

Analisis Penjualan Produk Terlaris di Toko Bangunan Pekanbaru Jaya Menggunakan Metode Clustering K-Means

Dihin Muriyatmoko¹, Dian Fikrianti², Fifid Rahma Ifalus³

^{1,2}Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor

³Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor

E-mail: [*¹dihin@unida.gontor.ac.id](mailto:1dihin@unida.gontor.ac.id), 2dianfikrianti@yahoo.co.id, 3fifidifalus42025@mhs.unida.gontor.ac.id

Abstrak – Toko bangunan Pekanbaru Jaya menghadapi kesulitan dalam mengidentifikasi produk yang paling diminati pelanggan, sehingga menghambat efisiensi pengelolaan stok dan strategi penjualan. Penelitian ini bertujuan untuk membantu toko menentukan produk terlaris dengan menganalisis data penjualan dua tahun terakhir menggunakan metode clustering K-Means. Parameter yang digunakan meliputi product, unit, quantity, dan total. Hasil analisis menghasilkan 5 cluster dengan nilai Davies Boulding Index (DBI) sebesar 0.428 dan Silhouette Score sebesar 0.69 yang menunjukkan kualitas pengelompokan yang baik. Setiap cluster memberikan wawasan tentang kelompok produk berdasarkan tingkat penjualan yang dapat digunakan untuk mengoptimalkan pengelolaan stok dan strategi pemasaran toko.

Kata Kunci — clustering, davies-buoldin index, k-means, silhoutte score

1. PENDAHULUAN

Kecanggihan jaringan komunikasi dan perkembangan teknologi informasi yang semakin pesat di era globalisasi telah memberikan dampak signifikan pada berbagai aspek kehidupan, termasuk dunia bisnis. Penerapan teknologi informasi memungkinkan aktivitas bisnis dilakukan dengan lebih efisien, sekaligus menghasilkan data dalam jumlah besar yang berpotensi untuk dianalisis guna mendukung pengambilan keputusan strategis[1]. Dalam era persaingan global yang semakin ketat, pengusaha dituntut untuk lebih adaptif dalam memahami kebutuhan konsumen dan menyusun strategi pemasaran yang efektif. Lanskap bisnis yang dinamis ini sangat dipengaruhi oleh selera konsumen yang terus berubah, serta tantangan lingkungan kompetitif yang semakin kompleks[2].

Toko Bangunan Pekanbaru Jaya berlokasi di Villa Fajar Indah II Jl. Barau-Barau Blok A-8 Kel. Sialang, kec. Tenayan raya, Pekanbaru, Riau ini menjual berbagai barang bangunan seperti tang, baut, cangkul, gergaji, palu dan produk lainnya yang dibutuhkan masyarakat untuk kebutuhan rumah tangga. Dalam operasionalnya, toko menggunakan sistem informasi untuk merekap data penghasilan secara efisien. Namun, sistem ini memiliki keterbatasan karena hanya mampu mencatat penghasilan tanpa dapat mengidentifikasi produk yang paling diminati pelanggan. Hal ini menjadi kendala bagi toko, mengingat produk yang dijual berasal dari pemasok di Solo, sehingga proses pengadaan barang membutuhkan perencanaan matang untuk efisiensi waktu dan biaya. Oleh karena itu diperlukan sistem yang dapat menganalisis data penjualan untuk mengidentifikasi produk yang paling diminati pelanggan. Informasi ini akan membantu toko mengelola stok dengan lebih baik, menghindari penumpukan barang, dan memastikan ketersediaan produk sesuai kebutuhan pelanggan.

Untuk mencapai tujuan tersebut, diperlukan penerapan metode data mining yang tepat guna mengekstrak informasi penting dari data penjualan. Salah satu metode yang dapat digunakan adalah *clustering K-Means*. *Clustering* merupakan teknik yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan kesamaan karakteristik[3]. *K-Means* adalah algoritma *clustering non-hierarkis* yang berfungsi untuk membagi data menjadi beberapa kelompok yang memiliki karakteristik serupa[4]. Dalam proses ini, data yang memiliki kemiripan atribut akan dikelompokkan bersama dalam satu *cluster*, sementara data yang berbeda akan dikelompokkan di *cluster* yang berbeda, sehingga terdapat variasi yang jelas antar *cluster*.

Penelitian yang dilakukan oleh Narayana Sakti Aji (2023) dengan judul "Penentuan Penjualan Barang Berdasarkan Pengelompokan Produk dengan *K-Means* Clustering Metode CRISP-DM Pada CV. Sembako Dina" bertujuan untuk meningkatkan penjualan barang dan manajemens tok berdasarkan pengelompokan produk yang optimal[5]. Padal penelitian ini menggunakan *Davies Index Bouldin* (DBI) untuk menentukan jumlah *cluster*. Dari hasil DBI terdapat 5 model *cluster* yang diuji, 2 *cluster* memiliki hasil -0.413, 3 *cluster* memiliki hasil -0.418, 4 *cluster* memiliki hasil -0.359, 5 *cluster* memiliki hasil -0.324. Dari hasil tersebut yang mendekati angka 0 yaitu 3 *cluster* yaitu *cluster* 1 dengan produk yang memiliki tingkat penjualan ditengah produk terlaris dan tidak laris,

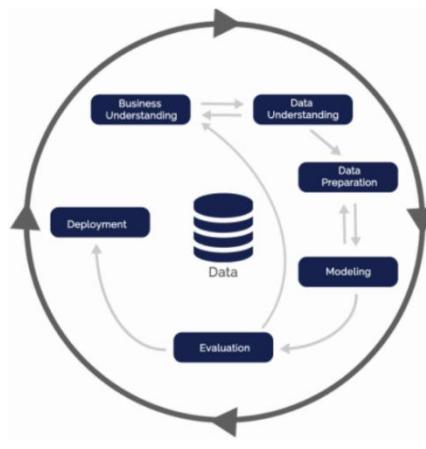
cluster 2 dengan produk lebih sedikit dibandingkan *cluster* atau produk tidak laris, *cluster* 3 dengan produk paling laris dibandingkan *cluster* 1 dan *cluster* 2.

Penelitian lain yang dilakukan oleh Tri Wahyudi (2023) dengan judul "Penerapan Algoritma *K-Means* Pada Data Penjualan Untuk Mendapatkan Produk Terlaris di Pt.Titian Nusantara Boga" bertujuan untuk mengelompokkan produk berdasarkan popularitasnya menggunakan metode CRISP-DM[6]. Penelitian ini menganalisis 793 data dan membentuk 3 *cluster*, *cluster* 1 mendapatkan hasil 128 produk, *cluster* 2 mendapatkan hasil 622 produk, dan *cluster* 3 mendapatkan hasil 43 produk. Penelitian ini melakukan uji validasi menggunakan *Davies Index Bouldin* (DBI) dengan hasil *cluster* 1 0.679, *cluster* 2 0.816, dan *cluster* 3 0.837. Dari hasil tersebut menjelaskan bahwa *cluster* 1 termasuk dalam kategori sangat laris, *cluster* 2 termasuk dalam kategori cukup laris dan *cluster* 3 termasuk dalam kategori kurang laris. Hasil tersebut bisa dikategorikan karena nilai DBI jika semakin mendekati nilai angka 0 maka kualitasnya semakin baik.

Tujuan dari *clustering* adalah untuk meminimalkan variasi dalam setiap *cluster* dan mengoptimalkan variasi antar *cluster* atau dikenal sebagai fungsi objektif dalam proses *clustering*. Dengan menggunakan sistem pengelompokan, perusahaan dapat dengan mudah mengidentifikasi barang bangunan yang diminati oleh pelanggan sehingga membantu dalam mengelola stok dengan lebih efisien, sehingga bahan bangunan yang tersedia di gudang tidak menumpuk atau mengalami kekurangan.

2. METODE PENELITIAN

Metode yang digunakan pada penelitian kali ini adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Sejak tahun 1996, CRISP-DM dikembangkan untuk menjadi proses industri bisnis yang paling umum digunakan perusahaan. Terdapat 6 proses tahapan dalam metode ini, yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. Alur cirlis DM dapat dilihat pada Gambar 1 berikut.



Gambar 1 Alur CRISP DM

2.1 Business Understanding

Tahap *business understanding* bertujuan untuk memahami tujuan bisnis dan mengubahnya menjadi masalah yang dapat diselesaikan melalui analisis data. Masalah yang dihadapi oleh pemilik Toko Bangunan Pekanbaru Jaya adalah pengelolaan stok barang yang masih dilakukan secara manual. Meskipun transaksi sudah dicatat menggunakan sistem komputerisasi, sistem tersebut tidak dapat menyimpan database dan hanya menghasilkan faktur sebagai bukti transaksi. Oleh karena itu, penelitian ini bertujuan untuk membantu mengatasi masalah tersebut dengan menggunakan teknik data mining untuk mengidentifikasi produk yang diminati pelanggan. Dengan ini toko dapat mengoptimalkan pengelolaan stok dan meningkatkan efektivitas penjualan.

2.2 Data Understanding

Data understanding merupakan tahap awal dalam proses data mining yang dimulai dengan pengumpulan, eksplorasi, dan evaluasi data untuk memahami karakteristik dan kualitasnya[7]. Tujuan dari tahap ini adalah memastikan data yang akan digunakan relevan, berkualitas tinggi dan siap untuk dianalisis lebih lanjut. Data yang digunakan dalam penelitian ini diperoleh dari pemilik Toko Bangunan Pekanbaru Jaya periode 2022-2023 dalam format excel. Variabel-variabel yang terdapat dalam data tersebut adalah sebagai berikut :

- a. Product
- b. Unit
- c. Quantity
- d. Total

Pada tahap ini data akan dieksplorasi lebih lanjut untuk memastikan konsistensi dan kelengkapan, serta mengidentifikasi jika ada data yang hilang atau tidak sesuai.

2.3 Data Preparation

Data preparation mencakup serangkaian tahapan yang diperlukan untuk mempersiapkan dataset akhir yang akan digunakan dalam proses *modelling*. Langkah pertama dalam tahap ini adalah pembersihan data yang melibatkan pemeriksaan data untuk mengidentifikasi dan memperbaiki kesalahan, duplikasi, atau nilai yang hilang. Setelah itu dilakukan transformasi data, yaitu mengubah format data atau menambahkan atribut baru sesuai dengan kebutuhan analisis.

2.4 Modelling

Pada tahap *modelling* data mining dilakukan dengan menggunakan algoritma yang telah ditetapkan untuk mencapai hasil yang sesuai dengan kebutuhan pengguna. Dalam penelitian ini, peneliti menggunakan metode *clustering* untuk mengelompokkan barang berdasarkan kuantitas dan harga. Langkah pertama adalah pemilihan algoritma *K-Means*. Cara algoritma ini mengelompokkan dapat dikatakan serupa dari sebuah data dengan cara mengukur jarak nilai dari satu data dengan data lain. Untuk menentukan jumlah cluster yang optimal, digunakan *Elbow Method*. Metode ini membantu memilih jumlah *cluster* dengan meminimalkan variasi dalam *cluster* dan menghindari *overfitting*.

2.5 Evaluation

Pada tahap *evaluation*, model yang telah dibangun dievaluasi untuk memastikan kualitas dan kinerjanya sesuai dengan tujuan proyek. Evaluasi ini penting untuk memvalidasi apakah model *clustering* yang digunakan dapat menghasilkan pengelompokan yang baik. Dalam penelitian ini digunakan *Davies Index Bouldin* (DBI) sebagai ukuran untuk mengevaluasi kualitas cluster[8] dan *Silhouette Score* untuk mengukur seberapa baik setiap titik data cocok dengan setiap *cluster*. DBI mengukur rasio antara jarak antar *cluster* dan ukuran *cluster*, dimana nilai yang lebih rendah menunjukkan pengelompokan yang lebih baik, yaitu jarak antar *cluster* lebih besar dan variasi dalam *cluster* lebih kecil. Nilai *Silhouette Score* berkisar antara -1 hingga 1, dimana nilai yang lebih tinggi atau mendekati 1 menunjukkan bahwa titik data lebih cocok di *cluster* yang digunakan

2.6 Deployment

Pada tahap *deployment*, hasil dari model *clustering* yang telah dibangun diterapkan untuk mengelompokkan data yang telah disiapkan ke dalam *cluster-cluster* yang telah ditentukan. Hasil *clustering* ini akan memberikan *insight* yang berharga bagi pemilik toko, seperti mengetahui produk mana yang diminati pelanggan. Dengan informasi ini, pemilik toko dapat merencanakan pengadaan barang dengan lebih efisien, mengurangi risiko penumpukan stok, dan memastikan ketersediaan produk yang sesuai dengan permintaan pasar. Selain itu, strategi pemasaran dapat lebih difokuskan pada produk-produk yang paling diminati, sehingga dapat meningkatkan profit dan efisiensi operasional toko.

3. HASIL DAN PEMBAHASAN

3.1 Data Understanding

Pada penelitian ini, data yang digunakan untuk pemodelan atau implementasi algoritma *K-Means* merupakan data rekap penjualan barang-barang bangunan di Toko Bangunan Pekanbaru Jaya selama periode 2 tahun, yaitu tahun 2022 hingga tahun 2023. Jumlah data yang digunakan sebanyak 2.455 entri data. *Table 1* berikut ini menyajikan sampel data dari rekap penjualan barang-barang bangunan di toko tersebut.

Table 1 Sampel Data Rekap Penjualan Barang-Barang Bangunan di Toko Bangunan Pekanbaru Jaya

Product	Quantity	Unit	Total
Cangkul ayam/buaya Tg	84	Bh	Rp 1.304.000
Gembok chrom ATS 50L	36	Bh	Rp 966.000
Klinting/gowok tg	80	bh	Rp 480.000
Kluntung tanggung	6	bh	Rp 192.000
Kunci laci ATR	1	Lsn	Rp 68.750
Mata bor besi HSS kng ATS 03.5 mm	10	pcs	Rp 31.750
Meteran karet magnet ATS 7.5m	18	Bh	Rp 545.004
Overpal XPTool ukir pth 3"	24	lsn	Rp 988.000
Palu Kambing SAR polos 0.50	12	bh	Rp 384.000
Palu Kambing XPTool magnet fiber 8 OZ	48	Bh	Rp 1.056.000

Setelah data sudah dikumpulkan, maka dilakukan eksplorasi terhadap tipe data dari setiap kolom yang ada untuk mengetahui gambaran lebih jelas mengenai struktur data secara keseluruhan. Berikut *Gambar 1* menunjukkan banyak data dan type data dari setiap kolom.

```
RangeIndex: 2455 entries, 0 to 2454
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   product     2455 non-null    object 
 1   unit        2455 non-null    object 
 2   total_cleaned  2455 non-null  int64  
 3   quantity_cleaned 2455 non-null  int64
```

Gambar 2 Hasil Eksplorasi Tipe data

Range Index menunjukkan bahwa data yang digunakan dalam *DataFrame* sebanyak 2445 dimana *indeks* dimulai dari 0 hingga 2454. Kemudian 4 kolom dalam *dataset* dengan nama *product* memiliki *type data* *object*, *unit* memiliki *type data* *object*, *total_cleaned* memiliki *type data* *int64*, dan *quantity_cleaned* memiliki *type data* *int64*. Kolom-kolom tersebut juga menunjukkan jumlah nilai *non-null* atau tidak ada nilai kosong dalam *dataset* tersebut maka *dataset* ini sudah dapat dianalisis lebih lanjut.

3.2 Data Preparation

Data Preparation adalah tahapan untuk memperbaiki masalah yang terdapat pada data sebelum data masuk ke tahap *modelling* sehingga menghasilkan *modelling* yang bagus. Berikut langkah-langkah dalam data preparation :

a. Cleaning Data

Pada tahap ini dilakukan pembersihan data pada kolom *total* dan kolom *quantity*. Kolom-kolom tersebut berisi data dalam bentuk angka dengan simbol mata uang, pemisah ribuan, serta desimal yang tidak relevan. Selain itu, terdapat nilai kosong yang dapat menghambat proses analisis selanjutnya. maka dilakukan pembersihan data untuk memastikan kolom tersebut memiliki format yang lebih sederhana dan dapat diolah.

b. Transformasi Data

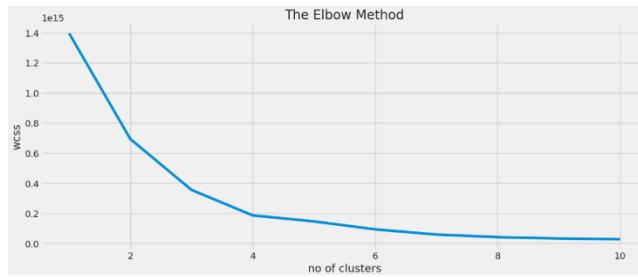
Dalam proses ini dilakukan pengolahan data pada kolom *unit*. Dari data yang diolah menunjukkan bahwa data tersebut memiliki satuan yang tidak konsisten. Untuk mengatasinya dilakukan proses normalisasi satuan melalui pendekatan mapping. *Table 2* berikut menunjukkan hasil transformasi data.

Table 2 Transformasi Data

bh	pcs	kg	kg
Bh		Kg	
pcs		rol	roll
Pcs		Rol	
lbr	sheet	lsn	
pak		Lsn	
Pak		ls	lesson
ktk		Ls	
Ktk	pack		
set	set	btl	bottle
psg	pair	Dus	
Psg		bks	box
ps			
Ps		Klg	

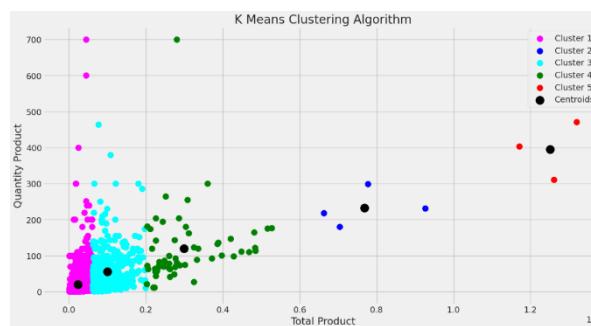
3.3 Modelling

Pada tahap ini dilakukan analisis data menggunakan algoritma K-Means, karena algoritma ini relatif mudah diimplementasikan dan cukup efisien dalam mengelompokkan data. K-Means bekerja dengan pendekatan iteratif sederhana untuk menemukan centroid dan menugaskan titik data ke cluster terdekat. Jumlah cluster ditentukan menggunakan metode Elbow.



Gambar 3 Grafik Elbow Methode

Gambar 2 menunjukkan grafik hasil dari Elbow Methode, dimana sumbu X menunjukkan jumlah kluster (*number of clusters*) sedangkan sumbu Y menunjukkan nilai *Within-Cluster Sum of Squares* (*WCSS*). Berdasarkan grafik tersebut, terlihat sekitar *cluster* ke-4 atau ke-5, menunjukkan bahwa jumlah *cluster* yang ideal untuk data ini kemungkinan adalah 4 atau 5. Dengan memilih jumlah *cluster* pada titik ini, kita dapat mencapai keseimbangan antara jumlah *cluster* yang cukup untuk memisahkan data dan menghindari *cluster* yang berlebihan atau tidak relevan.



Gambar 4 Visualisasi Data Clustering

Data diatas menunjukkan bahwa *cluster* dibagi menjadi 5 *cluster*. *cluster* pertama ditunjukkan oleh warna pink, *cluster* kedua ditunjukkan oleh warna biru, *cluster* ketiga ditunjukkan oleh warna hijau tosca, *cluster* keempat ditunjukkan oleh warna hijau gelap, dan *cluster* kelima ditunjukkan oleh warna merah.

3.4 Evaluation

Proses evaluasi dilakukan menggunakan pengujian *Silhouette Score*. Pengujian ini untuk mengukur seberapa baik kualitas *cluster* yang dikelompokkan[9]. Nilai *Silhouette Score* yang tinggi menandakan bahwa *clustering* tersebut memiliki hasil yang baik, sedangkan jika nilai *Silhouette Score*

yang rendah maka menunjukkan bahwa *clustering* tersebut memiliki penyebaran yang tidak homogen. Berdasarkan hasil *Silhouette Score* dari data yang telah diolah mendapatkan hasil 0.69, dimana hasil tersebut menunjukkan bahwa model *clustering* cukup baik. Secara umum, nilai diatas 0.5 dianggap baik, jadi hasil tersebut merupakan hasil yang positif. Selain itu pengujian juga dilakukan menggunakan *Davies Bouldin Index* (DBI) dengan hasil 0.428 yang menunjukkan bahwa *clustering* yang digunakan memiliki kualitas *clustering* yang baik dikarenakan dalam DBI jika nilai dibawah 1 maka hasilnya memiliki kualitas yang baik.

4. SIMPULAN

Berdasarkan hasil analisis dan hasil pengolahan data penjualan toko bangunan Pekanbaru Jaya maka dapat diambil kesimpulan sebagai berikut :

1. Metode *K-Means Clustering* dapat diterapkan pada pengolahan data penjualan Toko Bangunan Pekanbaru Jaya dalam mengelompokkan hasil penjualan.
2. Hasil pengolahan data tersebut dapat dihasilkan 5 *cluster*, *cluster* 0 berjumlah 1957, *cluster* 1 berjumlah 4, *cluster* 2 berjumlah 432, *cluster* 3 berjumlah 59, dan *cluster* 4 berjumlah 3.
3. Pengujian validitas menggunakan nilai *Silhouette Score* dengan hasil 0.69 dan hasil nilai Davies Bouldin Index (DBI) 0.428 yang menunjukkan bahwa hasil tersebut memiliki kualitas *cluster* yang baik.

5. SARAN

Saran untuk peneliti selanjutnya untuk dapat menggunakan metode lain dalam pengolahan datanya untuk bisa mendapatkan hasil yang memiliki kualitas lebih baik lagi.

DAFTAR PUSTAKA

- [1] A. Nugraha, O. Nurdian, and G. Dwilestari, “PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT,” *JATI J. Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, Nov. 2022, doi: 10.36040/jati.v6i2.5755.
- [2] M. Rizki and M. Mulyawan, “PENERAPAN METODE K-MEANS CLUSTERING PADA DATA PENJUALAN OPTIK CHANTIKA,” *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1303–1307, Sep. 2023, doi: 10.36040/jati.v7i2.6562.
- [3] F. M. Sarimole and L. Hakim, “Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang: Klasifikasi,” *J. Sains Dan Teknol.*, vol. 5, no. 3, pp. 846–854, Feb. 2024, doi: 10.55338/saintek.v5i3.2709.
- [4] Y. Suhanda, I. Kurniati, and S. Norma, “Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik,” *J. Teknol. Inform. Dan Komput.*, vol. 6, no. 2, pp. 12–20, Sep. 2020, doi: 10.37012/jtik.v6i2.299.
- [5] Narayana Sakti Aji, Fauzan Natsir, and Siti Istianah, “Penentuan Penjualan Barang Berdasarkan Pengelompokan Produk dengan K-Means Clustering Metode CRISP-DM Pada CV Sembako Dina,” *J. ZETROEM*, vol. 5, no. 2, pp. 119–126, Oct. 2023, doi: 10.36526/ztr.v5i2.3041.
- [6] T. Wahyudi, N. Sa’adah, and D. Puspitasari, “Penerapan Metode K-Means Pada Data Penjualan Untuk Mendapatkan Produk Terlaris Di PT. Titian Nusantara Boga,” vol. 5, no. 1, 2023.
- [7] K. Soewardy and G. M. G. Bororing, “ANALISIS TINGKAT KEPUASAN PENGGUNA APLIKASI MARKETPLACE TOKOPEDIA MENGGUNAKAN ALGORITMA C4.5,” *J. Inform. Dan Bisnis*, vol. 11, no. 1, pp. 38–47, Sep. 2022, doi: 10.46806/jib.v11i1.883.
- [8] 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. Inform. Technol. Sci. BITS*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2697.
- [9] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput. JIMA-Ilk.*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.