

PERBANDINGAN ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING UNTUK PENGELOMPOKAN WILAYAH BERDASARKAN PRODUKSI DAN HARGA CABAI DI INDONESIA

Nikita Siswadi, Jefri Jaya, Teny Handhayani, Irvan Lewenusa

Teknik Informatika, Universitas Tarumanagara
Jl. Letjen S. Parman No.1 Jakarta Barat, Indonesia
tenyh@fti.untar.ac.id

ABSTRAK

Cabai rawit merupakan komoditas hortikultura strategis yang berdampak besar terhadap inflasi dan ketahanan pangan nasional, namun masih menghadapi ketimpangan harga dan produktivitas antarwilayah di Indonesia. Penelitian ini bertujuan untuk menganalisis distribusi spasial harga dan produktivitas cabai rawit hijau menggunakan metode *K-Means* dan *Agglomerative Hierarchical Clustering*. Dataset diperoleh dari PIHPS Nasional (harga bulanan 2018–2024) dan Kementerian Pertanian (produktivitas 2018–2023), mencakup 34 provinsi dan lebih dari 300 kabupaten/kota. Evaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI) menunjukkan konfigurasi optimal pada lima klaster untuk produktivitas (*Silhouette* 0,64; DBI 0,38) dan tiga klaster untuk harga (*Silhouette* 0,348; DBI 1,010). Hasil menunjukkan konsistensi spasial antara kedua metode dalam mengidentifikasi wilayah prioritas. Jawa Timur, NTB, dan Jawa Barat masuk kategori produktivitas tinggi dan harga rendah, sedangkan Papua dan Maluku menunjukkan pola sebaliknya. Visualisasi peta dan grafik tren mengungkap daerah urban seperti Bandung dan Bekasi memiliki harga tinggi, sementara sentra produksi seperti Blitar dan Jember menunjukkan kestabilan harga. Temuan ini menegaskan perlunya pendekatan spasial dalam kebijakan pangan nasional, dengan fokus pada penguatan distribusi dan kapasitas produksi di wilayah tertinggal guna menciptakan ketahanan pangan yang inklusif dan berkelanjutan.

Kata kunci : Cabai rawit, *K-Means*, *Agglomerative Hierarchical*, produktivitas, harga cabai rawit, *Silhouette*

1. PENDAHULUAN

Cabai rawit merupakan komoditas hortikultura yang memiliki peran strategis dalam perekonomian Indonesia, baik dari sisi konsumsi domestik maupun potensi ekspor. Tingginya permintaan cabai rawit, baik di tingkat rumah tangga maupun industri makanan, menjadikannya sebagai komoditas yang berdampak signifikan terhadap harga pangan nasional dan tingkat inflasi [1][2]. Namun demikian, pengembangan produksi dan pemasaran cabai rawit di Indonesia masih menghadapi berbagai tantangan, antara lain fluktuasi harga yang tajam, ketimpangan distribusi antarwilayah, serta ketidakseimbangan antara produksi dan permintaan akibat keragaman geografis, iklim, dan karakteristik pasar di berbagai daerah. *Support Vector Regression* berhasil digunakan untuk memprediksi harga cabai di Indonesia [3].

Beberapa penelitian sebelumnya menunjukkan bahwa cabai Indonesia memiliki keunggulan komparatif dan kompetitif dalam pasar global. Namun, penelitian-penelitian tersebut masih terbatas pada satu daerah/kabupaten dan agroekosistem, seperti lahan basah (sawah)/lahan kering serta dataran rendah/dataran tinggi. Oleh karena itu, penelitian ini bertujuan untuk membahas dengan cakupan yang lebih luas antara provinsi serta antarwilayah di Indonesia [4].

Peningkatan produktivitas pertanian sangat erat kaitannya dengan pemasaran. Dengan adanya pasar yang dapat menampung produk-produk pertanian tersebut dengan harga yang menguntungkan, gairah petani untuk memproduksi meningkat. Pemasaran

merupakan salah satu subsistem yang sangat penting dalam sistem agribisnis, di mana barang dan jasa didistribusikan dari sentra produksi ke sentra konsumsi untuk memenuhi kebutuhan konsumen dan memberikan keuntungan bagi produsen. Mengembangkan sistem pemasaran yang efisien dapat mendorong kelancaran arus produksi cabai rawit dari produsen ke konsumen, yang pada akhirnya akan menggairahkan petani untuk memproduksi [4].

Salah satu metode yang dapat digunakan untuk menganalisis pola ini adalah teknik *clustering*, yang memungkinkan pengelompokan wilayah berdasarkan karakteristik produktivitas dan pemasaran yang serupa. Analisis berbasis data, khususnya dengan teknik *clustering* seperti *K-Means* [5][6][7][8] dan *Hierarchical* [9], sangat berguna dalam mengidentifikasi pola produksi dan pemasaran cabai rawit di Indonesia. Metode ini memberikan wawasan mengenai bagaimana wilayah-wilayah di Indonesia dapat digolongkan berdasarkan karakteristik tersebut, yang pada gilirannya dapat membantu pengambilan keputusan yang lebih tepat dalam hal distribusi dan kebijakan pemasaran.

Penelitian ini bertujuan melakukan pengelompokan wilayah menggunakan metode evaluasi *Silhouette Coefficient* dan *Davies Bouldin Index* (DBI) terhadap provinsi di Indonesia berdasarkan dataset produktivitas, dengan variabel seperti provinsi, kabupaten dan produktivitas cabai rawit. Dataset pemasaran, seperti harga cabai rawit hijau dari tahun 2018 hingga 2024, juga akan dianalisis. Output dari penelitian ini memberikan

informasi mengenai pengelompokan provinsi dan kabupaten di Indonesia berdasarkan data produktivitas dan harga cabai rawit hijau, yang diharapkan dapat meningkatkan efisiensi distribusi dan memberikan wawasan bagi kebijakan pemerintah dalam pengembangan cabai rawit di Indonesia.

2. TINJAUAN PUSTAKA

Clustering adalah teknik untuk membagi dataset yang ada menjadi beberapa kelompok, dengan tujuan agar elemen-elemen dalam satu kelompok memiliki kemiripan yang tinggi, sementara elemen-elemen antar kelompok memiliki perbedaan yang lebih besar. Analisis kluster memainkan peran penting dalam berbagai bidang seperti *data mining*, pengenalan pola, klasifikasi pola, penemuan informasi, kuantisasi vektor, dan kompresi data. Selain itu, *clustering* juga memiliki aplikasi yang tidak terhindarkan dalam sektor-sektor seperti pemasaran, fisika, biologi, geografi, dan geologi.

Algoritma *K-Means*, yang merupakan algoritma berbasis *centroid*, termasuk dalam kategori *clustering* non-hierarkis. Keunggulan utama dari algoritma *K-Means* terletak pada kesederhanaan dan efisiensinya. Dalam algoritma *K-Means*, jumlah kluster k dan nilai *centroid* pertama kali ditentukan [10]. *K-Means* membagi dataset yang ada ke dalam k kluster berdasarkan jarak terkecil antara titik data dengan *centroid*. Proses ini diulang dengan menghitung ulang posisi *centroid* setelah setiap iterasi, dan iterasi terus dilakukan hingga nilai *centroid* tidak berubah lagi [10].

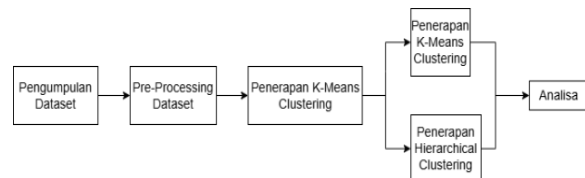
Metode *clustering hierarchical* yang paling umum digunakan adalah metode *agglomerative*, di mana pasangan objek atau kluster digabungkan menjadi kluster yang lebih besar hingga akhirnya membentuk satu kluster tunggal [11]. *Hierarchical clustering* merupakan metode pengelompokan data yang tidak memerlukan penentuan jumlah *cluster* secara eksplisit pada awal proses. Dalam penelitian ini, digunakan pendekatan *agglomerative hierarchical clustering*, di mana setiap objek dimulai sebagai satu *cluster* terpisah, dan secara iteratif digabungkan berdasarkan kriteria jarak tertentu hingga terbentuk jumlah *cluster* yang diinginkan [10].

Silhouette Coefficient adalah metode untuk mengukur kualitas clustering dengan menghitung seberapa baik suatu data berada dalam klusternya dibandingkan dengan kluster lainnya. Metode ini merupakan metode evaluasi *cluster* yang menggabungkan metode *cohesion* dan *separation*. *Cohesion* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* terdekatnya [12].

3. METODE PENELITIAN

Tahapan penelitian terdiri atas pengumpulan data, pra-pemrosesan data, clustering dan analisis. *Dataset* yang digunakan dalam penelitian ini diperoleh

dari dua sumber utama, yaitu Data Pusat Informasi Harga Pangan Strategis Nasional (PIHPS Nasional) dan Kementerian Pertanian Republik Indonesia. Data PIHPS Nasional mencakup laporan harian perkembangan harga cabai rawit hijau di setiap provinsi dan kabupaten di Indonesia, yang mencatat fluktuasi harga dari Januari 2015 hingga Februari 2025. Variabel yang digunakan dari data ini meliputi provinsi, kabupaten, serta harga cabai rawit hijau per bulan dari Januari 2018 sampai Desember 2024. Gambar 1 menampilkan Alur kerja penelitian



Gambar 1. Alur Kerja Penelitian

Sementara itu, data dari Kementerian Pertanian mencakup informasi tentang subsektor hortikultura, khususnya komoditas cabai rawit, dengan fokus pada produktivitas cabai rawit di tingkat kabupaten setiap provinsi di Indonesia, dari Januari 2015 hingga Desember 2025. Variabel yang digunakan dalam analisis ini meliputi provinsi, kabupaten, produktivitas dari tahun 2018 sampai 2023. Penggunaan kedua *dataset* ini memungkinkan analisis mendalam terhadap pola harga cabai rawit di pasar tradisional serta variasi produktivitas dan pemasaran cabai rawit di berbagai daerah di Indonesia.

Pra pemrosesan data merupakan langkah untuk mengubah data mentah menjadi format yang sesuai untuk digunakan dalam algoritma. Proses ini sangat penting karena data mentah biasanya tidak terstruktur dengan baik. Selain itu, algoritma seperti *K-Means* dan *Hierarchical* tidak dapat langsung bekerja dengan data mentah. Oleh karena itu, *pre-processing* diperlukan untuk mempermudah tahap analisis data dengan algoritma [13]. Penelitian ini menggunakan dua jenis data, yaitu data produktivitas cabai rawit dan data harga cabai rawit hijau, yang diperoleh dari berbagai kabupaten/kota di Indonesia. Data produktivitas mencakup periode tahunan 2018 hingga 2023, sedangkan data harga mencakup data bulanan dari 2018 hingga 2024.

Pembersihan format numerik dengan menghapus karakter non-angka dan mengganti koma menjadi titik untuk konversi ke tipe numerik float dilakukan pada dataset harga cabai. Rata-rata harga per wilayah dihitung sebagai representasi umum dari fluktuasi harga selama periode waktu tersebut. Nama kabupaten/kota kemudian dinormalisasi dengan menghapus awalan seperti “kab.” atau “kota”, serta dilakukan pemetaan manual untuk menangani variasi ejaan agar sesuai dengan data spasial (*shapefile*). Sementara itu, untuk data produktivitas, dilakukan agregasi produktivitas per tahun untuk setiap wilayah, diikuti dengan perhitungan rata-rata produktivitas

agregat sebagai fitur utama. Untuk seluruh data numerik, dilakukan proses standarisasi menggunakan *StandardScaler* guna memastikan semua variabel berada dalam skala yang sebanding dan tidak mendominasi dalam proses klusterisasi.

Metode *K-Means Clustering* diterapkan untuk mengelompokkan wilayah administratif di Indonesia berdasarkan dua indikator utama, yaitu produktivitas dan harga cabai rawit hijau, selama periode 2018–2024. Analisis dilakukan secara terpisah pada tingkat provinsi dan kabupaten/kota. Untuk produktivitas, digunakan rata-rata data tahunan sebagai fitur input, dengan jumlah klaster optimal ditetapkan sebanyak lima ($k = 5$) untuk provinsi dan tiga ($k = 3$) untuk kabupaten/kota. Klaster kemudian diberi label kualitatif seperti “Sangat Rendah” hingga “Sangat Tinggi”. Sementara itu, pada aspek harga, data bulanan dibersihkan dan dinormalisasi, lalu diklaster menggunakan *K-Means* dengan ($k = 3$) provinsi dan ($k = 2$) untuk kabupaten/kota, menghasilkan klasifikasi harga “Rendah”, “Sedang”, dan “Tinggi”.

Seluruh hasil klusterisasi divisualisasikan dalam bentuk peta tematik berbasis *shapefile* GADM, serta dilengkapi grafik batang dan daftar wilayah per kategori. Validasi jumlah klaster dilakukan dengan metrik evaluasi *Silhouette Score* dan *Davies-Bouldin Index*. Hasil analisis menunjukkan bahwa pendekatan ini efektif untuk mengidentifikasi pola spasial ketimpangan harga dan produktivitas, serta mendukung segmentasi wilayah yang berguna dalam perencanaan kebijakan pangan nasional. Dengan demikian, metode *K-Means* terbukti relevan sebagai alat eksplorasi data pertanian untuk mendukung pengambilan keputusan berbasis bukti.

Penelitian ini menerapkan metode *Hierarchical Clustering* untuk melakukan segmentasi wilayah administratif di Indonesia berdasarkan dua indikator utama harga dan produktivitas cabai rawit. Pada masing-masing indikator, dilakukan standarisasi data menggunakan *StandardScaler*, dan proses klusterisasi dilakukan dengan pendekatan *Agglomerative Clustering* menggunakan metode penggabungan *Ward linkage*, yang mempertimbangkan variansi antar-klaster. Untuk data harga, klusterisasi dilakukan pada tingkat provinsi dan kabupaten/kota. Di tingkat provinsi, jumlah klaster optimal ditetapkan sebanyak tiga ($k = 3$), menghasilkan kategori harga Rendah, Sedang, dan Tinggi. Sedangkan pada tingkat kabupaten/kota, digunakan dua klaster ($k = 2$), dengan kategori Rendah dan Tinggi berdasarkan nilai rata-rata harga.

Sementara itu, untuk data produktivitas, klusterisasi dilakukan pada tingkat provinsi dengan lima klaster ($k = 5$) yang dilabeli sebagai Sangat Rendah hingga Sangat Tinggi, dan pada tingkat kabupaten/kota dengan tiga klaster ($k = 3$). Evaluasi kualitas klaster dilakukan menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index*. Hasil klusterisasi divisualisasikan dalam bentuk peta tematik spasial, dan disertai dengan daftar wilayah per

kategori. Pendekatan ini memberikan pemetaan spasial berbasis data yang berguna untuk perumusan kebijakan peningkatan produktivitas dan stabilisasi harga pangan di tingkat lokal dan nasional.

Analisis klusterisasi dilakukan menggunakan dua pendekatan utama, yaitu *K-Means Clustering* dan *Hierarchical Clustering* dengan metode *Ward linkage*, yang diterapkan pada data produktivitas dan harga cabai rawit hijau baik di tingkat provinsi maupun kabupaten/kota. Sebelum proses klusterisasi, seluruh fitur numerik dinormalisasi menggunakan *StandardScaler* untuk memastikan skala data yang seragam. Hasil evaluasi menunjukkan bahwa pada data produktivitas provinsi, konfigurasi terbaik diperoleh saat jumlah klaster $k = 5$, dengan nilai *Silhouette Score* sebesar 0,64 dan *Davies-Bouldin Index* (DBI) sebesar 0,38 pada kedua algoritma. Untuk produktivitas kabupaten/kota, baik *K-Means* maupun *Hierarchical Clustering* menunjukkan performa terbaik pada $k=3$, dengan *Silhouette Score* sebesar 0,48 dan DBI masing-masing 0,68 dan 0,70, menunjukkan segmentasi wilayah yang cukup kompak dan terpisah jelas.

Sementara itu, untuk data harga cabai rawit hijau tingkat provinsi, pendekatan *Hierarchical Clustering* dengan $k = 3$ menunjukkan kinerja terbaik dengan *Silhouette Score* sebesar 0,348 dan DBI 1,010, sedikit lebih baik dibanding *K-Means* yang menghasilkan *Silhouette Score* 0,325 dan DBI 1,020 pada jumlah klaster yang sama. Di tingkat kabupaten/kota, *Hierarchical Clustering* juga lebih unggul dengan konfigurasi $k = 2$, menghasilkan *Silhouette Score* 0,32 dan DBI 1,16, dibandingkan dengan *K-Means* yang mencatat *Silhouette Score* 0,30 dan DBI 1,20. Hasil dari proses klusterisasi ini selanjutnya dikombinasikan dengan data spasial GADM (level 1 dan 2) dan divisualisasikan dalam bentuk peta tematik, yang menampilkan distribusi spasial wilayah berdasarkan kategori produktivitas dan harga. Visualisasi ini memfasilitasi pemetaan disparitas regional serta mendukung segmentasi wilayah untuk perumusan kebijakan yang lebih berbasis data dan geografis.

4. HASIL DAN PEMBAHASAN

Penelitian ini membandingkan dua metode *clustering K-Means* dan *Agglomerative Hierarchical* dalam mengelompokkan data Produktivitas Cabai Rawit dan Harga Cabai Rawit Hijau. Evaluasi performa dilakukan dengan $k = \{2, 3, 4, 5, 6\}$ berdasarkan dua metrik utama *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Nilai *Silhouette Score* yang tinggi dan DBI yang rendah menunjukkan performa *clustering* yang baik.

Eksperimen pertama dalam penelitian ini difokuskan pada pengelompokan 34 provinsi di Indonesia berdasarkan data produktivitas cabai rawit yang dihitung sebagai rata-rata produktivitas dari tahun 2018-2023 dan data harga cabai rawit hijau dihitung dari tahun 2018-2024. Eksperimen kedua

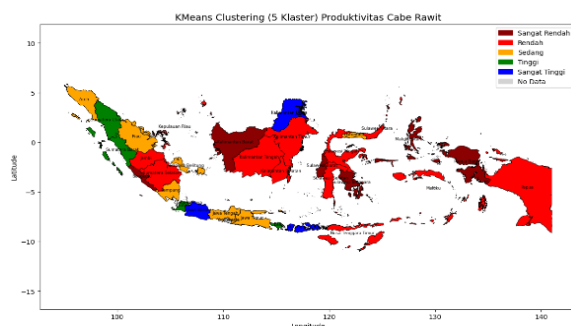
memperluas cakupan dan kedalaman analisis dengan menggunakan data tingkat kabupaten/kota. Visualisasi hasil ditampilkan dalam peta *geopandas* dan grafik.

4.1. Clustering Pada Dataset Provinsi

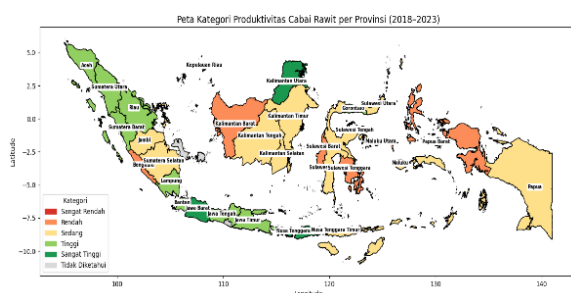
Hasil menunjukkan pada tabel 1 bahwa konfigurasi optimal untuk kedua metode adalah dengan jumlah kluster sebanyak 5, di mana nilai *Silhouette Score* tertinggi mencapai 0,64 dan DBI terendah sebesar 0,38. Ini menunjukkan bahwa kedua metode dapat menghasilkan kluster yang kompak dengan pemisahan yang baik antar kluster.

Tabel 1. Evaluasi *Clustering* Pada Dataset Produktivitas Cabai Rawit Tingkat Provinsi

Algoritma	Jumlah Cluster	Nilai Silhouette	Nilai Davies Bouldin Index
K-Means	2	0,54	0,63
K-Means	3	0,52	0,53
K-Means	4	0,62	0,40
K-Means	5	0,64	0,38
K-Means	6	0,63	0,42
Hierarchical	2	0,54	0,63
Hierarchical	3	0,52	0,45
Hierarchical	4	0,62	0,40
Hierarchical	5	0,64	0,38
Hierarchical	6	0,63	0,42



Gambar 2. Pemetaan Provinsi Hasil Produktivitas Cabai Rawit Menggunakan K-Means

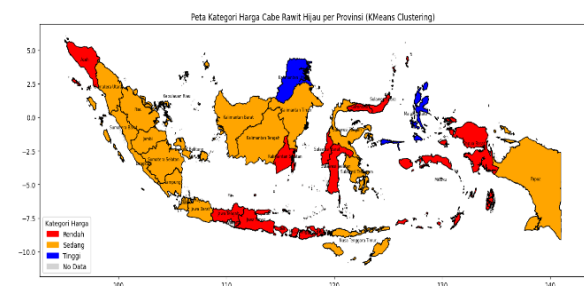


Gambar 3. Pemetaan Provinsi Hasil Produktivitas Cabai Rawit Menggunakan Hierarchical Clustering

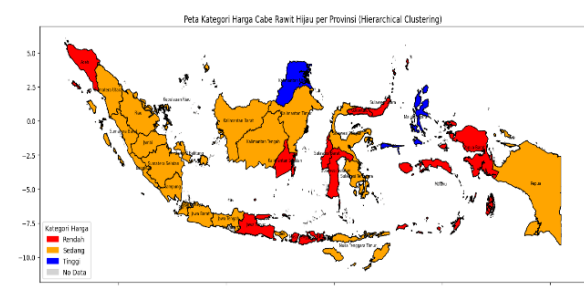
Pada tingkat provinsi, eksperimen pertama mengelompokkan 34 provinsi berdasarkan data produktivitas cabai rawit dan harga cabai rawit hijau. Hasil *clustering* menggunakan *K-Means* pada gambar 2 dan gambar 3 *Hierarchical Clustering* menunjukkan konsistensi yang tinggi dalam mengidentifikasi

provinsi-provinsi dengan produktivitas tinggi dan rendah. Kluster dengan produktivitas "Sangat Tinggi" (misalnya Nusa Tenggara Barat, Jawa Barat, Kalimantan Utara) teridentifikasi dengan baik, sementara wilayah dengan produktivitas rendah (terutama di bagian timur Indonesia seperti Papua dan Kalimantan Barat) juga terlihat jelas. Visualisasi spasial dari hasil *clustering* ini menunjukkan ketimpangan produktivitas yang signifikan antara Pulau Jawa dan wilayah Indonesia timur, dengan Pulau Jawa mendominasi provinsi dengan produktivitas tinggi.

Eksperimen kedua, yang melibatkan data harga cabai rawit hijau, Hasil menunjukkan pada tabel 2 bahwa konfigurasi optimal untuk kedua metode adalah dengan jumlah kluster sebanyak 3, di mana nilai *Silhouette Score* tertinggi mencapai 0,348 dan DBI terendah sebesar 1,010. Menunjukkan bahwa pada gambar 4 dan gambar 5 wilayah-wilayah pusat konsumsi (urban) seperti Bandung, Bekasi, dan Banjarmasin cenderung mengalami harga yang lebih tinggi, sedangkan wilayah dengan potensi produksi yang lebih baik, seperti Jawa Timur dan NTB, berada dalam kategori harga rendah. Hasil ini menunjukkan bahwa ketimpangan harga cabai rawit hijau lebih terkait dengan faktor logistik dan distribusi, dengan wilayah timur Indonesia lebih sering mengalami harga yang sangat tinggi atau rendah, sementara wilayah tengah dan barat lebih stabil. Dengan kedua metode *clustering*, *K-Means* dan *Hierarchical Clustering*, menghasilkan pola yang hampir serupa dalam hal identifikasi wilayah dengan harga tertinggi dan terendah, meskipun ada sedikit perbedaan dalam klasifikasi beberapa provinsi.



Gambar 4. Pemetaan Provinsi Harga Cabai Rawit Hijau Tingkat Provinsi Menggunakan K-Means



Gambar 5/ Pemetaan Provinsi Harga Cabai Rawit Hijau Tingkat Provinsi Menggunakan Hierarchical Clustering

Tabel 2. Evaluasi Metode *Clustering* Pada Dataset Harga Cabai Rawit Hijau Provinsi

Algoritma	Jumlah Cluster	Nilai Silhouette	Nilai Davies Bouldin Index
K-Means	2	0,302	1,178
K-Means	3	0,325	1,020
K-Means	4	0,321	0,966
K-Means	5	0,186	1,239
K-Means	6	0,162	1,187
Hierarchical	2	0,318	1,174
Hierarchical	3	0,348	1,010
Hierarchical	4	0,334	0,947
Hierarchical	5	0,140	1,237
Hierarchical	6	0,164	1,179

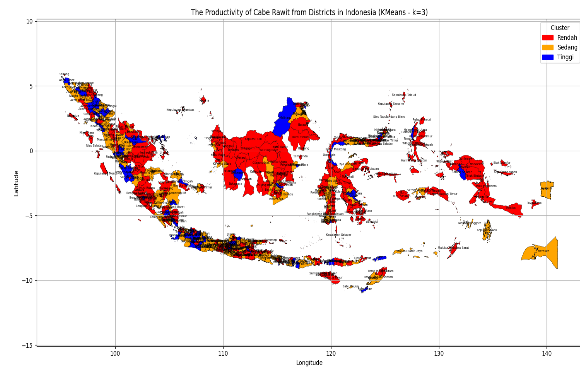
4.2. Clustering Dataset Kabupaten

Tujuan dari analisis ini adalah untuk mengidentifikasi distribusi spasial wilayah dengan produktivitas tinggi dan rendah, sehingga dapat digunakan sebagai dasar rekomendasi kebijakan pertanian berbasis kluster lokal.

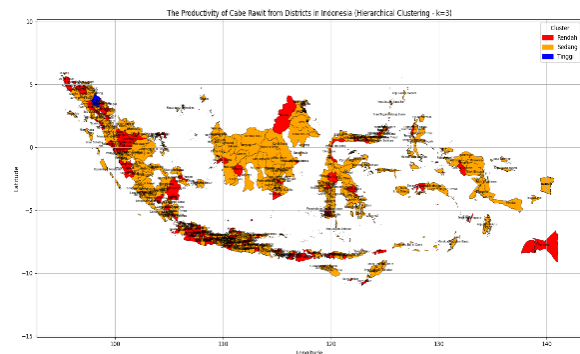
Berdasarkan hasil evaluasi pada tabel 3 pada penggunaan *K-Means*, konfigurasi dengan 3 kluster memberikan hasil terbaik dengan nilai *Silhouette Score* sebesar 0,48 dan *Davies-Bouldin Index* (DBI) sebesar 0,68. Hasil *clustering* pada gambar 6 ini menunjukkan bahwa kabupaten dengan produktivitas tinggi terkonsentrasi di wilayah Jawa dan sebagian Nusa Tenggara. Contoh kabupaten dengan produktivitas tinggi antara lain Aceh Tengah, Agam, Bandung Barat, Malang, dan Grobogan. Sementara itu, kabupaten dengan produktivitas rendah tersebar luas, terutama di wilayah timur Indonesia, seperti Papua dan Kalimantan. Kluster produktivitas sedang, seperti Padang dan Probolinggo, menunjukkan potensi untuk intervensi peningkatan produktivitas. Metode *Hierarchical* pada gambar 7 juga memberikan hasil serupa dengan 3 kluster pada tabel 3 yang menghasilkan *Silhouette Score* sebesar 0,48 dan DBI sebesar 0,70.

Tabel 3. Evaluasi *Clustering* Produktivitas Cabai Rawit Tingkat Kabupaten

Algoritma	Jumlah Cluster	Nilai Silhouette	Nilai Davies Bouldin Index
K-Means	2	0,48	0,95
K-Means	3	0,48	0,68
K-Means	4	0,37	0,88
K-Means	5	0,36	1,01
K-Means	6	0,31	1,24
Hierarchical	2	0,48	0,98
Hierarchical	3	0,48	0,70
Hierarchical	4	0,44	0,91
Hierarchical	5	0,28	1,14
Hierarchical	6	0,30	1,13

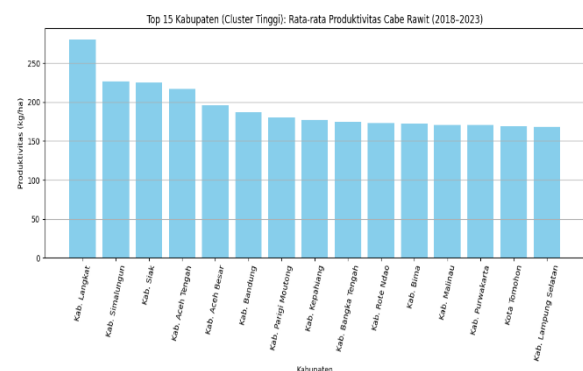


Gambar 6. Pemetaan Kabupaten Berdasarkan Produktivitas Cabai Rawit Menggunakan K-Means Clustering



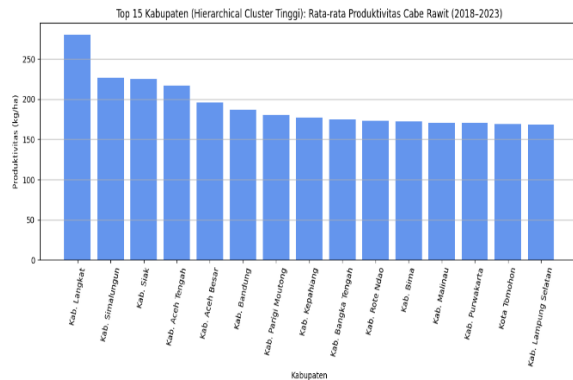
Gambar 7. Pemetaan Kabupaten Berdasarkan Produktivitas Cabai Rawit Menggunakan Hierarchical Clustering

Gambar 8 dan 9 menampilkan analisis 15 daerah tertinggi produktivitas cabai rawit mengungkapkan ketimpangan produktivitas antar kabupaten di Indonesia, berdasarkan algoritma K-Means dan *Hierarchical Clustering*. Kabupaten dengan produktivitas tinggi, seperti Langkat, Aceh Tengah, Simalungun, Siak, dan Malang, memiliki sistem pertanian yang efisien berkat teknologi dan infrastruktur yang baik. Sebaliknya, kabupaten dengan produktivitas rendah, seperti Bangli, Sumba Tengah, Buton Utara, Halmahera Selatan, dan Buru Selatan, menghadapi tantangan seperti keterbatasan teknologi dan infrastruktur.



Gambar 8. Grafik 15 Daerah Tertinggi Produktivitas Cabai Rawit Kabupaten Hasil K-Means Clustering

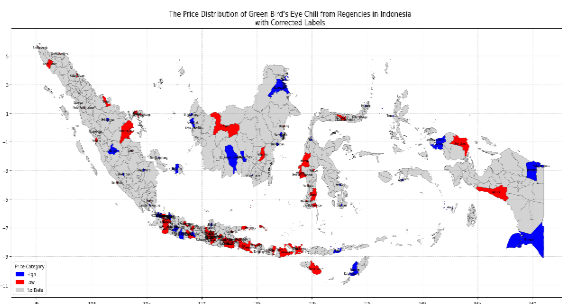
Daerah dengan peringkat 15 tertinggi produktivitas terbaik menjadi contoh model bagi kabupaten lain untuk meningkatkan hasil pertanian mereka. Analisis ini menunjukkan bahwa kebijakan pertanian harus fokus pada penguatan kapasitas dan infrastruktur di wilayah dengan produktivitas rendah, sementara kabupaten dengan produktivitas tinggi dapat dijadikan referensi untuk pengembangan lebih lanjut.



Gambar 9. Grafik 15 Daerah Tertinggi Produktivitas Cabai Rawit Kabupaten Hasil Hierarchical Clustering

Tabel 4. Evaluasi Clustering Data Harga Cabai Rawit Hijau Tingkat Kabupaten

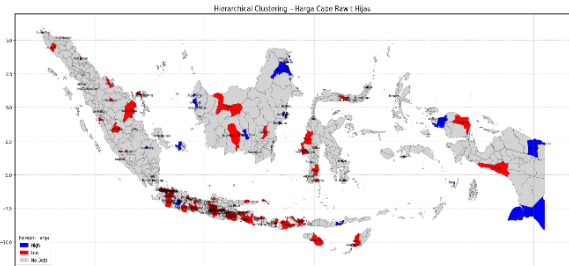
Algoritma	Jumlah Cluster	Nilai Silhouette	Nilai Davies Bouldin Index
K-Means	2	0,30	1,20
K-Means	3	0,29	1,13
K-Means	4	0,27	1,23
K-Means	5	0,26	1,19
K-Means	6	0,23	1,28
Hierarchical	2	0,32	1,16
Hierarchical	3	0,31	1,03
Hierarchical	4	0,29	1,19
Hierarchical	5	0,29	1,14
Hierarchical	6	0,23	1,34



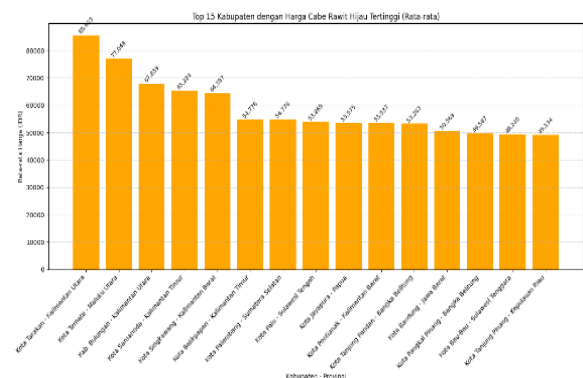
Gambar 10. Pemetaan Kabupaten Berdasarkan Harga Cabai Rawit Hijau Menggunakan K-Means Clustering

Untuk harga cabai rawit hijau, *K-Means* menunjukkan hasil terbaik pada tabel 4 pada konfigurasi dua klaster dengan *Silhouette Score* sebesar 0,30 dan DBI 1,20. Wilayah dengan harga tinggi terutama ditemukan di kota-kota urban dan pusat ekonomi seperti Bandung, Bekasi, dan

Balikpapan. Sebaliknya, harga rendah ditemukan di kabupaten dengan basis produksi hortikultura kuat seperti Blitar, Jember, Boyolali, dan Cilacap seperti pada gambar 10. *Hierarchical Clustering* dengan konfigurasi dua klaster menghasilkan hasil serupa, dengan harga tinggi di daerah urban seperti Bandung dan Jayapura, sementara harga rendah ditemukan di wilayah produsen seperti Jawa Timur dan NTB seperti pada gambar 11.



Gambar 11. Pemetaan Kabupaten Berdasarkan Harga Cabai Rawit Hijau Menggunakan Hierarchical Clustering



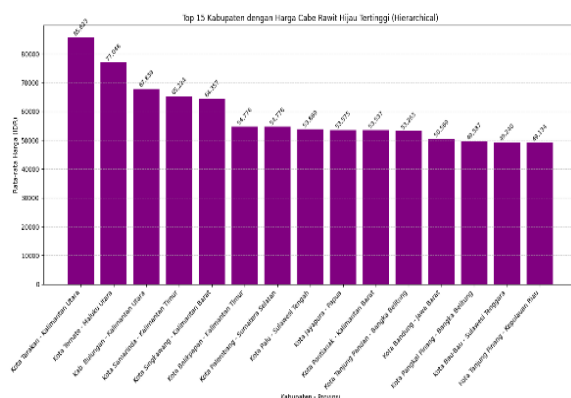
Gambar 12. Grafik 15 Daerah Tertinggi Harga Cabai Rawit Hijau Dataset Kabupaten Hasil K-Means Clustering

Secara keseluruhan, baik metode *K-Means* maupun *Hierarchical* menunjukkan konsistensi spasial dalam identifikasi wilayah dengan harga tertinggi dan terendah, khususnya di kategori peringkat 15 tertinggi dapat dilihat pada gambar 12 dan 13. Wilayah dengan harga tertinggi umumnya terletak di bagian timur dan tengah Indonesia, terutama Kalimantan, Maluku, dan Papua, yang ditandai oleh biaya logistik tinggi, keterbatasan infrastruktur distribusi, dan kemungkinan ketergantungan pada pasokan eksternal. Sebaliknya, kabupaten dengan harga terendah umumnya berasal dari wilayah dengan pusat produksi cabai dan jalur distribusi yang terintegrasi, seperti Sulawesi Selatan dan NTB.

Konsistensi hasil antara dua metode *clustering* meningkatkan keandalan temuan dan menegaskan bahwa disparitas harga cabai rawit hijau di Indonesia bukanlah hasil fluktuasi acak semata, melainkan berakar pada struktur geografis, logistik, dan tata niaga komoditas. Dengan demikian, kebijakan harga pangan nasional memerlukan pendekatan spasial yang adaptif,

dengan perhatian khusus pada wilayah-wilayah timur Indonesia yang rentan mengalami tekanan harga tinggi dan ketimpangan distribusi komoditas strategis.

Secara umum, kedua metode klusterisasi *K-Means* dan *Hierarchical Clustering* menunjukkan pola segmentasi wilayah yang konsisten dalam menganalisis data produktivitas dan harga cabai rawit hijau, baik pada tingkat provinsi maupun kabupaten/kota. Namun, terdapat perbedaan karakteristik penting antara keduanya. *K-Means* lebih efisien secara komputasi dan memberikan hasil yang stabil ketika jumlah kluster ditentukan sejak awal. Hal ini terlihat pada klusterisasi produktivitas provinsi, di mana *K-Means* pada $k=5$ mencapai performa tertinggi dengan *Silhouette Score* 0,64 dan DBI 0,38, menunjukkan separasi kluster yang sangat baik.



Gambar 13. Grafik 15 Daerah Tertinggi Harga Cabai Rawit Hijau Dataset Kabupaten Hasil Hierarchical Clustering

Di sisi lain, *Hierarchical Clustering* memiliki keunggulan dalam fleksibilitas struktur kluster dan tidak memerlukan penentuan jumlah kluster sejak awal. Metode ini menunjukkan hasil yang kompetitif, terutama pada analisis harga cabai rawit hijau provinsi, dengan nilai *Silhouette Score* 0,348 dan DBI 1,010 saat $k = 3$, sedikit lebih baik dari *K-Means* pada konfigurasi yang sama. *Hierarchical* juga memberikan gambaran lebih menyeluruh terhadap relasi antar wilayah, namun menjadi kurang efisien untuk dataset besar seperti di tingkat kabupaten.

Dengan demikian, *K-Means* cocok untuk eksplorasi cepat dan konsisten pada dataset besar, sementara *Hierarchical* lebih unggul dalam interpretabilitas struktur kluster dan cocok untuk analisis eksploratif pada skala yang lebih kecil. Pemilihan metode sebaiknya disesuaikan dengan tujuan analisis, ukuran data, dan kebutuhan visualisasi hierarki antar wilayah.

5. KESIMPULAN DAN SARAN

Penelitian ini menyajikan analisis spasial produktivitas dan harga cabai rawit hijau di Indonesia dengan pendekatan *unsupervised learning*, menggunakan dua metode klusterisasi utama, yaitu *K-Means* dan *Agglomerative Hierarchical Clustering*.

Analisis dilakukan pada dua tingkatan wilayah: provinsi dan kabupaten/kota, menggunakan data panel tahun 2018–2024. Evaluasi efektivitas klusterisasi dilakukan melalui dua metrik utama *Silhouette Score*, yang mengukur kekompakan dan separasi kluster, serta *Davies Bouldin Index* (DBI), yang menilai tingkat kesamaan antar kluster (semakin rendah, semakin baik).

Pada data produktivitas tingkat provinsi, hasil terbaik diperoleh saat jumlah kluster $k=5$, baik pada *K-Means* maupun *Hierarchical Clustering*, dengan *Silhouette Score* mencapai 0,64 dan DBI sebesar 0,38, menandakan segmentasi wilayah yang sangat baik. Untuk harga cabai rawit hijau di tingkat provinsi, klusterisasi optimal ditemukan pada $k = 3$, dengan *Hierarchical Clustering* mencatat performa tertinggi yaitu *Silhouette Score* 0,348 dan DBI 1,010, menunjukkan struktur kluster yang cukup stabil.

Di tingkat kabupaten/kota, analisis produktivitas menghasilkan hasil terbaik pada $k = 3$, dengan nilai *Silhouette* 0,48 dan DBI 0,68 pada *K-Means*, sedangkan untuk harga, hasil terbaik ditemukan pada $k = 2$, dengan *K-Means* mencatat *Silhouette Score* 0,30 dan DBI 1,20. Hasil-hasil ini menunjukkan bahwa meskipun pola kluster lebih kompleks di level kabupaten, pemisahan wilayah tetap menunjukkan struktur yang jelas dan dapat digunakan untuk keperluan perencanaan wilayah.

Dari hasil klusterisasi, ditemukan bahwa wilayah dengan produktivitas tinggi dan harga rendah umumnya berada di Pulau Jawa dan sebagian Sumatera (misalnya Jawa Timur, Sumatera Utara), yang mencerminkan efisiensi produksi dan konektivitas pasar yang baik. Sebaliknya, wilayah-wilayah di Indonesia Timur seperti Papua, Maluku, dan sebagian Kalimantan memiliki harga tinggi namun produktivitas rendah, mengindikasikan tantangan logistik, distribusi, dan akses teknologi.

Klusterisasi memungkinkan identifikasi wilayah prioritas intervensi, baik untuk peningkatan produktivitas, pengendalian harga, maupun penguatan infrastruktur. Temuan juga menunjukkan bahwa integrasi data spasial dan algoritma *machine learning* dapat memberikan pendekatan yang lebih presisi dan adaptif dalam pengelolaan komoditas strategis nasional seperti cabai rawit.

Hasil analisis menunjukkan bahwa wilayah seperti Papua, Maluku, dan Kalimantan mengalami kombinasi harga cabai rawit hijau yang tinggi dan produktivitas yang rendah. Oleh karena itu, perlu dilakukan pembangunan *cold storage* dan fasilitas pengolahan pascapanen di wilayah ini untuk mengurangi kerusakan hasil panen dan menekan ketergantungan pada pasokan dari luar daerah. Di sisi lain, daerah seperti Jawa Timur dan NTB yang tergolong dalam kluster produktivitas tinggi dan harga rendah dapat dijadikan sebagai hub distribusi nasional, khususnya untuk menyuplai wilayah urban seperti Bandung, Bekasi, dan Balikpapan yang secara konsisten menunjukkan harga tinggi.

Pemerintah juga dapat menetapkan insentif khusus berupa subsidi transportasi atau skema tol laut untuk mempermudah arus distribusi cabai rawit dari sentra produksi ke pusat konsumsi yang harganya tidak stabil. Selain itu, daerah dengan produktivitas sedang seperti Probolinggo dan Padang bisa menjadi target program intensifikasi melalui introduksi varietas unggul dan bantuan alat pertanian presisi, agar dapat naik kelas ke klaster produktivitas tinggi.

DAFTAR PUSTAKA

- [1] 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)," *VARIANCE: Journal of Statistics and Its Applications*, vol. 3, no. 2, pp. 48–60, Dec. 2021, doi: 10.30598/variancevol3iss2page48-60.
- [2] Syaiful Hasan Abdullah and Zaehol Fatah, "Analisis Produksi Cabai Rawit Indonesia Menggunakan Algoritma K-Means Clustering," *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, vol. 3, no. 1, pp. 66–74, Dec. 2024, doi: 10.59024/jiti.v3i1.1024.
- [3] T. Handhayani, I. Lewenusa, and M. Y. R. Arpipi, "Forecasting Volatile Fresh Chili Prices In Indonesia Using Support Vector Regression," in *2024 International Conference on Information Technology Research and Innovation (ICITRI)*, IEEE, Sep. 2024, pp. 281–286. doi: 10.1109/ICITRI62858.2024.10699118.
- [4] S. Saptana *et al.*, "Competitiveness and impact of government policy on chili in Indonesia," *Open Agric*, vol. 7, no. 1, pp. 226–237, Apr. 2022, doi: 10.1515/opag-2022-0083.
- [5] T. Handhayani and Z. Rusdi, "K-Means Using Dynamic Time Warping For Clustering Cities in Java Island According to Meteorological Conditions," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381899.
- [6] Y. A. Ishak, T. Handhayani, M. D. L. Sitorus, W. William, J. Pragantha, and I. Lewenusa, "Advanced Clustering Approach for Mapping Regions of Paddy Productivity in Indonesia Using Intelligent K-Means," in *IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, Manado: IEEE, Mar. 2024, pp. 269–274.
- [7] William, L. Bayuaji, N. J. Perdana, and T. Handhayani, "Mapping Indonesia's Regions Based on Carbon Emissions Using the K-Means Algorithm," in *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, IEEE, Jan. 2025, pp. 200–205. doi: 10.1109/ICoCSETI63724.2025.11020099.
- [8] G. Andrian, D. Arisandi, and T. Handhayani, "CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR DENGAN METODE K-MEANS DAN FUZZY C-MEANS," *INTI Nusa Mandiri*, vol. 18, no. 2, pp. 100–106, Feb. 2024, doi: 10.33480/inti.v18i2.5039.
- [9] T. Handhayani and I. Lewenusa, "An Analysis of Meteorological Data in Sumatra and Nearby using Agglomerative Clustering," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 2, pp. 234–241, Apr. 2024, doi: 10.29207/resti.v8i2.5663.
- [10] K. B., "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1600–1604, May 2020, doi: 10.30534/ijeter/2020/20852020.
- [11] M. Vichi, C. Cavicchia, and P. J. F. Groenen, "Hierarchical Means Clustering," *J Classif*, vol. 39, no. 3, pp. 553–577, Nov. 2022, doi: 10.1007/s00357-022-09419-7.
- [12] R. Y. Fa'rifah and D. Pramesti, "Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm," *Jurnal Varian*, vol. 5, no. 2, pp. 171–178, Apr. 2022, doi: 10.30812/varian.v5i2.1894.
- [13] F. Fathurrahman, S. Harini, and R. Kusumawati, "EVALUASI CLUSTERING K-MEANS DAN K-MEDOID PADA PERSEBARAN COVID-19 DI INDONESIA DENGAN METODE DAVIES-BOULDIN INDEX (DBI)," *Jurnal Mnemonic*, vol. 6, no. 2, pp. 117–128, Oct. 2023, doi: 10.36040/mnemonic.v6i2.6642.