

Penerapan Algoritma *K-Means* Untuk Mengklasifikasikan Penjualan Produk Dettol

Asih Andriani¹, Bella Cantika Sinaga², Dina Nur Hasana³

^{1,2,3} Mahasiswa Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal

¹ asihandriani982@gmail.com; ² bellacantikasinaga975@gmail.com; ³ dinanurhasana2003@gmail.com

* Email Koresponden

INFO ARTIKEL

Histori Artikel

Diterima: 26/Juli/2024

Ditinjau: 28/Juli/2024

Disetujui: 31/Juli/2024

ABSTRAK

PT. Everbright menjadi salah satu perusahaan distributor yang menjual berbagai macam produk yang salah satunya yaitu produk Dettol. Dalam mengoptimalkan stok dan pemasaran produk Dettol, perusahaan menghadapi kesulitan dalam mengidentifikasi penjualan produk yang laris serta yang kurang diminati oleh pelanggan. Data penelitian ini bersumber dari data transaksi penjualan sabun dettol pada bulan November sebanyak 77 data penjualan. Dengan penggunaan data mining, khususnya metode K-means Clustering, menjadi pendekatan yang relevan untuk memecahkan permasalahan tersebut. Tujuan dari penelitian ini adalah menghindari kelebihan persediaan yang tidak terjual serta memenuhi kebutuhan pelanggan yang beragam. Hasil dari penelitian ini menunjukkan visualisasi persebaran klaster produk sabun dettol dan pengelompokan berdasarkan tingkat penjualan. Terdapat data penjualan sebanyak 21 item barang yang paling laris terjual, 31 item barang yang laris terjual, dan 25 item barang yang kurang laris terjual sehingga perusahaan dapat melakukan persediaan barang berdasarkan yang laris terjual.

Kata Kunci: Sabun Dettol, Data Mining; K-means



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright ©2024 by Author. Published by PT Beranda Teknologi Academia

ABSTRACT

PT. Everbright has become one of the distributing companies selling various products, one of which is the Dettol product. In optimizing the stock and marketing of Dettol products, the company faces difficulties in identifying the best-selling products and those less favored by customers. This research data originates from the sales transaction data of Dettol soap in November, comprising 77 sales data. Through the use of data mining, particularly the K-means Clustering method, it becomes a relevant approach to solving this issue. The objective of this research is to avoid excess inventory that remains unsold while meeting the diverse needs of customers. The results of this study show the visualization of the distribution of Dettol soap product clusters and grouping based on sales levels. There were 21 best-selling item data, 31 well-selling item data, and 25 less popular item data, enabling the company to manage inventory based on the best-selling items.

Keywords: Dettol Soap, Data Mining, K-means

PENDAHULUAN

Perdagangan merupakan suatu jaringan global yang menghubungkan kebutuhan sehari-hari dengan barang-barang mewah melalui pertukaran yang beragam dan dinamis [1]. Tujuan dari perusahaan yaitu dengan menjalankan pemasaran dan pelayanan dengan baik dan tepat untuk mencapai target dari suatu perusahaan, serta menghindari sekecil mungkin hal-hal yang tidak efisien [2]. PT. Everbright merupakan sebagai salah satu PT distributor yang menjual berbagai macam produk seperti

kosmetik, sabun, baterai, susu, bahan dapur, minuman dan lain sebagainya. Salah satu produk yang banyak didistribusikan sebagai permintaan pasar yaitu sabun dengan merek Dettol. Sabun Dettol merupakan salah satu kebutuhan dalam pencegahan anti bakterial yang dapat melindungi kulit bakteri yang tidak terlihat. Sudah 80 tahun Dettol telah menjadi terdepan dalam perlindungan diri dari kuman penyakit yang menjadi produk kesehatan yang paling populer [3]. Tingginya jumlah permintaan sabun Dettol memungkinkan perusahaan dengan cermat dalam memonitor kebutuhan yang tersedia. Dengan begitu perusahaan dapat mengambil keputusan untuk meningkatkan ketersediaan barang, stok barang, dan kelengkapan barang guna menghindari kelebihan barang yang sama dan kurang diminati pelanggan

Namun terdapat permasalahan yang terjadi pada PT. Everbright dimana perusahaan tidak mengetahui penjualan dengan produk yang sangat laris, penjualan laris, dan penjualan kurang laris dan produk mana yang harus banyak disediakan di gudang. Penggunaan *data mining* menjadi proses yang digunakan yang memungkinkan dalam pengambilan keputusan dengan cepat dalam sebuah transaksi menjadi sebuah pengetahuan yang baru [4].

Data mining adalah proses semi otomatis yang menggunakan proses penggalian atau pencarian data dari tempat penyimpanan data yang besar seperti *database*, dalam pengolahan *data mining* sering kali menggunakan *machine learning*, matematika, statistik, dan teknik kecerdasan buatan untuk mengidentifikasi dan mengatur informasi yang berpotensi, berguna dan berwawasan luas [5][6]. Dalam kecerdasan buatan, matematika, teknik statistik dan *machine learning* merupakan suatu proses yang terdapat dalam *data mining*. Pada *data mining* terdapat berbagai metode yang digunakan seperti *naïve bayes*, *regresi linear*, *decision tree*, *K-means* dan *neural network* [7]. Pada masalah diatas adapun penggunaan metode yang dapat diambil yaitu metode *K-means*. Metode *K-means* adalah metode pengelompokan data berdasarkan karakteristik atau kategori yang sama, metode *K-means* digunakan untuk menentukan jumlah *cluster* (K), membuat *cluster* secara arbitrer, menempatkan data dalam *cluster* menggunakan jarak terpendek, menghitung pusat *cluster*, dan mengulangi langkah 2-4 hingga tidak ada data yang berpindah ke lokasi yang berbeda [8][9].

Penggunaan metode *K-means Clustering* dapat mengelompokkan dan memberikan informasi bagi PT. Everbright mengenai produk Dettol mana yang paling diminati oleh pelanggan agar produk tetap tersedia.

METODE

Penelitian ini menggunakan *machine learning* digunakan sebagai dasar keilmuan, mengolah dan menganalisis data dan menggunakan metode *K-means* sebagai hasil algoritma dalam pengelompokan data nantinya. *Machine learning* adalah salah satu aplikasi *AI (Artificial Intelligence)* untuk mencari dan menemukan pola yang unik dari kumpulan data yang berasal dari pengembangan sistem tanpa harus diprogram berulang kali [10]. *Machine learning* adalah mesin yang mempelajari tentang ilmu algoritma komputer yang dapat mengenali pola-pola beragam macam data untuk diolah [11]. Dari hasil penelitian sebelumnya dapat disimpulkan *Machine learning* adalah salah satu aplikasi dari kecerdasan buatan (*AI*) yang memanfaatkan mesin atau algoritma komputer untuk secara otomatis mencari, menemukan, dan memahami pola yang unik dari kumpulan data.

Berikut dibawah ini tahapan CRISP-DM (*Cross-Industry Standard Process for Data mining*). Langkah langkah pada *K-means Clustering* adalah pada tahap awal menentukan jumlah klaster yang diinginkan, yang direpresentasikan oleh nilai k. Setelah itu, langkah selanjutnya adalah melakukan inialisasi pusat klaster (*centroid*) sejumlah k. Inialisasi ini sering dilakukan secara acak dengan mengambil titik-titik awal dari data yang ada. Selanjutnya, perhitungan jarak antara setiap data *input* dengan masing-masing *centroid* dilakukan menggunakan rumus jarak *Euclidean*. Rumus ini mengukur jarak antara data dengan *centroid* nya dengan persamaan matematis yang dijelaskan. Berikut persamaan *Euclidian Distance* :

$$\sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \dots \dots \dots (1)$$

Ket : De = *Euclidian Distance*

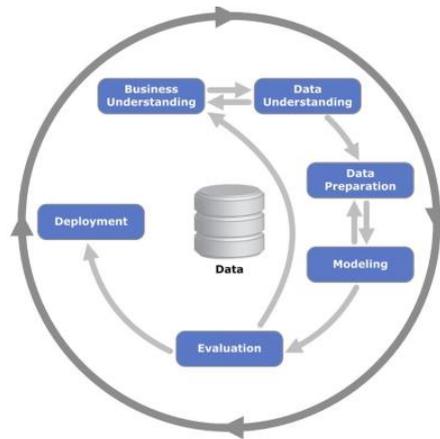


- i = banyaknya objek
- (x,y) = koordinat objek
- (s,t) = koordinat *centroid*

Setelah jarak dihitung, data kemudian diklasifikasikan ke dalam *cluster* yang memiliki *centroid* terdekat (dengan jarak terkecil). Langkah terakhir adalah memperbaharui nilai-nilai *centroid*. Nilai-nilai *centroid* baru dihitung dengan mengambil rata-rata dari seluruh data yang termasuk dalam klaster tersebut, sesuai dengan rumus yang disebutkan. Berikut rumus penentuan nilai *centroid* sebagai berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \dots \dots \dots (2)$$

- Ket : V_{ij} = *centroid* /rata-rata *cluster* ke-i untuk variabel ke-j
 N_i = jumlah data yang menjadi anggota *cluster* ke-i
 i,k = indeks dari *cluster*
 j = indeks dari variabel
 x_{kj} = nilai data ke-k yang ada di dalam *cluster* tersebut untuk variabel ke-j



Sumber [medium.easyread.co](http://medium.com/easyread.co)
 Gambar 1. Tahapan CRISP-DM

Data Understanding

Data penelitian ini meliputi kode produk, nama produk, berat barang, harga barang, stok barang, jumlah penjualan barang. Data yang digunakan adalah data transaksi penjualan selama 1 bulan yaitu pada bulan November 2023. Berikut ini gambaran data *understanding dataset* yang digunakan.

	Kategori Barang	Nama Barang	Berat	Harga	Stok	Terjual	Sisa	Presentasi
0	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	950 gr	65583	30	10	20	0.333333
1	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	410 gr	25536	24	17	7	0.708333
2	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	300 gr	24708	15	12	3	0.800000
3	Segar terus aktif terus Fresh	STAT Yuzu&Citrus	950 gr	65583	14	4	10	0.285714
4	Segar terus aktif terus Fresh	STAT Yuzu&Citrus	410 gr	25536	23	8	15	0.347826
...
72	Dettol Hand Sanitizer	Refresh	50 ml	9961	31	27	4	0.870968
73	Dettol Floor Cleancer	Pine	700 gr	9499	30	13	17	0.433333
74	Dettol Floor Cleancer	Lemon	700 gr	9499	26	17	9	0.653846
75	Dettol Floor Cleancer	Lavender	700 gr	9499	28	19	9	0.678571
76	Dettol Floor Cleancer	Apel	700 gr	9499	21	10	11	0.476190

77 rows x 8 columns

Gambar 2. Deskripsi Data

Data Preparation

Data Preparation adalah langkah penting dalam menangani perbaikan masalah yang ada pada data sebelum memasuki tahap pemodelan, memastikan bahwa hasil pemodelan yang dihasilkan menjadi lebih optimal [12]. Berikut dibawah ini proses dalam melakukan data preparation.

```
#melihat info data
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Kategori Barang  77 non-null     object
1   Nama Barang     77 non-null     object
2   Berat           77 non-null     object
3   Harga           77 non-null     int64
4   Stok            77 non-null     int64
5   Terjual         77 non-null     int64
6   Sisa            77 non-null     int64
7   Presentasi      77 non-null     float64
dtypes: float64(1), int64(4), object(3)
memory usage: 4.9+ KB
```

Gambar 3. Deskripsi Tipe Data

```
df.isnull().sum()

Kategori Barang    0
Nama Barang        0
Berat              0
Harga              0
Stok               0
Terjual            0
Sisa               0
Presentasi         0
dtype: int64
```

Gambar 4. Pengecekan Missing Value

Modelling

K-means Clustering adalah teknis pengelompokan data berdasarkan karakteristik yang sama dengan tujuan menemukan struktur tersembunyi dalam data tanpa label [13]. Teknik ini berprinsip pada penyusunan sejumlah K pusat massa atau rata-rata dari data yang ada. Pengelompokan data bertujuan untuk meminimalkan fungsi objektif yang telah ditetapkan, yang pada dasarnya berupaya mengurangi variasi di dalam setiap kelompok dan meningkatkan variasi antar kelompok [14].

Evaluation

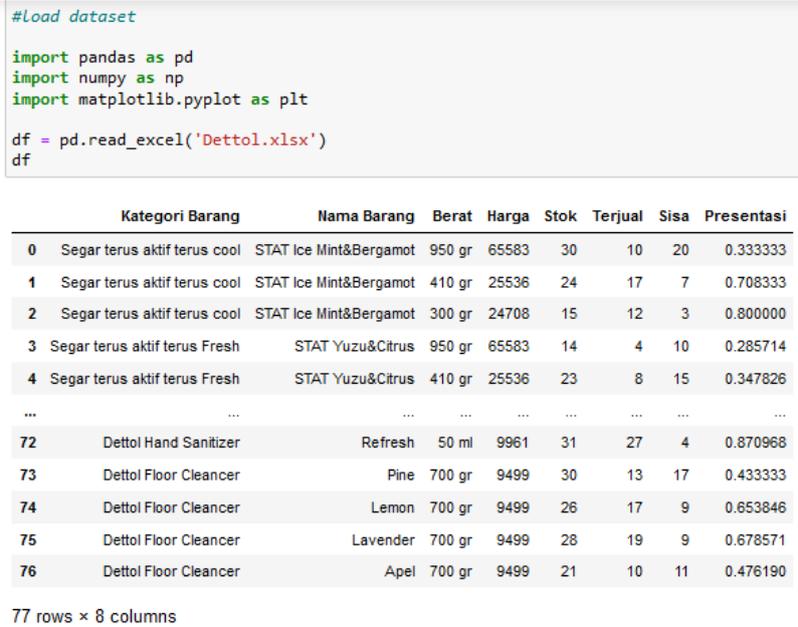
Dalam fase *evaluation*, langkah *evaluation* melibatkan pengujian terhadap kehandalan model *data mining* yang telah dikembangkan dan ditentukan apakah model tersebut cocok atau tidak untuk digunakan dalam proses bisnis [15].

Deployment

Hasil dari penerapan algoritma pada *data mining* di *interpretasikan* dalam bahasa pemrograman memudahkan PT. Everbright untuk membaca serta memahami informasinya dengan lebih cepat dan efektif.

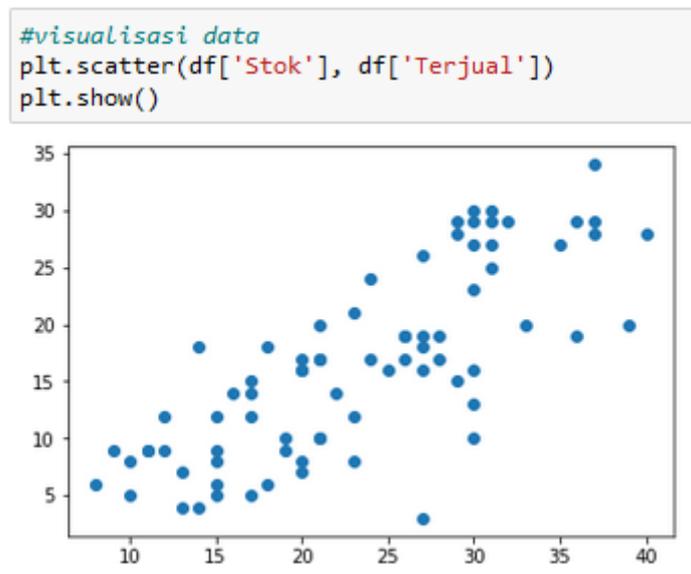
HASIL DAN PEMBAHASAN

Tahapan awal dalam penelitian ini adalah *import dataset* yang diperlukan dalam klasifikasi produk Dettol.



Gambar 5. *Dataset* Produk Dettol

Penentuan persebaran *cluster* dengan stok sebagai variabel x dan barang yang terjual sebagai variabel y.



Gambar 6. *Visualisasi* Data

Hasil persebaran titik *centorid* dan *cluster* dengan keterangan *cluster* 1 adalah penjualan kurang laris, *cluster* 0 adalah penjualan laris, dan *cluster* 2 adalah penjualan sangat laris.

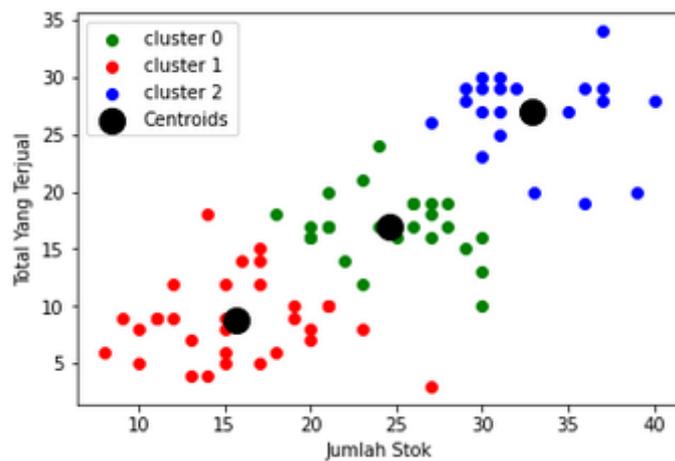
```
#visualisasi hasil cluster dan centroid
km = KMeans(n_clusters=3, random_state=0)
y_prediksi = km.fit_predict(df[['Stok','Terjual']])

df0 = df[df.cluster == 0]
df1 = df[df.cluster == 1]
df2 = df[df.cluster == 2]

plt.scatter(df0.Stok, df0['Terjual'], color='green', label = 'cluster 0')
plt.scatter(df1.Stok, df1['Terjual'], color='red', label = 'cluster 1')
plt.scatter(df2.Stok, df2['Terjual'], color='blue', label = 'cluster 2')

plt.scatter(km.cluster_centers[:, 0], km.cluster_centers[:, 1], s = 200, c = 'black', label = 'Centroids')

plt.xlabel('Jumlah Stok')
plt.ylabel('Total Yang Terjual')
plt.legend()
plt.show()
```



Gambar 7. Hasil Persebaran Cluster dan Centroid

Berikut dibawah ini tabel hasil cluster penentuan penjualan kurang laris, penjualan laris, dan penjualan sangat laris tiap produknya.

```
conditions = [
    (df['cluster']==0),
    (df['cluster']==1),
    (df['cluster']==2)]
choices = ['Penjualan Kurang Laris','Penjualan Laris','Penjualan Sangat Laris']
df['cluster'] = np.select(conditions, choices)
df
```

	Kategori Barang	Nama Barang	Berat	Harga	Stok	Terjual	Sisa	Presentasi	cluster
0	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	950 gr	65583	30	10	20	0.333333	Penjualan Kurang Laris
1	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	410 gr	25536	24	17	7	0.708333	Penjualan Kurang Laris
2	Segar terus aktif terus cool	STAT Ice Mint&Bergamot	300 gr	24708	15	12	3	0.800000	Penjualan Laris
3	Segar terus aktif terus Fresh	STAT Yuzu&Citrus	950 gr	65583	14	4	10	0.285714	Penjualan Laris
4	Segar terus aktif terus Fresh	STAT Yuzu&Citrus	410 gr	25536	23	8	15	0.347826	Penjualan Laris
...
72	Dettol Hand Sanitizer	Refresh	50 ml	9961	31	27	4	0.870968	Penjualan Sangat Laris
73	Dettol Floor Cleancer	Pine	700 gr	9499	30	13	17	0.433333	Penjualan Kurang Laris
74	Dettol Floor Cleancer	Lemon	700 gr	9499	26	17	9	0.653846	Penjualan Kurang Laris
75	Dettol Floor Cleancer	Lavender	700 gr	9499	28	19	9	0.678571	Penjualan Kurang Laris
76	Dettol Floor Cleancer	Apel	700 gr	9499	21	10	11	0.476190	Penjualan Laris

77 rows x 9 columns

Gambar 8. Hasil Cluster Tiap Produk

KESIMPULAN

Dengan penggunaan metode *K-means* pada data penjualan dettol, PT. Everbright dapat dengan mudah melihat hasil penjualan selama pada bulan November 2023. terdapat 3 warna persebaran *cluster* dengan ditandai dengan warna hijau (*cluster 0*), warna merah (*cluster 1*), dan warna biru (*cluster 2*). Untuk penjualan yang paling laris ditandai dengan warna biru, untuk penjualan yang laris ditandai dengan warna hijau, dan penjualan yang kurang laris ditandai dengan warna merah.

Dan hasil pengujian yang telah dilakukan terdapat 21 produk yang penjualan paling laris (*cluster 2*), 31 produk yang penjualan laris (*cluster 0*), dan 25 produk dengan penjualan kurang laris (*cluster 1*). Hasil dari pengelompokan tersebut diharapkan dapat menjadi pertimbangan pada PT. Everbright dalam menyediakan stok barang.

DAFTAR PUSTAKA

- [1] A. Tohawi, J. Iswanto, D. Fitrtotus, and S. Diyah, "Penerapan Etika Bisnis Islam dalam Perdagangan Bawang Merah di Pasar Sukomoro Nganjuk," vol. 05, no. 04, pp. 17814–17822, 2023.
- [2] S. Nurhaliza and T. N. Sari, "Pengaruh Kualitas Produk dan Citra Merek Terhadap Keputusan Pembelian Baterai ABC Pada PT. Everbright Medan," *J. Ilmu Manajemen, Ekon. dan Kewirausahaan*, vol. 1, no. 2, 2023.
- [3] D. Dirnaeni, "Pengaruh Kesan Kualitas, Citra Merek Dan Harga Terhadap Keputusan Pembelian Sabun Batang Anti Bakteri Dettol Di Kota Bekasi," *UG J.*, vol. 15, no. 01, pp. 8–18, 2021.
- [4] H. Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering," *J. Inform. Upgris*, vol. 5, no. 1, 2019, doi: 10.26877/jiu.v5i1.3091.
- [5] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 1–7, 2022.
- [6] 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.
- [7] M. Althaf and P. Perkasa, "Analisis Probabilitas Bencana Alam dengan Penerapan Data Mining Menggunakan K-Means dan Linier Regression," vol. 2, 2023.
- [8] N. Nugroho and F. D. Adhinata, "Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170–179, 2022, doi: 10.34148/teknika.v11i3.502.
- [9] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [10] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.
- [11] M. S. Dios Kurniawan, *Pengenalan Machine Learning dengan Python*. Elex Media Komputindo, 2022.
- [12] Y. Suhandi, 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, 2020, doi: 10.37012/jtik.v6i2.299.
- [13] E. M. Fitri, R. R. Suryono, and A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz," *J. Komputasi*, vol. 11, no. 2, pp. 157–168, 2023.
- [14] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi:



- 10.33365/jtk.v15i2.1162.
- [15] Muhammad Yamin Nurzaman and B. Nurina Sari, “Implementasi K-Means Clustering Dalam Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>