ITEJ
Information Technology Engineering Journals
eISSN : 2548-2157

Url : https://itej.uinssc.ac.id
Email : itej@syekhnurjati.ac.id

Stunting Risk Cluster Analysis In Petatal Plantation Village Using K-Means Clustering Approach

Dewi Andini Putri
Information System
Royal University
Asahan, Indonesia
dewii26182@gmail.com

Dewi Maharani
Information System
Royal University
Asahan, Indonesia
dewimaharani15@gmail.com

Ahmad Muhazir
Information System
Royal University
Asahan, Indonesia
ahmadmuhazir45@gmail.com

Abstract— Stunting remains a critical public health issue in rural communities, particularly in plantation-based villages where socioeconomic conditions, nutrition access, sanitation, maternal knowledge, and health service utilization may vary across households. This study aims to analyze stunting risk clusters in Petatal Plantation Village using the K-Means Clustering approach. The research applies a quantitative data mining method by grouping household or child-level data based on several risk indicators, including child age, nutritional status, birth weight, exclusive breastfeeding history, maternal education, household income, access to clean water, sanitation conditions, immunization status, and frequency of visits to health service facilities. The K-Means algorithm was used to classify the data into several clusters representing different levels of stunting risk. The clustering process involved data preprocessing, normalization, determination of the optimal number of clusters, model implementation, and interpretation of cluster characteristics. The results of the study are expected to identify distinct risk groups, such as low-risk, moderate-risk, and high-risk clusters. Households in the high-risk cluster are generally characterized by limited economic capacity, poor sanitation, low maternal nutrition awareness, inadequate dietary diversity, and irregular access to health services. Meanwhile, the moderate-risk cluster may show partial vulnerability, while the low-risk cluster reflects better nutritional and environmental conditions. This clustering analysis provides a data-driven basis for village authorities, health workers, and local stakeholders to design more targeted stunting prevention programs. Instead of applying a uniform intervention, the proposed approach supports priority-based decision-making according to the specific characteristics of each risk cluster. Therefore, K-Means Clustering can be considered an effective analytical tool for mapping stunting vulnerability and strengthening evidence-based public health intervention strategies in Petatal Plantation Village.

Keywords— stunting risk, cluster analysis, K-Means Clustering, data mining, Petatal Plantation Village, public health.

Article info: Date Submitted: 23/02/2026 | Date Revised: 03/03/2026 | Date Accepted: 13/04/2026

I. INTRODUCTION

The rapid development of information and communication technology has brought the world into the era of big data, where the volume, velocity, and variety of data increase significantly [1]. Data generated from various sources such as government agencies, the education sector, and the health field requires analytical methods capable of extracting valuable knowledge to support decision-making [2]. One of the most widely used

approaches for that purpose is data mining [3], namely the process of extracting hidden information from large datasets using statistical techniques, artificial intelligence (AI), and machine learning to generate new patterns and knowledge discovery [4].

In recent years, the application of data mining has not only been limited to the industrial and financial sectors, but has also rapidly developed in the field of public health[5]. This technology enables comprehensive health data analysis to find patterns that are not visible through conventional methods [6]. One of the algorithms that is popularly used is K-Means Clustering [7] an unsupervised learning method that is capable of grouping data into several clusters based on the similarity of characteristics [8]. This method is considered effective in supporting large-scale health data analysis to understand the distribution of disease risk, including in cases of stunting [9].

Stunting is a chronic nutritional problem characterized by a child's height being lower than the standard for their age due to long-term nutritional deficiencies [4]. The impact of stunting is not limited to physical aspects, but also affects cognitive development, productivity, and long-term health. The Indonesian government has established the reduction of stunting prevalence as a national priority through various programs such as the National Movement for Accelerating Stunting Reduction (GN-PPStunting). Nevertheless, stunting prevalence in some areas remains high, particularly in rural and plantation areas that have limited access to health services, nutrition education, and sanitation.

Although there are national policies and intervention programs, the distribution of stunting cases shows significant spatial variation and heterogeneity between villages/sub-districts, with some plantation or rural areas still exhibiting higher prevalence due to a combination of socio-economic factors, access to services, and local dietary patterns. At the village level, especially in plantation villages such as Petatal Plantation Village, detailed data on stunting patterns, for example family or household clusters at high risk, are still limited or have not been analyzed in depth. In addition, many stunting studies use statistical determinant analysis, but few studies utilize unsupervised data mining techniques (e.g., K-Means) to identify groups/clusters of at-risk toddlers at the village level. This leaves a practical gap between the data available at health posts/villages and the use of analytics for targeting interventions [10].

From the description above, the author has the idea to group toddler data based on the level of stunting risk using a data mining approach with the K-Means Clustering algorithm. Through this grouping process, it is expected that the characteristics of each child group, such as high, medium, and low-risk groups, can be identified, so that village authorities and health workers can design intervention programs that are more targeted according to the conditions of each group.

II. METHODOLOGY

The type of method in this research is quantitative and the model used for system development is waterfall [11]. This research began with the problem identification stage, which is understanding firsthand the stunting conditions that occur in Petatal Plantation Village. At this stage, the researcher conducted initial observations in the field, communicated with village officials and health workers, and reviewed available stunting case data. After the problems were identified, the researcher conducted a literature study by collecting various references from scientific journals, books, research articles, and official government reports related to stunting, its causal factors, and the application of data mining techniques, particularly the K-Means Clustering algorithm in the health sector. This stage aims to build a strong theoretical foundation and ensure that the methods used are relevant to the research objectives.

The next stage is the collection of data obtained from relevant agencies such as health centers or village offices, which contains information about the condition of toddlers, for

example age, weight, height, and nutritional status. The data that has been collected then goes through a data processing and cleaning process (data preprocessing) to remove incomplete, duplicate, or improperly formatted data, as well as performing normalization so that the data is ready to be analyzed. After that, the K-Means Clustering method is applied to group toddler data based on the risk level of stunting, for example into high, medium, and low-risk categories. The final stage is the analysis of results, which is interpreting the clustering results to understand the characteristics of each cluster, thereby providing an overview of the stunting condition in Plantation Petatal Village and helping relevant parties determine more appropriate handling measures [12].

III. RESULT AND DISCUSSION

A. Result

In the first stage, which is the analysis stage, it is a process of inspecting, cleaning, transforming, and modeling data with the aim of finding useful information, drawing conclusions, and supporting decision making [13]. In this study, data analysis was conducted to transform raw data of toddlers obtained from the Posyandu in Desa Perkebunan Petatal into information that has strategic value in identifying the risk level of stunting. The data were summarized into several variables presented in the following Table 1.

Table 1. Variable

No	Name	Variable
1	Age	K1
2	Body Weight	K2
3	Height	K3
4	BBL	K4
5	ASI	K5
6	Infection	K6

After determining the variables, the next step is to determine the output data. This output data serves as supporting information for the village authorities and health workers in knowing the risk level of stunting in toddlers based on the growth data patterns analyzed [14]. The data processing results are presented in the form of grouping (clustering) into three risk categories as shown in the following Table 2.

Table 2. Output Category

No	Cluster Name	Description
1	Cluster 1	High Risk of Stunting
2	Cluster 2	Midle Risk of Stunting
3	Cluster 3	Low Risk of Stunting

In the initial condition, before applying the K-Means Clustering method, the analysis of stunting risk in toddlers was conducted manually based on general observations of nutritional indicators such as weight, height, and age. This approach made it difficult to identify specific patterns or group toddlers based on similar risk characteristics, resulting in less accurate and less targeted intervention decisions. Health workers were only able to categorize toddlers broadly without a clear grouping structure supported by data analysis.

After applying the K-Means Clustering method, the analysis process becomes more systematic and data-driven [15]. The algorithm is able to group toddler data into three distinct clusters based on the similarity of their attributes, namely high-risk, medium-risk,

and low-risk groups. This clustering process allows for clearer identification of patterns in the data, where toddlers with similar nutritional and health characteristics are grouped together automatically.

The comparison between the two approaches shows that the use of K-Means Clustering provides a more objective and structured analysis compared to manual methods [16]. It not only simplifies the process of identifying risk groups but also improves the accuracy of classification, enabling health workers to prioritize interventions more effectively [17]. Therefore, the implementation of K-Means Clustering significantly enhances decision-making in stunting risk management by transforming raw data into meaningful information [18].

B. Design System

After determining the results of the program to be created, the next stage is system design. System design uses a use case diagram with the aim of being able to map system requirements in detail [19]. The use case diagram design is presented in the following figure 1.

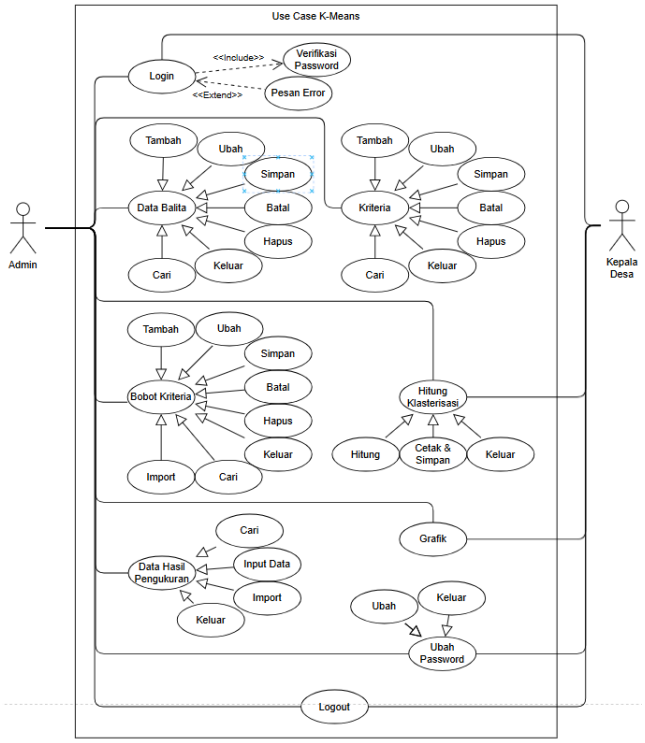

Figure 1. Use Case Diagram

C. Implementation System

After designing the system, the next step is to implement the design into the PHP programming language and MySQL database [20].

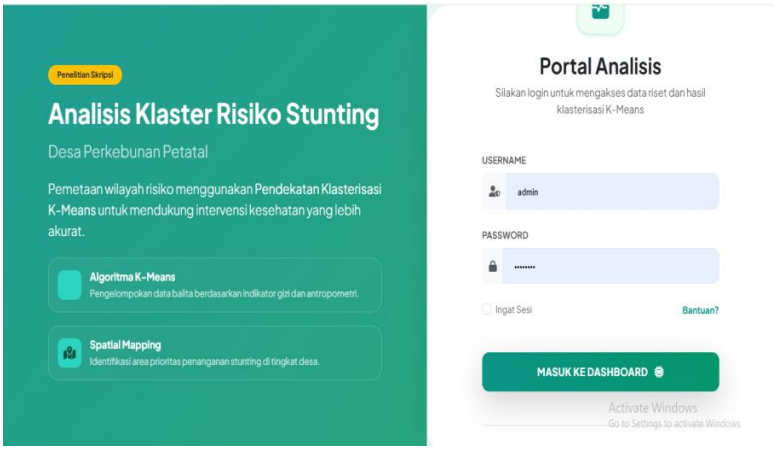

Figure 2. Login Page

After logging in, the admin will be directed to the main menu. In this K-Means system, there are several menus such as Toddler Data, Nutrition Criteria, Anthropometric Assessment, Clustering Process, and Reports.

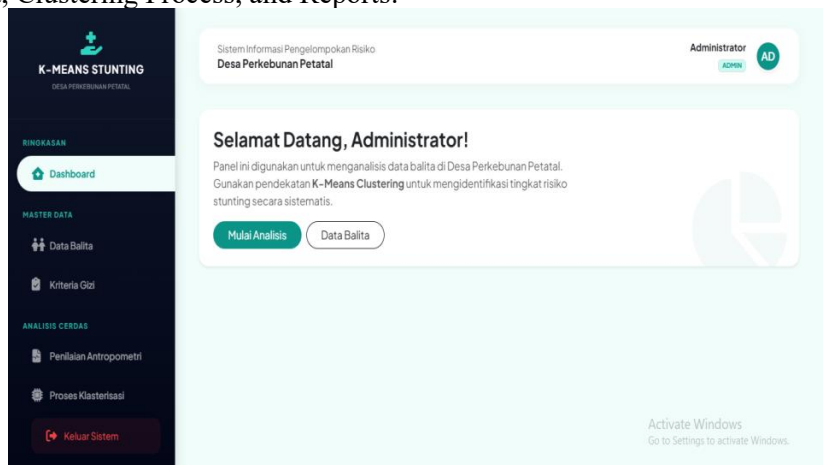

Figure 3. Admin Home Page

After that, the admin enters toddler data in the Toddler Data Form menu where this data is used to manage the toddler data that becomes the object of stunting risk analysis.

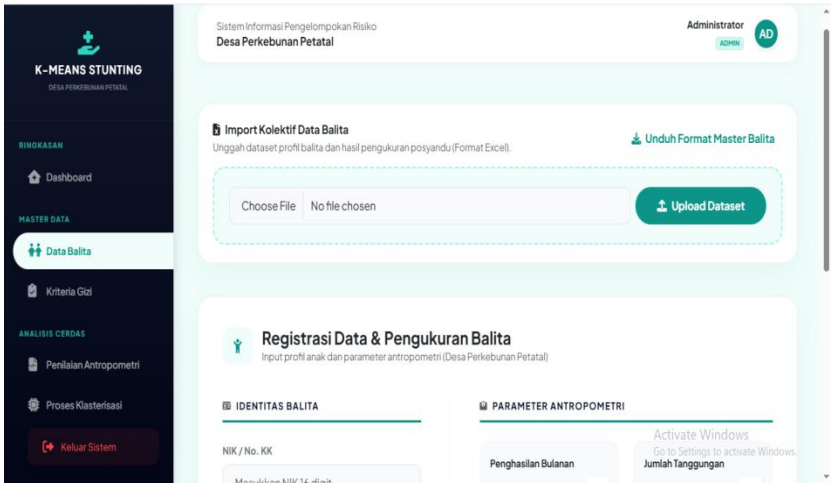

Figure 4. Toddler Form Data

Next, the admin enters the Criteria Data in the Criteria Data menu, which is used to manage the criteria. These criteria will be connected in the calculation of variables.

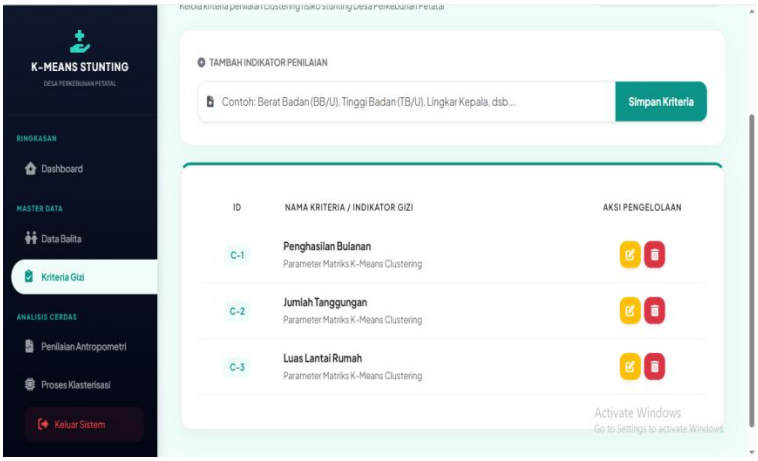

Figure 5. Nutritional Criteria

After entering the data, the next step is to carry out Anthropometric measurements in the Anthropometric Measurement menu.

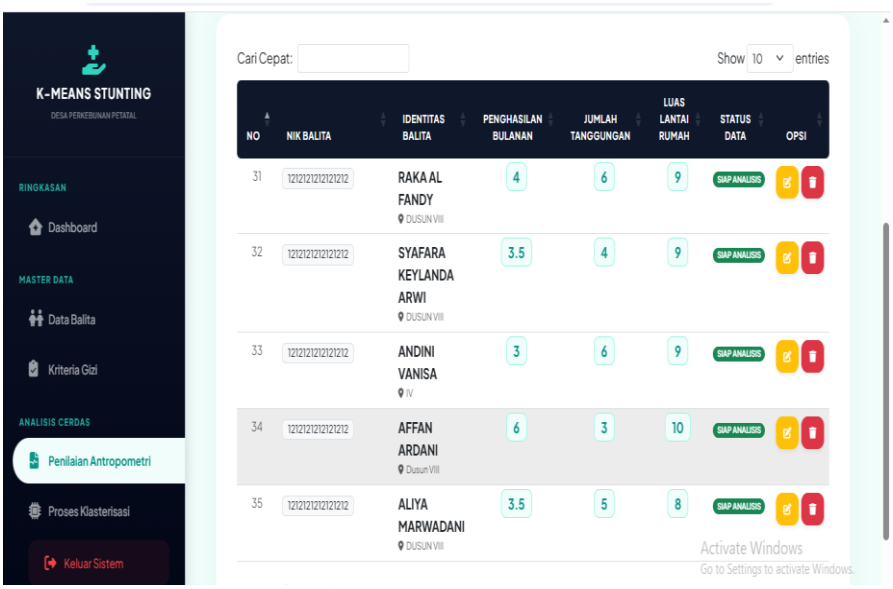

Figure 6. Anthropometric Measurement

Next, in the Clustering menu, it will calculate using the K-Means Clustering method to determine what clusters are produced from the data that has been previously inputted.

Nama Balita	Jarak ke Centroid (D)			Cluster Hasil		
	D1	D2	D3	C1	C2	C3
label2 Contoh	0.8441	0.5403	0.9719	-	✓	-
adfasdf	1.3483	0.7192	0.1944	-	-	✓
asdf	1.3316	0.7198	0.1959	-	-	✓
adfasdf	0.9531	0.3409	0.4498	-	✓	-
4 Tahun	0.2046	1.2399	1.4655	✓	-	-
5 Bulan	0.2291	0.9567	1.2222	✓	-	-
4 Bulan	0.2139	0.9485	1.1775	✓	-	-
3 Tahun	0.0752	1.1446	1.3528	✓	-	-
4 Tahun	0.2138	1.2482	1.4346	✓	-	-
4 Tahun	0.1348	1.1870	1.3820	✓	-	-
1 Tahun	0.0783	1.0358	1.2482	✓	-	-

Figure 7. Clustering Results

Figure 7 shows the clustering results obtained using the K-Means method, where the data are grouped into several clusters based on their similarity. Each cluster represents a group of data with similar characteristics, allowing for easier interpretation and analysis.

Based on the clustering results, it can be observed that each cluster has distinct characteristics that differentiate it from others. For instance, one cluster represents data with higher values, indicating a group with higher intensity or priority, while another cluster represents data with lower values. This grouping helps to identify patterns within the data that were previously difficult to observe using conventional methods.

Based on these results, several actions can be recommended. Data belonging to clusters with higher values can be prioritized for decision-making, such as targeted strategies, resource allocation, or promotional activities. Meanwhile, clusters with lower values can be further analyzed to identify potential improvements or optimization opportunities. Therefore, the clustering results provide valuable insights that support more effective and data-driven decision-making.

IV. Conclusion

Based on the results of research conducted on the Cluster Analysis of Stunting Risk in the Petatal Plantation Village using the K-Means Clustering algorithm, several conclusions can be drawn as follows: 1) The Web-Based Stunting Risk Analysis Information System was successfully designed and implemented using the PHP programming language and MySQL database. This system is capable of managing toddler data in a computerized manner, from the input, storage, to the presentation of analysis report results. 2) The clustering process is carried out using the variables of age, weight, and height with three clusters (k = 3). The system is able to perform centroid initialization, Euclidean distance calculation, data grouping, and iteration until reaching a convergent condition automatically, and 3) The results of grouping toddlers into the High-Risk category, Medium Risk and Low Risk can help healthcare workers and village authorities in determining intervention priorities. Thus, this system can serve as a decision-making tool in the stunting prevention program in Petatal Plantation Village.

References

[1] Sulistiani *et al.*, "Data Analysis of Student Monitoring Using the K-Means Clustering Method," *Indones. J. Mod. Sci. Technol.*, vol. 1, no. 2, pp. 50–57, 2025, doi: 10.64021/ijmst.1.2.50-57.2025.

[2] E. A. Wibowo and R. Aryanti, "Penerapan Metode Clustering K-Means Menggunakan RapidMiner untuk Klasifikasi Prestasi Siswa di Sekolah Swasta," *J. Inf. Technol. Informatics Eng.*, vol. 1, no. 1, pp. 20–24, 2025, [Online]. Available:

https://journal.jci.co.id/jitie/article/view/35

[3] I. Fazira and Z. F. Risawandi, "Klasterisasi Tingkat Kelayakan Provinsi Dalam Pembangunan Kawasan Industri Menggunakan Algoritma K-Means," vol. 10, no. 2, pp. 874–886, 2025.

[4] J. N. Iin, "Perbandingan Kmeans Dan Hierarchical Clustering Untuk Pemetaan Kawasan Rawan Stunting Di Kabupaten Kolaka," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 701–716, 2025.

[5] M. Fadhil, W. Fuadi, and Maryana, "Clustering Tingkat Kecanduan Game Mobile Legends Terhadap Keharmonisan Keluarga Menggunakan Metode K-Means," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 981–990, 2025.

[6] R. Rahmawati, W. Prihartono, and K. Cirebon, "Optimasi Stok Dengan Clustering Data Transaksi Penjualan Menggunakan Algoritma K-Means di Konter Agung Cell," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 2, 2025.

[7] Hasim Azari, Dwi Hartanti, and Aprilisa Arum Sari, "Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 379–389, 2024, doi: 10.29408/jit.v7i2.26016.

[8] J. Homepage *et al.*, "Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan," vol. 5, no. 1, pp. 350–358, 2025.

[9] A. Alawiyah, N. Aghnia, and F. F. Abdalah, "Implementasi Clustering Algoritma K-Means Pada Penjualan Beras Di CV Tangguh Bumi Perkasa," *J. Komisi (Jurnal Komput. dan Sist. Informasi)*, vol. 2, no. 2, pp. 17–23, 2025.

[10] M. Adelina Bui and A. Bahtiar, "Implementasi Metode Algoritma K-Means Clustering Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1451–1456, 2024, doi: 10.36040/jati.v8i2.8975.

[11] N. Bili, R. T. Abineno, and A. A. Pekuwali, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Peforma Siswa Pada Pembelajaran Bahasa Indonesia," in *Prosiding Seminar Nasional SATI*, 2024, pp. 523–537.

[12] W. Usino, "Klasterisasi Tingkat Kelayakan Provinsi Dalam Pembangunan Kawasan Industri Menggunakan Algoritma K-Means," vol. 3, no. September, pp. 324–333, 2024.

[13] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, doi: 10.58602/jima-ilkom.v3i1.26.

[14] D. D. Susilo, "Implementasi Data Mining Pada Pengelompokkan Data Pembeli Menggunakan Algoritma K-Means Pada PT. Otomotif," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, 2024.

[15] V. Afifah and D. Setyantoro, "Rancangan Sistem Pemilihan dan Penetapan Harga dalam Proses Pengadaan Barang dan Jasa Logistik Berbasis Web," *J. IKRA-ITH Inform.*, vol. 5, no. 2, pp. 108–117, 2021.

[16] S. Wijayanto and M. Yoka Fathoni, "Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jupiter*, vol. 13, no. 2, pp. 212–219, 2021.

[17] D. Alfatah, "Application of the K-Means Clustering Algorithm in Mapping the Regional Voter Strategy for the Legislative Candidates for the DPR RI," *J. Komputer, Inf. dan Teknol.*, vol. 1, no. 2, pp. 435–443, 2021, doi: 10.53697/jkomitek.v1i2.430.

[18] G. Fajriansyah, "Analisis Daftar Pemilih Tetap Pada Hasil Rekapitulasi Kpu Berdasarkan Usia Menggunakan Algoritma K-Means (Studi Kasus : Kota Bandar Lampung)," *Electr. - J. Rekayasa dan Teknol. Elektro*, vol. 15, no. 1, pp. 39–53, 2021, doi: 10.23960/elc.v15n1.2147.

[19] R. Farismana, "Penerapan K-Means Clustering Untuk Pemetaan Produktivitas Padi

Dan Prediksi Panen Di Kabupaten Indramayu," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 8, no. 3, p. 589, 2024, doi: 10.52362/jisamar.v8i3.1572.

[20] R. Bayu Pramana, "Potensi Masalah Pilkada Serentak 2020 dan Kekhawatiran Masyarakat Pemilih Muda," *KEMUDI J. Ilmu Pemerintah.*, vol. 5, no. 02, pp. 167–178, 2021, doi: 10.31629/kemudi.v5i02.2773.