



# Analisis Produksi Cabai Rawit Indonesia Menggunakan Algoritma K-Means Clustering

Syaiful Hasan Abdullah<sup>1</sup>, Zaehol Fatah<sup>2</sup>

<sup>1,2</sup> Teknologi Informasi <sup>2,3</sup> Fakultas Sains dan Teknologi, Universitas Ibrahimy Indonesia.

e-mail : [syaifulhasanabdullah12@gmail.com](mailto:syaifulhasanabdullah12@gmail.com)<sup>1</sup>

**Abstract.** Indonesia, with its fertile soil, holds significant potential in agriculture, primarily due to its geographic location in a tropical region characterized by high rainfall. These conditions support the optimal growth of various crops, establishing Indonesia as an agrarian nation. A large portion of the population relies on agriculture for their livelihood. One of the key agricultural commodities is bird's eye chili, known for its distinctive spiciness and frequent use in cooking. In addition to enhancing flavor with its intense heat, bird's eye chili also adds visual appeal to dishes and stimulates appetite. These characteristics make it an essential element in Indonesian cuisine. Bird's eye chili is one of Indonesia's primary agricultural commodities. Therefore, it is crucial to conduct an in-depth study of its production levels, including efforts to optimize output by analyzing the factors that influence its production. Data mining methods offer the ability to uncover hidden and interesting patterns within datasets. Furthermore, these methods can address user errors in data processing. One effective data mining approach for mapping or grouping similar data is clustering. Clustering provides a unique advantage over other data mining techniques as it can classify data without prior knowledge. This technique divides and groups data into clusters based on similar characteristics. Several algorithms are employed in clustering, with K-Means being one of the most popular.

**Keywords:** Bird's Eye Chili Production, Data Clustering, K-Means

**Abstrak.** Indonesia, dengan tanahnya yang subur, memiliki potensi besar dalam pertanian, terutama berkat letak geografisnya di wilayah tropis yang ditandai dengan curah hujan tinggi. Kondisi ini mendukung pertumbuhan berbagai jenis tanaman secara optimal, menjadikan Indonesia dikenal sebagai negara agraris. Sebagian besar penduduknya menggantungkan hidup pada sektor pertanian. Salah satu komoditas penting adalah cabai rawit, yang memiliki rasa pedas khas dan sering digunakan sebagai bahan masakan. Selain memberikan rasa pedas yang kuat, cabai rawit juga mempercantik tampilan hidangan dan mampu meningkatkan nafsu makan. Karakteristik ini menjadikan cabai rawit sebagai elemen penting dalam kuliner Indonesia. Cabai rawit adalah salah satu komoditas utama di Indonesia. Oleh karena itu, penting untuk melakukan kajian mendalam terkait tingkat produksinya, termasuk upaya mengoptimalkan hasil produksi melalui analisis berbagai faktor yang memengaruhinya. Metode data mining memungkinkan penggalian pola-pola tersembunyi yang menarik dalam kumpulan data. Selain itu, metode ini dapat digunakan untuk mengurangi kesalahan pengguna dalam proses pengolahan data. Salah satu pendekatan data mining yang efektif untuk memetakan atau mengelompokkan data serupa adalah klastering. Klastering memiliki kelebihan unik dibandingkan metode lain karena mampu mengklasifikasikan data tanpa memerlukan pengetahuan awal. Teknik ini membagi data menjadi kelompok-kelompok berdasarkan kemiripan karakteristik. Ada berbagai algoritma yang digunakan dalam klastering, dan salah satu yang paling populer adalah algoritma K-Means..

**Kata kunci:** Produksi Cabai Rawit, Pengelompokan Data, K-Means

## 1. LATAR BELAKANG

Cabai rawit (*Capsicum frutescens* L) merupakan salah satu komoditas pertanian yang memiliki peran penting dalam perekonomian dan kehidupan masyarakat Indonesia.[1] Sebagai bahan baku utama dalam berbagai masakan tradisional, cabai rawit memiliki permintaan yang tinggi di pasar domestik. Namun, produksi cabai rawit sering kali mengalami fluktuasi akibat berbagai faktor, seperti perubahan iklim, penggunaan lahan, serta metode pertanian yang kurang optimal. Produksi yang tidak stabil seringkali

menyebabkan gejolak harga di pasaran yang berimbas pada konsumen dan petani. Oleh karena itu, upaya untuk memahami pola produksi cabai rawit menjadi sangat penting dalam mendukung stabilitas pasokan dan harga.[2]

Seiring dengan perkembangan teknologi, analisis berbasis data semakin digunakan dalam bidang pertanian. Clustering telah menjadi instrumen yang valid untuk memecahkan masalah kompleks ilmu komputer dan statistic.[3] Teknik clustering mencakup beberapa algoritma, salah satu yang paling populer adalah K-Means Salah satu metode yang relevan untuk memahami pola produksi.[4] Algoritma ini dapat mengelompokkan data berdasarkan karakteristik yang mirip, sehingga memberikan wawasan baru dalam manajemen produksi. Dalam konteks ini, penelitian ini bertujuan untuk mengelompokkan data produksi cabai rawit di Indonesia pada tahun 2021-2023 dengan menggunakan algoritma clustering K-Means. Dengan pendekatan ini, diharapkan dapat diperoleh informasi yang lebih mendalam mengenai pola produksi berdasarkan wilayah atau faktor lainnya.

Penelitian ini akan memberikan kontribusi penting dalam pengambilan keputusan terkait kebijakan pertanian, khususnya untuk memaksimalkan produksi cabai rawit di berbagai wilayah di Indonesia. Pengetahuan yang dihasilkan dari penelitian ini dapat digunakan oleh petani, pengambil kebijakan, serta pihak-pihak yang terlibat dalam rantai pasokan cabai rawit untuk merancang strategi yang lebih efektif dalam menghadapi tantangan produksi. Beberapa penelitian sebelumnya telah mengidentifikasi faktor-faktor yang memengaruhi produksi cabai, seperti ketersediaan lahan, penggunaan pupuk, dan tenaga kerja.[5] Namun, penelitian yang memanfaatkan metode clustering untuk analisis produksi masih relatif terbatas, sehingga penelitian ini diharapkan dapat mengisi kekosongan tersebut dan memberikan wawasan baru dalam literatur terkait.

## **2. KAJIAN TEORITIS**

Bagian ini menguraikan teori-teori relevan yang mendasari topik penelitian dan memberikan ulasan tentang beberapa penelitian sebelumnya yang relevan dan memberikan acuan serta landasan bagi penelitian ini dilakukan. Jika ada hipotesis, bisa dinyatakan tidak tersurat dan tidak harus dalam kalimat tanya.

## **3. METODE PENELITIAN**

Dalam membuat penelitian ada beberapa tahapan yang dilakukan, tahapan penelitian dapat dilihat pada gambar dibawah ini.



**Gambar 1. Tahapan Penelitian.**

- a. Studi Literatur, pada tahap ini dilakukan pencarian, mengumpulkan dan menganalisa data dan metode yang akan digunakan pada penelitian ini.[6]
- b. Data yang digunakan pada penelitian ini diperoleh dari <https://www.bps.go.id/id/statistics-table/2/NjEjMg==/produksi-tanaman-sayuran.html>.
- c. Pengelompokan Data (clustering) dengan Algoritma K-Means  
Pengelompokan Data (*clustering*) adalah metode penting dalam analisis data yang digunakan untuk mengelompokkan berbagai objek ke dalam cluster atau kelompok tertentu berdasarkan karakteristik yang serupa. Pada proses clustering, objek-objek di dalam satu cluster memiliki kemiripan yang lebih tinggi dibandingkan dengan objek-objek yang berada di cluster lain.[7] Dengan melakukan pengelompokan ini, penelitian dapat memberikan wawasan yang lebih jelas kepada pengambil kebijakan atau petani tentang faktor-faktor yang memengaruhi produksi. Misalnya, apakah cluster dengan produksi rendah disebabkan oleh kondisi geografis, kurangnya teknologi, atau masalah distribusi. Hasil analisis ini dapat menjadi dasar untuk merumuskan strategi peningkatan produksi, alokasi sumber daya, atau pengembangan kebijakan agrikultur yang lebih efektif.

K-Means adalah metode pengelompokan data non hierarki yang membagi data menjadi dua atau lebih kelompok. Data dengan karakteristik serupa dikelompokkan bersama, sementara data dengan karakteristik berbeda ditempatkan di kelompok lain.[3] Metode k-means membagi data menjadi beberapa kelompok sehingga data dengan karakteristik yang sama berada pada cluster yang sama dan data dengan karakteristik yang berbeda berada pada cluster yang berbeda.[8] Algoritma K-Means merupakan salah satu dari algoritma yang banyak digunakan dalam pengelompokan karena kesederhanaan dan efisiensi dan diakui sebagai salah satu dari 10 algoritma data mining teratas oleh IEEE. dalam kasus produksi cabai rawit Indonesia, penting untuk memilih parameter yang tepat, seperti jumlah cluster dan fitur-fitur yang relevan

(contohnya volume produksi, atau penggunaan teknologi pertanian). Selain itu, evaluasi hasil clustering menggunakan metrik seperti silhouette coefficient atau elbow method sangat diperlukan untuk memastikan bahwa cluster yang dihasilkan mencerminkan pola yang bermakna dan dapat digunakan sebagai dasar pengambilan keputusan atau rekomendasi kebijakan agrikultur.



**Gambar 2. Algoritma K-Means**

#### d. Percobaan dan Pengujian

RapidMiner merupakan platform perangkat lunak yang andal dalam bidang data science dan machine learning. Perangkat ini menawarkan berbagai alat untuk proses persiapan data, pembuatan model, evaluasi, hingga penerapan. Dengan desain yang user-friendly, RapidMiner memungkinkan pengguna untuk merancang dan menguji berbagai model dengan mudah, bahkan tanpa keahlian pemrograman.[9]

## 4. HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Data

Dataset pada penelitian ini adalah data produksi cabai rawit di Indonesia yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia. Data tersebut diunduh dalam format Excel dengan ekstensi (.xls). Dataset mencakup data produksi cabai rawit dari seluruh provinsi di Indonesia dalam kurun waktu 2021-2023. Data ini merupakan data mentah yang berisi atribut seperti nama provinsi, kabupaten/kota, luas panen (hektar), produktivitas (kuintal/hektar), jumlah produksi (ton), serta data pendukung lainnya. Tidak semua data terisi lengkap, masih terdapat beberapa wilayah yang belum memiliki data produksi yang valid, sehingga perlu dilakukan proses pemilahan dan pengolahan data terlebih dahulu agar dapat digunakan dalam analisis lebih lanjut.

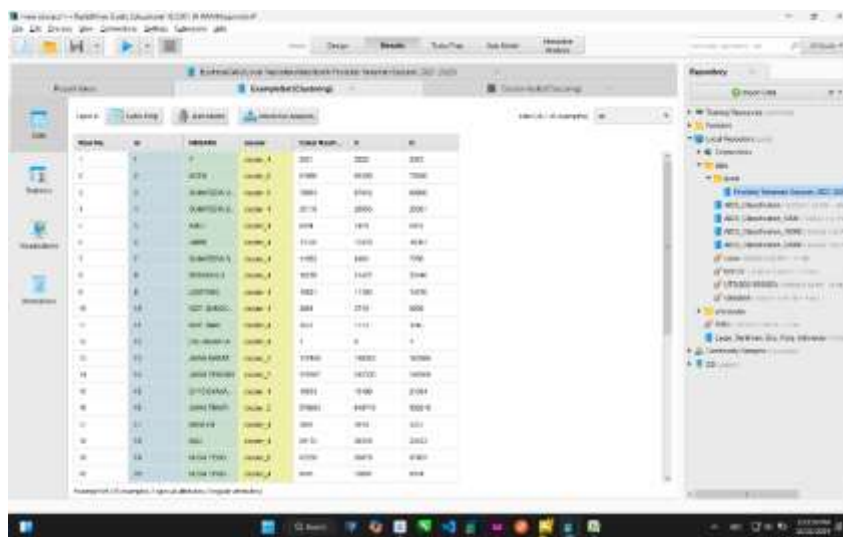
### 4.2 Praproses Data

Data yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia dilakukan pengecekan, pengolahan, dan disesuaikan dengan kebutuhan analisis menggunakan beberapa variabel. Dari 1.200 data produksi cabai rawit yang dikumpulkan dari 34 provinsi di Indonesia selama periode 2021-2023, hanya 850 data yang memenuhi kriteria validasi dan memiliki informasi yang lengkap untuk digunakan dalam proses clustering. Data yang telah melalui tahap praproses ini kemudian ditampilkan dalam tabel berikut.[10]

**Tabel 1. Tabel Data Produksi Cabai Rawit Indonesia**

NO	NEGARA	CABAI RAWIT (TON)		
		2021	2022	2023
1	Aceh	51686	65308	72920
2	Sumatera Utara	78663	87012	86880
3	Sumatera Barat	35118	26656	25081
..	.....			
32	Maluku Utara	3733	2873	3717
33	Papua Barat	3808	4107	3010
34	Papua	5054	4335	4556

### 4.3 Penerapan Clustering dan Algoritma K-Means



**Gambar 3. Pengelompokan Data Produksi Cabai Rawit Indonesia**

Gambar di atas menunjukkan hasil analisis clustering produksi cabai rawit Indonesia pada tahun 2021-2023 menggunakan algoritma k-means di aplikasi RapidMiner. Dataset terdiri dari nama wilayah, data produksi cabai rawit per tahun, dan hasil pengelompokan (cluster). Proses clustering mengelompokkan wilayah-wilayah dengan pola produksi yang serupa ke dalam beberapa cluster, seperti cluster\_0, cluster\_3, dan cluster\_4. Hasil ini membantu mengidentifikasi wilayah dengan tingkat produksi tinggi, sedang, atau rendah.

Analisis ini bertujuan untuk memahami pola distribusi produksi cabai rawit secara nasional. Sebagai contoh, wilayah seperti Aceh dan Sumatera Utara masuk ke dalam cluster tertentu berdasarkan tingkat produksinya. Informasi ini dapat digunakan untuk

merumuskan kebijakan strategis, seperti peningkatan produktivitas melalui subsidi, pelatihan petani, atau pengembangan infrastruktur di wilayah dengan produksi rendah. Dengan hasil clustering ini, penelitian memberikan wawasan penting dalam mendukung kebijakan dan pengelolaan produksi cabai rawit secara lebih efektif.

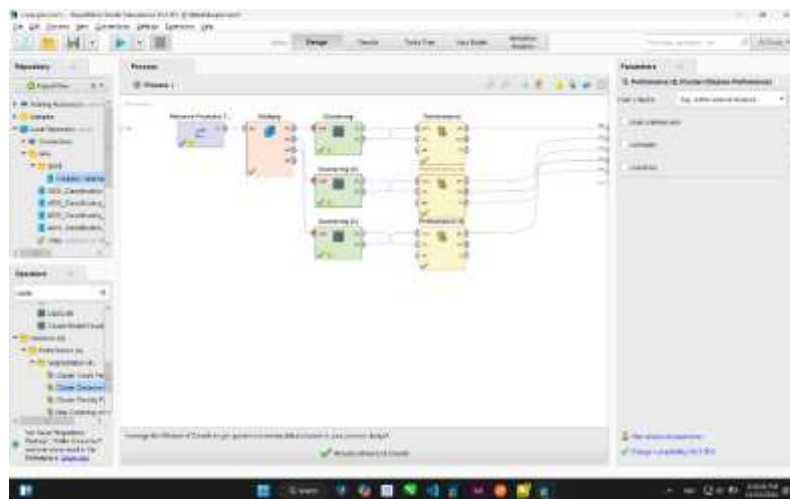


**Gambar 4. Cluster Model Produksi Cabai Rawit Indonesia**

Dari total 36 item yang dianalisis, mayoritas wilayah tergolong ke dalam Cluster 4 (29 item), yang menunjukkan pola produksi cabai rawit yang serupa di sebagian besar wilayah. Sementara itu, Cluster 0, 1, 2, dan 3 memiliki jumlah item yang jauh lebih sedikit, mencerminkan kelompok wilayah dengan pola produksi yang lebih spesifik atau unik.

Hasil ini mencerminkan distribusi produksi cabai rawit yang cenderung didominasi oleh kelompok tertentu, memungkinkan identifikasi wilayah dengan pola produksi yang menonjol. Pengelompokan ini penting untuk memahami pola produktivitas secara lebih terperinci dan untuk merumuskan kebijakan atau strategi yang sesuai di masing-masing cluster.

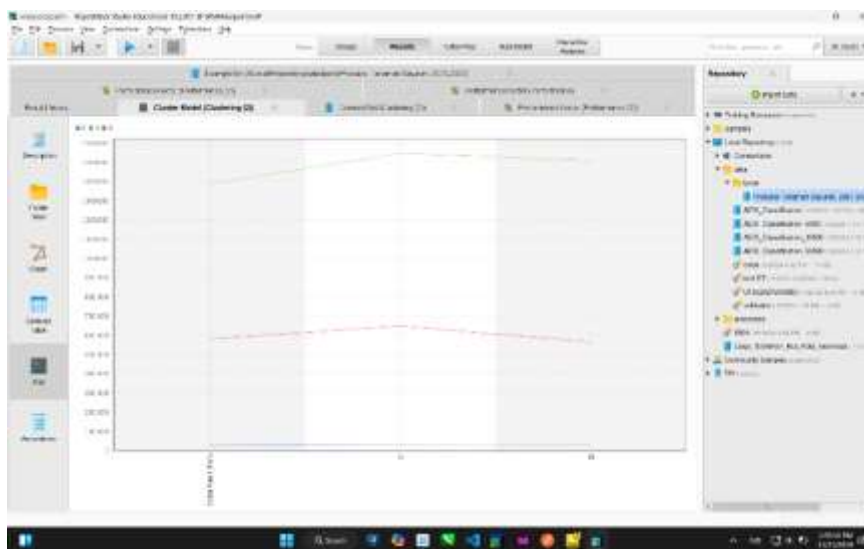
#### 4.4 Visualisai Hasil



### Gambar 5. Workflow Analisis Clustering Produksi Cabai Rawit

Gambar di atas menunjukkan alur proses analisis clustering produksi cabai rawit Indonesia menggunakan perangkat lunak RapidMiner. Proses ini dimulai dengan mengimpor dataset Produksi Tanaman Sayuran 2021-2023, yang kemudian digandakan menggunakan operator Multiply untuk memungkinkan eksperimen dengan parameter berbeda pada beberapa metode clustering. Tiga proses clustering diterapkan secara paralel dengan algoritma k-means, yang masing-masing menghasilkan cluster dengan konfigurasi berbeda.

Hasil clustering kemudian dievaluasi menggunakan operator Performance (Cluster Distance Performance), yang mengukur kinerja cluster berdasarkan jarak centroid antar cluster (rata-rata jarak dalam cluster). Parameter utama yang digunakan dalam evaluasi adalah Avg. within centroid distance, yang menilai seberapa kompak cluster yang dihasilkan. Proses ini memberikan informasi yang membantu memilih konfigurasi clustering terbaik, sehingga pola produksi cabai rawit dapat dipahami secara lebih mendalam. Alur ini memungkinkan peneliti untuk membandingkan hasil clustering dari berbagai konfigurasi secara efisien, mendukung analisis yang lebih akurat dan relevan dengan kebutuhan penelitian.



### Gambar 6. Visualisasi Hasil Cluster Model

Gambar di atas merupakan visualisasi hasil Cluster Model yang dihasilkan dari analisis clustering data produksi cabai rawit di Indonesia menggunakan algoritma k-Means. Grafik tersebut menunjukkan pola distribusi rata-rata dari data yang dikelompokkan ke dalam beberapa cluster berdasarkan kesamaan karakteristiknya. Sumbu x merepresentasikan nama variabel (dalam hal ini "Cabai Rawit"), sedangkan sumbu y merepresentasikan nilai produksi cabai rawit dalam ton.

Garis-garis dengan warna yang berbeda mewakili centroid dari masing-masing cluster, yang menggambarkan rata-rata nilai dalam kelompok tersebut. Dari visualisasi ini, dapat diidentifikasi adanya variasi signifikan antar cluster, di mana cluster tertentu menunjukkan produksi cabai rawit yang lebih tinggi dibandingkan cluster lainnya. Visualisasi ini mempermudah interpretasi pola clustering dan mendukung pengambilan keputusan, misalnya untuk memahami wilayah mana saja yang memiliki karakteristik produksi cabai rawit yang serupa atau berbeda, sehingga dapat digunakan sebagai dasar dalam perencanaan strategi peningkatan produksi cabai rawit secara nasional.

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengelompokkan pola produksi cabai rawit di Indonesia pada 2021-2023 menggunakan algoritma k-Means, sesuai dengan tujuan yang dinyatakan sebelumnya. Hasilnya menunjukkan pengelompokan wilayah berdasarkan pola produksi yang serupa, dengan sebagian besar wilayah tergabung dalam satu cluster utama.

Penelitian ini memberikan wawasan penting untuk mendukung kebijakan pertanian, seperti peningkatan produktivitas di wilayah dengan produksi rendah melalui subsidi atau pelatihan. Penerapan k-Means juga membuktikan efektivitasnya dalam mengidentifikasi pola data kompleks dan validasi menggunakan metrik jarak antar cluster memastikan hasil yang akurat.

## UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua saya, H. Sukaryo dan Hj. Zulaiha, atas doa dan dukungannya, serta kepada dosen pembimbing saya, Bapak Zaehol Fatah, M.Kom, atas bimbingan dan arahnya dalam penyusunan jurnal ini. Semoga segala kebaikan yang diberikan mendapatkan balasan terbaik dari Allah SWT.

## DAFTAR REFERENSI

- [1] N. Nurhikmah, I. Rosada, and I. Hasan, "Analisis Produksi Dan Pendapatan Usahatani Cabai Rawit Di Kelurahan Malakke, Kecamatan Belawa, Kabupaten Wajo," *Wiratani J. Ilm. Agribisnis*, vol. 2, no. 2, pp. 82–91, 2019, doi: 10.33096/wiratani.v2i2.37.
- [2] J. Piri, R. M. Juliana, and P. L. R. Yolanda, "Analisis Faktor-Faktor Yang Mempengaruhi Produksi Usahatani Cabai Di Desa Tambelang Kecamatan Maesaan Kabupaten Minahasa Selatan," *Agrirud*, vol. 4, no. 1, pp. 133–141, 2022.
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [4] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, "Klastering Sayuran Unggulan Menggunakan Algoritma K-Means," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 567–579, 2022, doi: 10.28932/jutisi.v8i3.5277.



- [5] A. D. Zahara T, N. S. Wisnujati, and E. Siswati, “ANALISIS PRODUKSI DAN PRODUKTIVITAS CABAI RAWIT (*Capsicum frutescens* L) di INDONESIA,” *J. Ilm. Sosio Agribis*, vol. 21, no. 1, pp. 18–29, 2021, doi: 10.30742/jisa21120211345.
- [6] R. Y. Hayuningtyas, I. Darwati, T. Informasi, U. N. Mandiri, U. Bina, and S. Informatika, “DOI: 10.37600/tekinkom.v7i1.1327,” vol. 7, pp. 25–32, 2024, doi: 10.37600/tekinkom.v7i1.1327.
- [7] M. B. Uddin and Z. Fatah, “Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas,” vol. 2, no. November, pp. 86–94, 2024.
- [8] L. Azzahra and Amru Yasir, “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–10, 2024, doi: 10.70340/jirsi.v3i1.88.
- [9] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.
- [10] I. Sentra, I. Berbasis, and G. Map, “G-Tech : Jurnal Teknologi Terapan,” vol. 8, no. 3, pp. 2112–2121, 2024.