

PENERAPAN METODE K-MEANS UNTUK MENGLASIFIKASI MINAT KONSUMEN TERHADAP PRODUK TOKO ONLINE

Marcelli Nara Bagdja, Ridho Taufiq Subagio, Viar Dwi Kartika

Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia
Jalan Kesambi No. 202, Drajat, Kec. Kesambi, Kota Cirebon, Indonesia
marcellia.bagdja.ti.20@cic.ac.id

ABSTRAK

Penentuan stok barang pada *online shop* merupakan salah satu tantangan utama yang dihadapi dalam memenuhi permintaan *customer* secara efisien. *Jiggushop* merupakan *online shop* yang masih menentukan stok barang secara manual. Ketidakakuratan dalam penentuan stok akan mengakibatkan kelebihan maupun kekurangan barang yang mengakibatkan kerugian. Penelitian ini bertujuan untuk membuat *website* klasifikasi barang menggunakan metode *K-Means clustering* untuk membantu proses pengelompokan barang. Proses *clustering* berdasarkan jumlah stok awal dan stok akhir untuk setiap produk. Aplikasi yang dibangun berupa *dashboard* yang menggunakan *framework streamlit* dan bahasa pemrograman *python* yang berfungsi untuk mengklasterisasi produk dengan tingkat permintaan tinggi, sedang, dan rendah. Hasil penelitian ini adalah sebuah aplikasi *website* yang berfungsi untuk mengklasterisasi produk kedalam 3 kelompok permintaan. Pengujian dilakukan dengan menggunakan *whitebox*, untuk menguji kode program yang menjalankan metode *K-Means*. Dari hasil pengujian *whitebox* didapatkan nilai *cyclomatic complexity* nya 1, yang menunjukkan tingkat kompleksitas yang rendah dan alur logika yang sederhana yang membuat aplikasi berjalan dengan efisien yang memungkinkan proses *clustering* yang cepat ketika menangani dataset yang besar.

Kata kunci : Stok Produk, K-Means, Clustering, Python

1. PENDAHULUAN

Di era perkembangan teknologi saat ini, online shop menjadi peluang bisnis yang sangat populer. Banyak orang yang beralih untuk berbelanja secara online karena dinilai lebih praktis. Berbelanja secara online dapat diakses darimana saja dan kapan saja, dan banyak orang memilih berbelanja secara online karena lebih banyak produk yang tersedia, serta mudah membandingkan produk. Online shop menyediakan preferensi customer dengan fitur rekomendasi produk berdasarkan riwayat pembelian customer, dan ulasan customer yang lebih positif. *Jiggushop* merupakan salah satu online shop yang berada di Kota Bandung yang ada sejak tahun 2017. *Jiggushop* ini menyediakan beberapa produk kecantikan seperti bodycare dan skincare. Dari masing-masing kategori yang ada ini memiliki produk yang cukup banyak, keseluruhan produk yang dimiliki oleh *Jiggushop* ini adalah 117 produk untuk kedua kategori tersebut. Dalam 3 bulan terakhir ini pola penjualan nya berubah-ubah karena pengaruh dari trend yang sedang berlangsung.

Tantangan yang harus dihadapi dalam mengelola *online shop* yaitu, penentuan stok barang yang tepat sesuai dengan target pasar. *Jiggushop* ini memiliki produk yang cukup banyak dan tingginya pemesanan dapat membuat pengelolaan stok barang menjadi rumit. Ada risiko yang akan terjadi jika stok barang kurang tepat yaitu, jika stok terlalu banyak dapat berisiko akan menumpuk di gudang dan jika stok terlalu sedikit akan dapat berisiko kehabisan barang dan kehilangan customer. Apabila menumpuk terlalu lama digudang akan mengakibatkan produk kadaluarsa karena memiliki masa *expired* yang dapat mengurangi kualitas produk. Untuk menentukan

jumlah stok barang yang tepat ini sulit, ketika harus memprediksi permintaan di masa depan. Perubahan trend belanja yang di pengaruhi oleh banyak faktor membuat prediksi menjadi sulit. Permintaan barang yang dijual di *online shop* seringkali berubah. Oleh karena itu, untuk penentuan stok barang yang tepat pada *online shop* ini dapat diterapkan nya *website dashboard* yang menggunakan metode *k-means* dengan mengklasterisasi untuk membantu penentuan stok barang yang tepat sesuai dengan target pasar yang terjadi. Dalam penggunaan metode algoritma *k-means* ini dapat mengelompokkan produk berdasarkan pola pembelian *customer*. Pada *online shop* ini metode *k-means* dapat mengoptimalkan persediaan produk yang paling diminati. Kelebihan dari metode ini yaitu dapat membagi data ke dalam kelompok berdasarkan jarak dari *centroid*. Penggunaan metode algoritma *k-means* juga mampu mengolah data dalam jumlah yang besar [1].

Metode yang sama dalam penelitian yang lain tentang penerapan metode *k-means* dalam penjualan produk Souq.com. Hasil dari penelitian tersebut ialah untuk memahami produk yang paling banyak diminati yang terbagi kedalam 3 kelompok dengan tujuan agar tidak terjadi penumpukan barang di gudang [2].

Dengan metode yang berbeda pada penentuan stok barang tentang penggunaan algoritma *apriori* untuk menentukan pola stok barang. Tujuan dari penelitian tersebut untuk mendapatkan hasil pola stok barang dengan menggunakan algoritma *apriori* [3].

Pada penelitian dengan metode dan bahasa pemrograman yang sama tentang data *mining k-means* yang menggunakan bahasa pemrograman *python* pada data pembayaran transaksi. Hasil dari penelitian

tersebut ialah pengelompokan data transaksi pembayaran menjadi beberapa klaster untuk menentukan strategi bisnis perusahaan [4].

2. TINJAUAN PUSTAKA

2.1. Online Shop

Online shop adalah toko *online* yang dilakukan melalui internet untuk bahan promosi nya dan menggunakan *website* untuk katalog nya [5].

Online shop merupakan toko *online* yang menjual berbagai macam produk. Tanpa memiliki toko, penjual pun dapat berjualan secara *online* dalam *platform e-commerce* [6].

Pada *online shop* ini *customer* dapat melihat produk yang dijual kemudian melakukan pemesanan dan pembayaran secara digital. Bisnis *online shop* ini mulai berkembang pesat seiring dengan berkembangnya teknologi.

2.2. Data Mining

Data mining merupakan proses menemukan pola dan informasi penting dari kumpulan data yang besar. Tujuan dari *data mining* yaitu untuk mengubah data mentah menjadi wawasan yang dapat digunakan untuk membuat keputusan yang lebih baik [3].

2.3. Clustering

Clustering bertujuan untuk mengelompokkan objek kedalam beberapa kelompok berdasarkan karakteristik yang sama [7].

2.4. K-Means

Algoritma *k-means* merupakan salah satu algoritma *cluster* yang banyak digunakan dalam *data mining* [8].

Algoritma *k-means* ini digunakan untuk mengelompokkan data pada *cluster* [2].

Algoritma *k-means* ini termasuk kedalam *unsupervised learning* karena tanpa *output* yang diinginkan dan tidak ada target variabel. Dalam algoritma ini perlu menentukan jumlah *cluster* *k*.

2.5. Klasterisasi

Klasterisasi biasanya digunakan sebagai tahap awal dalam metode *data mining* [9].

Klasterisasi merupakan suatu proses pengelompokan data kedalam *cluster* yang memiliki persamaan karakteristik tertentu [7].

Tujuan dari klasterisasi ialah untuk mengidentifikasi pola data.

2.6. Python

Python merupakan bahasa pemrograman yang bersifat *open source*. Bahasa pemrograman ini dapat digunakan pada pengembangan perangkat lunak yang bisa dijalankan melalui berbagai sistem operasi. Bahasa *python* ini didukung oleh banyak *library* yang didalam nya menyediakan fungsi analisis data dan fungsi *machine learning* [4].

2.7. Dataset

Dataset merupakan data yang dibutuhkan dalam mengolah data untuk penelitian ini [3].

2.8. Streamlit

Streamlit merupakan *framework* untuk pengembangan pembuatan aplikasi *web* untuk menganalisis data. Dengan menggunakan bahasa pemrograman *python* dapat diubah menjadi aplikasi *web* dengan menggunakan *streamlit*. *Framework streamlit* dapat memudahkan *developer* dalam membangun program dibidang data *science* dan *machine learning* yang berbasis *web* [10].

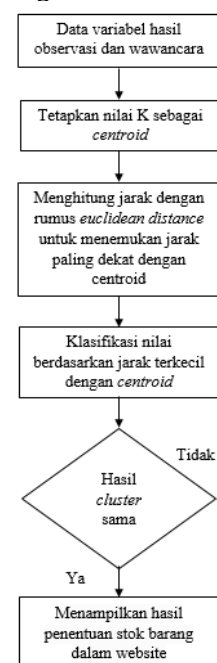
3. METODE PENELITIAN

3.1. Pengumpulan Data

Pengumpulan data dilakukan dengan 3 cara yaitu, observasi, dokumentasi, studi literatur. Observasi dilakukan terhadap *jiggushop* guna mengumpulkan data yang akan menunjang penelitian. Dokumentasi dilakukan terhadap *owner jiggushop* untuk mendapat informasi laporan penjualan. Studi literatur dilakukan dengan mempelajari jurnal yang berkaitan dengan penelitian ini.

Pada penelitian ini yang menggunakan *datamining* tahapan dari pemilihan data yang telah dilakukan melalui observasi, dokumentasi dan studi literatur kemudian dilakukan pemilihan atribut untuk analisis *clustering*. Pada tahap *pra processing* ini dilakukannya normalisasi data untuk atribut jumlah stok barang. Kemudian proses *clustering* menggunakan perhitungan *k-means*. Hasil dari proses tersebut merupakan sebaran klaster produk.

3.2. Metode Pengolahan



Gambar 1. Alur Perhitungan

Pada penelitian ini menggunakan algoritma *k-means* sebagai metode pengolahan dan analisis data. Setelah mendapatkan data dari *jiggushop* kemudian dilanjutkan dengan tahapan perhitungan *K-Means*.

Adapun rumus *euclidean K-Means* adalah sebagai berikut

$$D_{(i,j)} = \sqrt{(x_{1i} - x_{1j})^2 + \dots (x_{ki} - x_{kj})^2}$$

Dimana keterangannya sebagai berikut :

$D(i,j)$: jarak data ke I ke pusat *cluster*

X_{ki} : data ke I pada atribut data ke k

X_{kj} : data ke j pada atribut data ke k

Pada Gambar 1. diatas dapat dijelaskan untuk pengolahan dengan metode *k-means* setelah mendapat data untuk pengujian kemudian menetapkan nilai k sebagai *centroid*. Data yang digunakan ialah stok awal dan stok akhir pada produk penjualan. Setelah

menetapkan *centroid* kemudian menghitung jarak dengan rumus *euclidean distance* untuk menemukan jarak paling dekat dengan *centroid*. Kemudian mengklasifikasi nilai berdasarkan jarak terkecil dengan *centroid* dan akan didapati hasil untuk *cluster*. Jika hasil *cluster* lama dan baru belum sesuai maka harus mengulangi perhitungan dengan rumus awal, namun jika hasil *cluster* lama dan baru sudah sesuai maka perhitungan sudah selesai dan hasil sudah didapatkan.

4. HASIL DAN PEMBAHASAN

4.1. Perhitungan K-Means

Perhitungan *k-means* ini menggunakan *excel* sebagai tools membantu dalam menghitung dengan rumus yang sudah ada. Dataset ini berisikan nama produk, stok awal dan stok akhir yang akan di klasterisasi menjadi 3 klaster. Berikut langkah pengerjaan menggunakan metode *k-means* :

Tabel 1. Dataset

Data ke	Nama Produk	Stok Awal	Stok Akhir	Klaster 1	Klaster 2	Klaster 3
1	Iswhite masker badan	885	100	*		
2	Iswhite body serum	470	63	*		
3	Iswhite parfum	180	15		*	
4	Iswhite body scrub	210	63	*		
5	Beutytox whitemilk	130	15		*	
6	Hb bodymist	175	66		*	
7	Holigress body mist	1200	597	*		
8	Holigress macaroon	120	32		*	
9	Deodorant	120	47		*	
10	Milk recipe	115	13		*	
11	Secret clean propolis spray	100	47			*
12	Kala body mist	650	398	*		
13	Secret cleas propolis	65	18			*
14	Vaseline healthy	200	79		*	
15	Nivea extra	220	55	*		
16	Sheeby	290	51	*		
17	Kala roll on	1000	572	*		
18	La creame body	100	30		*	
19	Iswhite body wash serum	55	18			*
20	secret clean propolis antiseptic	210	80		*	
21	Beatetox gluta collagen	30	10			*
22	elumor oxter	100	24			*
23	Bromen	50	39			*
24	herborist juice	45	35			*
25	tmco lotion	26	3			*
26	coolvita collagen	130	2		*	
27	iswhite glow body	50	49			*
28	numilk tone up	10	5			*
29	breylee teeth	40	5			*
30	hanum handcream	120	40		*	
31	kala 30ml	250	26	*		
32	skinomie edp	120	40		*	
33	aha body serum	450	8	*		

Nilai *centroid* diambil dari rata-rata jumlah setiap klaster. Sehingga didapatkan nilai *centroid* sebagai berikut :

Tabel 2. Centroid Awal

C1	562,5	193,3
C2	143,3333	36
C3	51,90909	23

Kemudian melakukan perhitungan *euclidean distance* untuk menghitung jarak data ke masing-masing *centroid*.

Tabel 3. Perhitungan Jarak

	Klaster 1	Klaster 2	Klaster 3
1	112711,1		
2	25534,34		
3		1785,444	
4	141234,3		
5		618,7778	
6		1902,778	
7	569379,9		
8		560,4444	
9		665,4444	
10		1331,778	
11			2888,736
12	49558,34		
13			196,3719
14		5060,111	
15	136433,1		
16	94505,54		
17	334819,9		
18		1913,778	
19			34,55372
20		6380,444	
21			649,0083
22			2313,736
23			259,6446
24			191,7355
25			1071,281
26		1333,778	
27			679,6446
28			2080,372
29			465,8264
30		560,4444	
31	125645,5		
32		560,4444	
33	46992,34		
Hasil	1636815	22673,67	10830,91

Setelah mendapatkan hasil dari langkah sebelumnya, kemudian menggabungkan setiap data berdasarkan jarak terdekat dengan *centroid* pada masing-masing *cluster* yang akan menghasilkan *cluster* baru berdasarkan nilai minimum. Kemudian ulangi perhitungan *k-means* sampai mendapatkan hasil *centroid* lama dan *centroid* baru tidak berubah atau hasilnya sama. Berikut hasil akhir klasterisasi nya :

Tabel 4. Hasil Akhir

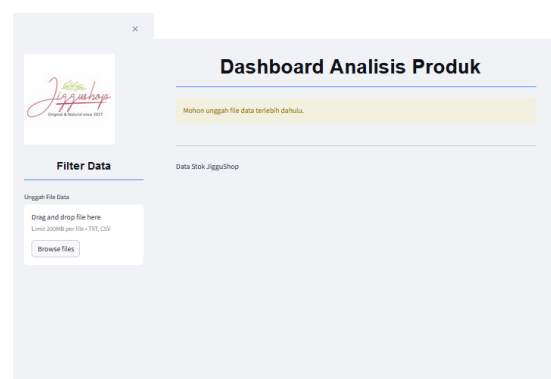
No	Nama Produk	Stok awal	Stok Akhir	klaster
1	Iswhite masker badan	885	100	1
2	Iswhite body serum	470	63	2
3	Iswhite parfum	180	15	2
4	Iswhite body scrub	210	63	2
5	Beutytox whitemilk	130	15	3
6	hanum beautycare bodymist	175	66	2
7	Holigress body mist	1200	597	1
8	Holigress macaroon	120	32	3

No	Nama Produk	Stok awal	Stok Akhir	klaster
9	Deodorant	120	47	3
10	Milk recipe	115	13	3
11	Secret clean propolis spray	100	47	3
12	Kala body mist	650	398	1
13	Secret cleas propolis wash	65	18	3
14	Vaseline healthy	200	79	2
15	Nivea extra	220	55	2
16	Sheeby	290	51	2
17	Kala roll on	1000	572	1
18	La creame body	100	30	3
19	Iswhite body wash serum	55	18	3
20	secret clean propolis antiseptic	210	80	2
21	Beatetox gluta collagen	30	10	3
22	elumor oster	100	24	3
23	Bromen	50	39	3
24	herborist juice	45	35	3
25	tmco lotion	26	3	3
26	coolvita collagen	130	2	3
27	iswhite glow body	50	49	3
28	numilk tone up	10	5	3
29	breylee teeth	40	5	3
30	hanum handcream	120	40	3
31	kala 30ml	250	26	2
32	skinomie edp	120	40	3
33	aha body serum	450	8	2

Dari perhitungan *k-means* diatas dapat disimpulkan untuk hasil perhitungan dibagi menjadi 3 klaster dengan keterangan klaster 1 paling diminati, klaster 2 jarang diminati, klaster 3 kurang diminati. Dari hasil itu didapatkan data produk untuk masing-masing klaster sebagai berikut :

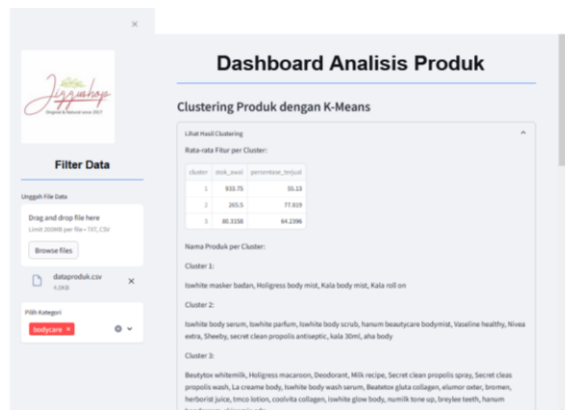
- Klaster 1 (Paling diminati) : Data ke 1,7,12,17
- Klaster 2 (Jarang diminati) : Data ke 2,3,4,6,14,15,20,31,33
- Klaster 3 (kurang diminati) : Data ke 5,8,9,10,11,13,18,19, 21,22,23,24,25,26,27, 28,29,30,32

4.2. Tampilan Dashboard



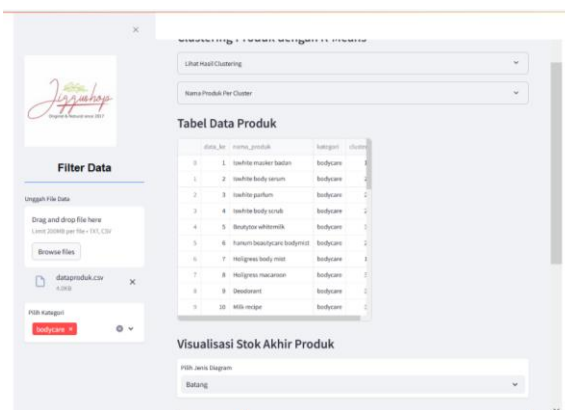
Gambar 2. Dashboard

Pada gambar 2 terdapat halaman *dashboard* yang mengharuskan *owner* untuk mengupload file dataset dengan format csv.



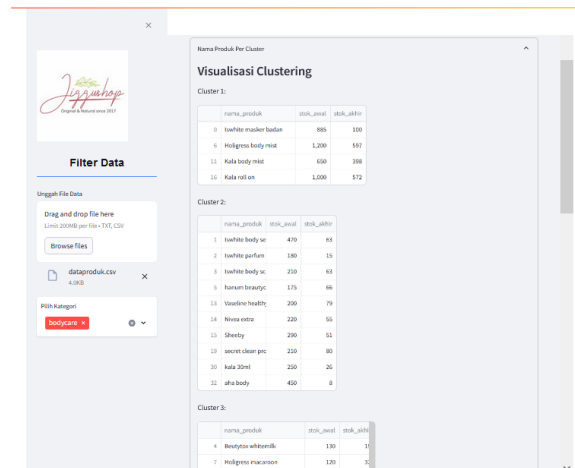
Gambar 3. Hasil Klasterisasi

Pada gambar 3 setelah memasukan *file* dataset kemudian dapat memilih fitur yang tersedia untuk melakukan klasterisasi. Halaman ini menampilkan ringkasan visualisasi dari dataset yang di impor. Fitur utama pada *dashboard* ini adalah filter kategori yang memungkinkan untuk menyaring data berdasarkan kategori yang ada. Setelah memilih kategori, sistem akan memproses klasterisasi dan menampilkan hasil berupa nilai *centroid* yang menjadi titik pusat dari setiap klaster. Hasil nya berupa pengelompokan produk kedalam 3 klaster dengan detail nama produk nya.



Gambar 4. Tabel Data

Pada gambar diatas menampilkan halaman tabel data hasil klasterisasi untuk setiap kategori yang dipilih. Tabel diatas meliputi nama produk yang mengidentifikasi produk secara spesifik.

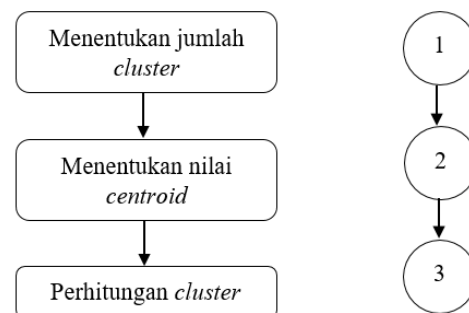


Gambar 5. Visualisasi

Pada gambar menampilkan tabel visualisasi klasterisasi dari setiap klaster. Setiap klaster dari proses ditampilkan dalam tabel terpisah yang meliputi nama produk, stok awal, dan stok akhir.

4.3. Pengujian

Pengujian ini menggunakan *whitebox* yang berfokus pada kode program untuk melakukan *k-means clustering* dimulai dengan menentukan jumlah klaster, menentukan nilai *centroid*, dan perhitungan *k-means*. Berikut ini *flowgraph k-means* :



Gambar 6. Flowgraph

Gambar diatas merupakan representasi visual dari proses *clustering* menggunakan *k-means*. *Flowgraph* diatas merupakan pengubahan data menjadi numerik yang terdiri dari 3 *node* yang diberi nomor 1 sampai 3, yang menggambarkan langkah-langkah utama dalam proses *klasterisasi*. Dari hasil *flowgraph* dapat menentukan nilai *cyclomatic complexity* untuk perhitungan *k-means*. Berikut perhitungan *cc* dalam klasterisasi beserta penjelasannya. Rumus *cyclomatic complexity* : $V(G)=E-N+2P$
 E : Jumlah *edge* atau panah = 2
 N : Jumlah *node* atau titik = 3
 P : Jumlah komponen yang terhubung = 1
 $V(G) = 2 - 3 + 2(1) = 1$

Jadi, *cyclomatic complexity* pada perhitungan *k-means* ini adalah 1. Hasil nilai *cyclomatic complexity* ini menentukan ada 1 jalur dalam program ini berarti

program ini sangat sederhana, tanpa adanya perancangan dan membuat program yang konsisten.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian ini didapatkan *website dashboard* yang mudah digunakan karena sederhana. *Dashboard* yang dihasilkan mampu mengklasterisasi setiap data baru. hasil klasterisasi dengan menerapkan *k-means* dapat membantu mengelompokkan produk berdasarkan pola penjualan dalam mengoptimalkan pengelolaan stok. Saran untuk penelitian selanjutnya diharapkan dapat mengembangkan strategi pengelolaan stok dan menambah fitur lain yang penting untuk mengantisipasi *trend* penjualan di masa depan.

DAFTAR PUSTAKA

- [1] H. Prastiwi, Jeny Pricilia, and Errissya Rasywir, "Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering," *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 1, pp. 141–148, 2022, doi: 10.33998/jakakom.2022.2.1.34.
- [2] F. Amin, D. S. Anggraeni, and Q. Aini, "Penerapan Metode K-Means dalam Penjualan Produk Souq.Com," *Appl. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 7–14, 2022, doi: 10.15408/aism.v5i1.22534.
- [3] I. Fitri Polorida Ginting and D. Saripurna, "Penerapan Data Mining Dalam Menentukan Pola Ketersediaan Stok Barang Berdasarkan Permintaan Konsumen Di Chykes Minimarket Menggunakan Algoritma Apriori," *J. Sains Manaj. Inform. dan Komput.*, vol. 20, no. 1, pp. 28–37, 2021.
- [4] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [5] S. Solihin and S. Zuhdi, "Pengaruh Kualitas Website dan Kemudahan Penggunaan Terhadap Keputusan Pembelian Online," *J. Inform. Kesatuan*, vol. 1, no. 1, pp. 13–22, 2021, doi: 10.37641/jikes.v1i1.403.
- [6] A. N. Safitri and F. A. Sianturi, "Analisa Metode Trend Moment Untuk Peramalan Penjualan Stok Barang Pada Toko Sun Oleh-Oleh," *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1.1, pp. 91–102, 2021.
- [7] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 242–248, 2022, doi: 10.33005/sitasi.v2i1.304.
- [8] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [9] I. Rizky Mahartika and A. Wibowo, "Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional," *Semin. Nas. Sist. Inf. dan Teknol.*, vol. 3, no. 1, pp. 87–91, 2019.
- [10] M. Fikri, "Klasifikasi Gerakan Yoga dengan Model Convolutional Neural Network Menggunakan Framework Streamlit," *Media Inform. Budidarma*, vol. 7, no. 1, pp. 509–519, 2023, doi: 10.30865/mib.v7i1.5520.