Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Analysis of Protein Consumption Data and Desired Dietary Patterns as a Basis for Provincial-Level Food Security Information Systems Using the K-Means Algorithm

Dian Fitri Islamati Munisah[1], Listanto Tri Utomo[2]

[1,2]*Information system, Universitas Merdeka Malang,Indonesia*
dianmunisah@gmail,com[1] ,listanto.tri@unmer.ac.id.com[2]

Abstract

Food security and nutritional status, particularly in efforts to eliminate stunting, are important issues in Indonesia. Stunting caused by chronic malnutrition is greatly influenced by low protein consumption, especially during the first 1,000 days of life. This study aims to analyze the relationship between average per capita protein consumption and the Food Consumption Pattern Score (FCPS) at the provincial level from 2021 to 2023, as well as to explore the role of information systems in supporting food security policies. Data were obtained from data.go.id and analyzed using descriptive statistics, Spearman's correlation, and K-Means clustering methods. Results showed a significant positive correlation between protein consumption and PPH scores ($\rho = 0.604$; $p < 0.001$), indicating that protein intake is closely related to dietary diversity. Cluster analysis yielded two main groups: a low cluster dominated by eastern Indonesia, and a high cluster including Yogyakarta and Jakarta. Although the national PPH score increased from 81.81 (2021) to 84.96 (2023), inter-regional disparities remain high. These findings underscore the need for cluster-based interventions and the use of information systems to support more informed decision-making. The limitations of the data, which are not yet fully curated, highlight the need for further studies considering socio-economic variables.

Keywords: *food security, stunting, protein consumption, Ideal Diet Pattern, clustering, K-Means.*

1. Introduction

Food security and nutritional status are important issues that are a concern for the Indonesian government, particularly in its efforts to eradicate stunting in toddlers.[1] Stunting, which is characterized by growth disorders due to chronic malnutrition, remains a serious problem in various regions.[2] A national survey shows that the prevalence of stunting in Indonesia in 2023 is 21.5%. However, this figure has decreased to 19.8% in 2024.[12] One of the main factors influencing stunting is an unbalanced diet and insufficient protein intake, particularly during the first 1,000 days of life (from conception to age 2). [3]

The Food Consumption Pattern Score (PPH) is an indicator used to measure the quality of people's dietary patterns based on the diversity and balance of the types of food consumed.[4] Additionally, the average protein intake per capita is an important parameter in assessing nutritional adequacy, particularly in supporting children's growth and development. The latest data shows that the average protein consumption of the Indonesian population per province in 2024 reached approximately 60.0 grams per capita (data is still under review). [5]

Understanding the relationship between protein consumption and PPH scores at the provincial level can provide a more comprehensive picture of food consumption quality and the risk of stunting. In today's digital age, effective data management and utilization are key to supporting public policy decision-making. Information systems play a crucial role in managing, analyzing, and presenting accurate food and nutrition data, including real-time data[11] used by policymakers to monitor regional disparities, set food intervention priorities, and design more targeted stunting eradication programs. Interactive and informative data visualization can accelerate the process of evaluating and planning food security and stunting eradication programs.[6] In this study, the K-Means algorithm will be used to group provinces based on protein consumption patterns and PPH scores, thereby identifying groups of provinces with different food security characteristics. [7]

This study aims to identify the relationship between average per capita protein consumption and the Provincial Food Consumption Pattern Score for the 2021-2023 period at the provincial level using K-Means cluster analysis, as well as to explore the role of information systems in supporting the monitoring and evaluation of food security policies. With this approach, it is hoped that the quality of food consumption can be improved and stunting rates in Indonesia can be reduced.

2. Research Method

2.1. Data Types and Sources

This study uses secondary data consisting of two main datasets, namely:

Average Protein Consumption per Capita by Province for 2021-2023, obtained from the official data portal data.go.id. This data includes estimates of average daily protein consumption per capita at the provincial level, calculated based on the protein content of various commodities consumed by the population. This data was selected because it is available nationally and reflects protein consumption based on government standards.

Food Consumption Pattern Score (PPH) for Districts/Cities for 2021-2023, obtained from the official publication data.go.id. The PPH score measures the quality of dietary patterns based on the diversity and balance of food types consumed at the district/city level. This data was selected because it is relevant in describing the quality of the population's dietary consumption. This dataset was then aggregated to the provincial level to align with the research analysis unit.

Both datasets were selected because they are open, credible, and issued by official government sources. The data covers almost all provinces in Indonesia, but there are some missing values (missing data) in some districts/cities, which have been addressed through data cleaning and aggregation processes.

2.2 Data Processing

Data on protein consumption per individual at the provincial level and the Desired Diet Pattern (PPH) score at the district/city level were obtained from various official sources. The data then underwent a cleaning process that included managing missing values, removing duplicate data, and correcting inconsistencies to ensure data quality and completeness. Given that the data comes from diverse administrative levels, district/city data is aggregated to the provincial level. It should be noted that all data used is still in the curation phase, so analysis results should be interpreted with caution and may be revised once final data becomes available.

2.3. Analysis Methods

Descriptive Analysis, conducted to determine the distribution and characteristics of per capita protein consumption and PPH scores in various regions. Descriptive statistics such as mean, median, and standard deviation are used to describe the data.

Correlation Analysis, Pearson or Spearman correlation is used to test the relationship between the average per capita protein consumption and PPH scores in adjusted regions.

Cluster Analysis, as an approach to group provinces in Indonesia based on similarities in protein consumption values and PPH scores. The algorithm applied is K-Means Clustering, which enables the formation of groups of provinces with similar characteristics. It is widely used on large datasets because it is computationally efficient, easy to implement, and its results are easy to interpret.[13] The optimal number of clusters is determined using the Elbow Method and Silhouette Score. The Elbow Method is used to find the optimal number of clusters based on inertia values, while the Silhouette Score is used to assess the quality of the separation of the formed clusters. The output from this process will be utilized to support further analysis of consumption patterns and food security.[8]

3. Results and Discussion

3.1 Data Collection and Pre-processing

The initial stage begins with the collection of secondary data obtained through the official government portal data.go.id. The datasets used include:

1. Food Pattern Scores per district/city in Indonesia for 2021-2024
2. Average per capita consumption (grams/capita/day) per province for the years 2021-2024

Although sourced from official agencies, the data obtained is not yet fully accurate. Therefore, data pre-processing steps were conducted, including: data transformation involving standardizing province naming formats, removing irrelevant columns such as “No” or empty data, converting values to numerical format, aggregating PPH scores from the district/city level to the provincial level, and filtering data by year (2021-2023) because in 2024 there is incomplete data, which would render the analysis invalid. The following is a heatmap of PPH scores by province and year (2021-2023) using the transformed data.

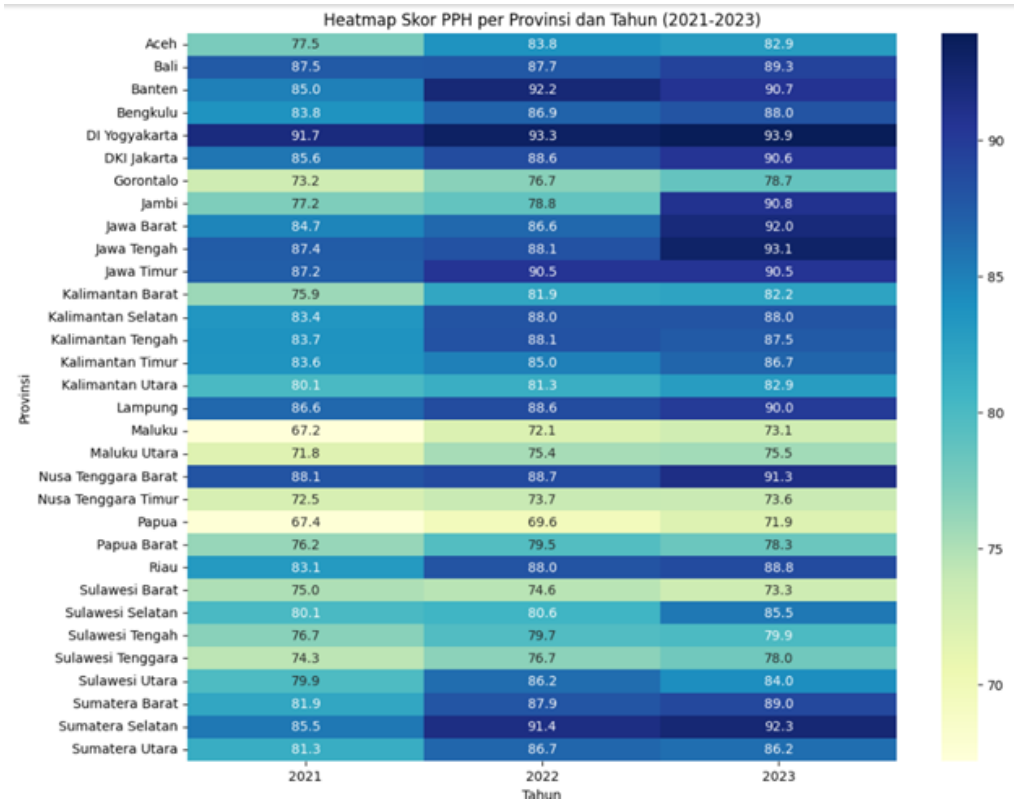

Fig. 1: Heatmap of PPH Scores by Province and Year (2021-2023)

The 2021-2023 PPH Score heatmap reveals three key insights:

1. There has been a consistent increase in the national food security pattern, with 85% of provinces showing an upward trend in scores, with the highest achievements in DIY (93.9) and Central Java (93.1), reflecting the effectiveness of nutrition interventions in Java-Bali.
2. Significant geographical disparities have been identified, with a 20.8-point difference between the highest and lowest scores (Maluku 73.1/NTT 73.6), indicating the need for a specific approach for the underdeveloped eastern regions of Indonesia.
3. There is an anomaly in West Sulawesi Province as the only province with a decrease in score (-1.7 points), suggesting the need to evaluate the region's food policy, while the extreme increase in Jambi (+13.6 points) should be studied as a model of success. This pattern underscores the importance of a performance-based differentiation strategy in national food security planning.

3.2 Descriptive Statistical Analysis

Descriptive analysis was conducted on aggregated provincial-level data that had undergone cleaning and pre-processing for the 2021-2023 period. The analysis results show that the national average protein consumption per capita over the past three years was around 61.05 grams/capita/day with a standard deviation of 5.61, while the national average Food Consumption Pattern (FCP) score was recorded at 83.90 with a standard deviation of 6.59. When examined annually, protein consumption remained relatively stable, though there were minor fluctuations, from 60.89 grams (2021), 60.86 grams (2022), to 61.32 grams (2023). Meanwhile, the PPH score shows an upward trend from 81.81 (2021), 83.57 (2022), to 84.96 (2023), indicating improvements in the quality and diversity of food consumption in most provinces.

The national PPH score range remains quite wide, with a minimum value of 67.18 and a maximum of 93.92, indicating significant disparities in consumption patterns between regions. Similarly, minimum protein consumption was recorded at 45.5 grams and maximum at 77.2 grams, reflecting variations in access to and sufficiency of protein-rich food sources across provinces. The geographical distribution of protein consumption and PPH scores visualized through (Figure 1 Heatmap of PPH Scores by Province and Year (2021-2023)) shows that provinces in Eastern Indonesia, such as Maluku, East Nusa Tenggara, Papua, and North Maluku, consistently have lower PPH scores compared to provinces in Java and Sumatra. These differences reflect spatial disparities in the diversity and quality of food consumption, which may be related to demand for nutritious food and development disparities between regions.[14]

The results of the normality test using Shapiro-Wilk show that both the protein consumption variable (p-value 0.0455) and the PPH score (p-value 0.0031) are not normally distributed, so further analysis needs to consider a non-parametric approach.

```
=== STATISTIK DESKRIPTIF ===
       Konsumsi Protein   Skor_PPH
count         96.000000  96.000000
mean          61.047917  83.030355
std            5.610760   6.596190
min           45.100000  67.181818
25%           58.100000  77.903708
50%           60.700000  83.910145
75%           64.125000  88.046346
max           77.200000  93.920000

=== STATISTIK PER TAHUN ===

Tahun 2021:
       Konsumsi Protein   Skor_PPH
count         32.000000  32.000000
mean          60.968750  80.472733
std            5.841091   6.181943
min           45.500000  67.181818
25%           58.100000  76.101511
50%           61.050000  81.631659
75%           63.775000  85.087500
max           77.200000  91.740000

Tahun 2022:
       Konsumsi Protein   Skor_PPH
count         32.000000  32.000000
mean          60.856250  83.656955
std            5.595617   6.336373
min           45.100000  69.624138
25%           57.700000  79.323951
50%           60.350000  86.412222
75%           64.325000  88.087857
max           74.800000  93.260000

Tahun 2023:
       Konsumsi Protein   Skor_PPH
count         32.000000  32.000000
mean          61.318750  84.961377
std            5.560891   6.630486
min           46.800000  71.910345
25%           58.925000  79.626282
50%           60.900000  87.113571
75%           64.350000  90.550219
max           76.100000  93.920000

=== UJI NORMALITAS ===
Konsumsi Protein - p-value: 0.0455
Skor PPH - p-value: 0.0031
```

Fig. 2: Descriptive Analysis Results

These findings are in line with literature stating that Indonesia's national PPH score is generally still below the ideal score of 100, indicating that food consumption diversity is not yet optimal across all regions. In addition, variations in protein consumption between provinces may reflect differences in access to animal and plant-based foods, as well as potential inequalities in national food distribution. Data from 2024 was not used in this analysis due to incomplete data that could affect the validity of the results.

3.3 Correlation Between Protein Consumption and PPH Scores

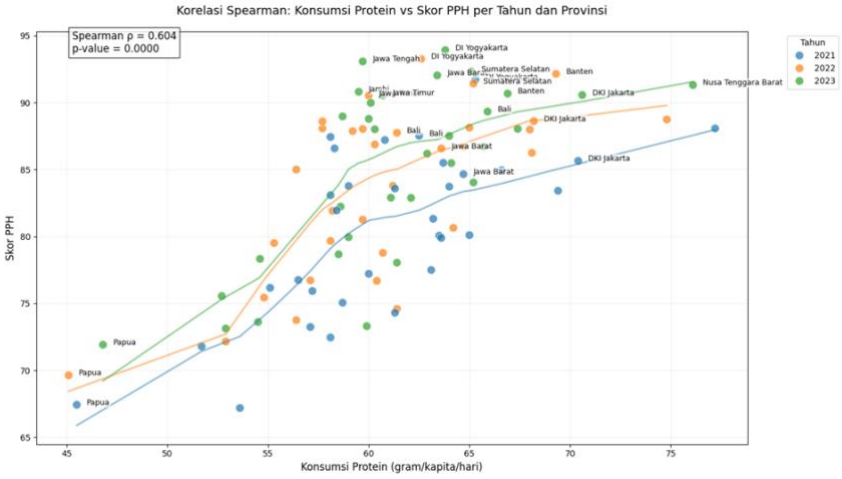

Fig. 3: Spearman Correlation Protein Consumption VS PPH Per Year and Province

The visualization of the relationship between protein consumption (grams/capita/day) and the Desired Diet Pattern (DDP) score at the provincial level during the 2021–2023 period can be seen in Figure 3. Based on the scatter plot, the following insights were obtained:

1. There is a fairly strong positive correlation between the two variables. The Spearman correlation test yielded a coefficient value of 0.604 with a p-value of 0.0000, indicating a statistically significant relationship. This means that the higher the protein consumption in a province, the better the quality and diversity of its dietary patterns.

2. Provinces such as DI Yogyakarta, DKI Jakarta, Bali, and Central Java consistently rank in the upper right quadrant of the graph, indicating high protein consumption accompanied by high PPH scores. This reflects the success of nutritional interventions and more equitable access to food in urban and western Indonesia.

3. Papua has consistently been the lowest-ranked province over the past three years, reflecting significant unresolved challenges. Conversely, provinces such as DI Yogyakarta, DKI Jakarta, Bali, and Central Java have consistently maintained high scores, potentially serving as models for provincial-level food and nutrition policies.

3.4 Cluster Analysis

As a follow-up step, cluster analysis was performed using the K-Means algorithm to group provinces based on protein consumption values and PPH scores.

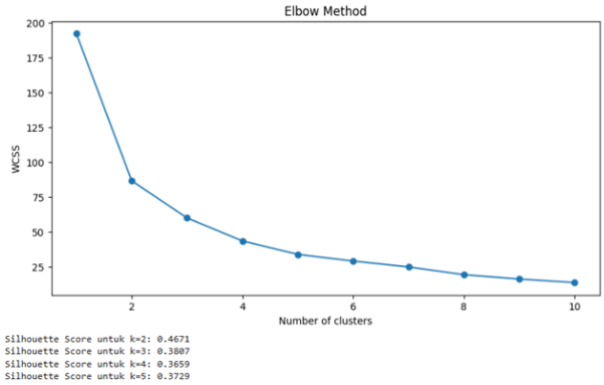

Fig. 4: Elbow Method for Determining the Optimal Number of Clusters

The exploration results using the Elbow Method show visualization and use the Silhouette Score to be more specific. Thus, the optimal number of clusters is k = 2, because the Silhouette Score is an evaluation metric used in cluster analysis to measure how well each data point is in its own cluster compared to other clusters. The Silhouette Score ranges from -1 to +1, with the following interpretations: values close to +1 indicate excellent cluster separation and that the data is in the appropriate cluster; values around 0 indicate that clusters overlap, making it difficult to clearly distinguish between clusters; while negative values (< 0) indicate that much of the data may be misplaced because it is closer to a different cluster. Thus, the higher the Silhouette Score, the better the quality of the clusters formed.[9]

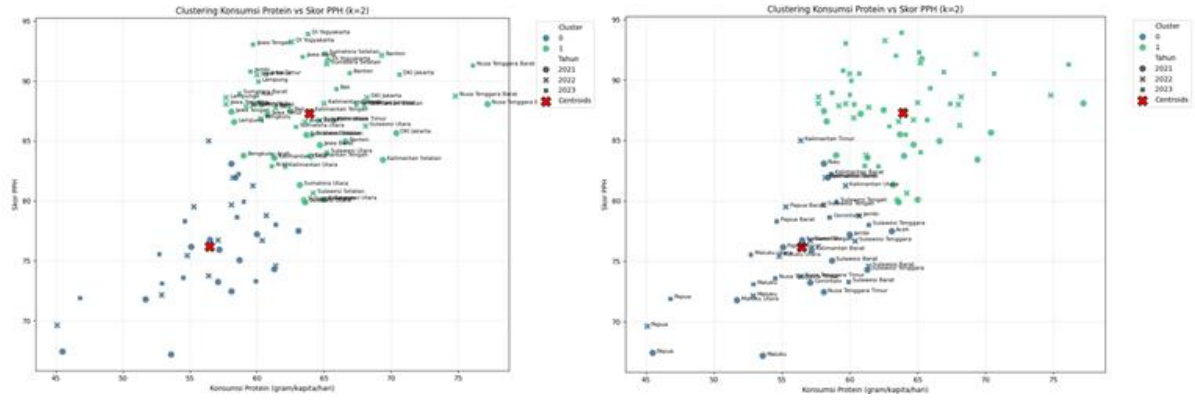

Fig. 5: Visualization of Clustering Results

From the visualization of Figure 5, which shows the results of clustering using the K-Means algorithm on protein consumption data (grams/capita/day) and the Desired Food Pattern (PPH) score at the provincial level in Indonesia during the period 2021–2023, several insights were generated:

1. Two main clusters have formed based on protein consumption levels and PPH scores, dividing the data into two groups: the low cluster (low consumption and low PPH scores) and the high cluster. This indicates significant performance differences between provincial groups in terms of nutrition and consumption patterns.
2. Provinces such as DI Yogyakarta, DKI Jakarta, and Bali consistently belong to the high-score cluster, indicating consistent food and nutrition performance year over year.

3. Provinces such as Papua, NTT, and Maluku mostly fall into the low-score cluster, highlighting regional disparities that were also evident in descriptive and correlation analyses.
4. The cluster center (centroid) helps map the direction of change by illustrating the average profile of each cluster. This is important for understanding the position and potential shift of provinces from one cluster to another over time. By displaying the centroid, readers can clearly see the differences in profiles between low and high clusters and understand the direction of data shifts if changes in protein consumption or PPH scores occur in the future. This approach is also useful for supporting cluster-based policy formulation, such as in determining priorities for interventions to increase protein consumption and improve dietary patterns in provinces that are still in clusters with low performance. The distribution of data across years, visualized through markers, also provides an overview of the dynamics of changes in protein consumption and PPH scores in various provinces during the observation period. [10]

5. Using different markers for 2021-2023, it can be seen that some provinces have changed position between years, indicating an improvement or decline in consumption performance and PPH scores.

4. Conclusion

This study examines the relationship between per capita protein consumption and the Desired Diet Pattern (DDP) score at the provincial level in Indonesia during 2021–2023 using descriptive analysis, visualization, and clustering. The results show significant variation between provinces, with a national trend of increasing DDP scores indicating an improvement in dietary patterns. There is a significant positive correlation protein consumption and PPH scores, confirming that increased protein consumption is closely associated with more diverse and higher-quality dietary patterns.

K-Means cluster analysis divided provinces into two main groups: a cluster with low scores and protein consumption (cluster 0), facing challenges in food distribution and access to nutrition, and a cluster with high scores and consumption (cluster 1), showing better food security. Although some other studies use three clusters, the use of two clusters in this study is sufficient to map disparities in dietary patterns across regions. This cluster-based approach is useful for mapping regional disparities and designing more targeted policy interventions.

The main implication of these findings underscores the importance of increasing protein consumption as a strategy for improving national dietary patterns. An effective data-based information system can be utilized to monitor disparities between regions and support more targeted food security and stunting eradication policies. Further research is recommended to include socio-economic variables to enrich a more comprehensive analysis. [15] Collaboration between the technology and food policy sectors is also suggested to develop a more effective and efficient system.

References

[1] Afifa Atira, B. N. (2023, September). Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk. *Jurnal Ilmiah Wahana Pendidikan*. doi:https://doi.org/10.5281/zenodo.8282638

[2] Analisis Hard dan Soft Clustering Untuk Pengelompokan. (2023, Oktober 4). *Jurnal Sistem dan Teknologi Informasi , 11*. doi:10.26418/justin.v11i4.68400

[3] Biro Komunikasi dan Informasi Publik, K. K. (2025, May 26). *SSGI 2024 : Prevalensi Stunting Nasional Turun Menjadi 19,8%*. Retrieved from Kemenkes: https://kemkes.go.id/id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198#:~:text=Survei%20nasional%20yang%20menjadi%20rujukan,19%2C8%25%20pada%202024.

[4] dio Caisar Darma, P. T. (2020). Ekonomika Gizi Dimensi Baru di Indonesia. Yayasan Kira Menulis. Retrieved from https://repository.unmul.ac.id/bitstream/handle/123456789/29127/FullBook%20Ekonomika%20Gizi.pdf?sequence=5&isAllowed=y

[5] Indonesia, B. K. (2023). Buku Kinerja Kementrian Kesehatan Republik Indonesia TAhun 2022-2023 Transformasi Kesehatan Mewujudkan Masyarakat Indonesia Sehat dan Unggul.

[6] Jivi Muzaqi Guntur, R. K. (2025). Algoritma K-Means untuk Meningkatkan Silhouette Score pada Pengelompokan Data Stok Bahan Manufaktur di PT. XYZ Kabupaten Majalengka. *Jurnal Ilmu Multidisiplin, 5*. Retrieved from https://jayapanguspress.penerbit.org/index.php/metta/article/view/4046/1916

[7] *Mengenal Stunting dan Gizi Buruk. Penyebab, Gejala, Dan Mencegah*. (2018, Januari 26). Retrieved from Kementrian Kesehatan Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat: https://promkes.kemkes.go.id/?p=8486

[8] Nur Aminudin, S. A. (2025, 5). Visualisasi Data Interaktif untuk Analisis Tren Stunting Pendek dan Sangat Pendek pada Balita di Kabupaten Pringsewu. *Jurnal Ilmu Multidisiplin*. Retrieved from https://jayapanguspress.penerbit.org/index.php/metta/article/view/4046/1916

[9] *Panduan Penghitungan Pola Pangan Harapan (PPH).* (2015, Desember). Retrieved from Badan Ketahanan Pangan Kementerian Pertanian 2015: https://repository.stikespersadanabire.ac.id/assets/upload/files/docs_1711520716.pdf?utm_.com

[10] Prihatini, H. d. (2011). FOOD DIVERSITY AND CONTRIBUTION TO ENERGY AND PROTEIN. *GAMBARAN KERAGAMAN MAKANAN DAN SUMBANGANNYA TERHADAP, 39*, 62-73.

[11] Putri Vania, B. N. (2023, November). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada ClusteringProduksi Padimenggunakan AlgoritmaK-Means. *Jurnal Ilmiah Wahana Pendidikan*. Retrieved from https://jurnal.peneliti.net/index.php/JIWP/article/view/5415/4459

[12] *Rata-rata Konsumsi Protein per Kapita Provinsi update Tahun 2024*. (2025, April 29). Retrieved from Satu Data Indonesia: 5. https://data.go.id/dataset/dataset/rata-rata-konsumsi-protein-per-kapita-provinsi-update-tahun-2024

[13] renaldo Fajar Nugraha Susilo, S. F. (2023). PENGGUNAAN ARTIFICIAL INTELLIGENCE (AI) DALAM MEMBANGUN SISTEM. *Jurnal Imagine, 3*.

[14] Siti Mutrofin, T. W. (2023). Perbandingan Kinerja Algoritma Kmeans dengan Kmeans Median. *Jurnal Informasi dan Teknologi, 5*, 88-91.

[15] *Tingkatkan Kualitas Konsumsi Pangan Nasional, NFA Gelar Bimtek Analisis Konsumsi Pangan Berbasis Pola Pangan Harapan (PPH) Bagi 38 Provinsi se-Indonesia*. (2023, 10 26). Retrieved from Badan Pangan Nasioal: https://badanpangan.go.id/blog/post/tingkatkan-kualitas-konsumsi-pangan-nasional-nfa-gelar-bimtek-analisis-konsumsi-pangan-berbasis-pola-pangan-harapan-pph-bagi-38-provinsi-se-indonesia