BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

Pemetaan Skor Pola Pangan Harapan Tingkat Provinsi Menggunakan K-Means Clustering dengan Inisialisasi Centroid Deterministik dan Validasi Multimetrik

Provincial Mapping of Desirable Dietary Pattern Scores Using K-Means Clustering with Deterministic Centroid Initialization and Multi-Metric Validation

Afriza Fliyani*, Wanayumini

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Asahan, Kisaran, Indonesia
Email: 1,*afrizaeliyanii@gmail.com, 2wanayumini@yahoo.co.id
Email Penulis Korespondensi: afrizaeliyanii@gmail.com
Received: **22/10/2026**, Revised: **30/10/2026**, Accepted: **31/10/2026**, Published**: 31/10/2026**

Abstrak−Ketahanan pangan tidak hanya ditentukan oleh ketersediaan pangan, tetapi juga oleh keragaman dan keseimbangan gizi konsumsi masyarakat yang di Indonesia diukur melalui Skor Pola Pangan Harapan. Perbedaan kondisi geografis, sosial, dan ekonomi antarprovinsi membuat capaian skor tersebut sangat beragam, sementara pemetaan yang objektif dan dapat diperbarui secara berkala masih terbatas. Penelitian ini bertujuan memetakan provinsi di Indonesia berdasarkan Skor Pola Pangan Harapan, menganalisis karakteristik setiap kelompok secara kewilayahan maupun antarwaktu, dan mengimplementasikan algoritma K-Means Clustering ke dalam sistem berbasis web. Data penelitian berupa 242 data skor provinsi tahun 2018–2024 dari portal Satu Data Indonesia. Data dinormalisasi dengan metode Min-Max, lalu dikelompokkan ke dalam tiga klaster menggunakan jarak Euclidean dengan centroid awal yang dipilih dari nilai tertinggi, median, dan terendah. Jumlah klaster diuji dengan metode Elbow, kualitas klaster diukur dengan Silhouette Score dan Davies-Bouldin Index, serta kestabilan hasil diuji melalui analisis sensitivitas terhadap 100 inisialisasi acak dan 100 inisialisasi K-Means++ disertai pengukuran runtime komputasi. Hasil penelitian menunjukkan bahwa algoritma konvergen pada iterasi kelima dan menghasilkan klaster Tinggi sebanyak 84 data dengan centroid 91,08, klaster Sedang sebanyak 109 data dengan centroid 83,43, dan klaster Rendah sebanyak 49 data dengan centroid 74,12. Nilai Silhouette Score sebesar 0,578 dan Davies-Bouldin Index sebesar 0,516 menunjukkan struktur klaster yang baik. Ketiga strategi inisialisasi hanya menghasilkan perbedaan keanggotaan pada dua data yang berada di batas klaster, sedangkan inisialisasi deterministik konvergen dalam 5 iterasi dengan runtime median 0,16 ms, lebih cepat daripada inisialisasi acak (0,29 ms) dan K-Means++ (0,30 ms). Jumlah provinsi berkategori Tinggi meningkat dari 5 provinsi pada tahun 2018 menjadi 22 provinsi pada tahun 2024, sedangkan Nusa Tenggara Timur dan Maluku konsisten berada pada kategori Rendah. Sistem web yang dibangun menghasilkan keluaran yang identik dengan perhitungan manual sehingga layak dimanfaatkan sebagai alat bantu penentuan prioritas kebijakan ketahanan pangan. Kontribusi penelitian ini meliputi (1) prosedur K-Means dengan inisialisasi deterministik yang dapat direproduksi dan teruji kestabilannya, (2) peta kategori Skor PPH seluruh provinsi selama tujuh tahun beserta batas skor antarkategori, dan (3) sistem berbasis web yang transparan dan dapat diperbarui sebagai alat bantu penentuan prioritas kebijakan ketahanan pangan.

Kata Kunci: K-Means Clustering; Pola Pangan Harapan; Ketahanan Pangan; Inisialisasi Centroid; Validasi Klaster

Abstract−Food security is determined not only by food availability but also by the diversity and nutritional balance of household consumption, which in Indonesia is measured by the Desirable Dietary Pattern score. Geographic, social, and economic differences among provinces lead to widely varying scores, while objective, updatable mappings remain limited. This study aimed to map Indonesian provinces based on their Desirable Dietary Pattern scores, to analyze the spatial and temporal characteristics of each group, and to implement the K-Means Clustering algorithm in a web-based system. The data consisted of 242 provincial score records for 2018–2024 obtained from the Indonesian One Data portal. The data were normalized with the Min-Max method and grouped into three clusters using Euclidean distance, with initial centroids taken from the highest, median, and lowest values. The number of clusters was examined with the Elbow method, cluster quality was assessed with the Silhouette Score and the Davies-Bouldin Index, and the stability of the results was tested through a sensitivity analysis of 100 random and 100 K-Means++ initializations together with a computational runtime measurement. The algorithm converged at the fifth iteration and produced a High cluster of 84 records (centroid 91.08), a Medium cluster of 109 records (centroid 83.43), and a Low cluster of 49 records (centroid 74.12). A Silhouette Score of 0.578 and a Davies-Bouldin Index of 0.516 indicated a good cluster structure. The three initialization strategies differed in the membership of only two records lying at cluster boundaries, while the deterministic initialization converged in 5 iterations with a median runtime of 0.16 ms, faster than random initialization (0.29 ms) and K-Means++ (0.30 ms). The number of provinces in the High category rose from 5 in 2018 to 22 in 2024, whereas East Nusa Tenggara and Maluku consistently remained in the Low category. The web system produced results identical to the manual calculation, making it a suitable tool for prioritizing food security policy. The contributions of this study are (1) a reproducible K-Means procedure with deterministic initialization whose stability has been tested, (2) a seven-year category map of the Desirable Dietary Pattern scores of all provinces with explicit score boundaries between categories, and (3) a transparent, updatable web-based system that supports the prioritization of food security policies.

Keywords: K-Means Clustering; Desirable Dietary Pattern; Food Security; Centroid Initialization; Cluster Validation

1. PENDAHULUAN

Ketahanan pangan merupakan kondisi terpenuhinya kebutuhan pangan dan gizi seluruh lapisan masyarakat hingga tingkat perseorangan, yang tercermin dari ketersediaan pangan yang cukup, aman, beragam, bergizi, merata, dan terjangkau [1].

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

Definisi tersebut menegaskan bahwa ketahanan pangan tidak cukup diukur dari jumlah pangan yang tersedia, tetapi juga dari mutu dan keragaman konsumsi masyarakat. Untuk menilai aspek mutu tersebut, pemerintah menetapkan Pola Pangan Harapan (PPH) sebagai instrumen yang mengukur jumlah dan komposisi pangan berdasarkan sembilan kelompok pangan, yaitu padi-padian, umbi-umbian, pangan hewani, minyak dan lemak, buah/biji berminyak, kacang-kacangan, gula, sayur dan buah, serta aneka bumbu dan bahan minuman. Hasil penilaian dinyatakan dalam Skor PPH dengan nilai ideal 100 dan dikategorikan ke dalam tingkatan baik, sedang, dan kurang [2].

Capaian Skor PPH di Indonesia masih menghadapi sejumlah tantangan. Kajian historis terhadap tren konsumsi pangan nasional menunjukkan bahwa penyelarasan pola makan masyarakat dengan PPH masih memerlukan sumber karbohidrat yang lebih beragam, konsumsi sayur dan buah yang lebih tinggi, serta pengurangan ketergantungan pada pangan tertentu [3]. Pada tingkat rumah tangga, Skor PPH juga dipengaruhi oleh pendapatan, tingkat pendidikan, dan pengetahuan gizi ibu rumah tangga [4]. Di sisi lain, Indonesia memiliki 38 provinsi dengan kondisi geografis, sosial, ekonomi, dan budaya yang sangat beragam. Analisis spasial terhadap ketahanan pangan memperlihatkan bahwa wilayah rawan pangan cenderung terkonsentrasi di kawasan timur Indonesia yang dicirikan oleh keterisolasian geografis, kinerja ekonomi yang lemah, dan keterbatasan infrastruktur, sedangkan kelompok wilayah dengan ketahanan pangan baik banyak ditemukan di kawasan barat, khususnya Pulau Jawa [5]. Kondisi tersebut menunjukkan perlunya pemetaan capaian Skor PPH antarprovinsi secara objektif agar intervensi kebijakan dapat diarahkan kepada wilayah yang paling membutuhkan.

Salah satu pendekatan yang sesuai untuk pemetaan tersebut adalah *clustering*, yaitu teknik *unsupervised learning* yang mengelompokkan objek sehingga objek dalam satu kelompok memiliki kemiripan tinggi dan berbeda dengan objek pada kelompok lain [6]. Di antara berbagai algoritma *clustering*, K-Means merupakan algoritma partisi yang paling banyak digunakan karena sederhana, efisien secara komputasi, dan hasilnya mudah diinterpretasikan [7], [8]. Tinjauan komprehensif terhadap perkembangan K-Means menunjukkan bahwa algoritma ini tetap relevan pada era data besar dan telah melahirkan berbagai varian untuk mengatasi kelemahannya, terutama sensitivitas terhadap pemilihan *centroid* awal dan penentuan jumlah klaster [8], [9]. Pengembangan seperti K-Means yang mampu mendeteksi jumlah klaster dan pencilan secara otomatis [10] serta perbaikan strategi inisialisasi berbasis K-Means++ [11] menunjukkan bahwa kualitas hasil K-Means sangat bergantung pada cara klaster awal dibentuk dan dievaluasi.

Sejumlah penelitian terdahulu menunjukkan keandalan K-Means pada berbagai jenis data numerik. Ramdhan et al. [12] menerapkan K-Means untuk mengelompokkan data persediaan barang ke dalam kelompok stok sedikit dan stok banyak dengan kinerja yang memadai, tetapi penelitian tersebut belum menguji jumlah klaster secara kuantitatif. Prastiwi et al. [13] memanfaatkan K-Means untuk menyusun strategi pengelolaan stok berdasarkan data transaksi penjualan minimarket dan menunjukkan bahwa algoritma ini tetap memberikan hasil yang bermakna pada data berskala kecil, meskipun cakupan datanya hanya dua puluh transaksi. Hendrastuty [14] menggunakan K-Means untuk mengevaluasi hasil pembelajaran siswa, sedangkan Sholeh dan Aeni [15] menegaskan pentingnya penggunaan metode Elbow, *Silhouette*, dan *Davies-Bouldin Index* secara bersamaan dalam mengevaluasi model K-Means karena masing-masing metrik menilai aspek klaster yang berbeda.

Pada bidang pangan dan gizi, terdapat lima penelitian terdahulu yang paling dekat dengan penelitian ini. Abdulah [5] memetakan ketahanan pangan 514 kabupaten/kota tahun 2019 menggunakan analisis autokorelasi spasial (*Moran's I* dan *Local Indicators of Spatial Association*) terhadap empat indikator, termasuk skor keragaman pangan, dan menemukan klaster rendah-rendah di kawasan timur Indonesia serta klaster tinggi-tinggi di kawasan barat, khususnya Pulau Jawa. Cipta et al. [16] mengelompokkan Skor PPH 507 kabupaten/kota tahun 2021–2024 menggunakan *agglomerative hierarchical clustering* dengan metode Ward dan memperoleh tiga zona dengan *Silhouette Coefficient* 0,40, dengan zona skor tinggi terkonsentrasi di Pulau Jawa dan zona skor rendah di kawasan timur Indonesia. Astria et al. [17] mengelompokkan 38 provinsi berdasarkan *Prevalence of Undernourishment* (PoU) tahun 2018–2024 yang dirata-ratakan per provinsi menggunakan K-Means dengan inisialisasi K-Means++ dan menemukan enam provinsi berisiko tinggi yang seluruhnya berada di Papua dan Maluku. Nilasari dan Safitri [18] mengelompokkan 35 provinsi berdasarkan lima indikator gizi anak dari Survei Status Gizi Indonesia (SSGI) 2024 untuk memprioritaskan intervensi *stunting*. Buleuw et al. [19] membandingkan K-Means dan metode Ward pada lima indikator ketahanan pangan 514 kabupaten/kota tahun 2024 dan menemukan bahwa K-Means unggul pada *Silhouette Score*, sedangkan Ward lebih baik pada sebagian besar ukuran validitas lainnya. Perbandingan kelima penelitian tersebut dengan penelitian ini disajikan pada Tabel 1.

Tabel 1. Perbandingan Penelitian Terdahulu dengan Penelitian Ini

Penelitian	Indikator dan Unit Analisis	Periode Data	Metode	Validasi	Luaran
Abdulah [5]	FSVA, skor keragaman pangan, skor variasi pangan, dan asupan kalori; 514 kabupaten/kota	2019	Autokorelasi spasial (*Moran's I*, LISA)	Uji signifikansi *Moran's I*	Peta klaster spasial ketahanan pangan
Cipta et al. [16]	Skor PPH; 507 kabupaten/kota	2021–2024	*Hierarchical clustering* (Ward)	*Silhouette* 0,40	Peta zona Skor PPH

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2720
This work is licensed under a Creative Commons Attribution 4.0 International License.

Penelitian	Indikator dan Unit Analisis	Periode Data	Metode	Validasi	Luaran
Astria et al. [17]	PoU; 38 provinsi	2018–2024 (dirata-ratakan)	K-Means (inisialisasi K-Means++)	Elbow, *Silhouette* 0,593, DBI 0,558	Peta prioritas kebijakan
Nilasari dan Safitri [18]	Lima indikator gizi anak; 35 provinsi	2024	*Hierarchical clustering* (Ward.D2)	*Silhouette*, Dunn, *Connectivity*, kofenetik	Prioritas intervensi *stunting*
Buleuw et al. [19]	Lima indikator ketahanan pangan; 514 kabupaten/kota	2024	K-Means dan Ward	*Silhouette*, DBI, Dunn, dan lainnya	Empat klaster; Ward terpilih
Penelitian ini	Skor PPH; 34–38 provinsi (242 data)	2018–2024 (per tahun)	K-Means (inisialisasi deterministik) dan uji sensitivitas inisialisasi acak dan K-Means++	Elbow, *Silhouette* 0,578, DBI 0,516, *runtime*	Matriks kategori antarwaktu dan sistem web

Tabel 1 memperlihatkan empat kesenjangan antara penelitian terdahulu dan penelitian ini. Pertama, dari sisi indikator, hanya Cipta et al. [16] yang menggunakan Skor PPH, sedangkan Abdulah [5] menggunakan skor keragaman pangan yang belum mengikuti bobot sembilan kelompok pangan PPH, dan tiga penelitian lainnya menggunakan indikator kecukupan konsumsi, gizi anak, atau ketahanan pangan umum yang tidak mengukur mutu dan keragaman konsumsi pangan. Kedua, dari sisi dimensi waktu, Abdulah [5], Nilasari dan Safitri [18], serta Buleuw et al. [19] hanya menggunakan data satu tahun, sedangkan Astria et al. [17] merata-ratakan data 2018–2024 sehingga perubahan kategori setiap provinsi dari tahun ke tahun tidak dapat diamati. Ketiga, dari sisi metode, kelima penelitian tersebut tidak melaporkan pengujian kestabilan hasil terhadap strategi inisialisasi *centroid* maupun pengukuran *runtime* komputasi, padahal kualitas hasil K-Means sangat dipengaruhi oleh pemilihan *centroid* awal [8], [9]. Keempat, dari sisi luaran, kelima penelitian tersebut berhenti pada tahap analisis dan tidak melaporkan pengembangan sistem yang dapat digunakan kembali oleh pemangku kepentingan ketika data terbaru tersedia.

Berdasarkan kesenjangan tersebut, penelitian ini mengisi celah berupa belum tersedianya pemetaan Skor PPH tingkat provinsi berbasis data tahunan dalam rentang waktu panjang yang menggunakan K-Means dengan validasi jumlah klaster, kualitas klaster, dan kestabilan inisialisasi secara lengkap, sekaligus diimplementasikan ke dalam sistem berbasis web yang menampilkan setiap tahapan perhitungan secara transparan. Penelitian ini bertujuan untuk (1) memetakan provinsi di Indonesia berdasarkan Skor PPH menggunakan algoritma K-Means Clustering, (2) menganalisis karakteristik dan pola setiap kelompok secara kewilayahan maupun antarwaktu, (3) menguji kestabilan hasil pengelompokan terhadap strategi inisialisasi *centroid* beserta efisiensi komputasinya, dan (4) mengimplementasikan algoritma tersebut ke dalam sistem berbasis web menggunakan PHP dan MySQL.

Kontribusi penelitian ini terdiri atas tiga aspek. Pertama, secara metodologis, penelitian ini menyajikan prosedur K-Means dengan inisialisasi *centroid* deterministik yang dapat direproduksi dan telah diuji kestabilannya melalui analisis sensitivitas terhadap inisialisasi acak dan K-Means++ serta pengukuran *runtime* komputasi. Kedua, secara empiris, penelitian ini menghasilkan peta kategori Skor PPH seluruh provinsi selama tujuh tahun beserta batas skor antarkategori yang eksplisit sehingga perubahan posisi setiap provinsi dapat ditelusuri. Ketiga, secara praktis, penelitian ini menghasilkan sistem berbasis web yang transparan dan dapat diperbarui sebagai alat bantu pemerintah daerah dan dinas ketahanan pangan dalam menentukan prioritas program diversifikasi pangan yang lebih adil dan tepat sasaran.

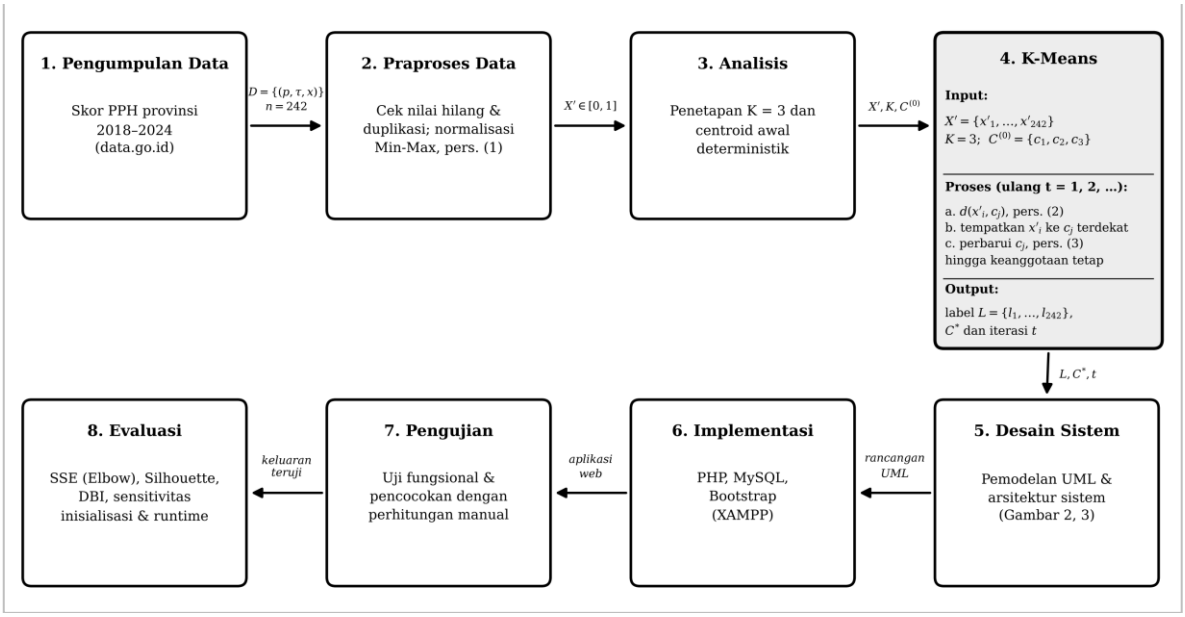
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan yang ditunjukkan pada Gambar 1. Tahapan diawali dengan pengumpulan data Skor PPH dari portal data terbuka pemerintah, kemudian dilanjutkan dengan praproses data yang meliputi pemeriksaan nilai yang hilang, penghapusan duplikasi, dan normalisasi Min-Max. Tahap analisis menetapkan jumlah klaster dan *centroid* awal, lalu tahap penerapan K-Means menghitung jarak Euclidean dan memperbarui *centroid* hingga konvergen. Hasil pengelompokan kemudian dimodelkan ke dalam rancangan sistem menggunakan *Unified Modeling Language* (UML), diimplementasikan menggunakan PHP, MySQL, dan Bootstrap, diuji secara fungsional, dan dievaluasi menggunakan metrik validasi klaster, analisis sensitivitas inisialisasi, dan pengukuran *runtime* komputasi. Gambar 1 juga memperlihatkan notasi masukan dan keluaran setiap tahap. Tahap pengumpulan data menghasilkan himpunan $D = \{(p, \tau, x)\}$ yang berisi provinsi p, tahun τ, dan Skor PPH x sebanyak $n = 242$ data, lalu tahap praproses mengubahnya menjadi $X' \in [0, 1]$. Blok K-Means menerima masukan X', $K = 3$, dan *centroid* awal $C^{(0)}$, kemudian menghasilkan keluaran berupa label klaster $L = \{l_1, \ldots, l_{242}\}$, *centroid* akhir C^*, dan jumlah iterasi t yang selanjutnya digunakan pada tahap perancangan, implementasi, pengujian, dan evaluasi.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2721
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421



Gambar 1. Tahapan Penelitian

2.2 Data Penelitian

Objek penelitian adalah data resmi Skor PPH provinsi di Indonesia yang diperoleh dari portal Satu Data Indonesia (data.go.id). Data mencakup periode 2018–2024, dengan setiap baris data merepresentasikan Skor PPH satu provinsi pada satu tahun. Tahun 2018–2023 masing-masing memuat 34 provinsi, sedangkan tahun 2024 memuat 38 provinsi akibat pemekaran wilayah Papua, sehingga total data yang diolah berjumlah 242 data. Penggunaan seluruh periode bertujuan agar pola pengelompokan lebih representatif dan memungkinkan pengamatan perubahan kategori setiap provinsi dari tahun ke tahun. Rekapitulasi data disajikan pada Tabel 2, yang menunjukkan bahwa Skor PPH berkisar antara 64,3 hingga 96,3 dan rata-ratanya cenderung meningkat selama periode pengamatan.

Tabel 2. Rekapitulasi Data Skor PPH Provinsi per Tahun

Tahun	Jumlah Provinsi	Skor Minimum	Skor Maksimum	Rata-rata
2018	34	64,3	95,1	82,16
2019	34	65,9	94,4	82,75
2020	34	67,9	94,3	82,61
2021	34	70,7	95,3	82,48
2022	34	73,7	96,2	85,86
2023	34	74,3	96,3	87,00
2024	38	68,0	95,7	86,29

Tabel 2 memperlihatkan bahwa jumlah data per tahun tetap sebanyak 34 provinsi pada tahun 2018–2023 dan bertambah menjadi 38 provinsi pada tahun 2024. Skor minimum meningkat dari 64,3 pada tahun 2018 menjadi 74,3 pada tahun 2023, kemudian turun menjadi 68,0 pada tahun 2024 karena masuknya provinsi hasil pemekaran di Papua yang skornya relatif rendah. Rata-rata skor naik dari 82,16 menjadi 87,00 pada tahun 2023 dan sedikit menurun menjadi 86,29 pada tahun 2024, sedangkan skor maksimum relatif stabil pada kisaran 94,3–96,3. Pola tersebut menunjukkan bahwa perbaikan mutu konsumsi pangan terutama terjadi pada provinsi dengan skor rendah dan menengah, sementara rentang skor antarprovinsi tetap lebar sehingga pengelompokan relevan untuk dilakukan.

2.3 Algoritma K-Means Clustering

Sebelum pengelompokan, data dinormalisasi ke rentang [0, 1] menggunakan *Min-Max Normalization* sebagaimana ditunjukkan pada persamaan (1), dengan $x_{\min} = 64{,}3$ dan $x_{\max} = 96{,}3$ sebagai nilai terendah dan tertinggi dari seluruh data. Normalisasi membuat seluruh nilai berada pada skala yang seragam sehingga perhitungan jarak tidak dipengaruhi oleh satuan data. Sebagai contoh penerapan, Skor PPH 72,7 dinormalisasi menjadi $(72{,}7 - 64{,}3)/(96{,}3 - 64{,}3) = 0{,}263$.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Jumlah klaster ditetapkan sebanyak K = 3 yang merepresentasikan kategori Skor PPH tinggi, sedang, dan rendah, sejalan dengan tiga kategori penilaian PPH [2]. Karena K-Means sensitif terhadap inisialisasi [8], [11], *centroid* awal dipilih secara deterministik dari data aktual, yaitu data dengan skor tertinggi (DI Yogyakarta 2023, 96,3), data dengan skor mendekati median (Lampung 2020, 84,5), dan data dengan skor terendah (Papua 2018, 64,3), sehingga hasil

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2722
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

penelitian dapat direproduksi. Selanjutnya, jarak setiap data terhadap setiap *centroid* dihitung menggunakan jarak Euclidean pada persamaan (2), dan setiap data ditempatkan pada klaster dengan jarak terkecil.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - c_{j,k})^2} \quad (2)$$

Dengan $d(x_i, c_j)$ adalah jarak data x_i ke *centroid* c_j dan n adalah jumlah atribut. Posisi *centroid* kemudian diperbarui menggunakan rata-rata seluruh anggota klaster pada persamaan (3), dengan $|C_j|$ adalah jumlah anggota klaster j. Proses penghitungan jarak, penempatan data, dan pembaruan *centroid* diulang hingga keanggotaan klaster tidak lagi berubah [7], [8]. Karena penelitian ini hanya menggunakan satu atribut ($n = 1$), persamaan (2) tereduksi menjadi selisih absolut $|x'_i - c_j|$; sebagai contoh, jarak data Aceh 2018 ($x' = 0{,}263$) ke *centroid* awal $C_2 = 0{,}631$ adalah $|0{,}263 - 0{,}631| = 0{,}369$.

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (3)$$

2.4 Evaluasi Klaster

Kesesuaian jumlah klaster diuji dengan metode Elbow, yaitu mengamati penurunan *Sum of Squared Error* (SSE) pada beberapa nilai K dan memilih titik ketika penurunan SSE mulai melandai membentuk siku [15], [20]. SSE dihitung sebagai jumlah kuadrat jarak setiap data terhadap *centroid* klasternya sebagaimana ditunjukkan pada persamaan (4).

$$\text{SSE} = \sum_{j=1}^{K} \sum_{x_i \in C_j} d(x_i, c_j)^2 \quad (4)$$

Kualitas klaster diukur menggunakan dua metrik evaluasi internal. *Silhouette Score* pada persamaan (5) membandingkan rata-rata jarak suatu data ke anggota klasternya sendiri, $a(i)$, dengan rata-rata jarak terkecil ke klaster lain, $b(i)$; nilainya berada pada rentang −1 hingga 1, dan nilai yang makin tinggi menunjukkan klaster yang makin baik [21], [22]. *Davies-Bouldin Index* pada persamaan (6) mengukur rasio antara sebaran di dalam klaster, S_i, dan jarak antar-*centroid*, $M_{i,j}$; nilai yang makin kecil menunjukkan klaster yang makin padat dan terpisah dengan baik [23].

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

$$\text{DBI} = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \frac{S_i + S_j}{M_{i,j}} \quad (6)$$

Centroid akhir yang masih berada pada skala ternormalisasi dikembalikan ke skala Skor PPH asli melalui denormalisasi pada persamaan (7) agar dapat diinterpretasikan sebagai nilai skor.

$$x = x'(x_{\max} - x_{\min}) + x_{\min} \quad (7)$$

Kestabilan hasil terhadap pemilihan *centroid* awal diuji melalui analisis sensitivitas yang membandingkan inisialisasi deterministik dengan dua strategi pembanding, yaitu inisialisasi acak yang memilih tiga data secara acak sebagai *centroid* awal dan inisialisasi K-Means++ yang memilih *centroid* berikutnya dengan peluang sebanding dengan kuadrat jarak ke *centroid* terdekat yang telah terpilih [9], [11]. Setiap strategi pembanding dijalankan sebanyak 100 kali dengan *seed* 1 sampai 100, lalu setiap hasil dibandingkan berdasarkan jumlah iterasi, kesamaan keanggotaan klaster terhadap hasil deterministik, *Silhouette Score*, dan *Davies-Bouldin Index*. Kompleksitas waktu K-Means adalah $O(n \cdot K \cdot t)$, dengan n jumlah data, K jumlah klaster, dan t jumlah iterasi [8]. Oleh karena itu, *runtime* juga diukur secara empiris menggunakan fungsi hrtime pada PHP 8.3 (CLI) sebagai nilai median dari pengulangan berulang untuk meredam fluktuasi beban prosesor. Untuk melihat skalabilitas, pengukuran dilakukan pula pada data yang diperbesar 10 dan 100 kali melalui replikasi dengan gangguan acak kecil sebesar ±0,005 pada skala ternormalisasi.

2.5 Perancangan dan Implementasi Sistem

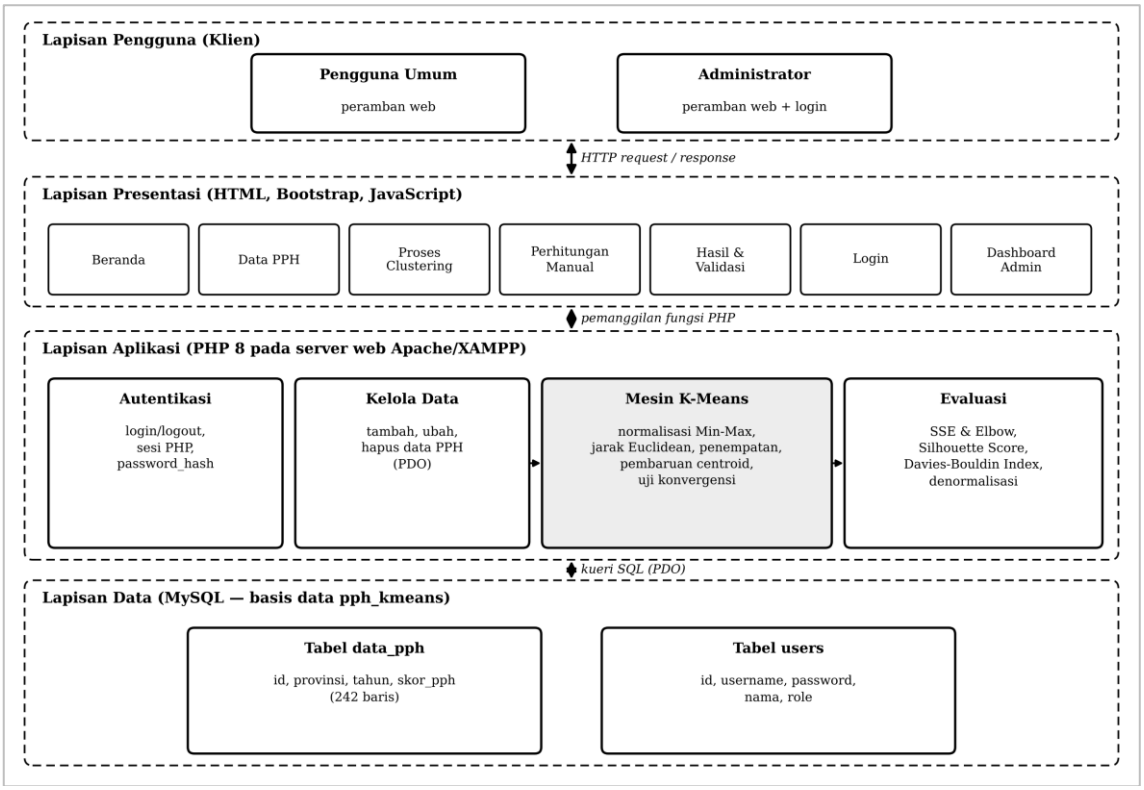
Sistem dirancang menggunakan UML yang meliputi *use case diagram*, *activity diagram*, *sequence diagram*, dan *class diagram*. Sistem diimplementasikan berbasis web menggunakan bahasa pemrograman PHP, basis data MySQL, dan *framework* Bootstrap pada lingkungan XAMPP. Algoritma K-Means ditulis langsung dalam PHP tanpa pustaka eksternal sehingga seluruh tahapan perhitungan berjalan di sisi *server* dan dapat ditampilkan kepada pengguna. Pengujian dilakukan secara fungsional terhadap seluruh fitur sistem serta dengan membandingkan keluaran sistem terhadap hasil perhitungan manual untuk memastikan bahwa algoritma diimplementasikan dengan benar. Rancangan arsitektur dan *use case* sistem masing-masing ditunjukkan pada Gambar 2 dan Gambar 3.

Arsitektur sistem pada Gambar 2 terdiri atas empat lapisan. Lapisan pengguna mencakup Pengguna Umum dan Administrator yang mengakses sistem melalui peramban web. Lapisan presentasi dibangun dengan HTML dan Bootstrap serta menyediakan halaman beranda, data PPH, proses *clustering*, perhitungan manual, hasil dan validasi, *login*, serta *dashboard* administrator. Lapisan aplikasi berisi modul autentikasi, modul pengelolaan data, mesin K-Means yang menjalankan normalisasi, penghitungan jarak, penempatan data, pembaruan *centroid*, dan uji konvergensi, serta modul evaluasi yang menghitung SSE, *Silhouette Score*, *Davies-Bouldin Index*, dan denormalisasi. Lapisan data menggunakan basis data MySQL yang menyimpan tabel data_pph dan tabel users. Hasil pengelompokan tidak disimpan secara

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

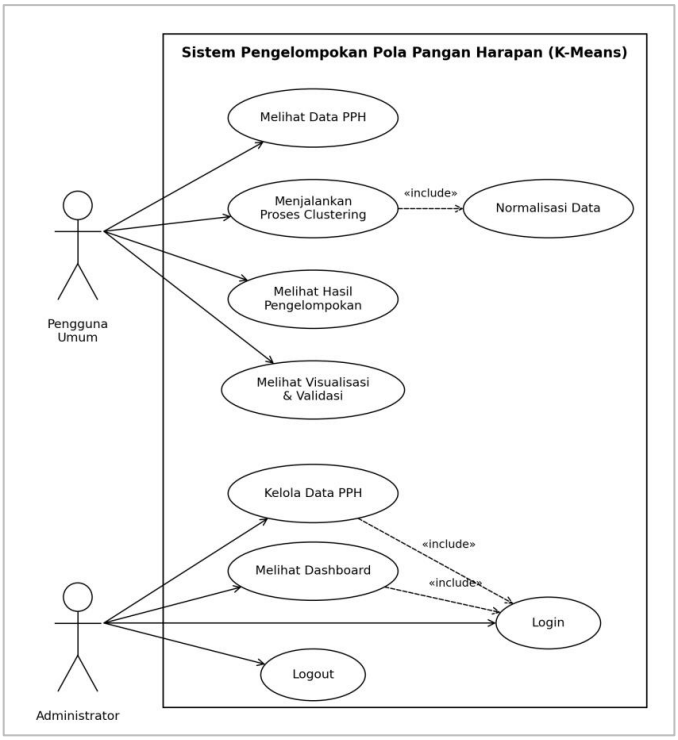
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

permanen, melainkan dihitung ulang dari basis data setiap kali proses dijalankan, sehingga setiap perubahan data langsung tercermin pada hasil.



Gambar 2. Arsitektur Sistem Berbasis Web

Fungsionalitas sistem dimodelkan dengan *use case diagram* pada Gambar 3 yang melibatkan dua aktor, yaitu Pengguna Umum dan Administrator. Pengguna Umum dapat melihat data PPH, menjalankan proses *clustering*, serta melihat hasil, visualisasi, dan validasi tanpa autentikasi. Administrator memiliki hak akses tambahan untuk mengelola data dan melihat *dashboard* setelah melakukan *login*. Relasi «include» pada Gambar 3 menunjukkan bahwa proses *clustering* selalu menyertakan tahap normalisasi data, sedangkan pengelolaan data dan *dashboard* selalu didahului proses autentikasi.



Gambar 3. Use Case Diagram Sistem

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2724
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

3. HASIL DAN PEMBAHASAN

3.1 Normalisasi Data dan Iterasi Pertama

Proses perhitungan diawali dengan normalisasi seluruh 242 data menggunakan persamaan (1). Sebagai contoh, data Aceh tahun 2018 dengan Skor PPH 72,7 menghasilkan nilai normalisasi (72,7 − 64,3)/(96,3 − 64,3) = 8,4/32 = 0,263, sedangkan data Sumatera Utara tahun 2018 dengan skor 84,8 menghasilkan nilai 20,5/32 = 0,641. Data dengan skor tertinggi menghasilkan nilai 1 dan data dengan skor terendah menghasilkan nilai 0, sehingga *centroid* awal C_1, C_2, dan C_3 masing-masing bernilai 1,000; 0,631; dan 0,000.

Pada iterasi pertama, jarak setiap data terhadap ketiga *centroid* awal dihitung menggunakan persamaan (2). Untuk data Aceh tahun 2018, jaraknya terhadap C_1 adalah $|0{,}263 - 1| = 0{,}738$, terhadap C_2 adalah $|0{,}263 - 0{,}631| = 0{,}369$, dan terhadap C_3 adalah $|0{,}263 - 0| = 0{,}263$. Karena jarak terkecil terdapat pada C_3, data tersebut ditempatkan pada klaster 3. Contoh hasil perhitungan untuk lima data pertama disajikan pada Tabel 3, yang memperlihatkan bahwa sebagian besar data dengan skor di sekitar nilai tengah langsung tertarik ke C_2 pada iterasi pertama.

Tabel 3. Contoh Perhitungan Jarak Euclidean pada Iterasi Pertama

No	Provinsi (Tahun)	Skor PPH	x'	d(C_1)	d(C_2)	d(C_3)	Klaster
1	Aceh (2018)	72,7	0,263	0,738	0,369	0,263	C_3
2	Sumatera Utara (2018)	84,8	0,641	0,359	0,009	0,641	C_2
3	Sumatera Barat (2018)	81,2	0,528	0,472	0,103	0,528	C_2
4	Riau (2018)	83,5	0,600	0,400	0,031	0,600	C_2
5	Jambi (2018)	85,3	0,656	0,344	0,025	0,656	C_2

Tabel 3 menunjukkan bahwa penempatan data ditentukan oleh jarak terkecil di antara ketiga *centroid*. Data Aceh 2018 memiliki jarak terkecil 0,263 ke C_3 sehingga masuk klaster 3, sedangkan empat data lainnya memiliki jarak terkecil ke C_2 dengan nilai antara 0,009 dan 0,103. Data Sumatera Utara 2018 memiliki jarak paling kecil ke C_2 (0,009) karena skornya (84,8) hampir sama dengan skor data acuan *centroid* awal C_2 (84,5). Setelah seluruh 242 data diproses dengan cara yang sama, iterasi pertama menempatkan 47 data pada C_1, 176 data pada C_2, dan 19 data pada C_3.

3.2 Pergerakan Centroid dan Konvergensi

Setelah seluruh data ditempatkan, *centroid* diperbarui menggunakan persamaan (3). Pada akhir iterasi pertama, klaster C_1 memiliki 47 anggota dengan rata-rata nilai ternormalisasi 0,895, sehingga nilai tersebut menjadi *centroid* C_1 pada iterasi kedua. Pergerakan *centroid* dan jumlah anggota klaster pada setiap iterasi disajikan pada Tabel 4. Secara rinci, jumlah nilai ternormalisasi anggota C_1 adalah 42,088 dari 47 data sehingga *centroid* barunya 42,088/47 = 0,895; jumlah anggota C_2 adalah 104,650 dari 176 data sehingga *centroid* barunya 104,650/176 = 0,595; dan jumlah anggota C_3 adalah 3,756 dari 19 data sehingga *centroid* barunya 3,756/19 = 0,198.

Tabel 4. Pergerakan Centroid dan Jumlah Anggota Klaster per Iterasi

Iterasi	C_1	C_2	C_3	$\|C_1\|$	$\|C_2\|$	$\|C_3\|$
1	1,000	0,631	0,000	47	176	19
2	0,895	0,595	0,198	75	124	43
3	0,849	0,599	0,290	83	110	49
4	0,838	0,599	0,307	84	109	49
5	0,837	0,598	0,307	84	109	49

Tabel 4 menunjukkan bahwa perubahan posisi *centroid* makin kecil dari iterasi ke iterasi. *Centroid* C_1 bergeser sebesar 0,105 pada iterasi kedua, 0,046 pada iterasi ketiga, 0,011 pada iterasi keempat, dan hanya 0,001 pada iterasi kelima. Pola serupa terjadi pada C_3 yang bergerak dari 0,000 ke 0,198, lalu ke 0,290, dan stabil pada 0,307. Perubahan keanggotaan terbesar terjadi pada iterasi kedua, ketika klaster C_2 menyusut dari 176 menjadi 124 anggota karena *centroid* C_1 dan C_3 yang awalnya berada pada nilai ekstrem bergerak mendekati pusat sebaran datanya. Pada iterasi kelima, keanggotaan klaster sama persis dengan iterasi keempat, yaitu 84, 109, dan 49 data, sehingga algoritma dinyatakan konvergen.

Setelah denormalisasi menggunakan persamaan (7), *centroid* akhir bernilai $C_1 = 0{,}837 \times 32 + 64{,}3 = 91{,}08$; $C_2 = 0{,}598 \times 32 + 64{,}3 = 83{,}43$; dan $C_3 = 0{,}307 \times 32 + 64{,}3 = 74{,}12$. Karena data hanya memiliki satu atribut, batas antarklaster pada kondisi konvergen berada tepat di titik tengah antara dua *centroid* yang berdekatan. Dengan demikian, Skor PPH di atas sekitar 87,3 termasuk klaster Tinggi, skor antara sekitar 78,8 hingga 87,3 termasuk klaster Sedang, dan skor di bawah sekitar 78,8 termasuk klaster Rendah. Batas yang eksplisit ini memudahkan pemangku kepentingan menafsirkan posisi suatu provinsi tanpa harus mengulang seluruh perhitungan. Secara lebih presisi, kedua batas tersebut bernilai 87,26 dan 78,77 sehingga skor 87,3 masih termasuk klaster Tinggi dan skor 78,8 masih termasuk klaster Sedang.

3.3 Penentuan Jumlah Klaster dengan Metode Elbow

Untuk memastikan kesesuaian jumlah klaster, SSE dihitung untuk K = 1 sampai K = 6 sebagaimana ditampilkan pada Tabel 5 dan Gambar 4. Penurunan SSE sangat tajam dari K = 1 (10,385) ke K = 2 (3,628), yaitu sebesar 6,757, lalu dari

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2725
This work is licensed under a Creative Commons Attribution 4.0 International License.

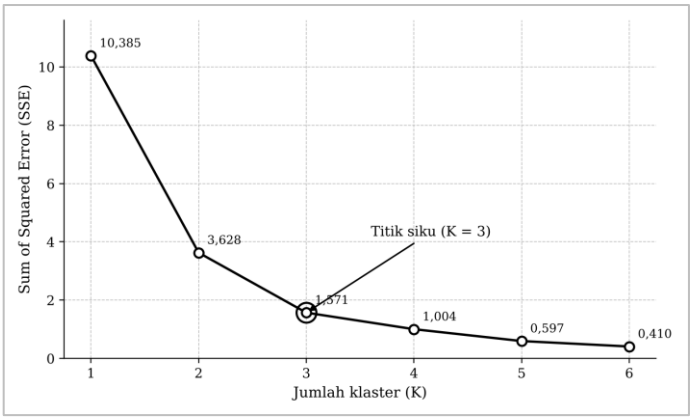
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

K = 2 ke K = 3 (1,571) sebesar 2,057. Setelah itu penurunan SSE mengecil secara nyata, yaitu 0,567 dari K = 3 ke K = 4, 0,407 dari K = 4 ke K = 5, dan 0,187 dari K = 5 ke K = 6. Titik siku terbentuk pada K = 3 karena penambahan klaster setelah titik tersebut tidak lagi memberikan pengurangan galat yang sebanding. Hasil ini mendukung penetapan tiga kategori tinggi, sedang, dan rendah yang sekaligus selaras dengan tiga tingkatan penilaian PPH dalam regulasi [2].

Tabel 5. Nilai SSE pada Metode Elbow

Jumlah Klaster (K)	1	2	3	4	5	6
SSE	10,385	3,628	1,571	1,004	0,597	0,410
Penurunan SSE	–	6,757	2,057	0,567	0,407	0,187

Tabel 5 memperlihatkan bahwa penurunan SSE dari K = 2 ke K = 3 masih sebesar 2,057 atau 56,7% dari SSE pada K = 2, sedangkan penurunan dari K = 3 ke K = 4 hanya 0,567 atau 36,1% dari SSE pada K = 3, dan penurunan berikutnya terus mengecil. Pola penurunan tersebut divisualisasikan pada Gambar 4, yang memperlihatkan perubahan kemiringan kurva paling tegas pada K = 3.



Gambar 4. Grafik Metode Elbow

3.4 Validasi Kualitas Klaster

Kualitas hasil pengelompokan dengan K = 3 ditunjukkan pada Tabel 6. Nilai *Silhouette Score* sebesar 0,578 berada di atas 0,5 yang menandakan bahwa sebagian besar data lebih dekat ke anggota klasternya sendiri dibandingkan ke klaster lain, sehingga struktur klaster tergolong baik [21], [22]. Nilai Davies-Bouldin Index sebesar 0,516 yang relatif kecil menunjukkan bahwa sebaran di dalam setiap klaster jauh lebih kecil daripada jarak antar-*centroid*, sehingga klaster-klaster yang terbentuk padat dan terpisah dengan jelas [23]. Kedua metrik tersebut saling menguatkan hasil metode Elbow bahwa K = 3 merupakan konfigurasi yang tepat untuk data penelitian ini.

Tabel 6. Hasil Validasi Kualitas Klaster

Metrik Evaluasi	Nilai	Interpretasi
Sum of Squared Error (SSE)	1,571	Makin kecil makin baik
Silhouette Score	0,578	Mendekati 1 makin baik
Davies-Bouldin Index	0,516	Makin kecil makin baik

Tabel 6 merangkum ketiga metrik evaluasi untuk K = 3. Perhitungan *Silhouette Score* dapat dicontohkan pada data Aceh 2018 (x' = 0,263) yang berada pada klaster Rendah. Rata-rata jaraknya ke anggota klaster sendiri adalah $a(i)$ = 0,103, sedangkan rata-rata jaraknya ke klaster Sedang dan Tinggi masing-masing 0,335 dan 0,575, sehingga $b(i)$ = 0,335 dan $s(i) = (0{,}335 - 0{,}103)/0{,}335 = 0{,}693$. Rata-rata $s(i)$ per klaster adalah 0,593 untuk klaster Tinggi, 0,563 untuk klaster Sedang, dan 0,588 untuk klaster Rendah, serta tidak ada satu pun data yang memiliki nilai $s(i)$ negatif. Nilai $s(i)$ terkecil terdapat pada data Sumatera Selatan 2020 (0,011) dan Sulawesi Barat 2019 (0,002), yaitu dua data yang skornya berada di sekitar batas antarklaster. Untuk *Davies-Bouldin Index*, rata-rata jarak anggota ke *centroid* klaster Tinggi, Sedang, dan Rendah adalah S_1 = 0,0680, S_2 = 0,0581, dan S_3 = 0,0851, sedangkan jarak antar-*centroid* adalah $M_{1,2}$ = 0,2392, $M_{1,3}$ = 0,5302, dan $M_{2,3}$ = 0,2909. Rasio terbesar untuk klaster 1, 2, dan 3 berturut-turut adalah 0,5272; 0,5272; dan 0,4922, sehingga DBI = (0,5272 + 0,5272 + 0,4922)/3 = 0,5155 atau 0,516.

3.5 Hasil Akhir Pemetaan Provinsi

Penamaan kategori ditentukan berdasarkan nilai *centroid* akhir pada skala Skor PPH asli. Ringkasan hasil akhir pengelompokan disajikan pada Tabel 7. Klaster Sedang merupakan kelompok terbesar dengan 109 data (45,0%), diikuti klaster Tinggi sebanyak 84 data (34,7%) dan klaster Rendah sebanyak 49 data (20,2%). Selisih *centroid* antara klaster Tinggi dan Rendah mencapai 16,96 poin, yang menggambarkan kesenjangan mutu konsumsi pangan yang cukup lebar antarwilayah.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2726
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

Tabel 7. Ringkasan Hasil Akhir Pengelompokan

Klaster	Kategori	Centroid Skor PPH	Jumlah Data	Persentase
1	Tinggi	91,08	84	34,7%
2	Sedang	83,43	109	45,0%
3	Rendah	74,12	49	20,2%
	Jumlah		242	100%

Tabel 7 juga menunjukkan bahwa rentang skor anggota setiap klaster tidak saling tumpang tindih, yaitu 87,3–96,3 untuk klaster Tinggi, 78,8–87,1 untuk klaster Sedang, dan 64,3–78,3 untuk klaster Rendah. Klaster Rendah memiliki rentang terlebar (14,0 poin) meskipun jumlah anggotanya paling sedikit, yang menandakan bahwa provinsi dengan capaian rendah lebih beragam dibandingkan provinsi pada dua klaster lainnya.

Karena data mencakup tujuh tahun, hasil pengelompokan juga memperlihatkan perubahan sebaran kategori dari waktu ke waktu sebagaimana disajikan pada Tabel 8. Jumlah provinsi berkategori Tinggi meningkat dari 5 provinsi pada tahun 2018 menjadi 22 provinsi pada tahun 2024, sedangkan jumlah provinsi berkategori Sedang menurun dari 22 menjadi 9 provinsi. Lonjakan terbesar terjadi pada tahun 2022, ketika jumlah provinsi berkategori Tinggi bertambah dari 8 menjadi 18 provinsi. Sebaliknya, jumlah provinsi berkategori Rendah relatif tidak berubah dan berada pada kisaran 5 hingga 10 provinsi setiap tahun.

Tabel 8. Sebaran Jumlah Provinsi per Kategori setiap Tahun

Kategori	2018	2019	2020	2021	2022	2023	2024
Tinggi	5	6	5	8	18	20	22
Sedang	22	22	21	16	10	9	9
Rendah	7	6	8	10	6	5	7
Jumlah	34	34	34	34	34	34	38

Berdasarkan Tabel 8, proporsi provinsi berkategori Tinggi meningkat dari 14,7% (5 dari 34 provinsi) pada tahun 2018 menjadi 57,9% (22 dari 38 provinsi) pada tahun 2024, sedangkan proporsi provinsi berkategori Rendah relatif stabil pada kisaran 14,7%–29,4% setiap tahun.

Matriks kategori setiap provinsi yang memiliki data lengkap tahun 2018–2024 disajikan pada Tabel 9, sedangkan empat provinsi hasil pemekaran yang baru tercatat pada tahun 2024 terbagi ke dalam kategori Sedang (dua provinsi) dan Rendah (dua provinsi). Dari 34 provinsi dalam Tabel 9, sebanyak 20 provinsi mengalami kenaikan kategori antara tahun 2018 dan 2024, 13 provinsi berada pada kategori yang sama, dan hanya Papua Barat yang mengalami penurunan dari Sedang menjadi Rendah. Sebanyak 17 provinsi konsisten berada pada kategori Tinggi selama tiga tahun terakhir (2022–2024).

Tabel 9. Matriks Kategori Klaster Tiap Provinsi Tahun 2018–2024

No	Provinsi	2018	2019	2020	2021	2022	2023	2024
1	Aceh	R	R	R	R	S	S	S
2	Sumatera Utara	S	S	S	S	T	T	T
3	Sumatera Barat	S	S	S	S	S	T	T
4	Riau	S	S	S	S	T	T	T
5	Jambi	S	S	S	S	T	T	T
6	Sumatera Selatan	S	S	T	T	T	T	T
7	Bengkulu	S	S	S	T	T	T	T
8	Lampung	S	S	S	S	T	T	T
9	Kepulauan Bangka Belitung	S	S	S	T	S	T	T
10	Kepulauan Riau	S	S	S	S	T	T	T
11	DKI Jakarta	T	T	S	S	T	T	T
12	Jawa Barat	S	S	S	S	T	T	T
13	Jawa Tengah	T	T	T	T	T	T	T
14	DI Yogyakarta	T	T	T	T	T	T	T
15	Jawa Timur	T	T	S	T	T	T	T
16	Banten	S	S	S	S	T	T	T
17	Bali	T	T	T	T	T	T	T
18	Nusa Tenggara Barat	S	T	T	T	T	T	T
19	Nusa Tenggara Timur	R	R	R	R	R	R	R
20	Kalimantan Barat	R	R	R	R	S	S	S
21	Kalimantan Tengah	S	S	S	S	T	T	T
22	Kalimantan Selatan	S	S	S	S	T	T	T
23	Kalimantan Timur	S	S	S	S	R	T	T
24	Kalimantan Utara	S	S	S	S	S	S	S
25	Sulawesi Utara	S	S	S	S	T	S	T

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2727
This work is licensed under a Creative Commons Attribution 4.0 International License.

No	Provinsi	2018	2019	2020	2021	2022	2023	2024
26	Sulawesi Tengah	S	S	S	R	S	S	S
27	Sulawesi Selatan	S	S	S	S	S	S	T
28	Sulawesi Tenggara	S	S	S	S	S	S	S
29	Gorontalo	R	S	R	R	R	R	R
30	Sulawesi Barat	S	S	R	R	R	R	S
31	Maluku	R	R	R	R	R	R	R
32	Maluku Utara	R	R	R	R	S	R	S
33	Papua Barat	S	R	R	R	S	S	R
34	Papua	R	S	S	R	R	S	R

Keterangan: T = Tinggi, S = Sedang, R = Rendah.

Tabel 9 juga memperlihatkan bahwa perubahan kategori tidak selalu bergerak satu arah. Kalimantan Timur merupakan satu-satunya provinsi yang pernah menempati ketiga kategori, yaitu Sedang pada tahun 2018–2021, Rendah pada tahun 2022, lalu Tinggi pada tahun 2023–2024. Fluktuasi serupa terlihat pada Papua, Papua Barat, dan Maluku Utara yang berpindah antara kategori Sedang dan Rendah dari tahun ke tahun. Pola ini menunjukkan bahwa capaian Skor PPH di sebagian provinsi kawasan timur belum stabil dan rentan terhadap perubahan kondisi pasokan maupun daya beli masyarakat. Sebaliknya, sebagian besar provinsi di Sumatera serta Kalimantan Tengah dan Kalimantan Selatan yang naik ke kategori Tinggi bertahan pada kategori tersebut hingga tahun 2024. Perbedaan stabilitas tersebut merupakan informasi penting bagi perencana kebijakan, karena provinsi dengan kategori yang berfluktuasi memerlukan pendekatan penguatan ketahanan yang berbeda dari provinsi yang konsisten tertinggal.

3.6 Analisis Sensitivitas Inisialisasi dan *Runtime* Komputasi

Untuk menguji apakah hasil pada subbab sebelumnya bergantung pada pilihan *centroid* awal, algoritma dijalankan ulang menggunakan inisialisasi acak dan K-Means++ masing-masing sebanyak 100 kali sesuai prosedur pada subbab 2.4. Ringkasan hasil ketiga strategi inisialisasi disajikan pada Tabel 10.

Tabel 10. Hasil Analisis Sensitivitas Strategi Inisialisasi Centroid

Strategi Inisialisasi	Run	Iterasi Rata-rata (Min–Maks)	Partisi 84/109/49	Partisi 83/109/50	Partisi Lain	Silhouette	DBI
Deterministik	1	5,00 (5–5)	1 (100%)	0 (0%)	0 (0%)	0,578	0,516
Acak	100	9,13 (3–14)	36 (36%)	64 (64%)	0 (0%)	0,578	0,515–0,516
K-Means++	100	7,38 (3–13)	52 (52%)	47 (47%)	1 (1%)	0,578	0,515–0,516

Tabel 10 menunjukkan bahwa ketiga strategi inisialisasi hanya menghasilkan dua konfigurasi partisi utama. Konfigurasi pertama, yaitu 84/109/49 dengan *centroid* 91,08; 83,43; dan 74,12, identik dengan hasil inisialisasi deterministik dan diperoleh pada 36% *run* acak serta 52% *run* K-Means++. Konfigurasi kedua, yaitu 83/109/50 dengan *centroid* 91,13; 83,51; dan 74,21, diperoleh pada 64% *run* acak dan 47% *run* K-Means++, sedangkan satu *run* K-Means++ sisanya menghasilkan partisi 84/108/50. Perbedaan antarkonfigurasi tersebut hanya terletak pada dua data, yaitu Sumatera Selatan 2020 (skor 87,3) yang berpindah dari klaster Tinggi ke Sedang dan Sulawesi Barat 2019 (skor 78,8) yang berpindah dari klaster Sedang ke Rendah. Kedua data tersebut hanya berjarak sekitar 0,04 dan 0,03 poin dari batas antarklaster 87,26 dan 78,77, serta memiliki nilai $s(i)$ terkecil pada subbab 3.4. Dengan demikian, 240 dari 242 data (99,2%) selalu menempati klaster yang sama apa pun strategi inisialisasinya.

Konfigurasi kedua memiliki SSE yang sedikit lebih kecil daripada hasil deterministik, dengan selisih kurang dari 0,0001 atau sekitar 0,005%. Artinya, inisialisasi deterministik tidak menjamin tercapainya optimum global, tetapi selisih tersebut terlalu kecil untuk mengubah kesimpulan pemetaan. Jika konfigurasi kedua digunakan, hanya dua sel pada Tabel 9 yang berubah, sedangkan seluruh temuan kewilayahan dan antarwaktu tetap sama. *Silhouette Score* seluruh *run* tetap 0,578 dan *Davies-Bouldin Index* berada pada rentang 0,515–0,516, sehingga kualitas klaster tidak sensitif terhadap strategi inisialisasi. Selain kestabilan hasil, efisiensi komputasi ketiga strategi diukur sebagaimana disajikan pada Tabel 11.

Tabel 11. Runtime Median K-Means per Strategi Inisialisasi (ms)

Ukuran Data (n)	Deterministik	Acak	K-Means++
242 (data penelitian)	0,16 (5,0)	0,29 (9,0)	0,30 (7,7)
2.420 (replikasi 10×)	1,91 (6,0)	3,60 (11,2)	3,62 (9,7)
24.200 (replikasi 100×)	22,65 (7,0)	64,59 (13,5)	43,69 (11,9)

Keterangan: angka dalam kurung adalah rata-rata jumlah iterasi; median dihitung dari 2.000, 300, dan 40 pengulangan untuk masing-masing ukuran data.

Tabel 11 menunjukkan bahwa pada data penelitian (n = 242), inisialisasi deterministik membutuhkan waktu median 0,16 ms, sekitar 1,8–1,9 kali lebih cepat daripada inisialisasi acak (0,29 ms) dan K-Means++ (0,30 ms). Keunggulan ini berasal dari jumlah iterasi yang lebih sedikit karena *centroid* awal sudah tersebar pada nilai tertinggi,

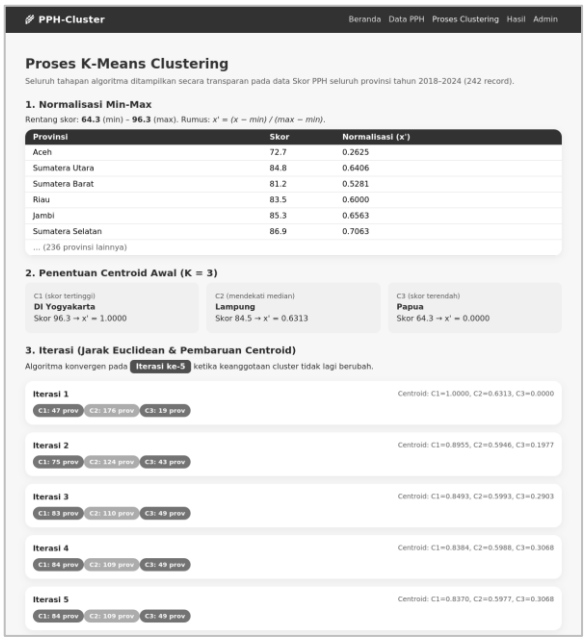
Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2719-2732
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1421

median, dan terendah, sedangkan K-Means++ juga memerlukan tambahan perhitungan jarak ketika memilih *centroid* awal. Ketika data diperbesar 10 dan 100 kali, *runtime* inisialisasi deterministik meningkat sekitar 12 dan 145 kali, mendekati pertumbuhan linear sesuai kompleksitas O($n \cdot K \cdot t$), dengan sedikit tambahan akibat jumlah iterasi yang naik dari 5 menjadi 7. Pada ukuran data terbesar, inisialisasi acak dan K-Means++ membutuhkan waktu sekitar 1,9–2,9 kali lebih lama daripada inisialisasi deterministik karena rata-rata iterasinya mencapai 11,9–13,5. Pada data penelitian, seluruh proses K-Means selesai dalam waktu kurang dari 1 ms, sementara komponen yang paling mahal adalah perhitungan *Silhouette Score* berkompleksitas O(n^2) yang membutuhkan sekitar 2,7 ms pada 242 data dan sekitar 265 ms pada 2.420 data, sedangkan metode Elbow untuk K = 1 sampai K = 6 membutuhkan sekitar 1,7 ms. Dengan waktu tanggap tersebut, sistem web dapat menjalankan ulang seluruh perhitungan setiap kali halaman diakses tanpa jeda yang berarti bagi pengguna.

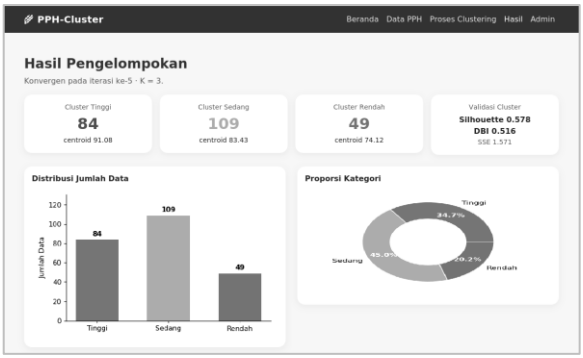
3.7 Implementasi Sistem

Rancangan pada Gambar 2 dan Gambar 3 diimplementasikan menjadi aplikasi web yang halaman utamanya ditunjukkan pada Gambar 5 dan Gambar 6. Halaman proses *clustering* pada Gambar 5 menampilkan seluruh tahapan algoritma secara transparan, mulai dari normalisasi Min-Max, penentuan *centroid* awal, hingga rincian setiap iterasi lengkap dengan posisi *centroid* dan jumlah anggota klaster. Halaman ini menunjukkan bahwa algoritma konvergen pada iterasi kelima dengan jumlah anggota dan posisi *centroid* yang sama dengan perhitungan manual pada Tabel 4.



Gambar 5. Tampilan Halaman Proses Clustering

Halaman hasil pada Gambar 6 menyajikan jumlah data setiap klaster beserta *centroid*-nya, nilai validasi, diagram batang dan diagram lingkaran distribusi klaster, serta daftar anggota setiap klaster. Pengujian fungsional menunjukkan bahwa seluruh fitur, yaitu beranda, data PPH, proses *clustering*, hasil, *login* administrator, dan *dashboard*, berjalan sesuai rancangan. Keluaran sistem identik dengan hasil perhitungan manual, yaitu 84, 109, dan 49 data dengan *Silhouette Score* 0,578, sehingga implementasi algoritma dinyatakan benar.



Gambar 6. Tampilan Halaman Hasil Pengelompokan

Implementasi algoritma pada sisi *server* mengikuti persamaan (1) sampai (7) secara langsung. Fungsi normalisasi membaca seluruh skor dari tabel data_pph, menentukan nilai minimum dan maksimum, lalu menghitung x' untuk setiap

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2729
This work is licensed under a Creative Commons Attribution 4.0 International License.

data. Fungsi iterasi menghitung jarak setiap data ke setiap *centroid*, menyimpan label klaster, dan memperbarui *centroid* sebagai rata-rata anggota, dan perulangan berhenti ketika tidak ada label yang berubah. Fungsi evaluasi kemudian menghitung SSE, *Silhouette Score*, dan *Davies-Bouldin Index* dari label akhir, lalu melakukan denormalisasi *centroid*. Setiap nilai antara, seperti jarak pada setiap iterasi dan rincian pembaruan *centroid*, ditampilkan pada halaman perhitungan manual sehingga pengguna dapat memverifikasi setiap langkah terhadap Tabel 3 dan Tabel 4.

3.8 Pembahasan

Hasil penelitian menunjukkan bahwa K-Means mampu memetakan Skor PPH provinsi ke dalam tiga kategori yang bermakna. Konvergensi yang cepat pada iterasi kelima serta pergeseran *centroid* yang terus mengecil menunjukkan bahwa inisialisasi deterministik dari nilai tertinggi, median, dan terendah memberikan titik awal yang stabil. Temuan ini sejalan dengan tinjauan Ikotun et al. [8] dan Harris dan de Amorim [9] yang menegaskan bahwa kualitas dan kecepatan konvergensi K-Means sangat dipengaruhi oleh pemilihan *centroid* awal. Keberhasilan K-Means pada data numerik ini juga konsisten dengan hasil Ramdhan et al. [12], Prastiwi et al. [13], dan Hendrastuty [14], meskipun penelitian ini menambahkan validasi jumlah klaster dan kualitas klaster yang tidak dilaporkan secara lengkap pada sebagian penelitian tersebut. Analisis sensitivitas pada subbab 3.6 memperkuat temuan tersebut karena inisialisasi deterministik konvergen lebih cepat daripada rata-rata inisialisasi acak dan K-Means++, sementara perbedaan hasilnya hanya terjadi pada dua data di batas klaster. Hal ini sejalan dengan Harris dan de Amorim [9] yang menyimpulkan bahwa strategi inisialisasi terbaik bergantung pada karakteristik data; pada data satu dimensi seperti Skor PPH, *centroid* awal yang tersebar pada nilai ekstrem dan median sudah memadai.

Dari sisi kewilayahan, hasil pemetaan memperlihatkan pola yang jelas. Jawa Tengah, DI Yogyakarta, dan Bali konsisten berada pada kategori Tinggi selama tujuh tahun, sedangkan Nusa Tenggara Timur dan Maluku konsisten berada pada kategori Rendah, dan Gorontalo berada pada kategori Rendah selama enam dari tujuh tahun. Pola ini selaras dengan temuan Cipta et al. [16] yang menempatkan wilayah di Pulau Jawa pada zona skor tinggi dan sebagian besar wilayah Indonesia bagian timur pada zona skor rendah, serta dengan Abdulah [5] yang menemukan kelompok rawan pangan terkonsentrasi di kawasan timur Indonesia. Astria et al. [17] juga menemukan bahwa seluruh provinsi berisiko tinggi ketidakcukupan konsumsi pangan berada di Papua dan Maluku. Kesamaan temuan dari indikator, tingkat wilayah, dan metode yang berbeda tersebut memperkuat kesimpulan bahwa kesenjangan mutu konsumsi pangan antara kawasan barat dan timur Indonesia bersifat struktural dan tidak hanya merupakan artefak metode analisis.

Dari sisi waktu, pergeseran sebagian besar provinsi dari kategori Sedang ke Tinggi, terutama sejak tahun 2022, menggambarkan perbaikan mutu konsumsi pangan nasional. Hal ini sejalan dengan kenaikan rata-rata Skor PPH dari 82,16 pada tahun 2018 menjadi 86,29 pada tahun 2024, meskipun sempat mencapai 87,00 pada tahun 2023. Perbaikan ini terutama dinikmati oleh provinsi di Sumatera dan Kalimantan yang sebelumnya berada pada kategori Sedang. Sebaliknya, provinsi yang menetap pada kategori Rendah menunjukkan bahwa perbaikan tersebut belum merata. Temuan Rossita et al. [3] mengenai perlunya diversifikasi sumber karbohidrat serta peningkatan konsumsi sayur, buah, dan protein nabati memberikan arah intervensi yang relevan, khususnya bagi provinsi yang konsisten berkategori Rendah dan bagi Papua Barat yang mengalami penurunan kategori. Provinsi-provinsi ini perlu diprioritaskan dalam program diversifikasi pangan berbasis sumber daya lokal sebagaimana diamanatkan dalam kebijakan ketahanan pangan nasional [1].

Dari sisi kualitas model, nilai *Silhouette Score* 0,578 dan Davies-Bouldin Index 0,516 menunjukkan klaster yang padat dan terpisah dengan baik, sedangkan metode Elbow mengonfirmasi K = 3 sebagai jumlah klaster yang optimal sesuai rekomendasi evaluasi multimetrik oleh Sholeh dan Aeni [15]. Nilai *Silhouette* ini lebih tinggi daripada nilai sekitar 0,40 yang dilaporkan Cipta et al. [16]. Perbandingan tersebut perlu dibaca dengan hati-hati karena penelitian ini hanya menggunakan satu atribut, yaitu skor total PPH, sedangkan data satu dimensi cenderung lebih mudah dipisahkan dibandingkan data multidimensi. Temuan Buleuw et al. [19] bahwa K-Means cenderung unggul pada *Silhouette Score* tetapi tidak selalu pada ukuran validitas lain juga menjadi pengingat bahwa satu metrik tidak cukup untuk menilai kualitas klaster. Oleh karena itu, penelitian ini menggunakan tiga ukuran secara bersamaan.

Keterbatasan utama penelitian ini adalah penggunaan skor total PPH sebagai satu-satunya variabel, sehingga hasil pemetaan menjelaskan seberapa tinggi capaian suatu provinsi tetapi belum menjelaskan kelompok pangan apa yang menyebabkan suatu provinsi tertinggal. Selain itu, meskipun inisialisasi deterministik telah dibandingkan dengan inisialisasi acak dan K-Means++, penelitian ini belum menguji varian K-Means yang menentukan jumlah klaster dan mendeteksi pencilan secara otomatis [10], dan pengukuran *runtime* dilakukan pada satu lingkungan perangkat keras sehingga nilai absolutnya dapat berbeda pada lingkungan lain. Meskipun demikian, implementasi sistem berbasis web memberikan nilai tambah berupa transparansi proses, karena setiap tahapan perhitungan dapat ditelusuri oleh pengguna dan keluarannya terbukti sama dengan perhitungan manual. Sistem ini dapat dimanfaatkan oleh dinas ketahanan pangan dan pemangku kepentingan lain sebagai alat bantu analisis yang mudah digunakan tanpa memerlukan perangkat lunak statistik khusus, serta dapat diperbarui setiap kali data PPH tahun terbaru tersedia.

4. KESIMPULAN

Penelitian ini menjawab keempat tujuan yang telah dirumuskan. Pertama, algoritma K-Means Clustering berhasil memetakan 242 data Skor Pola Pangan Harapan provinsi tahun 2018–2024 ke dalam tiga klaster setelah konvergen pada iterasi kelima, yaitu klaster Tinggi sebanyak 84 data dengan *centroid* 91,08, klaster Sedang sebanyak 109 data dengan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2730
This work is licensed under a Creative Commons Attribution 4.0 International License.

centroid 83,43, dan klaster Rendah sebanyak 49 data dengan *centroid* 74,12, dengan batas antarklaster pada skor 87,26 dan 78,77; jumlah klaster tersebut didukung metode Elbow dan kualitasnya tergolong baik dengan *Silhouette Score* 0,578 dan *Davies-Bouldin Index* 0,516. Kedua, secara kewilayahan Jawa Tengah, DI Yogyakarta, dan Bali konsisten berkategori Tinggi, sedangkan Nusa Tenggara Timur dan Maluku konsisten berkategori Rendah; secara antarwaktu jumlah provinsi berkategori Tinggi meningkat dari 5 provinsi pada tahun 2018 menjadi 22 provinsi pada tahun 2024, 20 dari 34 provinsi mengalami kenaikan kategori, dan hanya Papua Barat yang mengalami penurunan, sehingga perbaikan mutu konsumsi pangan nasional belum merata. Ketiga, hasil pengelompokan stabil terhadap strategi inisialisasi karena 240 dari 242 data selalu menempati klaster yang sama pada 200 *run* inisialisasi acak dan K-Means++, sedangkan inisialisasi deterministik konvergen dalam 5 iterasi dengan *runtime* median 0,16 ms atau sekitar 1,8–1,9 kali lebih cepat daripada kedua strategi pembanding. Keempat, sistem berbasis web yang dibangun dengan PHP dan MySQL menampilkan seluruh tahapan perhitungan secara transparan dan menghasilkan keluaran yang identik dengan perhitungan manual, sehingga dapat digunakan sebagai alat bantu penentuan prioritas program diversifikasi pangan, khususnya bagi provinsi yang konsisten berkategori Rendah. Kontribusi penelitian ini terletak pada tiga hal, yaitu prosedur K-Means dengan inisialisasi deterministik yang dapat direproduksi dan teruji kestabilannya, peta kategori Skor PPH seluruh provinsi selama tujuh tahun beserta batas skor antarkategori yang eksplisit, serta sistem berbasis web yang transparan dan dapat diperbarui sebagai alat bantu penentuan prioritas program diversifikasi pangan. Penelitian selanjutnya disarankan untuk menambahkan proporsi konsumsi sembilan kelompok pangan sebagai variabel agar penyebab ketertinggalan suatu provinsi dapat diidentifikasi, membandingkan K-Means dengan K-Medoids, DBSCAN, atau *hierarchical clustering*, memperluas data hingga tingkat kabupaten/kota, serta melengkapi sistem dengan fitur impor data, ekspor laporan, dan peta sebaran klaster.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: A. F., dan W.;
Metodologi: A. F., dan W.;
Perangkat Lunak: A. F.;
Validasi: W.;
Analisis Formal: Afriza Fliyani;
Investigasi: A. F.;
Sumber Daya: A. F.;
Kurasi Data: A. F.;
Penulisan Draf Awal: A. F.;
Penulisan, Peninjauan, dan Penyuntingan: A. F., dan W.;
Visualisasi: A. F.;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data Skor Pola Pangan Harapan yang digunakan dalam penelitian ini bersumber dari portal Satu Data Indonesia (data.go.id) dan dapat diakses secara terbuka. Data olahan dan hasil pengelompokan tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan dari pihak eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (*Not applicable*).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (*Not applicable*).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk membantu menyusun dan meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

REFERENCES

[1] Pemerintah Republik Indonesia, "Peraturan Pemerintah Nomor 17 Tahun 2015 tentang Ketahanan Pangan dan Gizi," Sekretariat Negara Republik Indonesia, Jakarta, 2015.
[2] Badan Pangan Nasional, "Peraturan Badan Pangan Nasional Republik Indonesia Nomor 11 Tahun 2023 tentang Pola Pangan Harapan," Badan Pangan Nasional, Jakarta, 2023.
[3] A. Rossita, T. Masui, K. Tsuchiya, and M. Sekiyama, "Assessing Food Consumption Trends in Indonesia: Historical Insights and Challenges for Adoption of the Desirable Dietary Pattern," *Jurnal Gizi dan Pangan*, vol. 21, no. 2, pp. 79–88, 2026, doi: 10.25182/jgp.2026.21.2.79-88.
[4] M. F. Reicilya, M. Mukson, and H. Setiyawan, "Analisis Faktor-Faktor yang Memengaruhi Pola Pangan Harapan pada Rumah Tangga di Kelurahan Ngijo Kota Semarang," *Mimbar Agribisnis: Jurnal Pemikiran Masyarakat Ilmiah Berwawasan Agribisnis*, vol. 10, no. 2, pp. 1652–1661, 2024, doi: 10.25157/ma.v10i2.13385.
[5] R. Abdulah, "Mapping Food Security in Indonesia: Geographic Clusters and Regional Disparities," *Indonesian Journal of Geography*, vol. 57, no. 3, pp. 440–447, 2025, doi: 10.22146/ijg.99419.
[6] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Siswa Baru," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 6, p. 1042, 2023, doi: 10.53513/jursi.v2i6.9148.
[7] A. E. Ezugwu *et al.*, "A Comprehensive Survey of Clustering Algorithms: State-of-the-Art Machine Learning Applications, Taxonomy, Challenges, and Future Research Prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104743, 2022, doi: 10.1016/j.engappai.2022.104743.
[8] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data," *Information Sciences*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.
[9] S. Harris and R. C. de Amorim, "An Extensive Empirical Comparison of K-Means Initialisation Algorithms," *IEEE Access*, vol. 10, pp. 58752–58768, 2022, doi: 10.1109/ACCESS.2022.3179803.
[10] P. Olukanmi, F. Nelwamondo, T. Marwala, and B. Twala, "Automatic Detection of Outliers and the Number of Clusters in K-Means Clustering via Chebyshev-Type Inequalities," *Neural Computing and Applications*, vol. 34, no. 8, pp. 5939–5958, 2022, doi: 10.1007/s00521-021-06689-x.
[11] J. Hämäläinen, T. Kärkkäinen, and T. Rossi, "Improving Scalable K-Means++," *Algorithms*, vol. 14, no. 1, p. 6, 2021, doi: 10.3390/a14010006.
[12] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, and K. Kaslani, "Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means," *Media Informasi Analisa dan Sistem (MEANS)*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.54367/means.v7i1.1826.
[13] H. Prastiwi, J. Pricilia, and E. Rasywir, "Implementasi Data Mining untuk Menentukan Persediaan Stok Barang di Mini Market Menggunakan Metode K-Means Clustering," *Jurnal Informatika dan Rekayasa Komputer (JAKAKOM)*, vol. 2, no. 1, pp. 141–148, 2022, doi: 10.33998/jakakom.2022.2.1.34.
[14] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, 2024, doi: 10.58602/jima-ilkom.v3i1.26.
[15] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56–65, 2023, doi: 10.30998/string.v8i1.16388.
[16] E. A. Cipta, B. Priyatna, F. Nurapriani, and A. L. Hananto, "Pola Konsumsi Pangan Skor PPH Wilayah di Indonesia Berdasarkan Gizi Tiap Daerah 2021–2024 Penerapan Metode Hierarchical Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 10, no. 3, pp. 4894–4900, May 2026, doi: 10.36040/jati.v10i3.18427.
[17] L. Astria, S. Meliani, A. Pratama, and A. Rohman, "Cluster Analysis of Provinces Based on the Prevalence of Undernourishment Using the K-Means Algorithm," *JOSAR (Journal of Students Academic Research)*, vol. 11, no. 2, pp. 334–346, 2026, doi: 10.35457/9wjmrq51.
[18] I. Nilasari and N. E. Safitri, "Provincial Classification of Child Malnutrition in Indonesia: A K-Means Clustering Approach for Evidence-Based Stunting Intervention Prioritization," *Journal of Computing and Smart Ecosystems*, vol. 2, no. 1, 2026, doi: 10.26714/jcase.v2i1.1144.
[19] Z. R. Buleuw, R. J. Situmeang, and D. J. Fitri, "Perbandingan Metode K-Means dan Ward pada Pengelompokan Ketahanan Pangan Kabupaten/Kota di Indonesia," *Jurnal Algoritme*, vol. 7, no. 1, pp. 133–144, 2026, doi: 10.35957/algoritme.v7i1.16194.
[20] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, p. 31, 2021, doi: 10.1186/s13638-021-01910-w.
[21] F. Batool and C. Hennig, "Clustering with the Average Silhouette Width," *Computational Statistics & Data Analysis*, vol. 158, p. 107190, 2021, doi: 10.1016/j.csda.2021.107190.
[22] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy*, vol. 23, no. 6, p. 759, 2021, doi: 10.3390/e23060759.
[23] A. M. Ikotun, F. Habyarimana, and A. E. Ezugwu, "Cluster Validity Indices for Automatic Clustering: A Comprehensive Review," *Heliyon*, vol. 11, no. 2, p. e41953, 2025, doi: 10.1016/j.heliyon.2025.e41953.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2732
This work is licensed under a Creative Commons Attribution 4.0 International License.