JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

Mapping Regional Clusters Based on Total Rescued Food in Indonesia Using K-Means Clustering

Muhammad Rafi Haidar Arsyad[*1], Wafi Arifin[2], dan Mohammad Riza Radyanto[3]
[1)] Data Science Study Program, Faculty Of Science And Technology, Universitas Sains Teknologi Ekonomi Digital Indonesia, Semarang, Indonesia
[2)] Informatics Study Program, Faculty Of Science And Technology, Universitas Sains Teknologi Ekonomi Digital Indonesia, Semarang, Indonesia
[3)] Industrial Engineering Study Program, Faculty Of Science And Technology, Universitas Sains Teknologi Ekonomi Digital Indonesia, Semarang, Indonesia
muhrafihaidar27@gmail.com [1]; wafiarifin0609@gmail.com [2], rizaradyanto@gmail.com[3]*
[*] *Corresponding author: muhrafihaidar27@gmail.com*

Abstract
This study maps regional clusters in Indonesia based on the total amount of rescued food using the K-Means algorithm. The data were obtained from the official dataset Jumlah Total Pangan yang Terselamatkan Update Bulan Januari 2026 published by the National Food Agency through data.go.id. The research followed an IMRAD-oriented workflow comprising data cleaning, data parsing, construction of regional aggregate features, determination of the optimal number of clusters using the elbow and silhouette methods, K-Means clustering, and visualization of cluster results. The analysis identified two optimal clusters, supported by a silhouette score of 0.809828 and a Davies-Bouldin Index of 0.113459. The findings reveal a clear disparity between regions with low and high rescued-food achievements, with West Java emerging as the most dominant region. This study contributes by utilizing an official rescued-food dataset and regional aggregate features as an analytical basis for monitoring, evaluating, and strengthening data-driven food rescue programs in Indonesia.
Keywords : *K-Means, Clustering, Rescued Food, Machine Learning, Aggregate Features*

Abstrak
Penelitian ini bertujuan memetakan klaster wilayah di Indonesia berdasarkan total pangan terselamatkan menggunakan algoritma K-Means. Data penelitian bersumber dari dataset resmi Jumlah Total Pangan yang Terselamatkan Update Bulan Januari 2026 yang dipublikasikan Badan Pangan Nasional melalui data.go.id. Tahapan penelitian disusun secara sistematis melalui pembersihan data, parsing, pembentukan fitur agregat wilayah, penentuan jumlah klaster optimal menggunakan metode elbow dan silhouette, pemodelan K-Means, serta visualisasi hasil klasterisasi. Hasil penelitian menunjukkan bahwa jumlah klaster optimal adalah dua klaster, dengan nilai silhouette score sebesar 0,809828 dan Davies-Bouldin Index sebesar 0,113459. Temuan ini menunjukkan adanya disparitas yang jelas antara wilayah dengan capaian pangan terselamatkan rendah dan tinggi, dengan Jawa Barat sebagai wilayah paling dominan. Kontribusi penelitian ini terletak pada pemanfaatan dataset resmi pangan terselamatkan dan fitur agregat

wilayah sebagai dasar analisis untuk mendukung monitoring, evaluasi, dan penguatan program penyelamatan pangan berbasis data di Indonesia.
Kata Kunci : K-Means, klasterisasi, ketahanan pangan, Pembelajaran Mesin, fitur agregat

1. INTRODUCTION

The management of surplus food has become an increasingly important issue in developing a sustainable food system in Indonesia [1]. In this context, the National Food Agency provides an open dataset entitled Jumlah Total Pangan yang Terselamatkan Update Bulan Januari 2026 through the Satu Data Indonesia portal [2]. The dataset defines rescued food as the total amount of food received from donors by food banks or food rescue actors [3]. A higher amount of received food indicates a greater potential for surplus food to be rescued and redistributed as edible food [4]. The availability of this official dataset is important because it provides an empirical basis for assessing food rescue achievements across regions in a more measurable and accountable manner [5].

From a policy perspective, this topic has become increasingly relevant because the Food Rescue Movement initiated by the National Food Agency in 2022 has expanded to 17 provinces [6]. The program has rescued more than 224 tons of food and distributed it to more than 456,000 beneficiaries [7]. This achievement shows that food rescue can no longer be viewed merely as a partial social activity [8]. Instead, it has become part of national food governance that requires support from data-driven analysis [9]. However, the available monthly data do not directly present regional typologies [10]. Therefore, policymakers need an analytical approach that can summarize similarities among regions into more interpretable groups [11].

The K-Means algorithm has been widely used to group regions based on food characteristics and socio-economic indicators [12]. Previous studies have shown that K-Means can classify provinces based on food crop production and map districts or cities based on indicators related to inflation formation [13]. These findings indicate that K-Means is relevant for simplifying diverse numerical data into several relatively homogeneous groups [14]. However, the use of official rescued food data as a basis for regional clustering remains limited. This condition creates a research gap that can be further developed [15].

This study proposes regional cluster mapping based on the total amount of rescued food in Indonesia using the K-Means approach. The novelty of this study lies in three aspects [16]. First, it utilizes an official dataset from the National Food Agency that specifically records rescued food indicators, rather than general indicators such as food production, inflation, or food security [17]. Second, this study does not rely solely on raw values but constructs aggregate regional features to represent the achievement level, intensity, and consistency of food rescue activities [18]. Third, the clustering results are directed toward producing a more operational regional typology as a basis for monitoring, evaluation, and priority setting in program interventions [19]. Therefore, this study is expected to contribute not only to the methodological development of cluster analysis using public data but also to strengthening decision-making in food rescue governance in Indonesia [20].

2. LITERATURE REVIEW

The data used in this study were obtained from the official dataset entitled “Jumlah Total Pangan yang Terselamatkan Update Bulan Januari 2026”, published through the Satu Data Indonesia portal and sourced from the National Food Agency. The dataset metadata explains that this indicator measures the total amount of food received from donors by food banks or food rescue practitioners before the selection and quality control process. A higher amount of food received indicates a greater potential for surplus food to be rescued and redistributed as consumable food. Therefore, this dataset is relevant for regional analysis because it provides information on rescued food achievements that can be compared across different regions in Indonesia.

Table 1.Variable Information

Variable Name	Description
Year	The year in which the rescued food data were observed.
Month	The month in which the data were observed.
Region_Code	The identification code of each region.
Region	The name of the region or province.
CPPD_Ton	The total amount of rescued food measured in tons.

Source : The dataset was obtained from the National Food Agency through the data.go.id portal.

This study applies the K-Means Clustering method to group regions based on similarities in rescued food characteristics. K-Means is an unsupervised clustering algorithm that partitions data into several groups by measuring the proximity of each data point to the cluster centroid. The algorithm aims to minimize inertia, also known as the within-cluster sum of squares, so that objects within the same cluster have relatively similar characteristics. In its implementation, the number of clusters must be determined in advance. Therefore, this study uses the elbow method and silhouette score to identify the most appropriate number of clusters. The silhouette score is used to evaluate the separation quality between clusters, where a value closer to 1 indicates a better clustering structure.

The research procedure was carried out systematically. The process began with importing the required libraries, uploading the dataset, cleaning and parsing the data, and checking data quality to ensure that the dataset was ready for analysis. The next stages involved constructing aggregate features for each region, selecting clustering features, determining the optimal number of clusters using the elbow method and silhouette score, and developing the K-Means clustering model. The resulting clusters were then labeled based on annual levels and visualized using two-dimensional PCA, monthly cluster profiles, and annual national monthly trends of total rescued food. These stages were designed to ensure that the clustering results were not only computationally accurate but also interpretable in supporting regional pattern analysis, as shown in Figure 1.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

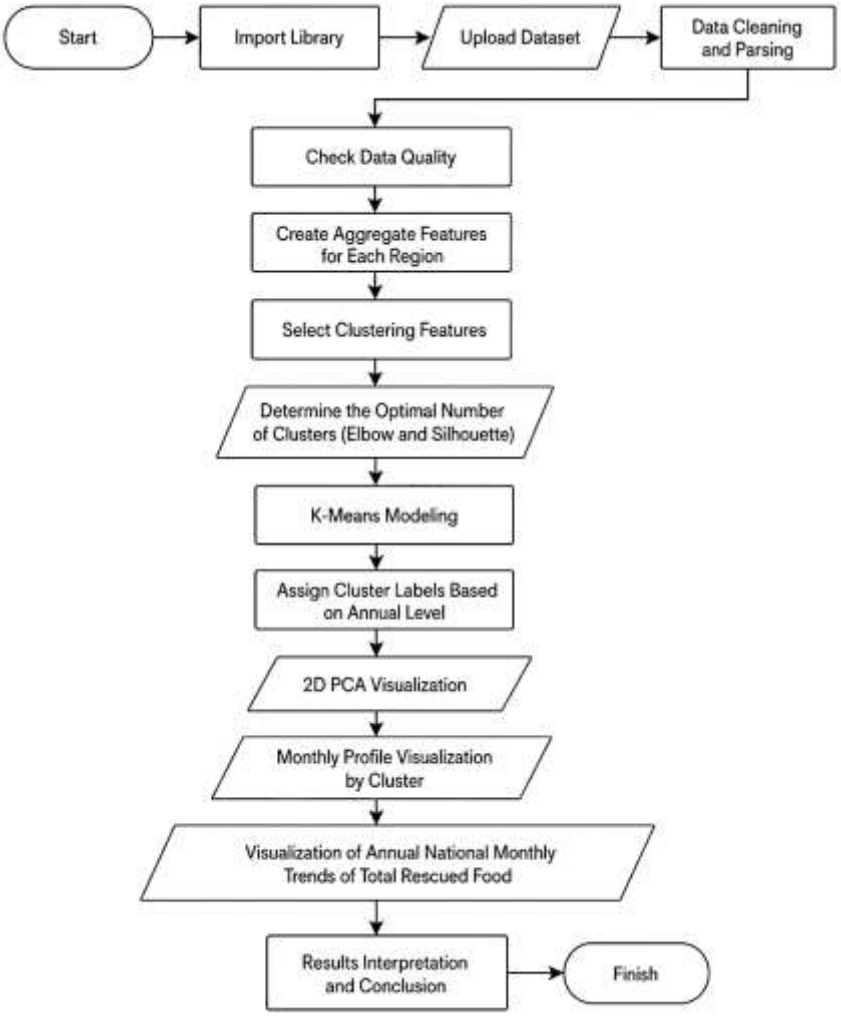

Figure 1. Research Flowchart.

Figure 1 illustrates the research stages, beginning with data preparation, feature construction, and the clustering process using K-Means, followed by visualization and interpretation of the results. This workflow was designed to ensure that the regional grouping based on the total amount of rescued food could be conducted systematically and produce information that is clear, structured, and easy to understand.

3. METHOD

This study employed an unsupervised machine learning approach, which is an analytical method used to identify similarity patterns in data without predefined class labels. The algorithm applied in this study was K-Means Clustering, as it is suitable for grouping regions based on the characteristics of total rescued food. The research data were obtained from the dataset Jumlah Total Pangan yang Terselamatkan Update Bulan Januari 2026, published by the National Food Agency through the data.go.id portal. All data processing procedures were conducted using Google Colab.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

3.1. K-Means Algorithm

K-Means is a clustering algorithm that partitions data into several groups based on the similarity of object characteristics. The main principle of this algorithm is to assign each data point to the nearest cluster center and then update the cluster centers iteratively until a stable clustering result is achieved. In this study, K-Means was used to group regions in Indonesia based on the pattern of total rescued food. This process produced regional clusters with low, moderate, or high rescued-food characteristics.

4. RESULTS AND DISCUSSION

In the machine learning analysis, the data cleaning and parsing process produced a consistent final dataset consisting of 456 rows and 8 columns. The dataset contained standardized variables, including year, date, region code, region name, CPPD_Ton value, month number, and month name. This process confirms that the initial data were successfully transformed into a more structured analytical format. As a result, each monthly observation for every region could be clearly identified and prepared for the next stages, including aggregate feature construction and clustering analysis, as shown in Figure 1.

Shape setelah cleaning: (456, 8)

	Tahun	Bulan	Kode_Wilayah	Wilayah	CPPD_Ton	Tanggal	Bulan_Num	Bulan_Nama
0	2025	31 Januari 2025	11	Aceh	1911	2025-01-31	1	January
1	2025	28 Februari 2025	11	Aceh	1871	2025-02-28	2	February
2	2025	31 Maret 2025	11	Aceh	1871	2025-03-31	3	March
3	2025	30 April 2025	11	Aceh	1871	2025-04-30	4	April
4	2025	31 Mei 2025	11	Aceh	1871	2025-05-31	5	May
5	2025	30 Juni 2025	11	Aceh	1871	2025-06-30	6	June
6	2025	31 Juli 2025	11	Aceh	1871	2025-07-31	7	July
7	2025	31 Agustus 2025	11	Aceh	1871	2025-08-31	8	August
8	2025	30 September 2025	11	Aceh	1871	2025-09-30	9	September
9	2025	31 Oktober 2025	11	Aceh	1871	2025-10-31	10	October
10	2025	30 November 2025	11	Aceh	1871	2025-11-30	11	November
11	2025	31 Desember 2025	11	Aceh	601	2025-12-31	12	December

Figure 1 presents the results of the data cleaning and parsing process for the total rescued food dataset.

After the data cleaning and parsing stages, the process continued with the construction of regional aggregate features, resulting in 38 rows and 18 columns. Each row represents one region with more informative annual statistical summaries. Features such as annual total, monthly average, monthly median, maximum value, minimum value, monthly standard deviation, and the number of recorded months were used to capture the achievement level, stability, and variation of rescued food in each region more comprehensively. Through this aggregate feature construction, the data, which were initially structured as monthly observations, were successfully transformed into a more concise and analytical numerical representation. This transformation makes the dataset more suitable for the feature selection and clustering modeling stages, as shown in Figure 2.

Shape fitur wilayah: (38, 18)

	Kode_Wilayah	Wilayah	total_tahunan	rata_rata_bulanan	median_bulanan	maksimum_bulanan	minimum_bulanan	std_bulanan	jumlah_bulan_tercatat
0	11	Aceh	21222	1768.500000	1871.0	1911	601	367.846930	12
1	12	Sumatera Utara	64366	5363.833333	5616.0	5727	2459	915.321785	12
2	13	Sumatera Barat	152426	12702.166667	12855.0	14186	9942	1432.347205	12
3	14	Riau	10212	851.000000	851.0	851	851	0.000000	12
4	15	Jambi	76368	6364.000000	6614.0	6614	3614	866.025404	12

Figure 2. Construction of Aggregate Features for Each Region

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

After the regional aggregate features were constructed, the next stage was to determine the optimal number of clusters by comparing several values of k using inertia, silhouette score, Davies-Bouldin Index, and Calinski-Harabasz Index. The evaluation results show that k = 2 is the most optimal number of clusters, as it produces the highest silhouette score of 0.809828 and the lowest Davies-Bouldin Index of 0.113459. These values indicate that the resulting cluster structure has strong internal compactness and excellent separation between clusters, as shown in Figure 3.

	k	inertia	silhouette_score	davies_bouldin_index	calinski_harabasz_index
0	2	190.557892	0.809828	0.113459	42.968128
1	3	122.661514	0.581760	0.832121	42.135657
2	4	82.172777	0.612537	0.606235	46.317552
3	5	53.942565	0.623056	0.320196	55.679106
4	6	25.733453	0.639176	0.341921	97.558067

```
Jumlah klaster terbaik berdasarkan silhouette score: 2
```

Figure 3. Determination of the Optimal Number of Clusters

In the next stage, the silhouette score visualization confirms that the highest value is obtained at k = 2, with a score of approximately 0.81. This result indicates that the two formed clusters have a high level of internal cohesion and are well separated from each other, as presented in Figure 4.

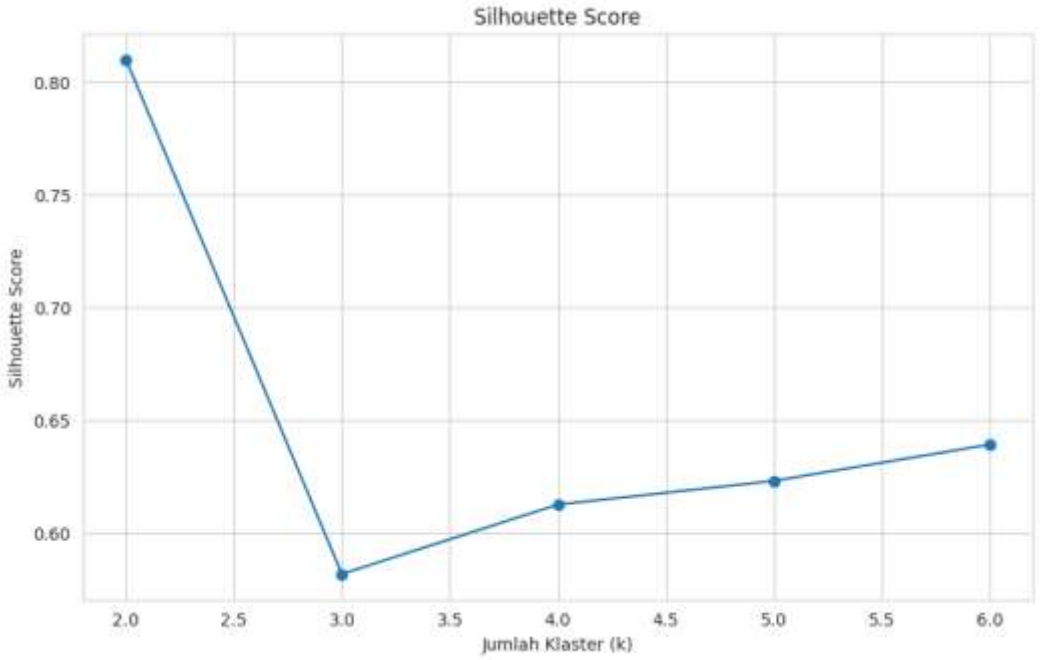

Figure 4. Cluster Visualization Based on Silhouette Score

The visualization of the average monthly CPPD_Ton for each cluster shows a clear difference between the low and high clusters. The high cluster consistently records values far above the low cluster across all months. However, the high cluster experiences a sharp decline in the second month before increasing again and remaining relatively stable at a high level until the end of the year. In contrast, the low cluster remains at a much lower level with limited fluctuation, as shown in Figure 5.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

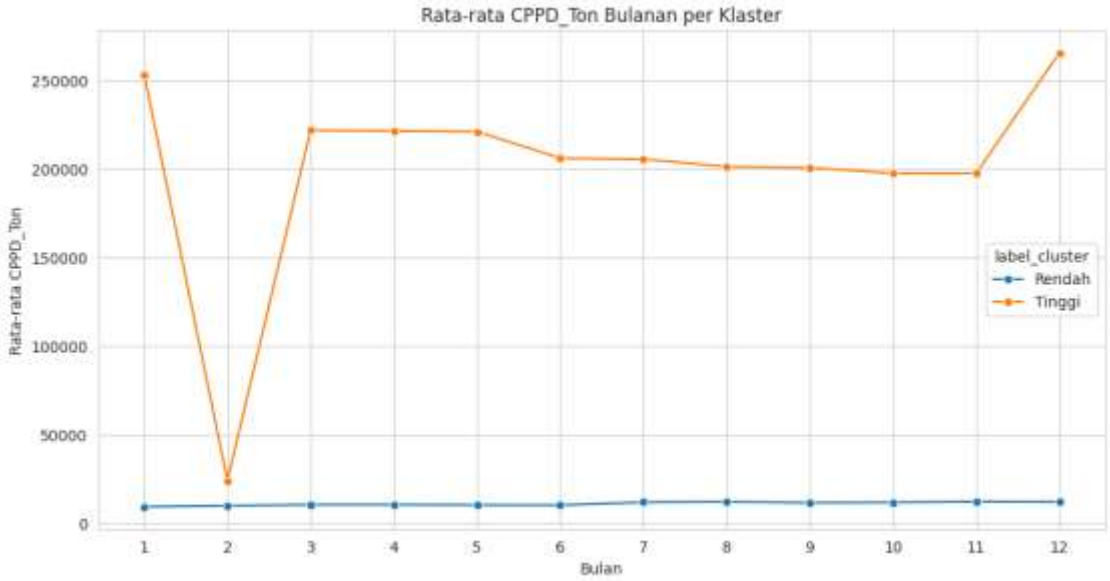

Figure 5. Visualization of Average Monthly CPPD_Ton by Cluster

The visualization of the national monthly trend of rescued food shows that the monthly achievement during the observation year fluctuates. The sharpest decline occurs in the second month, followed by an increase in the subsequent months. After this decline, the total amount of rescued food tends to move at a higher and relatively stable level from the middle of the year. The highest value is reached in the twelfth month, indicating stronger performance at the end of the observation period, as presented in Figure 6.

Figure 6. Monthly Trend of National Rescued Food Total

The regional cluster visualization using PCA shows that dimensionality reduction is able to separate the data structure into two relatively clear groups, namely the low cluster and the high cluster. Most regions are concentrated in the low cluster and are positioned relatively close to one another, indicating that regions in this group have more homogeneous characteristics. Meanwhile, West Java is positioned far apart along the main component axis and forms a separate high cluster, as shown in Figure 7.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

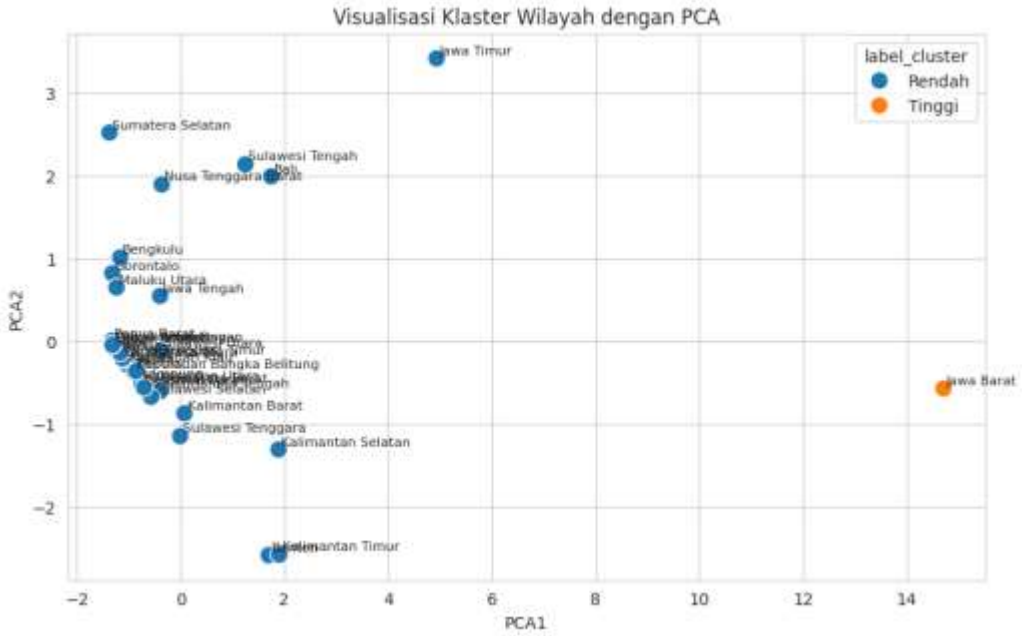

Figure 7. Regional Cluster Visualization Using PCA

The bar chart visualization of the top ten regions with the highest total rescued food shows that West Java occupies the most dominant position, with an achievement far exceeding other regions. This result confirms West Java's role as the main contributor to national food rescue performance. After West Java, regions such as East Java, East Kalimantan, Banten, and South Kalimantan occupy the next positions, although their values remain considerably lower than the highest-ranked region, as presented in Figure 8.

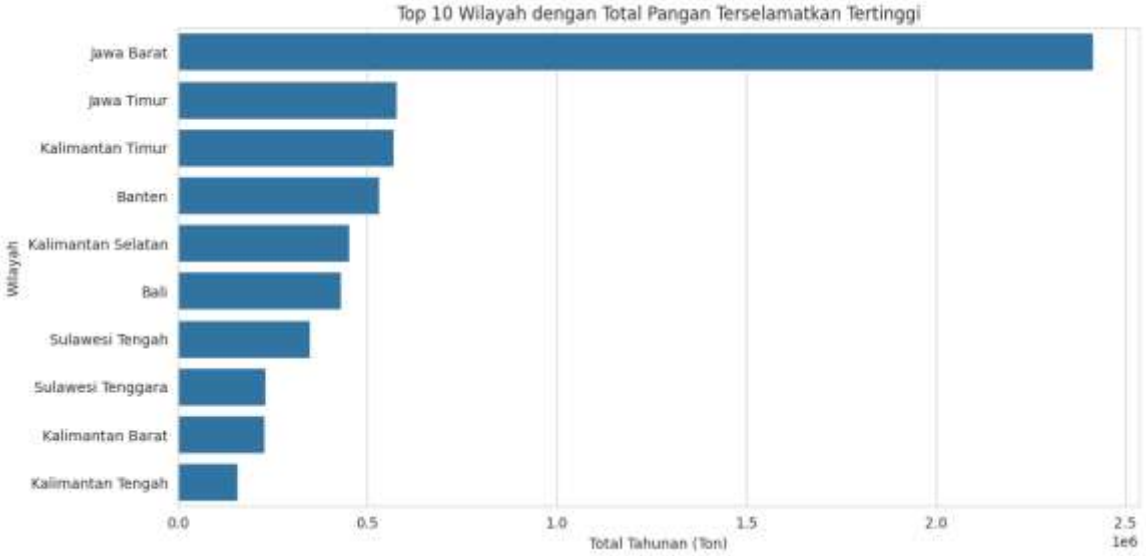

Figure 8. Top Ten Regions with the Highest Total Rescued Food

The radar chart visualization of cluster characteristics shows that the high cluster has the most dominant values in almost all main indicators, particularly annual total, monthly average, monthly standard deviation, linear trend, and national share percentage. This pattern reflects regions with large contributions, high intensity, and a dominant role in national rescued food achievement, as shown in Figure 9.

JOISTICS: Journal of Information Systems and Data Analytics
Copyright ©2025, JOISTICS :, p-ISSN: ……., e-ISSN: 3123-6715
On-line: https://journal.uyr.ac.id/index.php/JOISTICS

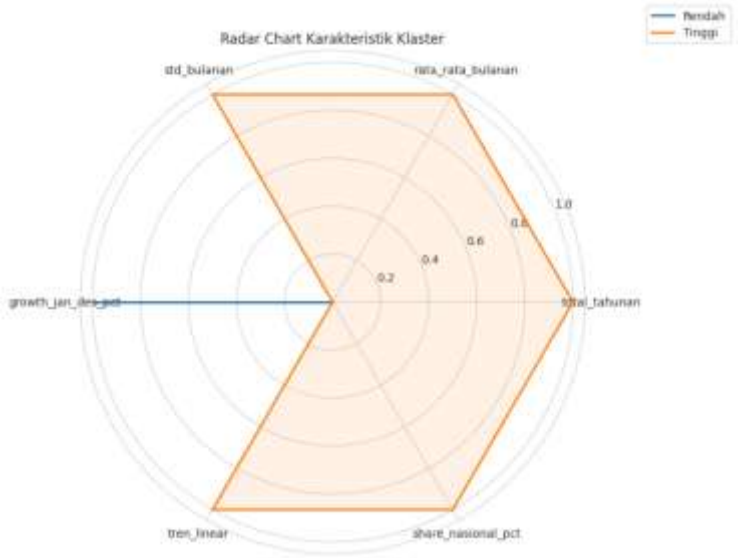

Figure 9. Radar Chart Visualization of Cluster Characteristics

5. CONCLUSION

This study demonstrates that the K-Means algorithm is effective for mapping regional clusters in Indonesia based on the total amount of rescued food. The optimal number of clusters was identified as k = 2, based on the evaluation results of the elbow method, silhouette score, Davies-Bouldin Index, and Calinski-Harabasz Index. The clustering results indicate two main groups: a low cluster and a high cluster. Most regions belong to the low cluster, while West Java emerges as the most dominant region and is clearly separated as the representative of the high cluster. These findings are supported by PCA visualization, monthly cluster profiles, national monthly trends, and a bar chart of the highest-performing regions, all of which consistently reveal disparities in rescued food achievements across regions. Therefore, this study provides a strong analytical contribution to supporting data-driven monitoring, evaluation, and prioritization of intervention programs for food rescue initiatives at the regional level.

ACKNOWLEDGEMENTS

The authors would like to thank the National Food Agency of Indonesia and the Satu Data Indonesia portal (data.go.id) for providing the official dataset used in this study. The authors also express their appreciation to Universitas Sains Technology Ekonomi Digital Indonesia for its academic support in the preparation of this article.

REFERENCES

[1] "DATA PANGAN TERSELAMATKAN BADAN PANGAN NASIONAL 2026".

[2] A. Idris and M. Hizbul, "Ketahanan Pangan, Air, Energi Dan Pertanian: Analisis Sebaran Spasial Dan Clustering Provinsi Di Indonesia," vol. 4, no. 5, 2023, doi: 10.55314/tsg.v4i5.597.

[3] N. A. HIDAYATULLAH, W. Prihartono, and F. Rohman, "CLUSTERING PENERIMA BANTUAN PANGAN BERBASIS ALGORITMA K-MEANS UNTUK MENINGKATKAN EFEKTIVITAS PROGRAM SOSIAL DI KOTA/KABUPATEN CIREBON," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5692.

[4] R. F. Sinaga, M. A. Prabukusumo, and J. Manurung, "Journal of Intelligent Decision Support System (IDSS) Comparison of k-means clustering with hierarchical agglomerative clustering for the analysis of food security of rice sector in Indonesia," 2025.

[5] Dini Andita Putri, Fitri Mudia Sari, and Chairini Wirdiastuti, "Application of K-Means Clustering for Grouping Plantation Production in West Pasaman Regency in 2024," *UNP Journal of Statistics and Data Science*, vol. 3, no. 4, pp. 482–487, Nov. 2025, doi: 10.24036/ujsds/vol3-iss4/426.

[6] Jafar Pahrudin and Sri Mulyeni, "Implementasi Algoritma K-Means Clustering dengan Python untuk Analisis Produksi Bawang Merah di Indonesia," *SOSIAL: Jurnal Ilmiah Pendidikan IPS*, vol. 3, no. 4, pp. 01–15, Oct. 2025, doi: 10.62383/sosial.v3i4.1228.

[7] C. A. Sinaga and P. M. Hasugian, "Comparison and Evaluation of Euclidean and Canberra Distances in the Adaptive K-Means Algorithm for Classifying the Food Security Status of Indonesian Provinces under a Creative Commons Attribution-NonCommercial 4.0 International License (CCBY-NC 4.0)," 2025, [Online]. Available: https://jurnal.seaninstitute.or.id/index.php/jukomi

[8] C. Syalom, U. Khaira, and D. Arsa, "CLUSTERING DAERAH RAWAN STUNTING DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS," *Jurnal Teknologi Informasi*, vol. 6, no. 1, 2025, doi: 10.46576/djtechno.

[9] T. A. Toto, "PERBANDINGAN ALGORITMA K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK PENGELOMPOKAN DAERAH PENGHASIL PADI DI INDONESIA," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.7251.

[10] S. Julianti, M. Widiartha, J. Raya, K. Unud, B. Jimbaran, and K. Selatan, "Algoritma K-Means untuk Clustering Provinsi di Indonesia Berdasarkan Kasus Stunting," *JNATIA*, vol. 2, no. 3, 2024.

[11] N. Ferawati Purba, A. Rumapea, and A. P. Silalahi, "Analisis Sentimen Dompet elektronik Pada Twitter Menggunakan Metode K-Means," 2024.

[12] S. Sugiono and R. D. Nugraheni, "Cashless Society: Cluster Analysis of Electronic Payment Users in E-Commerce," *Jejak*, vol. 16, no. 2, pp. 286–301, 2023, doi: 10.15294/jejak.v16i2.42451.

[13] N. Ferawati Purba, A. Rumapea, and A. P. Silalahi, "Analisis Sentimen Dompet elektronik Pada Twitter Menggunakan Metode K-Means," 2024.

[14] Y. Azarya and I. Budi, "Analisis Sentimen Berbasis Aspek Aplikasi Brimo berdasarkan Ulasan Pengguna di Google Playstore," *The Indonesian Journal of Computer Science*, vol. 14, no. 1, Feb. 2025, doi: 10.33022/ijcs.v14i1.4613.

[15] R. Bagus Trianto, A. Susilo Nugroho, E. Supriyadi, U. An Nuur Jl Gajah Mada No, K. Purwodadi, and K. Grobogan, "Klasterisasi Menggunakan Algoritma K-Means dan Elbow pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022," vol. 8, no. 1, p. 2023.

[16] "Analisis Sentimen Dan Pemodelan Topik Terhadap Aplikasi Pembelajaran Online Pada Platform Google Play (Studi Kasus: Quipper)," *Jurnal Ilmiah Komputasi*, vol. 23, no. 4, Dec. 2024, doi: 10.32409/jikstik.23.4.3663.

[17] N. A. Y. -, L. E. B. -, G. C. H. R. -, M. F. Z. -, A. -, and F. R. -, "ANALISIS PERBANDINGAN K-MEANS DAN DBSCAN DALAM PENGELOMPOKAN DATA TRAVEL REVIEW RATINGS MENGGUNAKAN EVALUASI SILHOUETTE INDEX DAN DAVIES-BOULDIN INDEX," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.6884.

[18] Safitri Juanita and R. D. Cahyono, "K-MEANS CLUSTERING WITH COMPARISON OF ELBOW AND SILHOUETTE METHODS FOR MEDICINES CLUSTERING BASED ON USER REVIEWS," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 1, pp. 283–289, Feb. 2024, doi: 10.52436/1.jutif.2024.5.1.1349.

[19] N. A. Azizah and T. Ernawati, "ANALISIS SENTIMEN ULASAN PELANGGAN TERHADAP PRODUK ROTI MACAN ARTISAN DI E-COMMERCE TOKOPEDIA MENGGUNAKAN METODE CLUSTERING," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 8, no. 3, p. 580, Aug. 2024, doi: 10.52362/jisamar.v8i3.1576.

[20] R. Khairuna, Nurdin, and Ar Razi, "ANALISIS PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK KLUSTERISASI TEKS ULASAN PADA APLIKASI COOKPAD," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 445–458, Jul. 2025, doi: 10.36341/rabit.v10i2.6131.