



Klasterisasi Potensi Budidaya Bawang Putih Berdasarkan Faktor Iklim Menggunakan PCA dan *K-Means*

Afida Nur Azizah¹, Achmad Mufliq², Awang Andhyka³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo
21422006.student@unusida.ac.id , 2achmadmufliq.it@unusida.ac.id , 3awang85.si@unusida.ac.id

Abstract

Indonesia's reliance on imported garlic of national demand is more than 90% a significant threat to food security. One key contributing factor is the mismatch between cultivation areas and the climatic conditions required for optimal garlic growth and bulb formation. This study aims to map climate suitability for garlic cultivation across 38 provinces in Indonesia using 2024 climate data, including average temperature, rainfall, number of rainy days, and sunlight duration. The research method integrates Principal Component Analysis (PCA) to reduce correlated climate variables, followed by *K-Means* clustering to group provinces based on climatic similarities. Evaluation using the Davies-Bouldin Index (0.7632) indicates that four clusters represent the optimal grouping structure. Cluster 3, comprising West Java, South Sulawesi, and Highland Papua, demonstrates the highest climatic suitability with an average temperature of 21.73°C. The PCA biplot visualization illustrates the contribution of each climate variable to cluster formation. These findings provide a data-driven basis for policy formulation in developing targeted domestic garlic production centers.

Keywords: garlic, agroclimate, PCA, *k-means*, food security

Abstrak

Ketergantungan Indonesia terhadap impor bawang putih yang melebihi 90% kebutuhan nasional menjadi ancaman serius bagi ketahanan pangan. Salah satu penyebab utamanya adalah ketidaksesuaian antara lokasi budidaya dengan karakteristik iklim yang diperlukan untuk pertumbuhan dan pembentukan umbi bawang putih secara optimal. Penelitian ini bertujuan memetakan kesesuaian iklim bagi budidaya bawang putih di 38 provinsi di Indonesia menggunakan data iklim tahun 2024 yang mencakup suhu rata-rata, curah hujan, jumlah hari hujan, dan lama penyinaran matahari. Metode penelitian menggabungkan *Principal Component Analysis (PCA)* untuk mereduksi variabel iklim yang berkorelasi, dilanjutkan dengan *K-Means Clustering* untuk mengelompokkan provinsi berdasarkan kemiripan kondisi iklim. Evaluasi menggunakan *Davies-Bouldin Index* sebesar 0,7632 menunjukkan bahwa empat klaster merupakan pengelompokan yang paling optimal. Klaster 3 yang terdiri atas Jawa Barat, Sulawesi Selatan, dan Papua Pegunungan menunjukkan tingkat kesesuaian iklim tertinggi dengan suhu rata-rata 21,73°C. Visualisasi biplot PCA menggambarkan kontribusi masing-masing variabel iklim dalam pembentukan klaster. Temuan ini dapat menjadi dasar ilmiah bagi perumusan kebijakan pengembangan sentra produksi bawang putih dalam negeri yang lebih terarah.

Kata kunci: bawang putih, agroklimat, PCA, *k-means*, ketahanan pangan

© 2026 Author
Creative Commons Attribution 4.0 International License



1. Pendahuluan

Sebagai salah satu bahan pangan pokok, bawang putih memiliki tingkat konsumsi yang tinggi di Indonesia. Akan tetapi, kontribusi produksi dalam negeri masih sangat rendah, sehingga kebutuhan nasional sebagian besar dipenuhi melalui impor yang mencapai lebih dari 90% [1]. Tingginya ketergantungan ini menimbulkan risiko terhadap stabilitas harga dan ketahanan pangan nasional, terutama ketika terjadi gangguan pasokan dari negara pengekspor.

Rendahnya hasil produksi bawang putih dalam negeri tidak terlepas dari kondisi iklim yang kurang sesuai pada sebagian wilayah penanaman. Secara agroklimat, bawang putih membutuhkan suhu sejuk serta ketersediaan air dan cahaya matahari yang seimbang untuk mendukung pembentukan umbi dengan baik, Ketidaksesuaian iklim ini menjadi salah satu penyebab utama rendahnya produktivitas [2][3].

Perkembangan teknologi pengolahan data membuka peluang pemanfaatan pendekatan data mining dalam bidang pertanian. Teknik klusterisasi memungkinkan pengelompokan wilayah berdasarkan kesamaan kondisi iklim, sehingga memudahkan analisis potensi dan perencanaan pengembangan pertanian [4].

Perkembangan teknologi *Artificial Intelligence (AI)*, terutama dalam bidang data mining, memberikan alternatif metode dalam mengolah dan menganalisis data pertanian yang bersifat kompleks dan berskala besar. Teknik klusterisasi sebagai bagian dari metode *unsupervised learning* [5] mampu mengungkap pola tersembunyi dalam data agroklimat. Melalui pendekatan ini, wilayah dengan kondisi iklim yang serupa dapat dikelompokkan secara sistematis sehingga mempermudah proses identifikasi kesesuaian lahan budidaya. Selain meningkatkan efisiensi analisis data, metode ini juga mendukung pengambilan keputusan yang lebih berbasis data dalam perencanaan pengembangan pertanian [6].

Dalam berbagai studi data mining, algoritma *K-Means* sering digunakan sebagai teknik klusterisasi untuk mengelompokkan data dengan karakteristik yang serupa. Penelitian yang dilakukan oleh Anisa pada tahun 2025 menghasilkan bahwa penerapan *K-Means* yang dikombinasikan dengan metode klasifikasi *Decision Tree* dapat menghasilkan performa klusterisasi yang sangat baik, dengan nilai *Silhouette Score* mencapai 0,8338 [7]

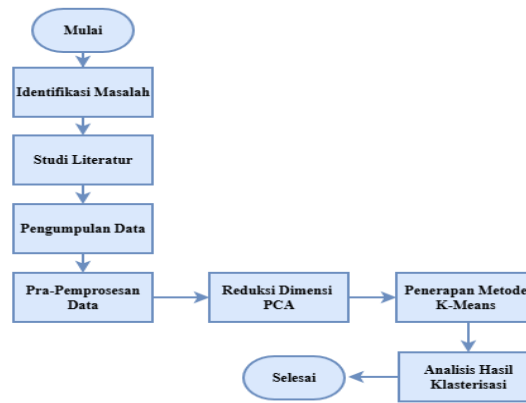
Penelitian sebelumnya juga telah menunjukkan bahwa kombinasi *PCA* dan *K-Means* efektif dalam pemetaan kesesuaian lahan pertanian. Selvida pada tahun 2025 [8], berhasil menerapkan metode tersebut pada wilayah dataran tinggi Sumatera Utara. Sementara itu, Yandi Anzari menerapkan algoritma *k-means* untuk klusterisasi pola iklim pada provinsi Jambi tahun 2020-2024. Metode *K-Means* diterapkan setelah standarisasi *Z Score*, dengan jumlah kluster optimal ditentukan menggunakan *Elbow Method*. Hasil penelitian menghasilkan tiga kluster utama yang merepresentasikan rezim iklim kering atau transisi, monsoonal normal, dan basah ekstrem, dengan nilai *Silhouette Score* sebesar 0,61 [9]

Meskipun demikian, sebagian besar penelitian terdahulu masih terbatas pada analisis spasial lokal dan belum mampu merepresentasikan dinamika variabilitas iklim secara nasional. Selain itu, masih terdapat keterbatasan penelitian yang mengintegrasikan anomali cuaca tahunan terbaru ke dalam algoritma klusterisasi, sehingga pemetaan yang tersedia berisiko tidak mencerminkan kondisi agroklimatologi terkini.

Berdasarkan celah penelitian tersebut, Penelitian ini memberikan kontribusi berupa pemetaan kesesuaian agroklimat bawang putih pada skala nasional menggunakan integrasi *PCA* dan *K-Means* berbasis data iklim terkini tahun 2024. Melalui kerangka kerja gabungan *PCA* dan *K-Means*, penelitian ini mampu mengidentifikasi wilayah potensial secara lebih objektif dan terukur. Hasil dari pemetaan ini diharapkan dapat menjadi rujukan teknis bagi perumusan kebijakan pengembangan sentra produksi bawang putih nasional dalam upaya penguatan ketahanan pangan dan pengurangan dependensi impor.

2. Metode Penelitian

Tahapan penelitian ini dirancang secara terstruktur dan berurutan. Penelitian dimulai dari identifikasi permasalahan, dilanjutkan dengan pengumpulan data iklim dan proses pra-pemrosesan guna memastikan kualitas data sebelum dilakukan analisis. Selanjutnya, data direduksi dimensinya menggunakan metode *Principal Component Analysis (PCA)* dan dikelompokkan menggunakan algoritma *K-Means* berdasarkan kemiripan kondisi iklim antarprovinsi. Evaluasi model dilakukan untuk menilai kualitas model. Hasil yang diperoleh kemudian dianalisis lebih lanjut guna menentukan wilayah potensi sebagai dasar penarikan kesimpulan dan penyusunan rekomendasi.



Gambar 1. Tahapan Penelitian

2.1. Identifikasi Masalah

Tahap identifikasi masalah dilakukan sebagai langkah awal untuk memahami permasalahan utama yang melandasi penelitian ini. Permasalahan tersebut berkaitan dengan ketergantungan terhadap kebutuhan impor bawang putih yang sangat tinggi dibandingkan dengan hasil produksi bawang putih dalam negeri yang masih rendah. Kondisi tersebut menunjukkan bahwa produksi bawang putih dalam negeri belum mampu berkembang secara optimal. Salah satu faktor utama yang menyebabkan rendahnya produksi tersebut adalah ketidaksesuaian antara lokasi budidaya yang belum sepenuhnya mempertimbangkan kesesuaian faktor iklim. Perbedaan karakteristik iklim antarwilayah di Indonesia, seperti suhu, curah hujan, dan intensitas penyinaran matahari, menyebabkan tingkat keberhasilan budidaya bawang putih menjadi bervariasi. Oleh karena itu, diperlukan pendekatan analitis berbasis data iklim untuk memetakan wilayah potensial secara lebih objektif dan sistematis sebagai dasar perencanaan pengembangan budidaya bawang putih.

2.2. Studi Literatur

Berdasarkan penelitian sebelumnya, diketahui bahwa ketahanan pangan bawang putih nasional masih menghadapi tantangan serius berupa tingginya ketergantungan impor dan lemahnya daya saing produk lokal [10]. Permasalahan tersebut menunjukkan perlunya upaya strategis dalam mengidentifikasi potensi pengembangan budidaya bawang putih di dalam negeri.

Beberapa penelitian terdahulu membuktikan bahwa metode data mining mampu mendukung pemetaan kesesuaian lahan pertanian secara objektif. Selvidia pada tahun 2025 [8] menerapkan PCA untuk mereduksi dimensi data dan K-Means untuk klusterisasi lahan bawang putih, sehingga hasil pengelompokan wilayah menjadi lebih tepat dan sesuai dengan karakteristik yang dimiliki.

Penelitian sebelumnya juga menerapkan metode *K-Means Clustering* yang diintegrasikan dengan *Particle Swarm Optimization (PSO)* untuk meningkatkan kualitas proses klusterisasi data. Hasil penelitian menunjukkan bahwa teknik *clustering* mampu mengelompokkan data berdasarkan kemiripan karakteristik sehingga dapat membantu proses analisis data secara lebih efektif [11]. Pada penelitian ini digunakan *Principal Component Analysis (PCA)* untuk mereduksi dimensi data sebelum dilakukan proses klusterisasi menggunakan *K-Means*.

Kajian agroklimatologi menegaskan bahwa faktor iklim memiliki pengaruh besar terhadap produktivitas karena dapat mempengaruhi keberhasilan budidaya bawang putih. Faktor iklim yang tidak sesuai berpotensi menurunkan produktivitas serta meningkatkan risiko kegagalan panen [3].

Perkembangan tanaman bawang putih dan proses pembentukan umbinya dipengaruhi secara signifikan oleh faktor lingkungan, terutama suhu dan intensitas penyinaran matahari. Kondisi photoperiod yang lebih lama dengan suhu yang lebih tinggi terbukti mampu mendukung pertumbuhan tanaman yang lebih optimal serta menghasilkan umbi dengan kualitas yang lebih baik [12].

Penelitian Yandi Anzari menerapkan algoritma k-means untuk klusterisasi pola iklim pada provinsi Jambi tahun 2020-2024 yang menunjukkan bahwa algoritma K-Means efektif digunakan untuk mengelompokkan wilayah berdasarkan kesamaan pola iklim setelah dilakukan standarisasi data [9]. Hasil klusterisasi tersebut dapat dimanfaatkan sebagai dasar dalam pemetaan zonasi wilayah pertanian.

Teknik spatial data mining dalam bidang pertanian dimanfaatkan untuk menganalisis faktor-faktor yang mempengaruhi produktivitas tanaman, seperti kondisi tanah, curah hujan, suhu, dan kelembaban. Melalui analisis data spasial, metode ini mampu menemukan pola tertentu sehingga dapat mengidentifikasi wilayah dengan

karakteristik lingkungan yang serupa dan membantu menentukan daerah yang potensial untuk pengembangan pertanian [13].

Hasil studi literatur menunjukkan bahwa metode PCA dan K-Means memiliki potensi besar dalam mendukung pemetaan wilayah pertanian. Penelitian ini memperluas cakupan penelitian sebelumnya dengan melakukan analisis pada seluruh wilayah Indonesia dengan menggunakan faktor iklim sebagai dasar dalam menentukan potensi budidaya bawang putih.

2.3. Pengumpulan Data

Penelitian ini menggunakan data iklim yang bersumber dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data diperoleh melalui situs resmi <https://dataonline.bmkg.go.id/>. Pengumpulan data dilakukan dengan mengunduh data dari stasiun-stasiun cuaca BMKG yang tersebar di Indonesia, mencakup empat variabel iklim utama, yaitu suhu rata-rata (°C) yang diukur setiap bulan, curah hujan (mm) sebagai akumulasi bulanan, jumlah hari hujan (hari) dengan kriteria curah hujan 1 mm per bulan, dan lama penyinaran matahari (jam) sebagai rata-rata bulanan [14].

Rentang waktu pengumpulan data meliputi periode Januari hingga Desember 2024, di mana data bulanan tersebut kemudian dihitung rata-ratanya untuk mendapatkan nilai tahunan. Cakupan wilayah penelitian mencakup 38 provinsi di Indonesia dengan representasi data dari berbagai stasiun cuaca di setiap provinsi. Data diperoleh dalam format Excel yang memiliki struktur kolom yang mencakup nama provinsi, nama stasiun, serta data bulanan untuk setiap variabel.

NO.	Provinsi	Nama Stasiun	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
1	Aceh	Stamet Maimun Saleh	27,12	27,97	28,38	29,45	29,21	29,24	29,52	27,84	27,47	27,53	27,09	26,56
		Stamet Malikussaleh	27,05	27,72	28,15	29,06	28,74	28,31	28,08	27,1	27,35	26,78	26,5	26,52
		Stamet Sultan Iskandar Muda	26,95	27,26	28,28	28,61	28,77	29,09	29,05	27,81	27,99	27,5	27,05	27,14
		Stamet Cut Nyak Dhien Nagan Raya	27,84	28,41	28,05	28,35	28,18	28,14	27,7	26,73	26,97	26,97	27,02	26,93
		Staklim Aceh	28,89	29,5	29,74	29,85	29,88	30,61	28,94	27,1	27,51	26,89	26,58	26,68
		Staklim Sumatera Utara	28,83	29,93	30,42	31,18	30,49	30,27	29,14	27,7	27,87	27,67	27,07	27,03
2	Sumatera Utara	Stamet Maritim Belawan	28,18	29,29	29,53	29,89	29,53	29,03	29,33	28,29	28,37	28,27	27,51	27,47
		Stamet Kualanamu	26,82	27,82	28,15	29,21	28,64	28,1	28,25	27,28	27,46	27,36	26,82	26,90
		Stasiun Geofisika Deli Serdang	28,55	29,65	30,1	30,46	29,78	30,12	28,56	26,79	27,14	26,90	26,34	26,47
		BBMKG Wilayah I	28,82	30,06	30,47	30,87	30,6	30,57	29,54	27,61	28,08	27,78	27,02	27,26
		Stamet Aek Godang	25,48	26,24	26,51	27,5	26,89	26,57	26,86	26,15	25,76	26,67	26,06	25,98
		Stamet F.L Tobing	27,2	27,73	27,23	28,08	27,68	27,29	27,82	26,65	26,53	27,05	27,03	26,65
3	Sumatera Barat	Stamet Binaka	27,32	28,06	27,47	28,1	27,92	27,38	27,23	26,52	26,44	26,52	26,60	26,13
		Stasiun Geofisika Gunung Sitoli							26,72	26,11	26,15	26,15	25,99	25,65
		Stamet Maritim Teluk Bayur	28,1	28,08	28,41	28,79	28,71	28,12	28,18	27,39	27,59	27,66	27,79	27,33
		Stamet Minangkabau	27,51	28,08	27,77	28,13	28,04	27,56	27,4	26,61	26,99	27,1	27,15	26,75
		Stasiun Geofisika Padang Panjang	23,90	24,02	24,5	24,52	24,85	24,25	24,5					

Gambar 2. Data Suhu Rata-rata

2.4. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk menyiapkan data sebelum dianalisis. Tahap ini meliputi pembersihan data, penanganan data yang tidak lengkap, penggabungan dan transformasi data, serta reduksi data awal menggunakan *python*. Tujuan dari tahapan ini adalah menghasilkan data yang bersih dan siap digunakan. Data kemudian dinormalisasi menggunakan metode *Z-score* untuk menyamakan skala variabel sehingga proses perhitungan jarak pada *K-Means* dapat dilakukan secara optimal.

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (1)$$

Keterangan :

- Z_{ij} : Nilai data hasil standarisasi
- X_{ij} : Nilai data ke-*i* pada variabel ke-*j*
- μ_j : Nilai rata-rata variabel ke-*j*
- σ_j : Simpangan baku variabel ke-*j*

2.5. Reduksi Dimensi PCA

Data iklim hasil pra-pemrosesan direduksi dimensinya menggunakan metode *Principal Component Analysis (PCA)*. Metode ini bertujuan untuk menyederhanakan data dengan membentuk sejumlah komponen utama yang mampu merepresentasikan sebagian besar variasi data, sehingga mendukung proses klusterisasi yang lebih efisien

dan optimal. Setelah data distandarisasi, matriks kovarians dibentuk untuk menggambarkan hubungan antar variabel. Matriks kovarians dihitung menggunakan persamaan :

$$S = \frac{1}{n-1} \cdot Z^T Z \quad (2)$$

Keterangan :

- S : Matriks kovarians
- Z : Matriks data hasil standarisasi (*Z-score*)
- Z^T : Transpose dari matriks Z
- n : Jumlah data atau observasi
- $n - 1$: Derajat kebebasan dalam perhitungan kovarians

Setelah matriks kovarians (S) diperoleh, langkah selanjutnya adalah menghitung nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*). Nilai eigen menunjukkan besarnya variasi data yang dapat dijelaskan oleh masing-masing komponen utama, sedangkan vektor eigen menunjukkan arah pembentukan komponen utama. Dihitung menggunakan rumus :

$$|S - \lambda I| = 0 \quad (3)$$

Keterangan :

- λ : Nilai eigen
- S : Matriks kovarians
- I : Matriks identitas
- $| \quad |$: Determinan Matriks

Menentukan komponen utama (*principal components*) dibentuk dari kombinasi linier variabel-variabel yang telah distandarisasi berdasarkan vektor eigen. Setiap komponen utama saling tidak berkorelasi dan disusun berdasarkan nilai eigen terbesar.

$$PC_k = a_k^1 Z^1 + a_k^2 Z^2 + \dots + a_{kp} Z_p \quad (4)$$

Keterangan :

- PC_k : Komponen utama ke- k
- a_{kp} : Koefisien vektor eigen
- Z_p : Variabel ke- p yang telah distandarisasi
- p : Jumlah variable

Komponen utama disusun berdasarkan nilai eigen terbesar dan dipilih hingga mampu menjelaskan sebagian besar variasi data. Komponen utama terpilih kemudian digunakan sebagai input pada proses klusterisasi K-Means.

2.6. Penerapan Metode *K-Means*

Metode *K-Means* digunakan untuk mengelompokkan data hasil PCA berdasarkan kemiripan karakteristik iklim. Jumlah kluster ditentukan melalui evaluasi menggunakan *Elbow Method*, *Silhouette Score*, *Davies Bouldin Index*, dan *Calinski Harabasz Index*. Proses klusterisasi dilakukan dengan menghitung jarak *Euclidean* antara data dan *centroid* hingga diperoleh pengelompokan yang optimal dan stabil [15]. Jarak *Euclidean* dihitung menggunakan persamaan :

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^P (x_{ik} - c_{jk})^2} \quad (5)$$

Keterangan :

- $d(x_i, c_j)$: Jarak data ke-i terhadap centroid kluster ke-j
- x_{ik} : Nilai variabel ke-k pada data ke-i
- c_{jk} : Nilai centroid kluster ke-j pada variabel ke-k
- P : Jumlah variabel (komponen utama hasil PCA)

Selanjutnya, centroid setiap kluster diperbarui berdasarkan nilai rata-rata seluruh data yang berada dalam kluster tersebut. Rumus perhitungan centroid dituliskan sebagai berikut:

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (6)$$

Keterangan :

- c_j : Centroid kluster ke-j
- $|C_j|$: Jumlah data dalam kluster ke-j
- x_i : Data ke-i dalam kluster C_j

Tujuan utama algoritma *K-Means* adalah meminimalkan jarak antara setiap data dengan *centroid* kluster tempat data tersebut berada. Tujuan ini dinyatakan dalam bentuk fungsi objektif, yaitu jumlah kuadrat jarak dalam kluster (*within-cluster sum of squares*), yang dirumuskan sebagai berikut:

$$J = \sum_{j=1}^K \sum_{x_i \in C_j} |x_i - c_j|^2 \quad (7)$$

Keterangan :

- J : Fungsi objektif yang diminimalkan
- K : Jumlah kluster
- C_j : Kluster ke-j
- c_j : Centeroid kluster ke-j

Proses klusterisasi dihentikan saat centroid telah mencapai kondisi konvergen atau jumlah iterasi maksimum terpenuhi. Algoritma K-Means menghasilkan kluster wilayah dengan karakteristik iklim yang relatif homogen dan diterapkan pada data hasil reduksi dimensi menggunakan PCA. Integrasi PCA dan K-Means diharapkan mampu menghasilkan kluster yang stabil dan representatif untuk mengidentifikasi potensi budidaya bawang putih di Indonesia.

2.7. Analisis Hasil Kluster

Hasil klusterisasi yang diperoleh dari metode K-Means selanjutnya dianalisis untuk mengetahui karakteristik masing-masing kluster berdasarkan faktor iklim. Setiap kluster dianalisis berdasarkan nilai rata-rata variabel iklim yang digunakan dalam penelitian, yaitu suhu rata-rata, curah hujan, jumlah hari hujan, dan lama penyinaran matahari. Analisis ini bertujuan untuk menggambarkan karakteristik iklim dominan pada masing-masing kluster.

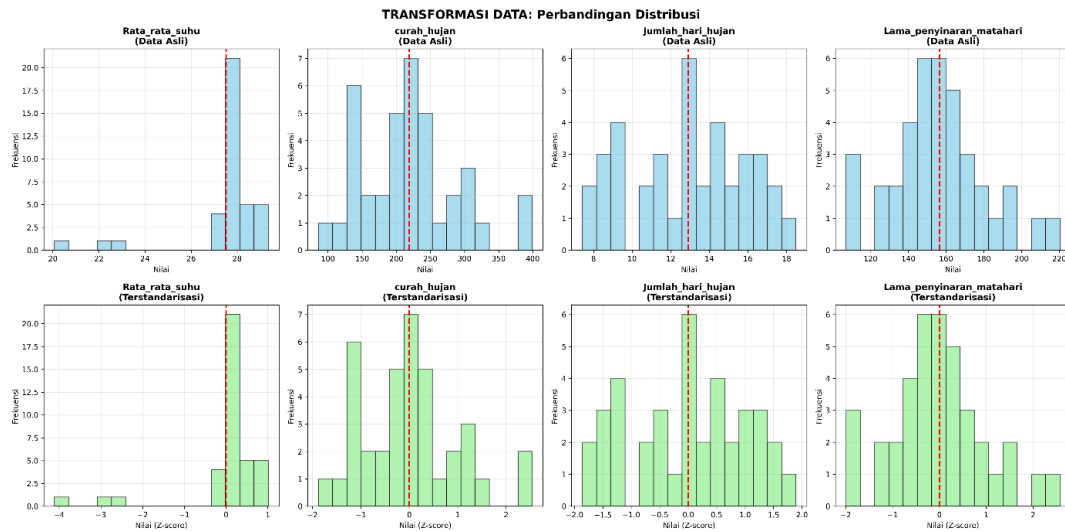
Karakteristik iklim pada setiap kluster dibandingkan dengan persyaratan tumbuh bawang putih yang diperoleh dari studi pustaka. Perbandingan ini digunakan untuk menilai kesesuaian dan potensi wilayah pada masing-masing kluster. Untuk memperjelas hasil analisis, klusterisasi divisualisasikan dalam bentuk gambar. Visualisasi tersebut memberikan gambaran sebaran wilayah berdasarkan kluster serta mendukung penentuan wilayah yang memiliki potensi pengembangan produksi bawang putih lebih tinggi.

3. Hasil dan Pembahasan

3.1. Hasil Pra-pemrosesan

Pada tahapan pra-pemrosesan data dilakukan sebagai tahapan awal untuk memastikan bahwa data iklim yang digunakan memenuhi kriteria kelayakan sebelum dianalisis lebih lanjut menggunakan metode PCA dan K-Means. Data yang digunakan dalam penelitian ini terdiri atas empat variabel iklim utama, yaitu suhu rata-rata, curah hujan, jumlah hari hujan, dan lama penyinaran matahari, yang mewakili kondisi iklim di berbagai wilayah Indonesia.

Hasil pemeriksaan awal menunjukkan bahwa seluruh variabel numerik berada dalam kondisi lengkap tanpa adanya nilai kosong maupun data ganda. Meskipun demikian, distribusi data pada masing-masing variabel menunjukkan perbedaan rentang dan variasi nilai yang cukup mencolok, terutama pada variabel curah hujan dan lama penyinaran matahari. Perbedaan skala antar variabel tersebut berpotensi mempengaruhi hasil analisis, khususnya pada metode berbasis perhitungan jarak seperti K-Means.



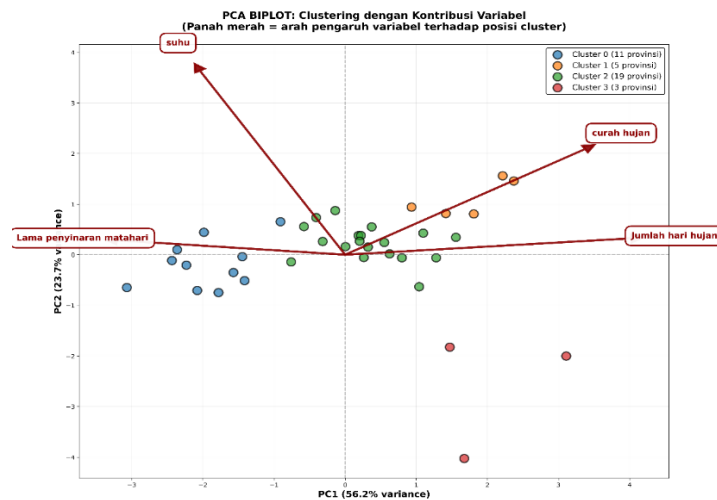
Gambar 3. Standarisasi Data

Untuk mengatasi perbedaan skala tersebut, dilakukan proses standarisasi data menggunakan metode Z-score. Proses ini bertujuan untuk menyetarakan skala seluruh variabel sehingga tidak ada satu variabel pun yang memiliki pengaruh dominan dalam proses analisis. Perbandingan distribusi data sebelum dan sesudah standarisasi, sebagaimana ditampilkan pada Gambar 3, menunjukkan bahwa data hasil transformasi memiliki nilai rata-rata yang mendekati nol dengan sebaran yang lebih seragam antarvariabel.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa data iklim telah berada dalam kondisi yang sesuai untuk dianalisis lebih lanjut. Data hasil pra-pemrosesan ini selanjutnya digunakan sebagai masukan pada tahap reduksi dimensi menggunakan PCA serta proses klusterisasi wilayah dengan algoritma K-Means.

3.2. Hasil Reduksi Dimensi Menggunakan PCA

Hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA) menunjukkan bahwa dua komponen utama pertama telah mampu merepresentasikan sebagian besar variasi data iklim. Komponen utama pertama (PC1) menjelaskan sebesar 56,2% variasi data, sedangkan komponen utama kedua (PC2) menjelaskan 23,7% variasi data. Dengan demikian, total variasi data yang dapat dijelaskan oleh dua komponen utama tersebut mencapai 79,9%, sehingga dinilai cukup representatif untuk menggambarkan pola hubungan antarvariabel iklim dan sebaran wilayah.



Gambar 4. Hasil PCA Biplot

Visualisasi hasil PCA dalam bentuk biplot, sebagaimana ditampilkan pada Gambar 4, memperlihatkan posisi setiap provinsi berdasarkan skor PC1 dan PC2, sekaligus arah dan kontribusi masing-masing variabel iklim terhadap pembentukan komponen utama. Panah pada biplot menunjukkan arah pengaruh variabel, di mana panjang dan arah panah merepresentasikan besarnya kontribusi variabel terhadap variasi data.

Pada biplot tersebut, variabel curah hujan dan jumlah hari hujan terlihat memiliki arah yang relatif sejalan dan dominan pada sumbu PC1 positif. Hal ini menunjukkan bahwa kedua variabel tersebut memberikan kontribusi yang kuat terhadap variasi data pada komponen utama pertama. Sementara itu, variabel suhu rata-rata memiliki arah yang berbeda dan cenderung berkontribusi pada variasi data di sumbu PC2, yang mengindikasikan peran suhu dalam membedakan karakteristik wilayah pada dimensi kedua.

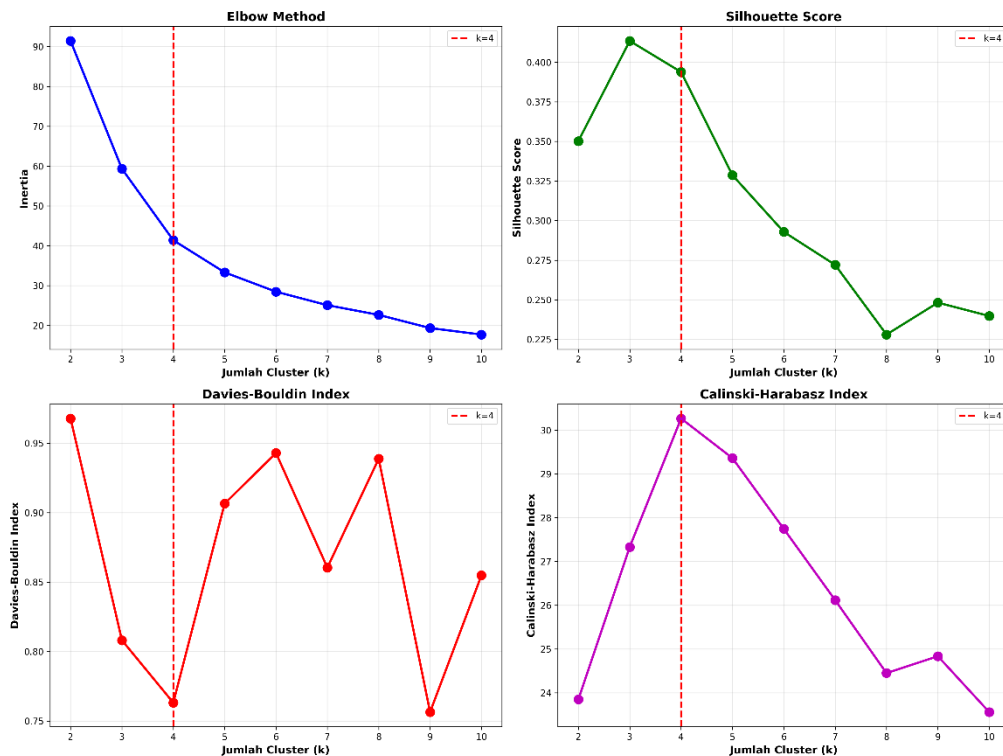
Variabel lama penyinaran matahari menunjukkan arah yang berlawanan dengan variabel curah hujan, yang mencerminkan adanya perbedaan pola kontribusi antara intensitas penyinaran dan faktor curah hujan dalam membentuk struktur data iklim. Pola ini terlihat dari arah panah penyinaran matahari yang mengarah ke sisi negatif PC1.

Sebaran titik pada biplot memperlihatkan bahwa wilayah di Indonesia terbagi ke dalam beberapa kelompok berdasarkan kedekatan posisi pada ruang dua dimensi PCA. Kelompok wilayah dengan karakteristik iklim yang serupa cenderung mengelompok pada area yang berdekatan, sedangkan wilayah dengan karakteristik yang berbeda tersebar pada posisi yang berjauhan. Visualisasi ini memberikan gambaran awal mengenai struktur pengelompokan data iklim sebelum dilakukan proses klusterisasi menggunakan metode K-Means.

Hasil PCA biplot ini selanjutnya digunakan sebagai dasar dalam tahap klusterisasi untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik iklim secara lebih efisien dan terstruktur.

3.3. Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster merupakan tahapan penting dalam proses klusterisasi karena berpengaruh langsung terhadap kualitas hasil pengelompokan data. Jumlah kluster yang terlalu sedikit dapat menyebabkan data dengan karakteristik berbeda berada dalam kelompok yang sama, sedangkan jumlah kluster yang terlalu banyak dapat menghasilkan pengelompokan yang kurang efisien. Oleh karena itu, pada penelitian ini penentuan jumlah kluster dilakukan dengan membandingkan beberapa metode evaluasi, yaitu *Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index (DBI)*, dan *Calinski-Harabasz Index (CHI)*, dengan pengujian jumlah kluster dari $k = 2$ sampai $k = 10$.



Gambar 5. Evaluasi Model

Berdasarkan pengujian menggunakan *Elbow Method*, penurunan nilai inertia terlihat cukup tajam pada penambahan jumlah kluster hingga $k = 4$. Setelah titik tersebut, grafik menunjukkan pola penurunan yang relatif landai. Pola ini menandakan bahwa $k = 4$ merupakan titik perubahan yang menunjukkan batas efektif jumlah kluster, karena penambahan kluster setelahnya tidak memberikan pengaruh yang signifikan terhadap peningkatan kualitas pengelompokan data.

Hasil evaluasi menggunakan *Silhouette Score* menunjukkan nilai tertinggi pada $k = 3$. Meskipun demikian, nilai Silhouette Score pada $k = 4$ masih menunjukkan tingkat kualitas kluster yang baik dan tidak mengalami penurunan yang terlalu jauh. Hal ini menunjukkan bahwa pemisahan antar kluster masih tergolong jelas dan struktur kluster tetap stabil pada jumlah kluster tersebut.

Pengujian menggunakan *Davies-Bouldin Index* menunjukkan bahwa nilai indeks terendah diperoleh pada $k = 4$. Nilai DBI yang lebih kecil mengindikasikan bahwa jarak antar kluster cukup jauh dan tingkat kemiripan data di dalam kluster relatif tinggi, sehingga hasil pengelompokan dapat dikatakan memiliki kualitas yang baik.

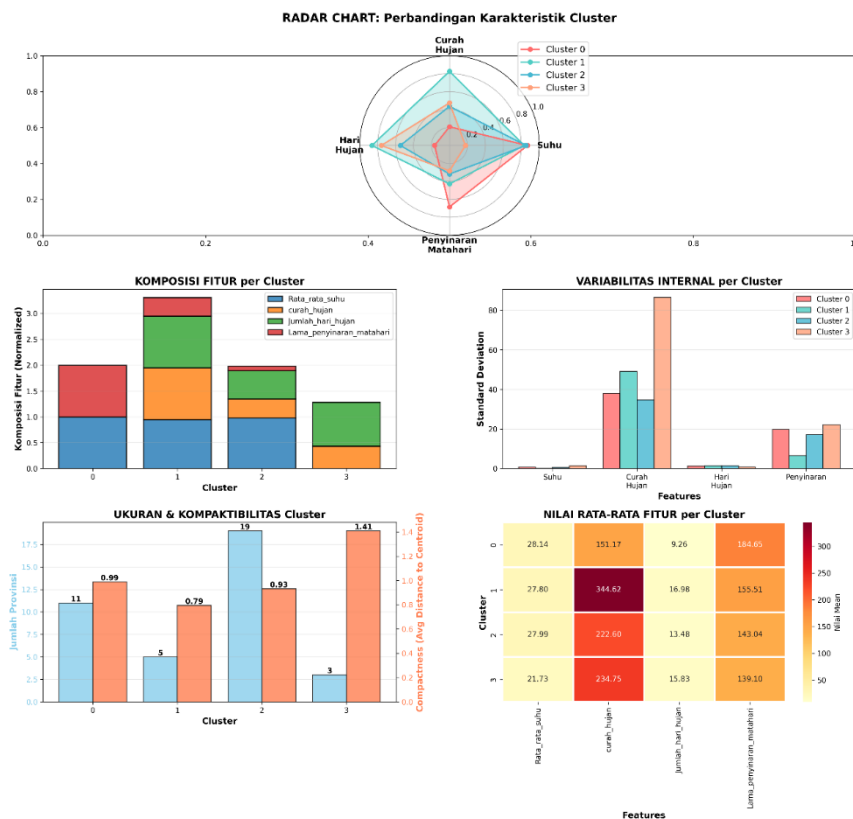
Selain itu, hasil pengujian menggunakan *Calinski Harabasz Index* menunjukkan bahwa nilai tertinggi juga dicapai pada $k = 4$. Nilai CHI yang tinggi menunjukkan bahwa variasi antar kluster lebih besar dibandingkan variasi di dalam kluster, sehingga struktur kluster yang terbentuk menjadi lebih jelas dan terpisah dengan baik.

Berdasarkan keseluruhan hasil evaluasi tersebut, jumlah kluster sebanyak empat kluster ($k = 4$) dipilih sebagai jumlah kluster yang paling optimal. Pemilihan ini didasarkan pada konsistensi hasil dari beberapa metode evaluasi yang menunjukkan bahwa jumlah kluster tersebut mampu memberikan keseimbangan antara tingkat homogenitas data dalam kluster dan perbedaan karakteristik antar kluster.

3.4 Analisis Hasil Kluster

Berdasarkan hasil penentuan jumlah kluster optimal, proses klusterisasi dilakukan menggunakan algoritma *K-Means* dengan jumlah kluster sebanyak empat kluster ($k = 4$). Klusterisasi dilakukan berdasarkan variabel iklim yang meliputi rata-rata suhu, curah hujan, jumlah hari hujan, dan lama penyinaran matahari. Setiap data dikelompokkan ke dalam kluster berdasarkan kedekatan jarak terhadap centroid menggunakan jarak *Euclidean*.

ANALISIS KARAKTERISTIK LENGKAP - 4 CLUSTER IKLIM



Gambar 6. Hasil K-Means

Hasil klasterisasi menunjukkan bahwa setiap klaster memiliki karakteristik iklim yang berbeda, baik dari segi intensitas curah hujan, jumlah hari hujan, maupun lama penyinaran matahari. Perbedaan tersebut ditunjukkan melalui nilai rata-rata fitur, variasi internal, serta tingkat kekompakan klaster.

Tabel 1. Daftar Provinsi Berdasarkan Hasil Klasterisasi K-Means Variabel Iklim Indonesia Tahun 2024

CLUSTER 0	CLUSTER 2
<p>(11 Provinsi) Suhu: 28,14°C Curah Hujan: 151,17 mm</p> <ol style="list-style-type: none"> 1. Aceh 2. DKI Jakarta 3. Banten 4. DI Yogyakarta 5. Jawa Timur 6. Bali 7. NTB 8. NTT 9. Gorontalo 10. Sulawesi Barat 11. Maluku 	<p>(19 Provinsi) Suhu: 27,09°C Curah Hujan: 222,60 mm</p> <ol style="list-style-type: none"> 1. Sumatera Utara 2. Riau 3. Jambi 4. Sumatera Selatan 5. Bengkulu 6. Lampung 7. Bangka Belitung 8. Kepulauan Riau 9. Banten 10. Kalimantan Barat 11. Kalimantan Selatan 12. Kalimantan Timur 13. Kalimantan Utara 14. Sulawesi Tengah 15. Sulawesi Tenggara 16. Maluku Utara 17. Papua Barat Daya 18. Papua 19. Papua Selatan
CLUSTER 1	CLUSTER 3
<p>(5 Provinsi) Suhu: 27,00°C Curah Hujan: 344,02 mm</p> <ol style="list-style-type: none"> 1. Sumatera Barat 2. Kalimantan Tengah 3. Sulawesi Utara 4. Papua Barat 5. Papua Tengah 	<p>(3 Provinsi) Suhu: 27,75°C Curah Hujan: 234,75 mm</p> <ol style="list-style-type: none"> 1. Jawa Barat 2. Sulawesi Selatan 3. Papua Pegunungan

Gambar 7. Daftar Provinsi Per-cluster

Berdasarkan hasil pengelompokan menggunakan algoritma K-Means, wilayah di Indonesia terbagi menjadi empat karakteristik iklim yang berbeda. Analisis mendalam terhadap setiap klaster adalah sebagai berikut:

- Cluster 3* (Potensi Tinggi), *cluster* ini merupakan wilayah yang paling ideal untuk budidaya bawang putih karena memiliki rata-rata suhu terendah, yaitu 21,73°C. Suhu sejuk ini sangat sesuai dengan syarat agroklimat tanaman bawang putih untuk mendukung pembentukan umbi secara optimal. Klaster ini mencakup 3 provinsi dengan tingkat kompaktibilitas (rata-rata jarak ke centroid) sebesar 1,41.
- Cluster 2* (Potensi sedang) merupakan kelompok dengan anggota terbanyak, yaitu 19 provinsi. Karakteristik iklimnya berada pada tingkat moderat dengan suhu rata-rata 27,99°C dan curah hujan 222,60 mm. Meskipun suhu rata-rata masih berada di atas kisaran ideal, beberapa wilayah dalam klaster ini masih berpotensi untuk dikembangkan dengan penerapan teknologi adaptif atau pemilihan varietas yang toleran terhadap suhu lebih tinggi.
- Cluster 1* (Curah Hujan Tinggi) memiliki karakteristik wilayah dengan curah hujan tertinggi mencapai 344,62 mm dan jumlah hari hujan sebanyak 16,98 hari per bulan. Terdapat 5 provinsi dalam kelompok ini yang cenderung memiliki kelembapan tinggi, yang jika tidak dikelola dapat meningkatkan risiko kegagalan panen akibat pembusukan.
- Cluster 0* (Suhu Panas & Kering) Terdiri dari 11 provinsi dengan suhu rata-rata tertinggi sebesar 28,14°C dan jumlah hari hujan paling sedikit (9,26 hari). Wilayah ini kurang ideal bagi bawang putih yang membutuhkan suhu sejuk.

Evaluasi kualitas klaster menunjukkan bahwa setiap klaster memiliki tingkat variasi internal yang relatif rendah, yang menandakan homogenitas data di dalam klaster cukup baik. Selain itu, nilai jarak rata-rata data terhadap centroid menunjukkan bahwa klaster yang terbentuk bersifat cukup kompak dan sesuai dengan prinsip kerja algoritma K-Means. Perbedaan nilai rata-rata antar klaster juga menunjukkan adanya pemisahan karakteristik iklim yang jelas. Hasil penelitian ini sejalan dengan penelitian Selvida pada tahun 2025 yang menunjukkan bahwa integrasi PCA dan K-Means mampu menghasilkan pengelompokan wilayah pertanian yang lebih stabil.

4. Kesimpulan

Berdasarkan analisis yang dilakukan, penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia menjadi empat klaster berdasarkan karakteristik iklim yang berbeda. *Cluster 3*, yang terdiri atas Jawa Barat, Sulawesi Selatan, dan Papua Pegunungan, merupakan wilayah dengan potensi tertinggi untuk budidaya bawang putih karena

memiliki suhu rata-rata terendah (21,73°C) yang sesuai dengan persyaratan agroklimat tanaman tersebut. *Cluster* 2 yang mencakup 19 provinsi memiliki kesesuaian sedang, sedangkan *Cluster* 1 dan 0 masing-masing dicirikan oleh curah hujan tinggi serta suhu panas dan kondisi kering yang kurang ideal untuk bawang putih. Hasil klasterisasi ini memberikan peta wilayah prioritas yang dapat mendukung pengambilan kebijakan pengembangan sentra produksi bawang putih dalam negeri. Untuk penelitian selanjutnya, disarankan untuk mengintegrasikan variabel non-iklim seperti jenis tanah, topografi, dan infrastruktur pertanian guna memperoleh pemetaan potensi yang lebih komprehensif.

Daftar Rujukan

- [1] M. R. D. Maharani, S. P. Wulandari, F. Aryawan, and D. K. Nipu, *Analisis Posisi Dan Tingkat Ketergantungan Impor Bawang Putih Indonesia Di Pasar Internasional*, vol. 21, no. 2. 2024. doi: 10.24929/fp.v21i2.3874.
- [2] A. Titisari, *Budi Daya Bawang Putih*. Pusat Perpustakaan dan Penyebaran Teknologi Pertanian, 2020.
- [3] Murmayani and Darwis, *Agroklimatologi*. 2024.
- [4] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, 2016, doi: 10.1186/s41044-016-0014-0.
- [5] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [6] J. Tamaela, E. Sedyono, and A. Setiawan, “Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara,” *J. Buana Inform.*, vol. 8, no. 3, pp. 151–160, 2017, doi: 10.24002/jbi.v8i3.1317.
- [7] A. Anisa, A. Turmudi Zy, and W. Hadikristanto, “Klasifikasi Kualitas Air dengan K-Means dan Decision Tree,” *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 5, no. 2, pp. 166–176, 2025, doi: 10.55382/jurnalpustakaai.v5i2.1114.
- [8] D. Selvida, A. F. Pulungan, and A. S. Huzafah, “Optimization of Garlic Cultivation Land Selection Using Pca and K-Means Approach in Spatial Intelligent System,” *Eastern-European J. Enterp. Technol.*, vol. 3, no. 2, pp. 54–64, 2025, doi: 10.15587/1729-4061.2025.325340.
- [9] Y. Anzari *et al.*, “PENERAPAN ALGORITMA K-MEANS UNTUK KLASTERISASI POLA IKLIM STUDI KASUS : PROVINSI JAMBI PERIODE 2020-2024,” vol. 6, no. 3, pp. 1037–1052, 2025, doi: 10.46576/djtechno.
- [10] B. Waryanto, T. Agustina, and D. Tjiptodharmono, “Analysis of farming efficiency and smart farming system development in supporting garlic self-sufficiency : A concept,” vol. 335, pp. 1–11, 2019, doi: 10.1088/1755-1315/335/1/012006.
- [11] D. Agustina, Y. I. Mukti, and S. Muntari, “Integrasi Particle Swam Optimization Menggunakan K-Means untuk Klasterisasi Pengangguran di Kota Pagar Alam,” *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 3, no. 1, pp. 34–41, 2023, doi: 10.55382/jurnalpustakaai.v3i1.543.
- [12] A. L. Cultivars *et al.*, “agronomy Influence of Di fferent Photoperiod and Temperature Regimes on Growth and Bulb Quality of Garlic,” 2019.
- [13] V. Kumbhar, A. Maru, and S. Kumari, “Spatial Data Mining-A tool for Spatial Decision Support System in Agriculture Management,” *J. Eng. Sci. Technol. Rev.*, vol. 15, no. 1, pp. 128–133, 2022, doi: 10.25103/jestr.151.16.
- [14] Kementerian Pertanian Republik Indonesia, *Statistik Pertanian*. 2024. [Online]. Available: https://satudata.pertanian.go.id/assets/docs/publikasi/STATISTIK_PERTANIAN_2024_c.pdf#page=11.28
- [15] D. Xu and Y. Tian, “A Comprehensive Survey of Clustering Algorithms,” *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, 2015, doi: 10.1007/s40745-015-0040-1.
