

## Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means

Cellia Auzia Nugraha<sup>1</sup>, Muhammad Afrizal Kesuma<sup>\*2</sup>, Oktari Indi Cahyani<sup>3</sup>, Masna Wati<sup>4</sup>,  
Haviluddin<sup>5</sup>

<sup>1,2,3,4,5</sup> Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman, Samarinda, Indonesia 75119

Email: <sup>1</sup>[celliaauzianugraha@gmail.com](mailto:celliaauzianugraha@gmail.com), <sup>2,\*</sup>[afrizalKesuma3.0@gmail.com](mailto:afrizalKesuma3.0@gmail.com), <sup>3</sup>[oktariaindi@gmail.com](mailto:oktariaindi@gmail.com),  
<sup>4</sup>[masnawati@fkti.unmul.ac.id](mailto:masnawati@fkti.unmul.ac.id), <sup>5</sup>[haviluddin@unmul.ac.id](mailto:haviluddin@unmul.ac.id)

**Abstrak.** Cabai rawit (*Capsicum frutescens* L.) merupakan komoditas penting di Indonesia dengan permintaan yang tinggi. Namun, harga cabai rawit sering mengalami fluktuasi yang signifikan akibat ketergantungan pada musim, cuaca, serta kendala distribusi. Tujuan penelitian ini yaitu untuk mengelompokkan harga cabai rawit berdasarkan provinsi di Indonesia menggunakan algoritma Principal Component Analysis (PCA) dan K-Means. Data yang digunakan merupakan harga cabai rawit dari 34 provinsi di Indonesia pada periode Januari 2018 hingga Desember 2024. Analisis pengelompokan dilakukan dengan 3 variasi jumlah kluster, yaitu 2, 3, dan 4 kluster. Pengujian jumlah kluster menggunakan metode Silhouette Coefficient menunjukkan bahwa jumlah kluster paling optimal adalah 4 dengan nilai sebesar 0,511. Hasil penelitian ini menunjukkan bahwa pengelompokan harga cabai rawit dengan metode PCA dan K-Means dapat membantu dalam memahami pola harga di berbagai provinsi. Selain itu, hasil pengelompokan ini diharapkan dapat menjadi dasar bagi perencanaan distribusi dan pengendalian harga yang lebih efektif.

**Kata Kunci :** Cabai Rawit, PCA, K-Means, Pengelompokan, Silhouette Coefficient.

**Abstract.** A cayenne pepper (*Capsicum frutescens* L.) is an important commodity in Indonesia with high demand. However, the cayenne pepper prices often experiences significant fluctuations due to dependence on seasonality, weather, and distribution constraints. This study purposes to classify the cayenne pepper prices based on provinces in Indonesia using the Principal Component Analysis (PCA) and K-Means algorithms. The data used in this study are the cayenne pepper prices from 34 provinces in Indonesia in the period January 2018 to December 2024. A clustering analysis was conducted with 3 variations in the number of clusters, consist of 2, 3, and 4 clusters. Based on the experiment, the cluster accuracy by using the Silhouette Coefficient method shows that the most optimal number of clusters is 4 with a value of 0.511. The results of this study indicate that the clustering of cayenne pepper prices using PCA and K-Means methods can help to understanding price patterns in various Provinces. In addition, the results of this clustering are expected to be the basis for more effective distribution planning and price control.

**Keyword :** Cayenne Pepper, PCA, K-Means, Clustering, Silhouette Coefficient.

### PENDAHULUAN

Cabai rawit (*Capsicum frutescens* L.) merupakan salah satu komoditas hortikultura yang memegang peran penting dalam konsumsi rumah tangga dan perekonomian Indonesia. Permintaannya relatif tinggi sepanjang tahun karena merupakan bahan pokok dalam banyak hidangan khas Nusantara [1]. Meski begitu, harga cabai rawit di Indonesia sering kali tidak stabil dan berbeda jauh antar daerah.

Data dari Pusat Informasi Harga Pangan Strategis (PIHPS) menunjukkan adanya ketimpangan harga yang mencolok antar provinsi. Contohnya, pada Juli 2021, harga cabai rawit di Maluku mencapai Rp67.500/kg, sedangkan rata-rata nasional hanya Rp41.244/kg [2]. Di sisi lain, provinsi seperti Jawa Timur yang merupakan daerah sentra produksi, memiliki harga yang jauh lebih rendah. Perbedaan harga ini tidak hanya mencerminkan ketidakseimbangan pasokan dan permintaan, tetapi juga menimbulkan tantangan dalam pemerataan akses dan stabilisasi harga.

Ketimpangan tersebut tidak dapat diatasi hanya dengan kebijakan umum tanpa pemahaman wilayah mana yang memiliki karakteristik harga yang serupa. Oleh karena itu, diperlukan pendekatan



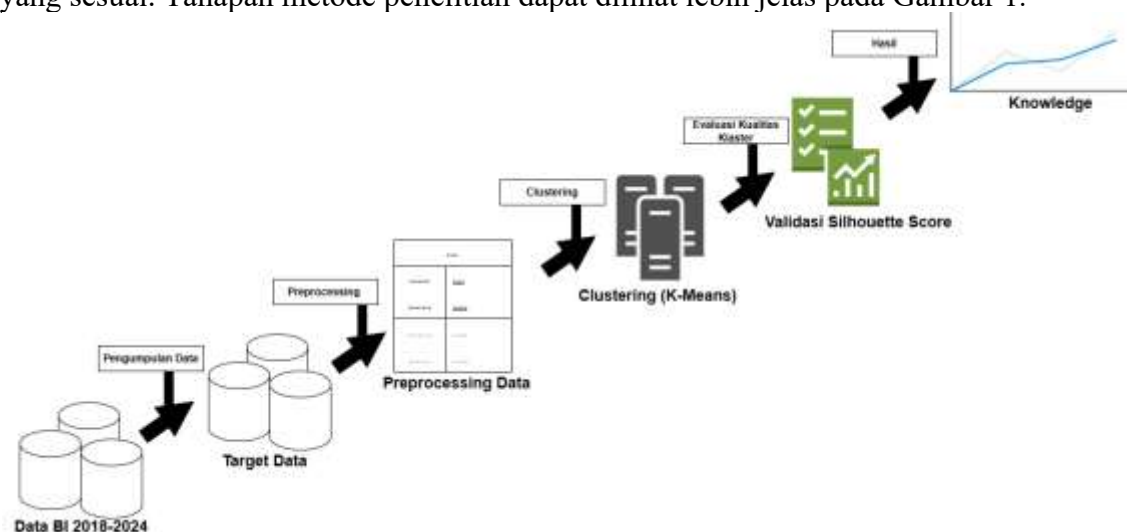
analisis data yang dapat membantu mengelompokkan provinsi berdasarkan pola harga cabai rawit yang dimiliki masing-masing provinsi tersebut.

Metode *clustering* menjadi salah satu solusi yang dapat digunakan untuk memetakan provinsi-provinsi di Indonesia berdasarkan kemiripan harga. Dengan menggunakan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data dan K-Means untuk melakukan pengelompokan. Penelitian yang menggunakan PCA dan algoritma K-Means telah dilakukan dalam banyak kasus. Rosyada dan Utari [3] menggabungkan PCA dengan K-Means untuk mereduksi dimensi variabel sebelum melakukan pengelompokan. Hasil dari penelitian ini menunjukkan bahwa penerapan PCA mampu memperbaiki kualitas pengelompokan dengan mengurangi *noise* pada data. Penelitian serupa oleh Putri dan Hayati [4] juga mendukung penggunaan PCA bersama K-Means, yang diterapkan untuk pengelompokan wilayah berdasarkan indikator pendidikan.

Penelitian ini bertujuan memberikan gambaran menyeluruh mengenai pola harga cabai rawit di berbagai provinsi [3], [5]. Hasil pengelompokan ini diharapkan dapat membantu dalam perencanaan distribusi, penentuan prioritas kebijakan, serta mendukung upaya pengendalian harga yang lebih terarah dan tepat sasaran [6].

## METODOLOGI PENELITIAN

Proses pengelompokan data dalam penelitian ini dilakukan menggunakan algoritma K-Means, yang merupakan salah satu metode clustering berbasis unsupervised learning. Sebelum menerapkan K-Means, dilakukan tahapan data preprocessing untuk memastikan data dalam kondisi yang optimal. Setelah itu, data yang telah diproses digunakan dalam pemodelan K-Means untuk menghasilkan kluster yang sesuai. Tahapan metode penelitian dapat dilihat lebih jelas pada Gambar 1.



**Gambar 1.** Tahapan Metode Penelitian

Gambar 1 menunjukkan alur kerja yang digunakan untuk menghasilkan pengetahuan dari data melalui analisis data menggunakan metode clustering K-Means. Proses dimulai dengan pengumpulan data dari Bank Indonesia (BI) selama periode 2018 hingga 2024. Data yang telah dikumpulkan kemudian disaring menjadi data target yang relevan. Selanjutnya, dilakukan tahap *preprocessing* untuk membersihkan dan mempersiapkan data agar siap dianalisis. Data yang telah diproses kemudian dianalisis menggunakan algoritma K-Means untuk mengelompokkan data ke dalam kluster-kluster yang memiliki kesamaan. Setelah proses *clustering*, metode Silhouette Score digunakan untuk mengevaluasi kualitas hasil kluster untuk memastikan pembagian kluster optimal. Terakhir, hasil dari proses ini dipahami dan digunakan untuk menghasilkan pengetahuan yang membantu dalam pengambilan keputusan.



## Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Pusat Informasi Harga Pangan Strategis Nasional Bank Indonesia. Data tersebut mencakup informasi mengenai harga cabai rawit di 34 provinsi Indonesia dari bulan Januari 2018 hingga Desember 2024. Data ini memiliki dimensi 34 x 68, dengan data keseluruhan berjumlah 2.312, dapat dilihat pada Tabel 1.

**Tabel 1.** Data Harga Cabe Rawit Tiap Provinsi Periode  
Januari 2018 – Desember 2024 (ribu/kg)

| No. | Provinsi             | 01/01/2018 | 01/02/2018 | 01/03/2018 | ... | 01/10/2024 | 01/11/2024 | 02/12/2024 |
|-----|----------------------|------------|------------|------------|-----|------------|------------|------------|
| 1   | Aceh                 | -          | 55.50      | 50.40      | ... | 47.90      | 40.40      | 38.40      |
| 2   | Sumatera Utara       | -          | 55.80      | 47.65      | ... | 40.60      | 34.35      | 33.75      |
| 3   | Sumatera Barat       | -          | 57.15      | 55.40      | ... | 44.65      | 39.65      | 37.90      |
| 4   | Riau                 | -          | 56.60      | 48.20      | ... | 43.40      | 38.30      | 33.50      |
| 5   | Kepulauan Riau       | -          | 49.00      | 39.70      | ... | 41.75      | 38.70      | 43.10      |
| 6   | Jambi                | -          | 43.35      | 42.95      | ... | 40.95      | 34.85      | 31.10      |
| 7   | Bengkulu             | -          | 43.25      | 35.65      | ... | 39.25      | 32.00      | 25.50      |
| 8   | Sumatera Selatan     | -          | 48.25      | 49.15      | ... | 54.60      | 50.95      | 37.00      |
| 9   | Kep. Bangka Belitung | -          | 42.75      | 49.70      | ... | 54.10      | 50.35      | 35.60      |
| 10  | Lampung              | -          | 39.40      | 38.15      | ... | 39.45      | 38.50      | 27.95      |
| 11  | Banten               | -          | 41.55      | 47.15      | ... | 43.70      | 33.35      | 31.30      |
| 12  | Jawa Barat           | -          | 41.30      | 44.65      | ... | 41.05      | 39.35      | 35.45      |
| 13  | DKI Jakarta          | -          | 46.25      | 51.25      | ... | 45.85      | 43.75      | 40.40      |
| 14  | Jawa Tengah          | 41.70      | 32.80      | 40.50      | ... | 32.70      | 30.95      | 24.25      |
| 15  | DI Yogyakarta        | -          | 36.50      | 42.15      | ... | 30.00      | 30.65      | 31.90      |
| 16  | Jawa Timur           | -          | 24.10      | 32.45      | ... | 26.55      | 23.50      | 20.80      |
| 17  | Bali                 | -          | 30.95      | 37.75      | ... | 39.20      | 30.15      | 26.65      |
| 18  | Nusa Tenggara Barat  | -          | 25.85      | 49.90      | ... | 26.40      | 23.60      | 22.05      |
| 19  | Nusa Tenggara Timur  | -          | 47.85      | 49.10      | ... | 51.30      | 53.20      | 52.05      |
| 20  | Kalimantan Barat     | -          | 55.35      | 64.85      | ... | 61.30      | 45.55      | 51.05      |
| 21  | Kalimantan Selatan   | -          | 37.30      | 46.40      | ... | 33.70      | 27.15      | 32.30      |
| 22  | Kalimantan Tengah    | -          | 42.55      | 60.65      | ... | 51.25      | 40.95      | 44.40      |
| 23  | Kalimantan Timur     | -          | 44.70      | 53.85      | ... | 49.20      | 42.80      | 45.65      |
| 24  | Kalimantan Utara     | -          | 55.65      | 54.40      | ... | 83.45      | 78.45      | 85.00      |
| 25  | Gorontalo            | -          | 32.00      | 55.00      | ... | 71.70      | 45.00      | 62.10      |
| 26  | Sulawesi Selatan     | -          | 23.60      | 29.10      | ... | 25.70      | 27.60      | 30.05      |
| 27  | Sulawesi Tenggara    | -          | 28.65      | 32.35      | ... | 45.15      | 53.45      | 44.40      |
| 28  | Sulawesi Tengah      | -          | 35.25      | 51.25      | ... | 48.15      | 46.90      | 61.60      |
| 29  | Sulawesi Utara       | -          | 33.25      | 44.00      | ... | 72.95      | 59.35      | 57.55      |
| 30  | Sulawesi Barat       | -          | 20.45      | 26.40      | ... | 29.60      | 27.70      | 31.90      |
| 31  | Maluku               | -          | 35.85      | 40.05      | ... | 70.35      | 56.90      | 57.85      |
| 32  | Maluku Utara         | -          | 46.25      | 51.25      | ... | 77.50      | 85.00      | 82.50      |
| 33  | Papua                | -          | 34.65      | 48.75      | ... | 57.50      | 45.55      | 53.35      |
| 34  | Papua Barat          | -          | 37.50      | 48.75      | ... | 91.25      | 60.00      | 71.25      |

Data pada Tabel 1 memiliki simbol – yang menandakan bahwa ada data yang hilang, sehingga dibutuhkan tahapan data preprocessing sebelum melakukan pemodelan.



## Data Preprocessing

Sebelum data diolah, dilakukan serangkaian proses data preprocessing. Manfaat dari proses ini adalah untuk meningkatkan kualitas hasil pengelompokan data. Adapun penjelasan tahapan *data preprocessing* yang dilakukan adalah sebagai berikut.

1. Pembersihan Data (*Data Cleaning*), tahapan ini dilakukan untuk menangani nilai yang hilang pada data.
2. Reduksi Dimensi Data, tahapan ini dilakukan untuk mengatasi multikolinearitas pada data.

## Algoritma Principal Component Analysis (PCA)

PCA adalah metode yang digunakan untuk mereduksi dimensi dataset dengan merangkum variabel berdimensi tinggi menjadi lebih sedikit tanpa kehilangan informasi utama. Selain digunakan untuk reduksi dimensi, PCA juga dapat membantu menguji apakah variabel dalam dataset memiliki keterkaitan atau bersifat independen, sehingga mempermudah analisis lebih lanjut [7]. PCA umumnya digunakan ketika terjadi multikolinearitas pada data, yaitu ketika dua atau lebih variabel independen saling berkorelasi secara signifikan. Pengecekan multikolinearitas dapat dilakukan dengan beberapa cara, salah satunya adalah dengan visualisasi heatmap correlation berdasarkan koefisien Pearson. Heatmap adalah teknik visualisasi yang menyajikan kekuatan hubungan antar variabel dalam bentuk warna [8].

Dalam analisis korelasi, *heatmap correlation* digunakan untuk memetakan hubungan antar variabel berdasarkan koefisien Pearson, dengan nilai berkisar antara -1 hingga 1. Jika korelasi antarvariabel terlalu tinggi ( $|r| > 0.8$ ), dapat terjadi kolinearitas yang memengaruhi analisis lebih lanjut. Hal ini sejalan dengan penelitian sebelumnya [9], yang menyatakan bahwa nilai  $r$  yang mendekati -1 menunjukkan adanya hubungan linier negatif yang kuat, sedangkan nilai  $r$  yang mendekati 1 menunjukkan adanya hubungan linier positif yang kuat. Untuk mengatasinya, dapat diterapkan PCA.

## Algoritma K-Means Clustering

Algoritma K-Means adalah metode pengelompokan data non-hierarki dalam kategori *Unsupervised Learning* yang membagi data menjadi dua atau lebih kelompok berdasarkan prinsip partisi. Data dengan karakteristik serupa dikelompokkan bersama, sementara yang berbeda ditempatkan di kelompok lain, dengan setiap kelompok memiliki titik pusat (*centroid*) yang mewakili karakteristiknya [10], [11], [12]. Dalam penelitian ini, metode *Euclidean* digunakan untuk menentukan keanggotaan kluster menggunakan Persamaan 1.

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]} \quad (1)$$

Dimana,  $d$  adalah jarak Euclidean;  $(x_1, y_1)$  adalah koordinat titik pertama;  $(x_2, y_2)$  adalah koordinat titik kedua

Setelah seluruh data dikelompokkan, posisi centroid diperbarui berdasarkan rata-rata anggota kluster, dan proses ini berulang hingga hasilnya konvergen.

## Metode Silhouette Coefficient

Validasi hasil pengelompokan dalam penelitian ini dilakukan menggunakan *Silhouette Coefficient*, yang mengukur seberapa baik suatu objek berada dalam kluster yang tepat dibandingkan dengan kluster lainnya [13]. Nilai *Silhouette Coefficient* dihitung menggunakan Persamaan 2.

$$SC_{(p)} = \frac{b_p - a_p}{\max(a_p, b_p)} \quad (2)$$

Dimana  $a_p$  merupakan rata-rata selisih harga objek  $p$  dengan objek lain dalam kluster yang



sama, sedangkan  $b_p$  adalah rata-rata selisih harga objek  $p$  dengan kluster terdekat lainnya. Metode *Silhouette Coefficient* bernilai -1 hingga 1, di mana nilai mendekati 1 menandakan kluster sesuai, 0 menunjukkan objek berada di perbatasan kluster, dan nilai negatif menandakan kluster tidak tepat. Validasi dalam penelitian ini dilakukan dengan rata-rata *Silhouette Score* setelah penerapan PCA dan K-Means untuk menilai keakuratan pengelompokan harga cabai rawit berdasarkan pola perubahan harga di tiap Provinsi.

## HASIL DAN PEMBAHASAN

Data yang sudah dilakukan tahapan preprocessing kemudian dikelompokkan dengan menggunakan algoritma K-Means. Pengelompokan dilakukan dengan variasi jumlah kluster, yaitu 2 kluster, 3 kluster, dan 4 kluster untuk menentukan pengelompokan yang paling optimal.

### Hasil Preprocessing Data

Berikut adalah hasil data yang diperoleh setelah dilakukan tahapan preprocessing untuk penanganan nilai yang kosong dan reduksi dimensi dengan menggunakan PCA dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Preprocessing Data dengan PCA

| Provinsi | KU1         | KU2        | KU3        | KU4        |
|----------|-------------|------------|------------|------------|
| V1       | -0.67319112 | 4.0165376  | 0.9080508  | -1.9890839 |
| V2       | -0.72009385 | 3.8441734  | 0.4202255  | -1.8548207 |
| V3       | -0.83510353 | 5.2253747  | -0.2962472 | -2.3394179 |
| V4       | -0.72356024 | 4.3536793  | 3.0158861  | -2.1451299 |
| V5       | -0.69341062 | 4.4942990  | 3.6322835  | -1.7917429 |
| V6       | -0.72847373 | 4.3238271  | 5.1193254  | -2.5259774 |
| V7       | -0.82845980 | 3.8018942  | 2.9306363  | -3.5958469 |
| V8       | -0.70261255 | 3.8672614  | 4.0893523  | -3.1354614 |
| V9       | -0.42914448 | 2.4604232  | 6.0541090  | -1.7126301 |
| V10      | -0.69547059 | 3.3927827  | 5.8113440  | -2.2464232 |
| V11      | -0.66271344 | 3.7420162  | 6.2521452  | -1.8443502 |
| V12      | -0.64573377 | 3.5602865  | 6.3887262  | -1.9027838 |
| V13      | -0.69338346 | 3.6228506  | 6.2098607  | -2.1699523 |
| V14      | -0.70413316 | 2.8334111  | 5.9077742  | -2.4842034 |
| V15      | -0.60772026 | 3.3184017  | 6.5120739  | -1.8243768 |
| V16      | -0.68166737 | 3.0779842  | 6.4669520  | -2.1947732 |
| V17      | -0.66344646 | 3.0747047  | 6.3401612  | -2.0800367 |
| V18      | -0.55294179 | 2.4885139  | 7.0675438  | -1.5477764 |
| V19      | -0.07317513 | -1.3515869 | 5.3451684  | -1.9714531 |
| V20      | -0.37360374 | 4.1878046  | 4.7535944  | -0.5918129 |
| V21      | -0.49850679 | 1.8792355  | 6.6705341  | -1.6833309 |
| V22      | -0.39842401 | 1.7403298  | 6.4836138  | -1.6513896 |
| V23      | -0.33210846 | 1.2603662  | 6.4432717  | -0.8839502 |
| V24      | -0.15248101 | 0.7949673  | 6.3136586  | -2.5353108 |
| V25      | -0.26410167 | -3.9735479 | 4.1621775  | -3.8617398 |
| V26      | -0.50535966 | -0.0450472 | 6.6071102  | -2.5839460 |
| V27      | -0.26979691 | -3.3920021 | 3.7458407  | -2.6229715 |
| V28      | -0.19785737 | -2.4111289 | 4.6173199  | -2.4849250 |
| V29      | -0.15067029 | -4.2485924 | 3.2511908  | -3.7681643 |
| V30      | -0.22584266 | -1.1798211 | 5.9807838  | -2.6906426 |



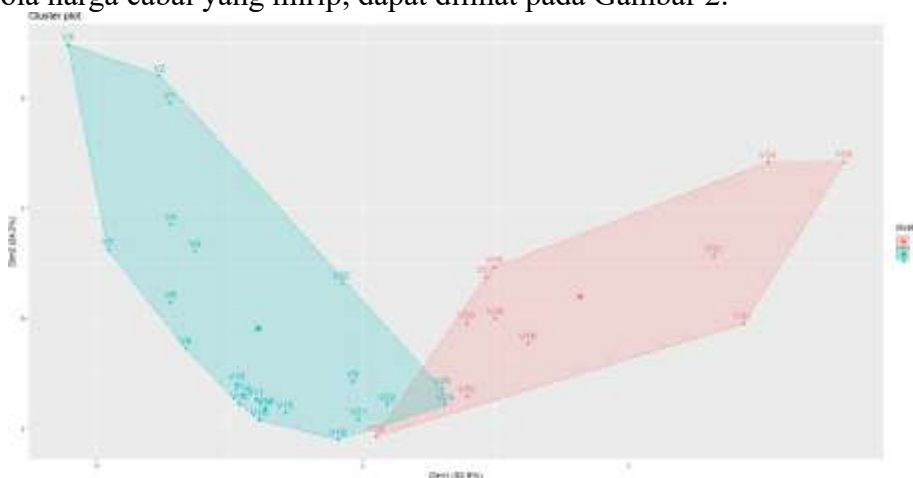
| Provinsi | KU1        | KU2        | KU3       | KU4       |
|----------|------------|------------|-----------|-----------|
| V31      | 0.22486818 | -1.7028279 | 6.2406188 | 1.5460376 |
| V32      | 0.34602598 | -1.8060154 | 4.6921633 | 0.2127392 |
| V33      | 0.37989301 | 0.2264535  | 4.7332115 | 5.4707989 |
| V34      | 0.21857458 | -1.0436531 | 4.0502641 | 3.7357139 |

### Hasil Pengelompokan dengan Variasi Jumlah Klaster

Data yang telah dilakukan tahapan preprocessing tersebut kemudian dimodelkan dengan menggunakan algoritma K-Means dengan jumlah klaster yang berbeda. Hasil dari penerapan algoritma K-Means dengan jumlah klaster yang berbeda menunjukkan pola pengelompokan yang beragam.

#### a. Pengelompokan dengan 2 Klaster

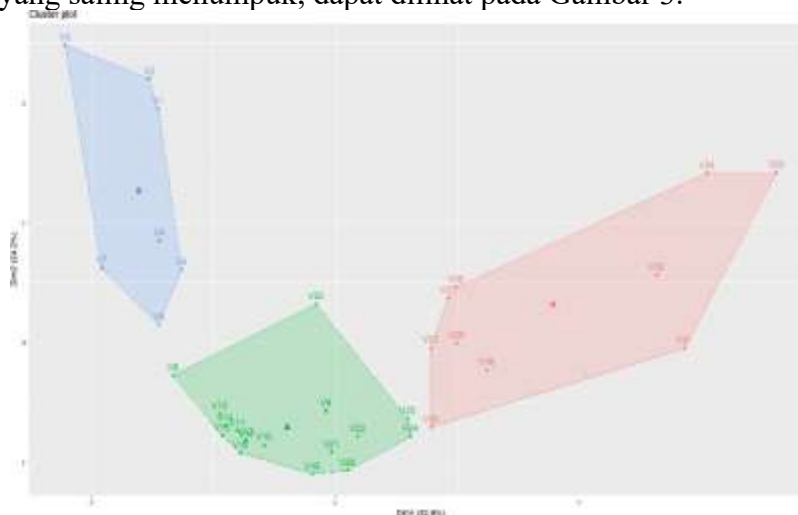
Pada pengelompokan 2 klaster, provinsi-provinsi terpisah ke dalam dua kelompok besar berdasarkan pola harga cabai yang mirip, dapat dilihat pada Gambar 2.



**Gambar 2.** Hasil Pengelompokan dengan 2 Klaster

#### b. Pengelompokan dengan 3 Klaster

Pada pengelompokan 3 klaster, pemisahan lebih spesifik terjadi dan tidak ada lagi hasil pengelompokan yang saling menumpuk, dapat dilihat pada Gambar 3.



**Gambar 3.** Hasil Pengelompokan dengan 3 Klaster

#### c. Pengelompokan dengan 4 Klaster

Pada pengelompokan 4 klaster, perbedaan antara kelompok semakin jelas, memungkinkan analisis lebih mendalam terhadap kriteria pengelompokan, dapat dilihat pada Gambar 4.





**Gambar 4.** Hasil Pengelompokan dengan 4 Klaster

Gambar 4 memperlihatkan hasil visualisasi dari dua komponen utama. Dimensi 1 merefleksikan nilai dari Komponen Utama 1, yang menggambarkan rata-rata harga cabai rawit di setiap provinsi. Di sisi lain, Dimensi 2 mencerminkan nilai dari Komponen Utama 2, yang menunjukkan tingkat fluktuasi atau ketidakstabilan harga cabai rawit seiring berjalannya waktu di masing-masing Provinsi. Sehingga dapat dianalisis bahwa terdapat 4 klaster yang masing-masing memiliki karakteristik sebagaimana tercantum dalam Tabel 3.

**Tabel 1.** Karakteristik Tiap Klaster

| Klaster | Karakteristik                                                        |
|---------|----------------------------------------------------------------------|
| 1       | Rata-rata harga cabai menengah dan fluktuasi harga rendah            |
| 2       | Rata-rata harga cabai relatif rendah dan fluktuasi harga sedang      |
| 3       | Rata-rata harga cabai cukup tinggi dan fluktuasi harga sangat tinggi |
| 4       | Rata-rata harga cabai sangat tinggi dan fluktuasi harga tinggi       |

### Hasil Uji Silhouette Coefficient

Untuk menentukan jumlah klaster terbaik, dilakukan pengujian berdasarkan Silhouette Coefficient dengan menggunakan rumus Persamaan 2, yang mengukur seberapa baik objek dalam klaster tersebut terpisah dari klaster lain, dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil Uji Silhouette Coefficient

| Klaster | Silhouette Score |
|---------|------------------|
| 2       | 0.4629001        |
| 3       | 0.4558199        |
| 4       | 0.5113710        |

Berdasarkan Tabel 4, terlihat bahwa jumlah klaster 4 memiliki nilai silhouette tertinggi, yang artinya jumlah klaster paling optimal untuk dilakukan pengelompokan.

### Visualisasi Hasil Klaster

Setelah menentukan jumlah klaster terbaik, hasil pengelompokan divisualisasikan dalam bentuk peta geografis untuk memberikan gambaran yang lebih jelas mengenai distribusi harga cabai di berbagai Provinsi dapat dilihat pada Gambar 5 dan hasil pengelompokan dapat dilihat pada Tabel



5.

**Gambar 5.** Visualisasi Hasil Klaster**Tabel 5.** Hasil Pengelompokan dengan 4 Klaster

| Klaster | Provinsi                                                                                                                                                                                                                            | Jumlah | Warna     |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----------|
| 1       | Kepulauan Bangka Belitung, Lampung, Banten, Jawa Barat, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Selatan | 15     | Oranye    |
| 2       | Riau, Kepulauan Riau, Jambi, Bengkulu, Sumatera Selatan, Kalimantan Barat                                                                                                                                                           | 6      | Hijau     |
| 3       | Aceh, Sumatera Utara, Sumatera Barat                                                                                                                                                                                                | 3      | Biru Laut |
| 4       | Nusa Tenggara Timur, Gorontalo, Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Utara, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat                                                                                        | 10     | Ungu      |

## KESIMPULAN

Penerapan kombinasi algoritma PCA dan K-Means dalam pengelompokan harga cabai rawit berdasarkan provinsi di Indonesia telah dipresentasikan. Pengujian dilakukan dengan variasi jumlah klaster  $k = 2$ ,  $k = 3$ , dan  $k = 4$ , dimana hasil evaluasi menggunakan metode Silhouette Coefficient menunjukkan bahwa 4 klaster merupakan jumlah paling optimal dengan nilai sebesar 0,5113710. Penggunaan 3 klaster menghasilkan nilai Silhouette Coefficient sebesar 0,4558199, sedangkan 2 klaster menghasilkan nilai 0,4629001.

Berdasarkan nilai Silhouette Coefficient tersebut, dapat disimpulkan bahwa pengelompokan dengan 4 klaster memiliki kualitas pengelompokan yang paling baik dalam menempatkan provinsi-provinsi dengan karakteristik harga cabai rawit yang serupa ke dalam satu klaster. Hasil pengelompokan ini dapat membantu dalam memahami pola distribusi harga cabai rawit di berbagai provinsi Indonesia, sehingga dapat menjadi dasar bagi pemangku kebijakan dalam merencanakan distribusi dan mengendalikan fluktuasi harga yang sering terjadi akibat faktor musim, cuaca, dan kendala distribusi. Metode PCA dan K-Means terbukti dapat menjadi alternatif analisis untuk mengidentifikasi pola harga komoditas strategis seperti cabai rawit.



## DAFTAR PUSTAKA

- [1] A. A. Cahya and R. H. Br Bangun, “Karakteristik Petani dan Kelayakan Usahatani Cabai Besar (*Capsicum Annum L*) dan Cabai Rawit (*Capsicum Frutescens L*) di Sumatera Utara,” *Agricore J. Agribisnis dan Sos. Ekon. Pertan. Unpad*, vol. 5, no. 1, pp. 49–58, 2020, doi: 10.24198/agricore.v5i1.27139.
- [2] C. F. Palembang and S. P. Palembang, “Pengelompokan Tingkat Harga Cabai Rawit Berdasarkan Provinsi Di Indonesia Menggunakan Algoritma K-Means Clustering (per Juli 2020-Juli 2021),” *Var. J. Stat. Its Appl.*, vol. 3, no. 2, pp. 48–60, 2021, doi: 10.30598/variancevol3iss2page48-60.
- [3] I. A. Rosyada and D. T. Utari, “Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering,” *Jambura J. Probab. Stat.*, vol. 5, no. 1, pp. 6–13, 2024, doi: 10.37905/jjps.v5i1.18733.
- [4] N. S. Putri and M. N. Hayati, “Pengelompokan Kabupaten/Kota di Kalimantan Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means dengan Optimasi Principal Component Analysis Grouping Regencies/Cities in Kalimantan Based on Educational Indicators Used K-Means Method with Principal ,” vol. 15, no. November, pp. 128–138, 2024, doi: 10.30872/eksponensial.v15i2.1373.
- [5] A. Novita, Ii. Ernawati, and N. Chamidah, “Klasterisasi Provinsi Di Indonesia Berdasarkan Produktivitas Komoditas Pangan Menggunakan Algoritma K-Means,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 462–471, 2022.
- [6] K. I. Hafsari, M. T. Al Hijrah, T. K. Wijayanti, and R. Kurniawan, “Implementasi Algoritma K-Means pada Pengelompokan Ketahanan Pangan di Indonesia Menurut Kabupaten / Kota,” vol. 2024, no. Senada, pp. 52–63, 2024.
- [7] S. Rahmi, P. Sirait, and E. S. Panjaitan, “Optimasi Klasifikasi Bayesian Network Melalui Reduksi Attribute Menggunakan Metode Principal Component Analysis,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 955–962, 2020, doi: 10.30865/mib.v4i4.2370.
- [8] A. Bengnga and R. Ishak, “Implementasi Seleksi Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Correlation Matrix with Heatmap,” *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 2, pp. 169–174, 2022, doi: 10.37905/jjee.v4i2.14403.
- [9] P. Putu, N. Ardhaneswari, I. W. C. Suwitra, and J. J. I. S. Siwirabuda, “Analisis Korelasi Pearson Dalam Menentukan Hubungan Harga Dengan Volume Penjualan Wardah Matte Lip Cream Pada Platform E-Commerce Shopee,” vol. 02, no. 02, pp. 151–156, 2024.
- [10] E. R. ARINI, “K-Means dan K-Medoids Clustering Pada Kasus Tindak Pidana Khusus,” *J. Sci. Nusantara*, vol. 4, no. 2, pp. 79–87, 2024, doi: 10.28926/jsnu.v4i2.1519.
- [11] N. Januari, S. H. Abdullah, Z. Fatah, T. Informasi, F. Sains, and U. Ibrahimy, “Analisis Produksi Cabai Rawit Indonesia Menggunakan Algoritma K- Means Clustering,” vol. 3, no. 1, pp. 66–74, 2025.
- [12] A. Hoerunnisa, G. Dwilestari, F. Dikananda, H. Sunana, and D. Pratama, “Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 103–110, 2024, doi: 10.36040/jati.v8i1.8249.
- [13] R. Fikri, A. Mushardiyanto, M. N. Laudza’Banin, K. Maureen, and H. Patria, “Pengelompokan Kabupaten/Kota di Indonesia Berdasarkan Informasi Kemiskinan Tahun 2020 Menggunakan Metode K-Means Clustering Analysis,” *Semin. Nas. Tek. dan Manaj. Ind.*, vol. 1, no. 1, pp. 190–199, 2021, doi: 10.28932/sentekmi2021.v1i1.76.

