Mapping Voter Segmentation Based on Demographic Characteristics Using the K-Means Clustering Method

Dedy Rahman Harahap, Rian Farta Wijaya

Abstract

This study aims to identify voter demographic patterns through voter segmentation using the K-Means Clustering algorithm based on Permanent Voter List (Daftar Pemilih Tetap/DPT) data from Pematangsiantar, North Sumatra, Indonesia. The research adopts a quantitative approach to analyze demographic characteristics and identify similarities among voters across administrative areas. The dataset consists of 75,078 valid voter records collected from 20 Excel workbooks representing 20 urban villages across eight districts. The demographic attributes include the number of registered voters, average age, gender composition, and age-group distribution. The research methodology consists of data collection, data preprocessing, feature selection, data transformation, standardization, K-Means implementation, and cluster evaluation. The optimal number of clusters is determined using the Silhouette Score by comparing different clustering configurations. The experimental results indicate that two clusters provide the optimal segmentation, achieving a Silhouette Score of 0.4893. The resulting clusters represent different demographic profiles among the analyzed urban villages, particularly in terms of voter population, age composition, and gender distribution. Furthermore, the visualization results demonstrate demographic variations across administrative areas and provide a structured overview of voter characteristics. These findings indicate that K-Means Clustering can transform administrative electoral records into meaningful demographic information that supports regional voter analysis and the planning of voter education activities. The study also provides an analytical foundation for understanding voter distribution based on demographic similarities. However, the findings are limited to the available DPT records and do not directly measure actual voter turnout, political preferences, or individual voting behavior. Future research may incorporate additional electoral datasets and alternative clustering techniques to improve segmentation analysis and provide more comprehensive insights.
..

***Keywords**:* Voter Segmentation, Demographic Characteristics, K-Means Clustering, Permanent Voter List, Electoral Data.

[1]Dedy Rahman Harahap
[1]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: ekyr422@gmail.com [1]

[2]Rian Farta Wijaya
[2]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: rianfartawijaya@dosen.pancabudi.ac.id [2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

The development of information technology has provided opportunities for utilizing electoral data as a source of information to support data-driven decision-making. One important aspect of electoral data management is the analysis of voter characteristics based on demographic information. Voter segmentation can help identify population patterns, understand the distribution of registered voters, and provide a more structured overview of the demographic composition of an electoral area. In Indonesia, the Permanent Voter List (Daftar Pemilih Tetap/DPT) is an important source of electoral information that can be further analyzed using data mining techniques. The utilization of this data is particularly relevant for improving the effectiveness of electoral data analysis and supporting voter-related planning [1] [2]

In the context of Pematangsiantar, North Sumatra, the availability of Permanent Voter List data across different districts and urban villages provides an opportunity to conduct demographic-based voter segmentation. The dataset used in this study consists of electoral records organized by administrative areas, covering eight districts (*kecamatan*) and their respective urban villages (*kelurahan*). Such data can be utilized to identify similarities and differences in voter characteristics across geographical areas. However, voter data that is primarily maintained for administrative and electoral purposes requires further analytical processing to generate meaningful information about demographic distribution and voter composition [3] [4] .

One method that can be applied to address this problem is K-Means Clustering, an unsupervised machine learning algorithm that groups data based on similarities among selected attributes. The algorithm works by determining a number of cluster centers (*centroids*) and assigning each data point to the nearest centroid based on distance calculations. Through an iterative process, K-Means updates the cluster centers until the grouping results become relatively stable. This method is suitable for demographic segmentation because it can identify groups of areas or aggregated voter profiles with similar characteristics without requiring predefined class labels. The resulting clusters can subsequently be interpreted to describe different demographic patterns across electoral areas [5] [6] .

Based on these considerations, this study proposes "Mapping Voter Segmentation Based on Demographic Characteristics Using the K-Means Clustering Method", focusing on Permanent Voter List data from Pematangsiantar. The research aims to group voter data based on relevant demographic attributes, identify the characteristics of each cluster, and analyze the distribution of voter segments across administrative areas. The clustering results are expected to provide information regarding demographic patterns and the concentration of voter characteristics, which may support electoral data analysis and voter education planning. Furthermore, this study is expected to demonstrate the application of K-Means Clustering as a data mining approach for transforming administrative electoral data into more informative analytical insights [7] [8] .

Important methodological note: Since DPT data may contain personal information, the analysis should exclude names, NIK, full addresses, and other personally identifiable information. Use anonymized or aggregated demographic attributes for clustering to ensure data privacy [9] .

Literature Review

Voter segmentation is an analytical approach used to group voters based on similarities in demographic characteristics, such as age, gender, and geographical distribution. Demographic analysis plays an important role in understanding the composition and distribution of registered voters within an electoral area. In Indonesia, the Permanent Voter List (Daftar Pemilih Tetap/DPT) serves as an important source of information for identifying voter characteristics and supporting electoral administration. Through systematic data analysis, voter records can be transformed into meaningful information that describes demographic patterns across different administrative regions [11] [12] .

K-Means Clustering is an unsupervised machine learning algorithm widely used for grouping datasets based on similarities between observations. The algorithm partitions data into a predefined number of clusters by calculating the distance between data points and cluster centroids. The clustering process involves selecting the number of clusters (K), initializing centroids, assigning observations to the nearest centroid, and updating centroid positions iteratively until the clustering results converge. The effectiveness of K-Means depends on data quality, feature selection, normalization, and the determination of an appropriate number of clusters. Therefore, preprocessing and cluster evaluation are essential to obtain reliable segmentation results [14] [15] .

Previous studies have demonstrated the application of K-Means Clustering in demographic analysis, population segmentation, and regional data classification. These applications indicate that clustering can reveal hidden patterns and similarities that may not be easily identified through conventional descriptive analysis. In this study, K-Means Clustering is applied to Permanent Voter List data from Pematangsiantar to identify voter segments based on demographic characteristics and their distribution across administrative areas. The research focuses on developing demographic profiles for each cluster to provide a clearer understanding of voter composition and support data-driven electoral planning [16].

Research Methodology

This study employs a quantitative research approach using data mining techniques to perform voter segmentation based on demographic characteristics. The research data are obtained from the Permanent Voter List (Daftar Pemilih Tetap/DPT) provided by the General Elections Commission (KPU) in Pematangsiantar, North Sumatra, Indonesia. The dataset contains demographic information organized by administrative areas, including districts and urban villages. The research focuses on identifying demographic patterns and grouping voter characteristics based on relevant attributes while ensuring that personally identifiable information is excluded from the analysis.

The research process consists of several stages, including data collection, data preprocessing, feature selection, data transformation, and K-Means Clustering implementation. Data preprocessing involves checking data completeness, removing duplicate records, handling missing values, and selecting relevant demographic variables such as age, gender, and regional distribution. Numerical attributes are normalized to ensure that differences in measurement scales do not influence the clustering process. Subsequently, the K-Means algorithm is implemented by determining the number of clusters (K), initializing centroids, calculating distances using the Euclidean Distance method, and assigning observations to their nearest cluster centers until the clustering process converges.

The final stage involves evaluating and interpreting the clustering results to identify the demographic characteristics of each voter segment. The clustering performance is assessed

using the **Silhouette Score** to measure the quality of cluster separation and cohesion. The resulting clusters are then analyzed based on their demographic profiles and distribution across administrative areas in Pematangsiantar. The findings are expected to provide a structured representation of voter segmentation and support a better understanding of demographic patterns for electoral data analysis and voter education planning.

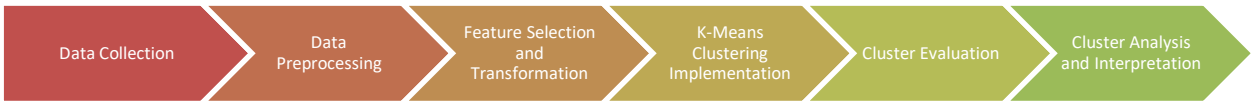

Figure 1. Research Methodology

1. Data Collection
   The initial stage involves collecting Permanent Voter List data from KPU Pematangsiantar. The dataset contains demographic information, including voter age, gender, and administrative area. These data serve as the primary source for identifying demographic characteristics and developing voter segmentation.
2. Data Preprocessing
   The collected data are processed to ensure consistency and quality before analysis. This stage includes removing duplicate records, handling missing values, correcting inconsistent formats, and excluding personally identifiable information. The objective is to produce a clean and structured dataset.
3. Feature Selection and Transformation
   Relevant demographic attributes are selected based on their contribution to voter segmentation. Variables such as age, gender, and regional distribution are prepared for the clustering process. Categorical variables are encoded into numerical representations, while numerical attributes are normalized to ensure balanced contributions during distance calculations.
4. K-Means Clustering Implementation
   The K-Means algorithm is applied to group voters based on similarities in their demographic characteristics. The process involves determining the number of clusters (K), initializing centroids, calculating Euclidean distances, assigning data points to the nearest centroids, and updating cluster centers iteratively until convergence is achieved.
5. Cluster Evaluation
   The resulting clusters are evaluated to determine their quality and separation. The Silhouette Score is used to measure the similarity of observations within the same cluster compared to observations in other clusters. The evaluation results help determine whether the selected number of clusters provides meaningful demographic segmentation.
6. Cluster Analysis and Interpretation
   The final stage involves analyzing the characteristics of each cluster based on demographic attributes and administrative distribution. Each cluster is interpreted according to its dominant characteristics, such as age composition, gender distribution, and regional concentration. The results are presented through tables and visualizations to provide an informative overview of voter segmentation and support electoral data analysis and voter education planning.

Results

This section presents the empirical results of voter segmentation based on demographic characteristics using the K-Means Clustering algorithm. The analysis was conducted on 75,078 valid voter records extracted from 20 uploaded Excel workbooks, representing 8 districts and 20 urban villages in Pematangsiantar, North Sumatra. To ensure that the clustering process

describes area-level demographic patterns rather than individual voting behavior, the urban village was used as the unit of analysis. Each observation therefore represents an aggregated demographic profile of one urban village. The features included the number of registered voters, average age, male-voter proportion, and the proportions of six age groups. Direct personal identifiers were not included in the analytical features.

The analysis began by consolidating the voter records from the available workbooks and validating the age and gender fields. Records were retained when the gender field contained a valid code (L or P) and the recorded age was between 17 and 110 years. After validation, individual-level records were aggregated by urban village to calculate the demographic indicators required for clustering. Numerical features were standardized before the K-Means algorithm was applied, because the variables have different scales and units. The optimal number of clusters was examined by comparing Silhouette Scores for K values ranging from 2 to 6.

Table 1. Summary of the Analyzed Dataset

Dataset component	Result
Valid voter records	75,078
Excel workbooks processed	20
Districts represented	8
Urban villages represented	20
Clustering unit	Aggregated urban-village profile
Demographic features used	9
Age range included	17 years and above
Clustering algorithm	K-Means

Table 1 summarizes the scope and structure of the dataset used in this study. A total of 75,078 valid voter records were extracted from the 20 uploaded workbooks and organized into 20 urban-village profiles across eight districts. The number of records refers to rows successfully read from the submitted files after applying the age and gender validity checks. Accordingly, this figure describes the dataset provided for the analysis and should not be interpreted as confirmation that the uploaded files represent the complete Permanent Voter List for every administrative area in Pematangsiantar. The use of aggregated village profiles also means that the clustering results describe similarities between areas, not classifications of individual voters.

Table 2. Distribution of Voters by Gender and Age Group

Demographic category	Number of voters	Percentage
Male (L)	36,038	48.00%
Female (P)	39,040	52.00%
Age 17–25	15,862	21.13%
Age 26–35	14,969	19.94%
Age 36–45	15,057	20.06%
Age 46–55	11,873	15.81%
Age 56–65	9,481	12.63%
Age >65	7,836	10.44%
Total	75,078	100.00%

Table 2 describes the demographic composition of the valid voter records according to gender and age group. The dataset contains 36,038 male voters (48.00%) and 39,040 female voters (52.00%). The largest age category is 17–25 years, accounting for 15,862 records (21.13%). The overall mean age is 41.86 years. These descriptive statistics provide an initial

picture of the demographic structure represented in the submitted files. However, the gender and age distributions should be interpreted as characteristics of registered voters in the dataset, not as measures of voter attendance or electoral participation.

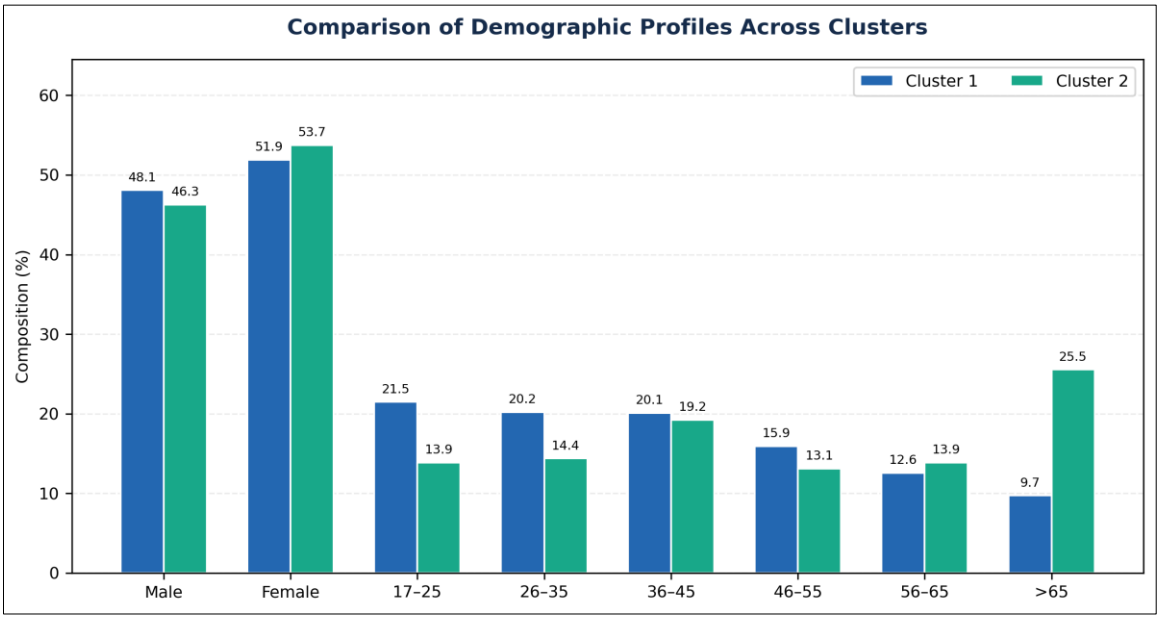

Figure 1. Comparison of Demographic Profiles Across Clusters

Figure 1 compares the demographic composition of the two clusters using the percentage of male and female voters and the percentage distribution across six age categories. Each bar represents the proportion of a demographic category among all voter records belonging to the relevant cluster. This percentage-based presentation allows the clusters to be compared even though they contain different numbers of urban villages and voter records. The chart provides a direct view of the demographic dimensions that may distinguish one cluster from another. It should be read together with the numerical cluster profile in Table 5, because percentage composition alone does not describe the total size of each cluster.

Table 3. Voter Characteristics by Urban Village

District	Urban village	Voter records	Mean age	Male	Female	Cluster
Siantar Barat	Dwikora	2,041	49.98	46.35%	53.65%	Cluster 2
Siantar Barat	Proklamasi	1,373	47.63	46.10%	53.90%	Cluster 2
Siantar Barat	Simarito	4,977	41.61	48.80%	51.20%	Cluster 1
Siantar Marihat	Pardamean	2,736	41.07	49.23%	50.77%	Cluster 1
Siantar Marihat	Sukamaju	3,061	42.91	47.63%	52.37%	Cluster 1
Siantar Marihat	Sukamakmur	1,085	41.98	46.64%	53.36%	Cluster 1
Siantar Marimbun	Naga Huta	2,965	41.06	47.59%	52.41%	Cluster 1
Siantar Marimbun	Naga Huta Timur	1,745	41.11	47.85%	52.15%	Cluster 1

Siantar Martoba	Naga Pitu	3,701	40.77	49.26%	50.74%	Cluster 1
Siantar Martoba	Pondok Sayur	5,925	40.62	49.05%	50.95%	Cluster 1
Siantar Martoba	Tambun Nabolon	5,422	39.29	48.40%	51.60%	Cluster 1
Siantar Selatan	Karo	2,987	44.24	45.70%	54.30%	Cluster 1
Siantar Selatan	Toba	1,579	43.77	44.71%	55.29%	Cluster 1
Siantar Sitalasari	Bah Sorma	3,170	41.25	51.58%	48.42%	Cluster 1
Siantar Sitalasari	Gurilla	1,933	40.81	49.20%	50.80%	Cluster 1
Siantar Sitalasari	Setia Negara	6,230	40.78	43.50%	56.50%	Cluster 1
Siantar Timur	Kebun Sayur	3,285	40.88	48.22%	51.78%	Cluster 1
Siantar Timur	Tomuan	7,173	41.33	47.94%	52.06%	Cluster 1
Siantar Utara	Martoba	7,215	41.31	49.48%	50.52%	Cluster 1
Siantar Utara	Melayu	6,475	44.38	48.80%	51.20%	Cluster 1

Table 3 provides a more detailed view of the demographic characteristics of each urban village. The number of voter records indicates the scale of the submitted DPT data in each area, while mean age and gender proportions describe the demographic composition. The final column shows the cluster assignment produced by K-Means. These details make it possible to trace the cluster membership back to the administrative areas included in the dataset and to examine whether areas assigned to the same cluster share similar demographic patterns. Differences in voter counts should not be interpreted as differences in turnout, since the dataset contains registration records rather than attendance records.

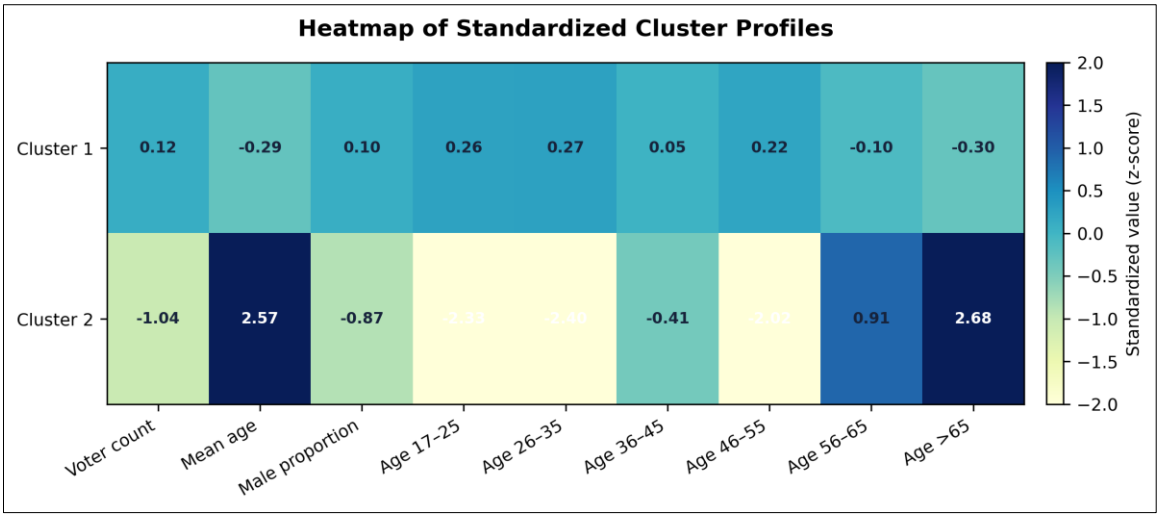

Figure 2. Heatmap of Standardized Cluster Profiles

Figure 2 presents a heatmap of the mean standardized feature values for each cluster. Standardization transforms the original variables into z-scores, allowing features with different measurement scales to be examined within a common visual range. Positive values indicate that the average feature value for a cluster is above the overall mean across the analyzed urban villages, whereas negative values indicate a value below that mean. The heatmap therefore helps identify which characteristics contribute most visibly to the differences between the cluster profiles. For example, the voter-count and age-related features can be compared alongside gender composition without allowing their original measurement scales to dominate the visualization.

Table 4. Silhouette Scores for Different Numbers of Clusters

Number of clusters (K)	Silhouette Score
2	0.4893
3	0.2585
4	0.2172
5	0.2030
6	0.2081

Table 4 reports the Silhouette Scores obtained for K values from 2 to 6. The highest score was obtained at K=2, with a value of 0.4893. The scores for K=3, K=4, K=5, and K=6 were 0.2585, 0.2172, 0.2030, and 0.2081, respectively. Since the two-cluster solution produced the highest score among the tested alternatives, K=2 was selected for the final segmentation. The Silhouette Score evaluates how close each observation is to its own cluster compared with other clusters; a higher value generally indicates more cohesive and better-separated groups. This metric is used here to compare clustering configurations, not to measure predictive accuracy.

Table 5. Demographic Profile of the Final Clusters

Cluster	Urban villages	Voter records	Mean age	Male share	Female share
Cluster 1	18	71,664	41.62	47.98%	52.02%
Cluster 2	2	3,414	48.80	46.23%	53.77%

Cluster 1 consists of 18 urban villages and 71,664 voter records. The average of the village-level mean ages in this cluster is 41.62 years. The average male proportion is 47.98%, while the average female proportion is 52.02%. The urban villages assigned to this cluster are Simarito, Pardamean, Sukamaju, Sukamakmur, Naga Huta, Naga Huta Timur, Naga Pitu, Pondok Sayur, Tambun Nabolon, Karo, Toba, Bah Sorma, Gurilla, Setia Negara, Kebun Sayur, Tomuan, Martoba, Melayu. These values summarize the demographic profile of the areas grouped together by the algorithm. The cluster label itself is an algorithmic identifier and does not imply a ranking, quality level, or category of political participation.

Cluster 2 consists of 2 urban villages and 3,414 voter records. The average of the village-level mean ages in this cluster is 48.80 years. The average male proportion is 46.23%, while the average female proportion is 53.77%. The urban villages assigned to this cluster are Dwikora, Proklamasi. These values summarize the demographic profile of the areas grouped together by the algorithm. The cluster label itself is an algorithmic identifier and does not imply a ranking, quality level, or category of political participation.

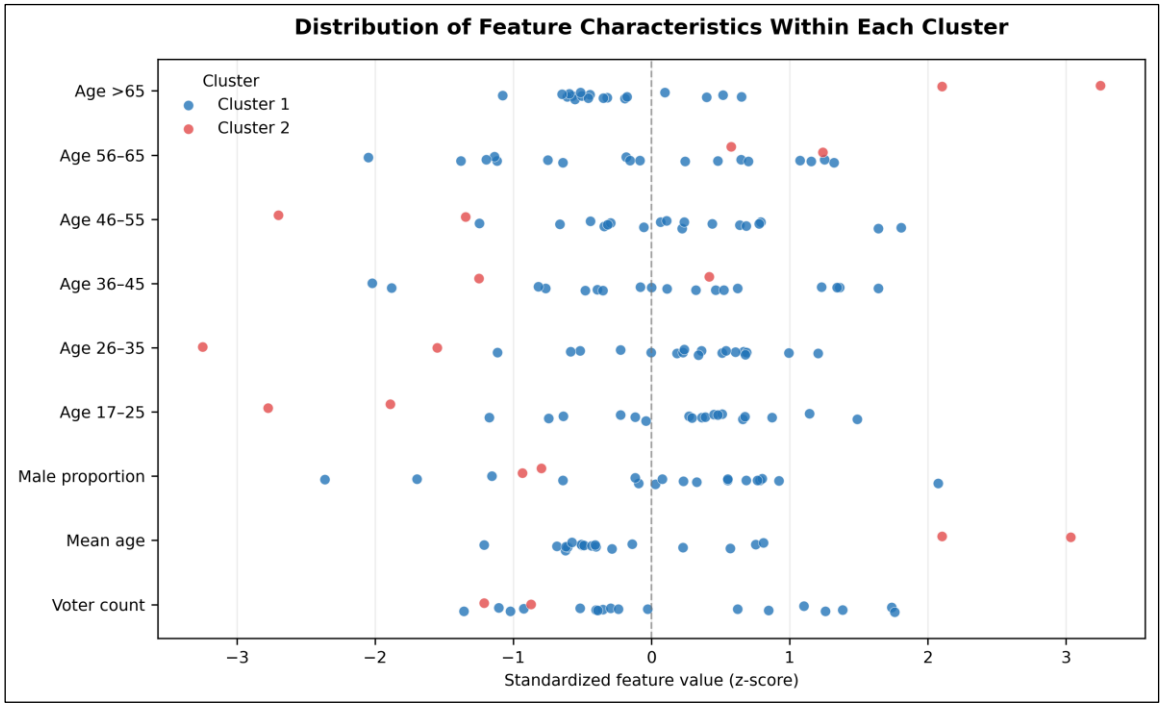

Figure 3. Distribution of Feature Characteristics Within Each Cluster

Figure 3 displays the distribution of standardized feature values for the urban villages assigned to each cluster. Each point represents one urban village, and the horizontal position indicates the standardized value of a particular feature. The dashed vertical line at zero represents the overall mean of that feature before standardization. Points located to the right of zero indicate above-average values, while points to the left indicate below-average values. This plot complements the heatmap by showing the spread of individual village profiles within each cluster, rather than presenting only the cluster-level averages. It is a feature-distribution visualization and should not be interpreted as a SHAP plot or as evidence of causal influence.

Taken together, the results indicate that K-Means divided the 20 urban-village profiles into two groups based on the combined pattern of voter-record volume, mean age, gender composition, and age-group proportions. The resulting groups provide a simplified representation of demographic similarities across the areas included in the submitted files. The comparison of cluster profiles suggests that segmentation should be interpreted through the combined set of features rather than through a single variable such as voter count or gender. This is important because two areas with similar numbers of registered voters may still have different age structures, while areas with different population sizes may share similar demographic proportions.

From an applied perspective, the segmentation can support descriptive electoral data analysis and help the relevant institution understand how demographic composition varies across administrative areas. Such information may be useful as an initial reference when preparing general voter-information or voter-education activities that take local demographic profiles into account. Nevertheless, the DPT dataset records voter registration and demographic attributes; it does not record whether an individual attended the polling station or cast a ballot. Therefore, the present findings cannot be used to infer actual voter turnout, political preferences, candidate support, or individual voting behavior. Any analysis of participation rates would require additional official records on the number of registered voters and the number of voters who exercised their voting rights.

The findings are also subject to data-coverage and methodological limitations. The analysis is based on the 20 urban-village workbooks supplied for this study, and citywide generalization requires confirmation that the complete set of relevant DPT files has been

included. Furthermore, the clustering unit is the urban village, resulting in 20 observations for the K-Means model. Although the Silhouette Score supports the selection of two clusters among the tested alternatives, the interpretation of the groups remains dependent on the selected features, the preprocessing procedure, and the submitted dataset. Future analysis may include additional verified administrative files and other anonymized area-level demographic indicators to provide a broader description of voter characteristics.

Conclusion

This study successfully applied the K-Means Clustering algorithm to segment voter characteristics based on demographic data from the Permanent Voter List (DPT) in Pematangsiantar, North Sumatra. The analysis involved 75,078 valid voter records from 20 urban villages, using demographic attributes consisting of voter population, average age, gender composition, and age-group distribution. Based on the Silhouette Score evaluation, the optimal number of clusters was identified as two clusters, achieving a score of 0.4893. The clustering results demonstrate that voter characteristics vary across administrative areas, providing a structured representation of demographic similarities and differences among the analyzed urban villages.

The findings indicate that K-Means Clustering can be utilized as an effective analytical approach for identifying demographic patterns and supporting area-based voter segmentation. The resulting clusters can provide useful information for electoral data analysis and the planning of voter education activities based on demographic characteristics. However, this study is limited to the available DPT dataset and does not measure actual voter turnout or individual voting behavior. Future research is recommended to incorporate more comprehensive electoral datasets, additional demographic variables, and alternative clustering algorithms to improve segmentation quality and provide broader insights into voter characteristics.

References

[1] H. W. Dhany, F. Izhari, and S. Komputer, “Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN),” 2024. [Online]. Available: www.ejournal.marqchainstitute.or.id/index.php/JICT

[2] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, “Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm,” *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[3] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, and A. Ahyanuardi, “Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment,” *JURNAL INFOTEL*, vol. 15, no. 2, pp. 55–61, Jun. 2023, doi: 10.20895/infotel.v15i2.934.

[4] M. S. Novelan and S. Aryza, “OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD

DISTRIBUTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 917–925, Jun. 2025, doi: 10.33480/jitk.v10i4.6467.

[5] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[6] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, and K. B. Seminar, "The Smart Governance Framework and Enterprise System's Capability for Improving Bio-Business Licensing Services," *Informatics*, vol. 10, no. 2, Jun. 2023, doi: 10.3390/informatics10020053.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, and P. Wanny, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.

[8] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA," *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[9] M. Iqbal and S. Efendi, "Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 1, pp. 11–20, May 2023, doi: 10.21512/comtech.v14i1.8243.

[10] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions," Jul. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9071177.

[11] W. Cetera, W. Gogołek, A. Żołnierski, and D. Jaruga, "Potential for the use of large unstructured data resources by public innovation support institutions," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00610-6.

[12] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100719.

[13] A. Rejeb *et al.*, "The Internet of Things (IoT) in healthcare: Taking stock and moving forward," Jul. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.iot.2023.100721.

[14] M. Overton, S. Larson, L. J. Carlson, and S. Kleinschmit, "Public data primacy: the changing landscape of public service delivery as big data gets bigger," *Global Public Policy and Governance*, vol. 2, no. 4, pp. 381–399, Dec. 2022, doi: 10.1007/s43508-022-00052-z.

[15] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, and R. E. Putera, "JIEE: Effectiveness of the OSSRBA (Online Single Submission Risk Based Approach) System in Business Licensing Services in Payakumbuh City DPMPTSP," vol. 3, pp. 22–40, 2023.

[16] W. Jin, "Challenges and Innovative Countermeasures Faced by Public Administration in the Context of Big Data and Internet of Things," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8949365.