan di atas, Ci merepresentasikan *centroid* untuk fitur ke-1, M menunjukkan jumlah data dalam satu kluster, sedangkan i mengacu pada fitur ke-i dalam kelompok tersebut. Jika masih terjadi perubahan posisi data antar kluster, maka proses iterasi akan terus dilanjutkan ke tahap berikutnya

K-Means merupakan salah satu metode non hirarki *clustering* yang mencoba membagi data yang ada menjadi satu atau lebih *cluster*. Algoritma mengelompokkan data ke dalam suatu *cluster* berdasarkan kemiripan atau kesamaan karakteristik data sedangkan data dengan karakteristik yang tidak mirip atau berbeda dikelompokkan ke dalam *cluster* yang lain[13]. Tahap ini merupakan proses penyelesaian dengan menerapkan metode *K-Means Clustering* menggunakan rumus yang telah ditentukan sebelumnya. Seperti yang telah dijelaskan pada Bab Pendahuluan, tujuan dari penelitian ini adalah untuk melakukan pengelompokan data wilayah toko penjualan air mineral kemasan pada PT.Tirta Alpin Makmur (Amoz) menggunakan *K-Means Clustering.* Proses perhitungan data secara keseluruhan menggunakan *software MS-Excel* untuk mengurangi dampak *human eror* yang terjadi saat perhitungan manual.

Tabel 3 Data *Centroid* Awal

Centroid	**Nama Wilayah**	**Jumlah Toko**	**Jumlah Penjualan (Pallet/bulan)**	**Frekuensi Pemesanan (kali/bulan)**
C1	Desa Batu Gemuk	5	12	7
C2	Desa Namo Pakam	6	14	8
C3	Desa Ujung Labuhan	8	19	10

a. Menghitung Jarak Ke Masing- masinng *Centroid*

Tahap selanjutnya yaitu menghitungn jarak ke masing- masing *centroid*, untuk menghitung jarak antara data dengan *centroid* digunakan rumus *Eucliden Distance* sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{r}(xi - yi)^2}$$

dimana:
d(*x*,*y*) = jarak data *x* ke *y*
xi = Nilai fitur ke-i dari *x*
yi = Nilai fitur ke-I dari *y*
r = Jumlah fitur dalam vektor

Berikut contoh perhitungan jarak dari data wilayah 1 (WL1) iterasi ke-1 terhadap pusat *Cluster* :

$C1_{WL1} = \sqrt{(5-5)^2 + (12-12)^2 + (7-7)^2}$
$= 0{,}00$
$C2_{WL1} = \sqrt{(5-6)^2 + (12-14)^2 + (7-8)^2}$
$= 2{,}45$
$C3_{WL1} = \sqrt{(5-8)^2 + (12-19)^2 + (7-10)^2}$
$= 8{,}19$

Berikut contoh perhitungan jarak dari data wilayah 2 (WL2) iterasi ke-1 terhadap pusat *Cluster* :

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

$C1_{WL2} = \sqrt{(5-6)^2 + (12-14)^2 + (7-8)^2}$
$= 2{,}45$
$C2_{WL2} = \sqrt{(6-6)^2 + (14-14)^2 + (8-8)^2}$
$= 0{,}00$
$C3_{WL2} = \sqrt{(8-6)^2 + (14-19)^2 + (8-10)^2}$
$= 5{,}74$

Berikut contoh perhitungan jarak dari data wilayah 6 (WL6) iterasi ke-1 terhadap pusat *Cluster* :

$C1_{WL6} = \sqrt{(7-5)^2 + (18-12)^2 + (10-7)^2}$
$= 7{,}00$
$C2_{WL6} = \sqrt{(7-6)^2 + (18-14)^2 + (10-8)^2}$
$= 4{,}58$
$C3_{WL6} = \sqrt{(7-8)^2 + (18-19)^2 + (10-10)^2}$
$= 1{,}41$

Dan seterusnya dilanjutkan perhitungan untuk data selanjutnya. Lalu kemudian akan didapatkan jarak. Dapat dilihat pada tabel dibawah ini.

Tabel 4 Data Nilai Jarak ke *Centroid* iterasi 1

No	Nama Wilayah	C1	C2	C3	Jarak terdekat	Cluster
1	Desa Batu Gemuk	0,00	2,45	8,19	0,00	*cluster 1*
2	Desa Batu Mbelin	2,45	0,00	5,74	0,00	*cluster 2*
3	Desa Batu Penjemuran	5,39	3,00	3,16	3,00	*cluster 2*
4	Desa Batu Rejo	1,73	1,00	6,63	1,00	*cluster 2*
5	Desa Bekukul	3,16	5,48	11,18	3,16	*cluster 1*
6	Desa Cinta Rakyat	7,00	4,58	1,41	1,41	*cluster 3*
7	Desa Gunung Berita	1,00	2,24	7,87	1,00	*cluster 1*
8	Desa Gunung Kelawas	3,16	5,48	11,18	3,16	*cluster 1*
9	Desa Jaba	1,00	2,24	7,87	1,00	*cluster 1*
10	Desa Jati Kesuma	6,16	3,74	2,24	2,24	*cluster 3*
…	…	…	…	…	…	...
30	Desa Kuta Tengah	3,00	1,00	5,48	1,00	*cluster 2*

Untuk membandingkan hasil perhitungan, jarak terdekat antara data pusat *cluster* akan dipilih, jarak ini menunjukkan bahwa data tersebut termasuk dalam kelompok dengan pusat *cluster* terdekat. Tabel berikut menunjukkan pengelompokan data.

Tabel 5 Pengelompokan data hasil jarak *Centroid* iterasi 1

No	Nama Wilayah	C1	C2	C3
1	Desa Batu Gemuk	**1**		
2	Desa Batu Mbelin		**1**	
3	Desa Batu Penjemuran		1	
4	Desa Batu Rejo		1	
5	Desa Bekukul	1		
6	Desa Cinta Rakyat			**1**

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

7	Desa Gunung Berita	1		
8	Desa Gunung Kelawas	1		
9	Desa Jaba	1		
10	Desa Jati Kesuma			1
…	…	…	…	…
30	Desa Kuta Tengah		1	

Setelah diperoleh anggota dari masing-masing *cluster*, tahap berikutnya adalah menghitung *centroid* baru dengan menggunakan persamaan berikut:

$$Cj = \frac{1}{Nk}\sum_{l=1}^{Nk} x_{jl}$$

Dimana :
Cj = *Centroid* baru
Xjl = Anggota *Cluster* pada atribut
Nk = Jumlah data dalam *cluster*

Dengan menggunakan persamaan tersebut, diperoleh nilai *centroid* terbaru untuk atribut pertama, yaitu:

$C1_{k1} = 5 + 5 + 6 + 5 + 6 + 5 + 5 + 5 + 6 + 5 + 5 + 5 + 5 + 6 + 5 + 5 + 5 = \frac{89}{17} = 5{,}24$

$C1_{k2} = 12 + 9 + 12 + 9 + 12 + 11 + 9 + 11 + 12 + 11 + 9 + 9 + 10 + 12 + 13 + 11 + 11 = \frac{183}{17} = 10{,}76$

$C1_{k3} = 7 + 6 + 7 + 6 + 7 + 7 + 6 + 7 + 7 + 7 + 6 + 6 + 7 + 7 + 8 + 7 + 7 = \frac{115}{17} = 6{,}76$

Langkah perhitungan yang sama diterapkan hingga data pada *centroid* C2 dan C3, sehingga diperoleh rincian hasil sebagai berikut:

Tabel 6 Nilai *Centroid* Baru iterasi ke-1

Centroid	Jumlah Toko	Jumlah Penjualan	Frekuensi Pemesanan
C1	5,24	10,76	6,76
C2	6,44	14,00	8,11
C3	8,00	18,75	10,00

Karena terdapat perubahan nilai *centroid* dan hasil pengelompokan iterasi ke-1 maka akan dilakukan iterasi ke-2. Langkah selanjutnya adalah mengulangi perhitungan yang sama dengan langkah sebelumnya, namun, kali ini nilai pusat data baru *(centroid)* digunakan untuk perhitungan baru.

Tabel 7 Data Nilai Jarak ke *Centroid* iterasi ke-2

No	Nama Wilayah	C1	C2	C3	Jarak terdekat	Cluster
1	Desa Batu Gemuk	1,28	2,71	7,97	1,28	*cluster 1*
2	Desa Batu Mbelin	3,55	0,46	5,53	0,46	*cluster 2*
3	Desa Batu Penjemuran	6,33	2,69	2,93	2,69	*cluster 2*
4	Desa Batu Rejo	2,67	1,10	6,41	1,10	*cluster 2*
5	Desa Bekukul	1,94	5,62	10,96	1,94	*cluster 1*
6	Desa Cinta Rakyat	8,12	4,46	1,25	1,25	*cluster 3*
7	Desa Gunung Berita	1,47	2,33	7,65	1,47	*cluster 1*
8	Desa Gunung Kelawas	1,94	5,62	10,96	1,94	*cluster 1*
9	Desa Jaba	1,47	2,33	7,65	1,47	*cluster 1*
10	Desa Jati Kesuma	7,18	3,49	2,02	2,02	*cluster 3*

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

…	…	…	…	…	…	…
30	Desa Kuta Tengah	3,89	0,57	5,25	0,57	*cluster 2*

Perbandingan jarak hasil perhitungan pada iterasi kedua dilakukan untuk menentukan kedekatan tiap data terhadap pusat *cluster*. Proses pengelompokan data ditampilkan dalam tabel di bawah ini :

Tabel 8 Pengelompokan data hasil jarak *Centroid* iterasi ke-2

No	Nama Wilayah	C1	C2	C3
1	Desa Batu Gemuk	1		
2	Desa Batu Mbelin		1	
3	Desa Batu Penjemuran		1	
4	Desa Batu Rejo		1	
5	Desa Bekukul	1		
6	Desa Cinta Rakyat			1
7	Desa Gunung Berita	1		
8	Desa Gunung Kelawas	1		
9	Desa Jaba	1		
10	Desa Jati Kesuma			1
…	…	…	…	…
30	Desa Kuta Tengah		1	

kemudian setelah mengetahui anggota tiap *cluster*, persamaan berikut ini digunakan untuk menemukan *centroid* baru:

$$Cj = \frac{1}{Nk}\sum_{l=1}^{Nk} x_{jl}$$

Dimana :
Cj = *Centroid* baru
Xjl = Anggota *Cluster* pada atribut
Nk = Jumlah data dalam *cluster*
Dengan menggunakan persamaan tersebut, diperoleh nilai *centroid* terbaru untuk atribut kedua, yaitu:

$C2_{k1} = \frac{6+8+6+7+6+6+6+5+6+7}{10} = 6{,}30$

$C2_{k2} = \frac{14+16+13+15+13+14+14+13+13+14}{10} = 13{,}90$

$C2_{k3} = \frac{8+9+8+8+8+8+8+8+8+8}{10} = 8{,}10$

Langkah perhitungan yang sama diterapkan hingga data pada *centroid* C2 dan C3, sehingga diperoleh rincian hasil sebagai berikut:

Tabel 9 Nilai *Centroid* Baru iterasi ke-2

Centroid	Jumlah Toko	Jumlah Penjualan	Frekuensi Pemesanan
C1	5,25	10,63	6,69
C2	6,30	13,90	8,10
C3	8,00	18,75	10,00

Karena terdapat perubahan nilai *centroid* dan hasil pengelompokan iterasi ke-2 maka akan dilakukan iterasi ke-3. Langkah selanjutnya yaitu dengan mengulangi perhitungan yang sama dengan langkah sebelumnya, untuk perhitungan yang baru ini, pusat *cluster* (*centroid*) ini digunakan nilai pusat data baru. Berikut adalah hasil kelompok data dari perhitungan jarak data dengan pusat *cluster* yang baru pada iterasi ke-3 dapat dilihat pada tabel dibawah ini

Tabel 10 Data Nilai Jarak ke *Centroid* iterasi ke-3

No	Nama Wilayah	C1	C2	C3	Jarak terdekat	Cluster

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

1	Desa Batu Gemuk	1,43	2,55	7,97	1,43	*cluster 1*
2	Desa Batu Mbelin	3,70	0,33	5,53	0,33	*cluster 2*
3	Desa Batu Penjemuran	6,47	2,85	2,93	2,85	*cluster 2*
4	Desa Batu Rejo	2,82	0,95	6,41	0,95	*cluster 2*
5	Desa Bekukul	1,78	5,49	10,96	1,78	*cluster 1*
6	Desa Cinta Rakyat	8,27	4,57	1,25	1,25	*cluster 3*
7	Desa Gunung Berita	1,60	2,22	7,65	1,60	*cluster 1*
8	Desa Gunung Kelawas	1,78	5,49	10,96	1,78	*cluster 1*
9	Desa Jaba	1,60	2,22	7,65	1,60	*cluster 1*
10	Desa Jati Kesuma	7,32	3,65	2,02	2,02	*cluster 3*
…	…	…	…	…	…	…
30	Desa Kuta Tengah	4,02	0,71	5,25	0,71	*cluster 2*

Jarak hasil perhitungan iterasi ke-3 akan dilakukan perbandingan dan dipilih jarak terdekat antara data pusat *cluster*, pengelompokan data dapat dilihat pada tabel dibawah ini :

Tabel 11 Pengelompokan data hasil jarak *Centroid* iterasi ke-3

No	Nama Wilayah	C1	C2	C3
1	Desa Batu Gemuk	1		
2	Desa Batu Mbelin		1	
3	Desa Batu Penjemuran		1	
4	Desa Batu Rejo		1	
5	Desa Bekukul	1		
6	Desa Cinta Rakyat			1
7	Desa Gunung Berita	1		
8	Desa Gunung Kelawas	1		
9	Desa Jaba	1		
10	Desa Jati Kesuma			1
…	…	…	…	…
30	Desa Kuta Tengah		1	

Kemudian untuk penentuan *centroid* baru setelah diketahui anggota tiap- tiap *cluster*, digunakan persamaan berikut:

$$Cj = \frac{1}{Nk}\sum_{l=1}^{Nk} x_{jl}$$

Dimana :
Cj = *Centroid* baru
Xjl = Anggota *Cluster* pada atribut
Nk = Jumlah data dalam *cluster*
Jadi dengan menggunakan persamaan tersebut, didapat *centroid* baru untuk atribut ketiga , yaitu :

$C3_{k1} = \frac{7+8+9+8}{4} = 8{,}00$

$C3_{k2} = \frac{18+17+21+19}{4} = 10{,}75$

$C3_{k3} = \frac{10+9+11+10}{4} = 10{,}00$

Lakukan perhitungan yang sama ke data ke C1, dan C2. Sehingga didapatkan hasil detail sebagai berikut :

Tabel 12 Nilai *Centroid* Baru iterasi ke-3

Centroid	Jumlah Toko	Jumlah Penjualan	Frekuensi Pemesanan

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

C1	5,25	10,63	6,69
C2	6,30	13,90	8,10
C3	8,00	18,75	10,00

Jarak hasil perhitungan iterasi ke-3 akan dilakukan perbandingan dan dipilih jarak terdekat antara data pusat *cluster* jarak ini menunjukkan bahwa data tersebut termasuk dalam kelompok dengan pusat *cluster* terdekat. Pengelompokan Data berdasarkan jarak terdekat dilakukan setelah mendapatkan hasil dari iterasi ke-3 jarak *centroid*. Pengelompokan data dapat dilihat pada tabel dibawah ini :

Tabel 13 Pengelompokan data hasil jarak *Centroid* iterasi ke-3

No	Nama Wilayah	C1	C2	C3
1	Desa Batu Gemuk	1		
2	Desa Batu Mbelin		1	
3	Desa Batu Penjemuran		1	
4	Desa Batu Rejo		1	
5	Desa Bekukul	1		
6	Desa Cinta Rakyat			1
7	Desa Gunung Berita	1		
8	Desa Gunung Kelawas	1		
9	Desa Jaba	1		
10	Desa Jati Kesuma			1
…	…	…	…	…
30	Desa Kuta Tengah		1	

Dari seluruh hasil perhitungan yang dilakukan pada iterasi ke-3 ini nilai dan anggota *centroid* tidak berubah. Maka kesimpulan pengelompokan akan diperoleh pada iterasi ke-3. Setelah proses iterasi selesai, langkah selanjutnya adalah analisis hasil pengelompokan data. Berdasarkan pengelompokan data pada iterasi terakhir didapatkan hasil dataset yang tergabung di *cluster* 1 sebagai kelompok data 'Wilayah Permintaan Rendah (WPR)', dan dataset yang tergabung pada *cluster* 2 sebagai kelompok data 'Wilayah Permintaan Sedang (WPS)', dan dataset yang tergabung pada *cluster* 3 sebagai kelompok data 'Wilayah Permintaan Tinggi (WPT)'.

2.4 Analisis Hasil Cluster

Setelah memperoleh nilai jarak, langkah berikutnya adalah melakukan analisis klaster. Proses ini dimulai dengan mengevaluasi validitas hasil, mengidentifikasi klaster dengan nilai *indeks* terbaik, mendeskripsikan masing-masing klaster berdasarkan atribut yang terbentuk, mengamati pola hubungan antar data hasil *klasterisasi*, serta menyusun kesimpulan dari analisis tersebut. Adapun hasil analisis *K-Means Clustering* terhadap data sampel yang digunakan disajikan sebagai berikut.

Tabel 13 Kelompok Wilayah Permintaan Rendah (WPR) -C1

No	Nama Wilayah	Jumlah Toko	Jumlah Penjualan (pallet/bulan)	Frekuensi Pemesanan (kali/bulan)
1	Desa Batu Gemuk	5	12	7
2	Desa Bekukul	5	9	6
3	Desa Gunung Berita	6	12	7
4	Desa Gunung Kelawas	5	9	6
5	Desa Jaba	6	12	7
6	Desa Lubang Ido	5	11	7
7	Desa Namo Landur	5	9	6
8	Desa Namo Mbaru	5	11	7
9	Desa Tangkahan Baru	6	12	7

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

10	Desa Namo Pinang	5	11	7
11	Desa Rimo Mungkur	5	9	6
12	Desa Rumah Mbacang	5	9	6
13	Desa Salang Tungir	5	10	7
14	Desa Sudirejo	6	12	7
15	Desa Tanjung Selamat	5	11	7
16	Desa Uruk Gedang	5	11	7

Tabel 14 Kelompok Wilayah Permintaan Sedang (WPS) -C2

No	Nama Wilayah	Jumlah Toko	Jumlah Penjualan (pallet/bulan)	Frekuensi Pemesanan (kali/bulan)
1	Desa Batu Mbelin	6	14	8
2	Desa Batu Penjemuran	8	16	9
3	Desa Batu Rejo	6	13	8
4	Desa Kuta Tualah	7	15	8
5	Desa Kwala Simeme	6	13	8
6	Desa Namo Batang	6	14	8
7	Desa Namo Pakam	6	14	8
8	Desa Suka Mulia Hilir	5	13	8
9	Desa Suka Mulia Hulu	6	13	8
10	Desa Kuta Tengah	7	14	8

Tabel 15 Kelompok Wilayah Permintaan Tinggi (WPT) -C3

No	Nama Wilayah	Jumlah Toko	Jumlah Penjualan (pallet/bulan)	Frekuensi Pemesanan (kali/bulan)
1	Desa Cinta Rakyat	7	18	10
2	Desa Namorambe	9	21	11
3	Desa Jati Kesuma	8	17	9
4	Desa Ujung Labuhan	8	19	10

3. HASIL DAN PEMBAHASAN

Aplikasi Data Mining ini dilengkapi antarmuka sistem terdiri yang dari halaman *login*, halaman menu utama, halaman data wilayah, halaman data *centroid* awal, halaman proses *clustering*, dan halaman laporan hasil *clustering*.

1. Tampilan Halaman Login
   Halaman login digunakan sebagai pintu masuk ke dalam sistem, memberikan lapisan keamanan untuk mencegah akses dari pengguna yang tidak berwenang ,berikut adalah tampilan dari halaman login.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Gambar 3 Tampilan Halaman Login

2. Tampilan Halaman Menu Utama
Halaman Menu Utama berperan sebagai pusat navigasi utama bagi pengguna setelah berhasil login ke dalam sistem..

Gambar 4 Tampilan Halaman Menu Utama

3. Tampilan Halaman Data Wilayah
Halaman Data Wilayah berfungsi sebagai wadah untuk mengelola dan menampilkan data wilayah yang menjadi referensi utama dalam proses pengelompokan toko penjualan air mineral kemasan.

Gambar 5 Tampilan Halaman Data Wilayah

4. Tampilan Halaman Tambah Data Wilayah
Halaman Tambah Data Wilayah adalah bagian penting dari sistem yang memungkinkan pengguna untuk memasukkan data baru terkait wilayah yang akan digunakan dalam proses pengelompokan toko

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Gambar 6 Tampilan Halaman Tambah Data Wilayah

5. Tampilan Halaman Data *Centroid* Awal
   Halaman Data *Centroid* Awal pada sistem ini menampilkan data awal yang digunakan sebagai titik pusat (*centroid*) dalam proses perhitungan algoritma *K-Means Clustering.*

Gambar 7 Tampilan Halaman Data *Centroid* Awal

6. Tampilan Halaman Tambah *Centroid* Awal
   Halaman Tambah *Centroid* Awal berfungsi untuk memungkinkan pengguna yang berwenang untuk menentukan nilai *centroid* pertama yang akan digunakan sebagai titik acuan untuk *clustering.*

Gambar 8 Tampilan Halaman Tambah *Centroid* Awal

7. Tampilan Halaman Proses *K-Means Clustering*
   Halaman Proses *K-Means Clustering* merupakan fitur utama dalam sistem yang menampilkan hasil dari proses pengelompokan data wilayah penjualan air mineral kemasan menggunakan algoritma *K-Means Clustering*.

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 5, September 2025, Hal 1243 - 1257
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Gambar 9 Tampilan Halaman Proses *K-Means Clustering*

8. Tampilan Halaman Hasil *K-Means Clustering*
   Tampilan halaman hasil *K-means clustering* menyajikan hasil akhir dari proses *clustering* wilayah toko penjualan air mineral kemasan pada PT. Tirta Alpin Makmur (Amoz)..

Gambar 10 Tampilan Halaman Hasil *K-Means Clustering*

5. KESIMPULAN

Berdasarkan hasil penelitian dan analisis terhadap pengelompokan wilayah toko penjualan menggunakan algoritma K-Means Clustering pada PT. Tirta Alpin Makmur (Amoz), dapat disimpulkan bahwa algoritma K-Means Clustering berhasil diterapkan untuk mengelompokkan wilayah toko penjualan air mineral kemasan di perusahaan tersebut. Melalui pengelompokan berdasarkan atribut seperti jumlah toko, jumlah penjualan, dan frekuensi pemesanan, wilayah-wilayah tersebut dapat diklasifikasikan ke dalam beberapa kategori, yaitu wilayah dengan permintaan rendah, sedang, dan tinggi. Selain itu, sistem berbasis web yang dirancang untuk mengimplementasikan pengelompokan wilayah toko penjualan dengan algoritma K-Means Clustering juga telah berhasil dibangun. Sistem ini memungkinkan proses pengelompokan dilakukan secara otomatis dengan hasil clustering yang mudah dipahami, serta dapat diakses oleh pengguna untuk keperluan evaluasi dan pengambilan keputusan lebih lanjut dalam distribusi produk.

UCAPAN TERIMAKASIH

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya penelitian dapat terselesaikan dengan baik. Dalam proses penelitian ini, penulis banyak menerima dukungan, bantuan, dan bimbingan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat penulis ingin menyampaikan ucapan terima kasih kepada:

- Bapak Mukhlis Ramadhan SE., M.Kom, selaku dosen pembimbing I yang telah memberikan bimbingan dan arahan sehingga penelitian ini dapat terselesaikan.
- Ibu Rina Mahyuni S.Pd., M.S, selaku dosen pembimbing II yang selalu memberikan masukan, saran, dan dorongan semangat dalam proses penelitian ini.
- PT. Tirta Alpin Makmur, yang telah memberikan izin, serta data yang diperlukan sehingga penelitian ini dapat terlaksana dengan baik.

Akhir kata, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga penelitian ini dapat memberikan manfaat bagi penulis maupun pihak-pihak lain yang membutuhkan.

DAFTAR PUSTAKA

[1] N. F. Adani, A. F. Boy, S. Kom, M. Kom, and R. Syahputra, “Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan STMIK Triguna Dharma **

Program Studi Sistem Informasi, STMIK Triguna Dharma *** Program Studi Sistem Informasi, STMIK Triguna Dharma," *Jurnal CyberTech*, vol. x. No.x, [Online]. Available: https://ojs.trigunadharma.ac.id

[2] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "TINJAUAN PUSTAKA SISTEMATIS: PENERAPAN DATA MINING TEKNIK CLUSTERING ALGORITMA K-MEANS," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, pp. 52–61, Oct. 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[3] P. Pattasang, R. A. Hadiguna, K. Penulis, and : Pattasang, "RANCANGAN USAHA AIR MINUM DALAM KEMASAN MENGGUNAKAN MEREK MINERAL SANTRY (DI PONDOK PESANTREN INSAN MANDIRI BATAM)," vol. 2, no. 2, 2021, doi: 10.38035/jmpis.v2i2.

[4] M. Fajar Fauzan, A. Irma Purnamasari, and G. Dwilestari, "PENERAPAN DATA MINING UNTUK MENGANALISIS PENJUALAN AIR MINUM DALAM KEMASAN SELAMA MASA PANDEMI COVID-19," 2023.

[5] D. J. Lubis and M. B. Tamam, "Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor," vol. 12, pp. 7–20, 2022, doi: 10.36350/jbs.v12i1.

[6] Y. Pratiwi and M. Mulyawan, "Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup berdasarkan Tingkat Provinsi," *Blend Sains Jurnal Teknik*, vol. 1, no. 4, pp. 284–294, Mar. 2023, doi: 10.56211/blendsains.v1i4.233.

[7] M. Rizqi Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *Jurnal Teknologi Ilmu Komputer*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.

[8] D. Mutiara *et al.*, "Pemilihan Pelanggan Potensial Dengan Melakukan Pemetaan Area Dengan Metode Algoritma K-NN dan K-Means Di Yamaha Nusantara Motor Purwokerto," 2021.

[9] S. Marliska Hutabarat, A. Sindar, and S. Pelita Nusantara Jl Iskandar Muda No, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 2, no. 2, 2019.

[10] F. Munawar, A. Nasution, P. Studi Sistem Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal, "SISTEMASI: Jurnal Sistem Informasi Pengelompokan Wilayah Berdasarkan Kubik Air menggunakan Algoritma K-Means pada Perumda Air Minum Tirta Silaupiasa Kabupaten Asahan Regional Grouping Based on Cubic Water Using K-Means Algorithm at Perumda Air Minum Tirta Silaupiasa Asahan Regency." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[11] J. Homepage, D. Syaputri, P. Herwina Noprita, and S. Romelah, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implemnetation of K-Means Algorithm for Economic Distribution Clustering Base on Demographics of Population Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," vol. 1, pp. 1–6, 2021.

[12] J. Homepage and I. Ramadhani, "IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Implementation Of K-Means Algorithm For Palm Oil Productivity Data Clustering Implementasi Algoritma K-Means Untuk Klustering Data Produktivitas Kelapa Sawit".

[13] A. Septiarini, I. A. Thaher, and N. Puspitasari, "Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means Clustering," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 131–141, Jul. 2022, doi: 10.34010/komputika.v11i2.5518.