https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 02, 2025

JU-KOMI
E-ISSN: 2963-0460

Comparison and Evaluation of Euclidean and Canberra Distances in the Adaptive K-Means Algorithm for Classifying the Food Security Status of Indonesian Provinces

Cinthya Agatha Sinaga[1] , Paska Marto Hasugian[2]
[1]Universitas Katolik Santo Thomas Medan, Jl.Setia Budi No.479F Tanjung Sari, Medan, Indonesia,
[2]Program studi Sains Data, fakultas Ilmu Komputer komputer, univesitas Katolik Santo Thomas

ARTICLE INFO

Keywords:
K-Means Adaptive,
Classification,
Food Security Status,
Euclidean Distance.
Canberra Distance

ABSTRACT

Food security issues in Indonesia are a major concern because they affect the sustainability of people's livelihoods and regional disparities. This study was conducted to classify food security conditions between provinces based on two main indicators, namely the Food Security Index and the Percentage of Adequate Food Consumption. The method used is the K-Means Adaptive algorithm with a comparison of two types of distance measurements, namely Euclidean and Canberra. The selection of centroids is done gradually using a probabilistic approach to improve the stability of the clustering results. Before conducting a comprehensive test, the method is first tested using sample data to see the characteristics of each distance function. Subsequently, all data were analyzed using Python programming, and the results were evaluated using the Silhouette Score metric. The analysis results showed that the Canberra distance function provided better clustering quality than the Euclidean function with a value of 0.415. This approach is expected to serve as a reference for more accurate and informative regional-based food security analysis.

Email :
cinthyaagathaa@gmail.com

INTRODUCTION

Food security is a strategic national issue that greatly determines economic, social, and political stability in Indonesia. Good food security reflects the extent to which a country is able to provide access to sufficient, safe, and nutritious food for its entire population. (FAO, 2021). To monitor food security at the regional level, the Indonesian government uses the Food Security Index (IKP), which summarizes various indicators such as availability, accessibility, and utilization of food. The IKP has been used as the basis for public policy decisions regarding food aid distribution, village development, and mapping of food-insecure areas (Badan Ketahanan Pangan, 2022).

With the development of information technology, data-based analytical approaches are now being used to deepen the analysis of food security conditions. One method that is widely used in mapping and grouping regional data is the clustering algorithm, especially K-Means, which is considered effective and easy to implement (Agus Lestari, Paranita Kartika, and Nur Budiman 2021).

However, the effectiveness of the K-Means algorithm is highly dependent on two important factors, namely the selection of initial centroids and the method of measuring the distance between data points. Incorrect selection of centroids can result in unrepresentative clusters, while the use of inappropriate distance metrics can obscure the actual data structure (Sulistiyawati and Supriyanto 2021).

In this context, the Adaptive K-Means approach was developed to improve the random initialization weakness in standard K-Means. Adaptive K-Means systematically selects initial centroids based on probabilistic distribution, thereby producing more stable and accurate clustering (Anto et al. 2024). In addition to the adaptive approach to centroid selection, the type of distance function used in the clustering algorithm also greatly affects the final results. Two distance metrics that are often compared in data analysis research are Euclidean Distance and Canberra Distance, each with different mathematical characteristics (Kurniawan et al. 2022).

Euclidean Distance is the most commonly used metric in data processing, mainly due to its simplicity and intuitive geometric interpretation. However, this metric tends to be less sensitive to small values and does not take into account the ratio between values, so it can be less effective for data with varying scales (Informasi 2017). Conversely, Canberra Distance is more suitable when the main focus is on relative differences rather than absolute ones. Because it calculates the ratio between the difference and the sum of values, this metric can be more sensitive to small changes and is often used in cases of social or economic data classification (Kurniawan et al. 2022).

Based on the importance of these two aspects, this study aims to compare the performance of the K-Means Adaptive algorithm with two types of distance functions, namely Euclidean and Canberra, in clustering provinces in Indonesia based on the Food Security Index and food consumption adequacy levels. This analysis aims to provide a more objective and accurate picture of food security status between regions.

The test results will be evaluated using the **Silhouette Score**, a metric widely relied upon in clustering analysis because it considers internal consistency and separation between clusters. (Yulisasih et al. 2024). This study is expected to contribute to selecting the most appropriate clustering approach for food security index data based on provinces, while supporting evidence-based policy making in the context of food security.

METHODS

This study uses secondary data obtained from the official website of the Indonesian government, https://data.go.id, which presents data on the Food Security Index (IKP) and the Percentage of Adequate Food Consumption by province. These two variables were chosen because they reflect the two main pillars of food security: supply capacity and the level of consumption fulfillment of the community. The initial data consists of 38 entries corresponding to the number of provinces in Indonesia, each containing FSI values and food consumption percentages in numerical form.

The initial stage involves data cleaning to ensure data validity prior to analysis. Entries containing missing values in one or both variables were checked. These incomplete entries were deleted to maintain the consistency of the analysis results, leaving 34 entries after cleaning. Only two numerical features were used in the processing, while administrative information such as province names was ignored.

After the clean data was obtained, it was grouped into six clusters based on food security levels, namely (1) Very Secure, (2) Secure, (3) Moderately Secure, (4) Moderately Vulnerable, (5) Vulnerable, and (6) Very Vulnerable. The grouping was performed using the K-Means Adaptive algorithm, where the initial centroid selection was not done randomly but using a probabilistic approach. This procedure begins with the manual selection of one centroid, followed by five other centroids determined based on the weighted probability of the squared distance of each data point from the existing centroids. This approach was adopted to minimize errors due to random initialization, as commonly occurs in conventional K-Means.

To assess the performance of the distance measurement method, two popular formulas in clustering were used, namely Euclidean Distance and Canberra Distance. Before being applied comprehensively, each method was first tested manually using 24 sample data representing all cluster classes. Manual testing was performed using Excel software so that each distance calculation and cluster formation process could be observed transparently and in detail. This initial testing served as the basis for assessing the behavior of each formula before proceeding to the comprehensive testing stage.

Further testing was conducted on the entire dataset using the Python programming language. This automation process enabled more efficient and accurate computation of all data entries, including iterative centroid updates, cluster formation, and visualization of results. For each distance method, the same initial centroid was used to ensure more objective comparison of results. Visualization was performed in the form of a two-dimensional scatter plot to visually understand the clustering patterns at each iteration.

The final step in this method is to evaluate the quality of the clustering results using the Silhouette Score metric, which measures cohesion within clusters and separation between clusters. The Silhouette Score value is calculated for each data point and averaged as an indicator of the overall performance of the model. By comparing the Silhouette Score values from each distance method, this study evaluates which method is more optimal in representing the structure of food security between provinces, while also providing a strong foundation for data-driven decision-making.

RESULTS AND DISCUSSION

Data Description

The data used in this study is secondary data taken from the official government website, data.go.id. Of the total 38 provinces, only 34 entries were used in the analysis process after going through a data cleaning stage, in which entries with empty or incomplete data on one or both variables were deleted to maintain consistency of results. The following is the data that has been cleaned:

Table 1. Dataset After Cleaning

No	IKP (Percent)	Food Consumption Adequacy (Percent)
1	73,94	90,9
2	77,49	92,46
3	84,32	91,12
4	70,42	89,07
5	74,94	89,42
	…………………	
30	71,99	93,47
31	62,68	68,34
32	61,44	71,56
33	51,36	78,09
34	40,21	73,97

Manual Testing of Sample Data

The next centroid selection in the K-Means Adaptive method is performed using a probability distribution, where the probability of selecting each data point as a centroid depends on its squared distance to the nearest centroid that has already been selected. For the initial centroid value (C1), it is determined randomly, then determining the centroid 2-6. The formula used is:

$$\mathrm{D(Xi)}^2 = \sum ||\ \mathrm{Xi} - \mathrm{Cj}\ ||^2 \quad \Big| \quad P(xi) = \frac{D(Xi)^2}{\sum_{i=1}^{n} D(Xi)^2}$$

Using this formula, the following results are obtained for Centroid 1-6:

Table 2. Centroid values 1-6 for sample testing

IKP	KKP	Centroid
85,13	96,47	**C1**
40,21	73,97	**C2**
62,68	68,34	**C3**
82,97	94,01	**C4**
66,29	90,45	**C5**
61,44	71,56	**C6**

https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 02, 2025
JU-KOMI
E-ISSN: 2963-0460

With the centroid set, the process continues with testing the sample data using two distance calculation methods, namely Euclidean and Canberra, to observe the clustering results manually.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \qquad d(p,q) = \sum_{i=1}^{n} \frac{|p_i - q_i|}{|p_i| + |q_i|}$$

*The left side is the Euclidean Distance formula; the right side is the Canberra Distance formula.

C1	C2	C3	C4	C5	C6	Cluster
37.89864	97.746401	25.71951	9.150549737	7.66537186	35.03785687	C5
2.68244	61.41034491	5.694900205	16.87938283	25.98159059	26.2517027	C5
9.41097	47.32667958	31.42803	3.188781882	18.02414384	58.10120585	C4
16.48645	33.775735	22.12702	13.48125695	4.754547195	18.8784272	C5
8.38028	33.88156568	28.38519	9.28270241	8.7117085	20.8215785	C5
31.32595	39.35610153	28.18715	6.590022472	8.56415187	25.7830587	C5
33.33778	38.81044784	24.28706	10.33314886	7.30076438	22.09314981	C5
1.59113	45.39648552	28.91661	4.706186737	16.3291657	27.61089236	C5
36.55812	33.78031327	22.30821	13.59773344	4.074125167	18.82582623	C5
34.71862	30.8816591	22.48277	17.85067358	0.0	19.50268187	C5
50.35898	0.0	35.09837	3.277555488	38.7784336	54.37625161	C1
3.27713	47.2203871	32.72805	0.0	20.4027121	3.450587029	C3
1.58203	42.60868427	29.8282	4.582321786	32.5061945	27.1567832	C6
34.50434	57.34676189	21.48536	11.25140241	30.7458029	19.9252655	C5
9.19418	43.21388186	28.8165	6.652132075	30.5873982	28.79022687	C4
33.81485	19.80987315	25.44917	7.592127052	30.6236464	25.5227775	C5
3.69069	47.5852505	32.48850	3.12440475	37.7153982	30.92045612	C4
33.79634	18.75784945	25.5031	7.58682136	35.9135537	23.73848026	C4
30.18649	58.85872163	24.46507	8.8108238	31.4353803	23.89356489	C4
15.47813	37.3861622	26.7094	9.8008672	6.45061237	24.31770715	C5
35.90027	23.58458072	0.0	32.72062872	22.4027131	3.450587029	C3
34.57261	50.96287024	4.59507	10.95157144	0.0	35.02877516	C1
38.44785	15.85863728	30.28262	10.1823244	53.03029575	0.0	C6
50.73898	0.0	25.16458	47.7250871	38.8095581	21.3865305	C2

Gambar 1. Results of Iteration 1 Euclidean Distance

C1	C2	C3	C4	C5	C6	Cluster
0.18007357	0.3981715358	0.224091333	0.014367908	0.057034624	0.211371301	C5
0.06828548	0.427835227	0.235657416	0.042404143	0.088885198	0.242949289	C4
0.07539082	0.458084567	0.290686027	0.03568871	0.125485265	0.277230435	C4
0.13443525	0.365087135	0.189846067	0.108789029	0.027389789	0.177110813	C5
0.1011835	0.398165031	0.222780078	0.075874921	0.060793068	0.38903358	C5
0.08728076	0.415684576	0.241599559	0.056778918	0.074834795	0.270930054	C4
0.10798157	0.396058938	0.218662277	0.082284363	0.052547065	0.23753427	C5
0.05368913	0.439085205	0.270069808	0.027858067	0.11571814	0.257178006	C4
0.13441991	0.365085167	0.189844537	0.108775004	0.034891018	0.177119446	C5
0.15662843	0.343135743	0.187231512	0.133856894	0.0	0.154598446	C5
0.0	0.09396457	0.322585566	0.025764236	0.156628129	0.309867592	C1
0.02516424	0.466414192	0.297402174	0.0	0.131093884	0.288861595	C3
0.08028784	0.435022788	0.263480239	0.034593403	0.096789579	0.250766713	C4
0.11908877	0.37926236	0.205455584	0.09149626	0.085058537	0.192555687	C5
0.06838754	0.428285008	0.255390201	0.042952753	0.088728595	0.242711695	C4
0.08852265	0.408068031	0.235680926	0.062756216	0.070238437	0.222893871	C4
0.02601478	0.466514114	0.297197789	0.01044325	0.130715023	0.28448874	C3
0.08038809	0.408389068	0.235884020	0.062502008	0.085775161	0.223103079	C4
0.09980491	0.402037648	0.231855857	0.065205832	0.108648088	0.220580425	C5
0.09324483	0.359987024	0.274437558	0.03137013	0.057640983	0.211841416	C5
0.32258557	0.257598086	0.0	0.29142174	0.187231512	0.030380772	C3
0.38058755	0.22541247	0.030310772	0.284881508	0.159568848	0.0	C6
0.25722187	0.148059338	0.165840133	0.327826731	0.180227534	0.13299851	C6
0.49039846	0.0	0.257950091	0.468434102	0.345113745	0.32514407	C2

Gambar 2 Results of Iteration 1 Canberra Distance

To apply the K-Means Adaptive algorithm in the clustering process, there are several core stages that are repeated until stable results are obtained. These stages include calculating distances, determining cluster membership, and updating centroid positions. In general, these steps can be explained as follows: (1) Calculate the Euclidean distance between each data point and all centroids; (2) Assign each data point to the cluster with the smallest distance; (3) Recalculate the position of the centroid based on the average of the points in the cluster; (4) Repeat the process until there are no changes in the clusters or the iteration limit is reached.

Automatic Clustering Results and Visualization

The next step was to test the entire data set using the Python programming language on the Google Colab platform. Visualization in the form of a two-dimensional scatter plot showed the cluster structure formed based on each distance function. This study grouped the data into six clusters representing village categories, namely Very Resilient, Resilient, Moderately Resilient, Moderately Vulnerable, Vulnerable, and Very Vulnerable. The initial centroid value (C1) was set randomly by the user, with this study taking the value from row 11, namely (83.15, 96.47). Subsequent centroids are automatically calculated by the program using a probabilistic approach, ensuring that the cluster center selection process is controlled and objective. The final results

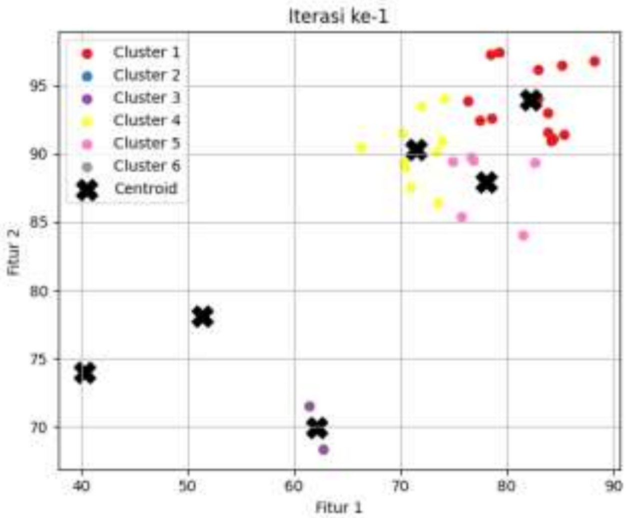
 JU-KOMI
E-ISSN: 2963-0460

show the amount of data distributed in each cluster, which serves as the basis for analyzing village characteristics.

The following are the results of data clustering iterations visualized to illustrate the distribution patterns of each cluster. The visualization is presented by comparing Euclidean distance and Canberra distance to highlight differences in measuring the proximity between data points. Each graph displays the distribution of data points relative to the centroid, while also showing the variations in distance that influence cluster formation. By comparing these two distance methods, the analysis becomes more in-depth in assessing the accuracy of clustering based on data characteristics.

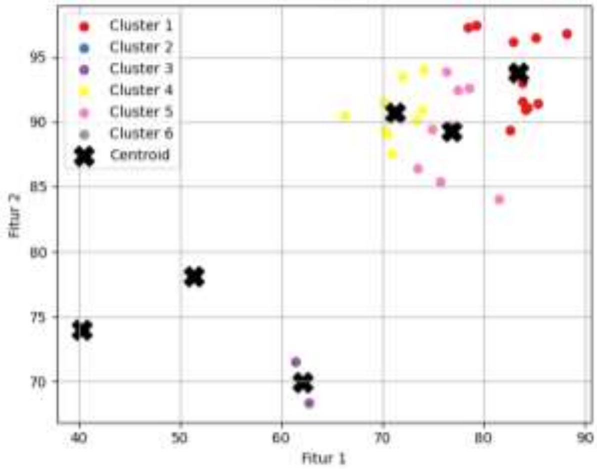
1. Results and Visualization with Euclidean Distance
   Based on the probability approach operated with Python programming, all selected centroids are displayed as follows:
   C1: [85.13 96.47], C2: [40.21 73.97], C3: [62.68 68.34], C4: [66.29 90.45], C5: [81.47 84.01], C6: [51.36 78.09]
   Once the Centroid 1-6 values are obtained, the program will calculate using the Euclidean formula and provide a visualization of each iteration as well as calculate the amount of data included in each cluster. The following are the results of the K-Means Adaptive iteration (5 iterations).
   ITERATION 1 :

Data	Centroid 1	Centroid 2	Centroid 3	Centroid 4	Centroid 5	Centroid 6
Data 1	12.49964	37.7404	25.213099	7.662324	10.206518	25.968595
Data 2	8.628424	41.613441	28.304893	11.980243	9.346991	29.820694
Data 3	5.419793	47.24652	33.318833	18.962642	8.933071	35.444011
Data 4	16.585469	21.118972	28.213722	24.354457	17.153496	21.914012
Data 5	16.453648	37.966068	28.587449	2.514848	12.533492	25.766795
Data 6	4.332777	41.906879	30.812201	12.795234	9.946475	25.17967
Data 7	7.531113	43.661991	26.118261	18.109676	12.442779	25.788226
Data 8	12.907478	40.698551	26.081181	9.220446	8.237456	31.718479
Data 9	6.570732	43.571838	28.706226	12.106276	12.352425	19.95182
Data 10	5.72082	32.091286	20.991817	12.204623	16.948242	38.492863
Data 11	12.028283	35.788021	22.907993	2.914818	10.989407	15.498462
Data 12	3.78024	27.258953	18.886208	10.825127	14.986684	35.209253
Data 13	5.601217	39.650738	31.452225	17.670958	8.479824	35.472167
Data 14	5.601217	39.650738	31.452225	17.670958	8.479824	35.472167
Data 15	5.601217	39.650738	31.452225	17.670958	8.479824	35.472167
Data 16	18.463673	35.555763	28.032452	19.698524	13.158222	35.986162
Data 17	16.867813	43.765598	33.965069	26.881248	18.559383	31.765829
Data 18	8.76548	43.560123	27.943783	14.769517	8.249215	33.781527
Data 19	7.611763	42.557545	25.046579	12.729421	6.966757	22.631888
Data 20	19.179844	31.858604	24.00985	5.042759	11.121834	23.692984
Data 21	16.686422	47.221397	34.348206	28.329563	18.926573	36.099598
Data 22	15.988642	39.887905	26.495377	14.676517	6.266757	36.398841
Data 23	19.480644	47.239891	35.405125	30.617235	20.911005	25.517199
Data 24	12.988242	37.086052	27.460356	13.659152	9.695091	25.578189
Data 25	18.976044	45.279339	33.368622	29.078537	17.689554	37.476505
Data 26	18.486084	39.87318	26.465517	14.635752	6.456274	32.491366
Data 27	19.486044	47.264343	34.505527	30.565723	20.841505	38.684155
Data 28	19.986042	25.986462	25.466572	12.499753	7.865703	32.658465
Data 29	18.768044	29.98597	24.486187	6.486572	5.871246	19.984021
Data 30	18.378842	33.688553	27.578669	13.884553	9.486153	14.486058
Data 31	23.166632	7.484347	3.745065	5.882206	23.469655	12.0
Data 32	12.0	11.366573	7.450564	3.058062	12.960257	12.0
Data 33	0.0	23.166184	38.580456	38.460375	44.460375	8.986388
Data 34	50.239098	0.0	23.166184	38.580456	44.460375	11.818637



ITERATION 5:

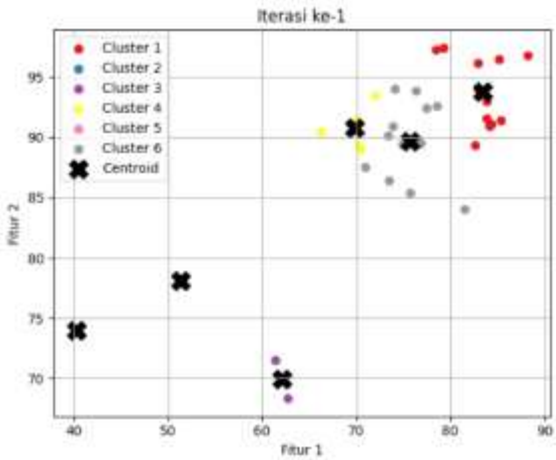
	Centroid 1	Centroid 2	Centroid 3	Centroid 4	Centroid 5	Centroid 6
Data 1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 34	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]



2. Results and Visualization with Canberra Distance
   The following are the results of iterations using Canberra distance, where the centroid value is determined in the same way as for Euclidean distance (4 iterations).
   ITERATION 1 :

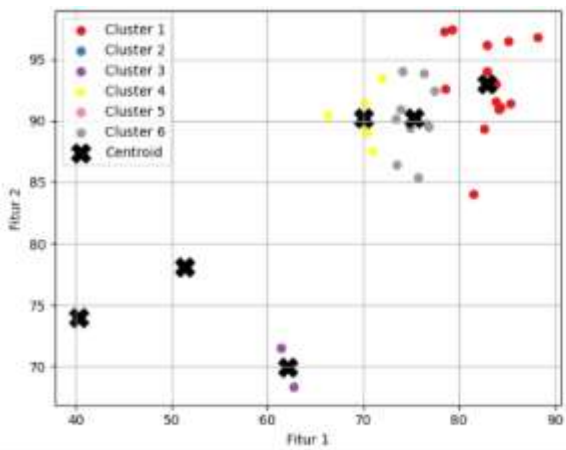
	Centroid 1	Centroid 2	Centroid 3	Centroid 4	Centroid 5	Centroid 6
Data 1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data 33	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]
Data 34	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]

https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 02, 2025
JU-KOMI
E-ISSN: 2963-0460



ITERATION 4:

	Centroid 1	Centroid 2	Centroid 3	Centroid 4	Centroid 5	Centroid 6
Data 1	0.864825	0.398175	0.217599	0.811132	0.258811	0.69475
Data 2	0.881919	0.427935	0.25917	0.869715	0.297651	0.658775
Data 3	0.872524	0.380868	0.250983	0.877276	0.319998	0.69796
Data 4	0.892753	0.404166	0.232040	0.788247	0.27337	0.692561
Data 5	0.872522	0.386045	0.238627	0.781413	0.276442	0.69796
Data 6	0.88733	0.396044	0.223692	0.885431	0.272814	0.69796
Data 7	0.885821	0.394164	0.21041	0.816083	0.269973	0.69796
Data 8	0.315576	0.505438	0.138040	0.844331	0.222712	0.671169
Data 9	0.115222	0.405672	0.181496	0.821455	0.279381	0.69192
Data 10	0.87752	0.386045	0.238627	0.781452	0.276442	0.69796
Data 11	0.868813	0.37247	0.230486	0.789169	0.278751	0.687861
Data 12	0.88026	0.393214	0.220658	0.883372	0.272226	0.69286
Data 13	0.865262	0.361944	0.246792	0.769083	0.282177	0.689121
Data 14	0.872586	0.380868	0.230486	0.789169	0.278751	0.687861
Data 15	0.861273	0.357241	0.221953	0.812669	0.279786	0.685266
Data 16	0.864862	0.359576	0.184875	0.840686	0.274555	0.683212
Data 17	0.883843	0.404164	0.210951	0.882604	0.277623	0.69286
Data 18	0.885265	0.393214	0.220658	0.883372	0.272226	0.69286
Data 19	0.872584	0.380868	0.230486	0.789169	0.278751	0.687861
Data 20	0.872522	0.386045	0.238627	0.781412	0.276442	0.69796
Data 21	0.864862	0.359576	0.184875	0.840686	0.274555	0.683212
Data 22	0.861273	0.357241	0.221953	0.812669	0.279786	0.685266
Data 23	0.88026	0.393214	0.220658	0.883372	0.272226	0.69286
Data 24	0.864862	0.359576	0.184875	0.840686	0.274555	0.683212
Data 25	0.88026	0.393214	0.220658	0.883372	0.272226	0.69286
Data 26	0.88733	0.396044	0.223692	0.885431	0.272814	0.69796
Data 27	0.872524	0.386045	0.238627	0.781412	0.276442	0.69796
Data 28	0.865262	0.361944	0.246792	0.769083	0.282177	0.689121
Data 29	0.861273	0.357241	0.221953	0.812669	0.279786	0.685266
Data 30	0.864862	0.359576	0.184875	0.840686	0.274555	0.683212
Data 31	0.872522	0.386045	0.238627	0.781412	0.276442	0.69796
Data 32	0.861273	0.357241	0.221953	0.812669	0.279786	0.685266
Data 33	0.864862	0.359576	0.184875	0.840686	0.274555	0.683212
Data 34	0.400396	0.0	0.141602	0.309209	0.138859	0.378468



Based on the results of the K-Mens Adaptive iteration and visualization using Euclidean and Canberra distances, it is evident that the Euclidean distance yields relatively large values because the calculation is absolute relative to the data value scale. The greater the difference between the data values and the centroid, the larger the Euclidean distance value recorded in the table. Therefore, Euclidean distance is sensitive to extreme values or outliers, so in the clustering process, differences between clusters often appear more pronounced at the beginning of the iteration and gradually decrease as the centroid approaches its optimal position. In contrast, the Canberra distance produces smaller and more stable values than the Euclidean distance. Its advantage lies in its sensitivity to small values, because if the values of both variables are close to zero, the Canberra distance will increase dramatically. That is why in the initial iterations, the

Canberra distance values may vary more sharply in data with values close to zero, but overall it is more moderate than the Euclidean distance in capturing similarity patterns.

Evaluation of the Application of Euclidean and Canberra Distance with the Silhouette Score Formula

Silhouette Score is an evaluation metric used to measure how well a data point is collected in its cluster. The formula for Silhouette Score for a data point is:

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}$$

The Silhouette Score itself measures how similar a point is to its own cluster compared to other clusters. This value ranges from -1 to 1, with values close to 1 indicating good clustering, and values close to or less than 0 indicating that the data may be misclassified or located near the boundary between clusters. This evaluation was conducted on 34 data points with an initial centroid value or C1 equal to the centroid when performing distance testing, namely 85.13 and 96.47.

Silhouette Score Euclidean Distance		
Index	**Cluster**	**Silhouette Score**
0	3	0,14274778
1	4	0,321544334
2	0	0,528834796
3	3	0,532549781
4	4	0,089281801
5	3	0,137435009
6	3	0,235623737
7	0	0,18127514
8	3	0,549839927
9	3	0,461753577
10	0	0,570079007
11	0	0,555778789
12	0	0,542408038
13	0	0,509359203
14	0	0,539537329
15	0	0,251568942
16	0	0,510584746
17	0	0,161742236
18	3	0,371677792
19	4	-0,072460855
20	3	0,556541938
21	0	0,536712211
22	4	0,15907743
23	4	0,244126641
24	4	0,092051811
25	4	0,407888552
26	0	0,586370178
27	4	0,389322236

Silhouette Score Canberra Distance		
Index	**Cluster**	**Silhouette Score**
0	5	0,318416446
1	5	0,317598763
2	0	0,527084848
3	3	0,555045848
4	5	0,460952341
5	5	0,298011484
6	5	0,170726538
7	0	0,31596155
8	3	0,573375532
9	3	0,430720469
10	0	0,590608415
11	0	0,556993413
12	0	0,524938826
13	0	0,515918685
14	0	0,545091583
15	0	0,208655819
16	0	0,508956356
17	0	0,071094426
18	3	0,39736451
19	5	0,165943393
20	3	0,538966326
21	0	0,575024143
22	0	-0,165643889
23	5	0,405838424
24	5	0,422869463
25	5	0,538874625
26	0	0,579073248
27	5	0,549478444

28	4	0,186105553
29	3	0,408173476
30	2	0,769043028
31	2	0,712704322
32	5	0
33	1	0

28	0	0,043576934
29	3	0,073973917
30	2	0,800981948
31	2	0,751822904
32	4	0
33	1	0

The average Silhouette Score based on Euclidean distance of 0.3579 indicates moderate clustering quality, reflecting clusters that are partially formed but still have many points close to the inter-cluster boundaries. Individual silhouette scores vary widely, ranging from approximately -0.072460855 to 0.769043028, indicating differences in the level of cohesion and separation between clusters. This condition indicates that there is still data that is not accurately placed in its cluster, so the separation between clusters is not yet fully optimal.

Meanwhile, the use of the Canberra distance resulted in an average Silhouette Score of 0.415, which is slightly higher, indicating a relatively better cluster structure, although not yet perfect. Most Silhouette values range from -0.165643889 to 0.800981948, indicating that much of the data is grouped fairly appropriately, though there are still negative values indicating that some points are closer to other clusters. This suggests that while the Canberra distance method can enhance cluster clarity, the potential for data misclassification or the presence of outliers still requires caution.

CONCLUSION

Based on the results of implementing the K-Means algorithm in clustering provincial resilience status using Provincial Food Security Index (IKP) data consisting of two main features, namely the Food Security Index (IKP) and the Percentage of Adequate Food Consumption, the quality of clustering was evaluated using the Silhouette Score. The evaluation results show that the use of Canberra distance produces an average Silhouette Score of 0.415, while the use of Euclidean distance produces a score of 0.3865. A Silhouette Score value closer to 1 indicates that Canberra distance provides better clustering results than Euclidean distance in the context of the IKP data used. This means that the distances between points within clusters are more consistent and better separated between clusters when using the Euclidean distance. Thus, it can be concluded that in grouping food security status based on the IKP, the choice of distance metric has a significant influence on the clustering results, and Canberra Distance is more suitable for use in this context than Euclidean Distance. This is important to consider in data-driven policy-making in the field of rural development.

REFERENCE

Aditya, Agil, Ivan Jovian, and Betha Nurina Sari. 2020. “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019.” Jurnal Media Informatika Budidarma 4(1): 51. doi:10.30865/mib.v4i1.1784.

Agus Lestari, Wiwit, Kurnia Paranita Kartika, and Saiful Nur Budiman. 2021. “Klasterisasi Siswa Berdasarkan Hasil Belajar Menggunakan K-Means Berbasis Web (Studi Kasus : Tk. Prima Insan Sholeh Talun).” JATI (Jurnal Mahasiswa Teknik Informatika) 6(1): 9–16. doi:10.36040/jati.v6i1.4261.

Anto, Andri, Tri Susilo, Novi Lestari, Program Studi Informatika, Universitas Bina Insan, Program Studi, Rekayasa Sistem, et al. 2024. “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS.” 16(2): 84–93.

Azhar, Anisa Laila, Suliyanto Suliyanto, Nur Chamidah, Elly Ana, and Dita Amelia. 2023. “Pemodelan Indeks Ketahanan Pangan Di Indonesia Berdasarkan Pendekatan Regresi Logistik Ordinal Data Panel Efek Acak.” Jurnal Ketahanan Nasional 29(2): 166. doi:10.22146/jkn.86511.

Cadavid, Luis, Vivek Arulnathan, and Nathan Pelletier. 2024. "Food Security and Food Sovereignty: A Review of Commonly Used Indicators and Consideration of Environmental Sustainability Aspects." Sustainability (Switzerland) 16(24). doi:10.3390/su162411034.

Desa, D I, and Bapinang Hulu. 2024. "Model Klasterisasi Data Penduduk Menggunakan Algoritma K-Means Untuk Mengetahui Prioritas Penerima Bantuan Sosial Di Desa Bapinang Hulu." 7: 774–84. doi:10.37600/tekinkom.v7i2.1588.

Informasi, Fakultas Teknologi. 2017. "IMPLEMENTASI CLUSTERING K-MEANS DALAM DOMAIN WAVELET MENGGUNAKAN ADAPTIF SOFT-THRESHOLDING UNTUK DENOISING."

Kurniawan, Deny, Dedi Triyanto, Mochamad Wahyudi, Sistem Informasi, Teknik Informatika, Universitas Bina, Sarana Informatika, Jl Kramat, and Raya No. 2022. "Comparison of Euclidean Distance, Camberra Distance, and Chebychev Distance in K-Means Algorithm Based on Dbi Evaluation." Jurnal Mantik 5(36): 2830–38.

Mulyani, Heti, Ricak Agus Setiawan, and Halimil Fathi. 2023. "Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data)." Journal of Information Technology and Its Utilization 6(2): 45–50. doi:10.56873/jitu.6.2.5243.

Salasa, Andi Rachman. 2021. "Paradigma Dan Dimensi Strategi Ketahanan Pangan Indonesia." Jejaring Administrasi Publik 13(1): 35–48. doi:10.20473/jap.v13i1.29357.

Sulistiyawati, Ari, and Eko Supriyanto. 2021. "Implementasi Algoritma K-Means Clustring Dalam Penetuan Siswa Kelas Unggulan." Jurnal Tekno Kompak 15(2): 25. doi:10.33365/jtk.v15i2.1162.

Syam, Febrizal Alfarasy. 2017. "Jurnal Ilmu Komputer Dan Bisnis, Volume 8, Nomor 1, Mei 2017." Jurnal Ilmu Komputer dan Bisnis 8(Sunjana 2010): 1841–46.

Yulisasih, Baiq Nikum, Herman Herman, Sunardi Sunardi, and Herman Yuliansyah. 2024. "Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation." ILKOM Jurnal Ilmiah 16(3): 330–42. doi:10.33096/ilkom.v16i3.2325.330-342.