



Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 1, No. 2, 2023



A Statistical Clustering Approach: Mapping Population Indicators Through Probabilistic Analysis in Aceh Province, Indonesia

Novi Reandy Sasmita^{1,*}, Moh Khairul¹, Hizir Sofyan¹, Rumaisa Kruba¹, Selvi Mardalena¹, Arriz Dahlawy¹, Feby Apriliansyah², Muliadi Muliadi², Dimas Chaerul Ekty Saputra³, Teuku Rizky Noviandy⁴ and Ahmad Watsiq Maula⁵

¹ Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah, Banda Aceh 23111, Indonesia: novireandys@usk.ac.id (N.R.S); mhd_khairu@mhs.unsyiah.ac.id (M.K); hizir@usk.ac.id (H.Z); rumaisakruba@usk.ac.id (R.K); selvimardalena@usk.ac.id (S.M); arriz_d20@mhs.unsyiah.ac.id (A.D);

² Department of Regional and Urban Planning, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia: feby.aprilians@gmail.com (F.A); muliadi_abi@usk.ac.id (M.M)

³ Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand: dimaschaerulekty.s@kkumail.com (D.C.E.S)

⁴ Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia: trizkynoviandy@gmail.com (T.R.N)

⁵ Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine Public Health and Nursing, Universitas Gadjah Mada, Sleman 55281, Indonesia: ahmad.watsiq.m@ugm.ac.id (A.W.M)

* Correspondence: novireandys@usk.ac.id

Article History

Received 26 October 2023
 Revised 10 December 2023
 Accepted 19 December 2023
 Available Online 24 December 2023

Keywords:

Population indicator
 Probabilistic Fuzzy C-Means
 Spearman test
 Aceh Province
 Shapiro Wilk test

Abstract

The clustering, one of statistical analysis, can be used for understanding population patterns and as a basis for more targeted policy making. In this ecological study, we explored the population dynamics across 23 districts/cities in Aceh Province. The study used the Aceh Population Development Profile Year 2022 data, focusing on the total population, in-migrants, out-migrants, fertility, and maternal mortality as variables. The study employed descriptive statistics to ascertain the data distribution, followed by the Shapiro-Wilk test to evaluate normality, which is crucial for selecting the appropriate statistical methods. The Spearman test was used to determine correlations between the total population and the variable as indicators. Probabilistic Fuzzy C-Means (PFCM) method is used for clustering. To optimize clustering, the silhouette coefficient was calculated using the Euclidean Distance and the elbow method, with the results analyzed using R-4.3.2 software. This study's design and methods aim to provide a nuanced understanding of demographic patterns for targeted policy-making and regional development in Aceh, Indonesia. Based on the data normality test results, only fertility (p -value = 0.45), while the other variables are not normally distributed. Spearman test was used, and the results showed that only in-migrants (p -value = 1.78×10^{-6}) and out-migrants (p -value = 2.30×10^{-6}) correlated to the Aceh Province population. Using the population variable and the two variables associated with it, it was found that 4 is the best optimum number of clusters, where clusters 1, 2, 3, and 4 consist of three districts/city, nine districts/city, four districts/city and seven districts/city respectively.



Copyright: © 2023 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License. (<https://creativecommons.org/licenses/by-nc/4.0/>)

1. Introduction

The population of Indonesia has continued to increase annually. These differences are particularly evident when comparing urban and rural areas and between major islands such as Java, Sumatra, Kalimantan, and other regions. Factors such as internal migration, economic growth, government policies, and natural conditions play a role in shaping the country's complex population patterns. According to data from the United Nations, Indonesia had a population of 277 million people by the year 2023, ranking among the top three countries with the highest population in the world [1].

Aceh Province, located at the western tip of Indonesia, experiences unique population dynamics influenced by migration, immigration, birth rates, and mortality. The population of Aceh's 23 districts/cities shows variations that reflect diverse socio-economic, cultural, and geographical conditions. Based on data from the Statistics of Aceh Province (BPS Aceh) in 2023, Aceh Province ranks 13th in terms of population size in Indonesia, with a total of 5,407,855 individuals residing across its various districts and cities [2]. A comprehensive analysis of these population dynamics is important for more effective and sustainable regional development planning.

In the literature, clustering methods such as K-means and Hierarchical Clustering have been widely used for demographic analysis. In the literature, clustering methods such as K-means have been widely used for demographic analysis. Some of the studies applied the K-Means method as a clustering technique on the socio-economic distribution of the community based on demographic population [3]. A study in DKI Jakarta was also conducted to analyze population density clustering [4]. Furthermore, K-means is also applied as a clustering method for clustering population migration data of each sub-district in the Rembang Regency [5].

Another clustering research is the application of Time Series Clustering in Poverty Clustering in Indonesia [6]. This study is one of Hierarchical Clustering. Although many studies have applied K-Means and Hierarchical Clustering, both methods have limitations in handling uncertainty and ambiguity in population data [7, 8].

Although many population studies have been conducted using clustering methods, there are still few that focus on clustering that considers calculation factors that have uncertainty values for districts/cities based on population indicators. Fuzzy clustering can be an alternative in clustering analysis [9], especially for population data in Aceh Province. This clustering is important for understanding population patterns and as a basis for

more targeted policy-making. In addition, not many studies have examined the correlation between population indicators such as migration, immigration, birth rates, and maternal mortality rates to population growth in depth.

There is a gap in a study that specifically examines the population in Aceh Province, which is more flexible in handling data uncertainty. Furthermore, previous studies tend to focus on descriptive analyses of the population without comprehensively examining the correlation between various demographic indicators, especially in the context of Aceh, which has unique social and cultural dynamics.

This study introduces a new approach using the Probabilistic Fuzzy C-means (PFCM) method to cluster districts in Aceh based on population indicators. PFCM is one of the fuzzy clustering methods. The reason for using PFCM in this study is that PFCM is an advanced clustering technique that offers a nuanced approach to grouping data by allowing each point to belong to multiple clusters with varying degrees of membership. This method acknowledges the real-world ambiguity and overlap between categories, providing a more flexible and representative model of complex datasets than traditional hard clustering. PFCM's fuzzy approach is especially useful when data points are not distinctly separable into single categories.

Moreover, PFCM enhances robustness to outliers and noise by incorporating a probabilistic membership function, which can assign lower significance to data points that do not fit well into any cluster, thus preserving the integrity of the clustering structure. Its ability to handle diverse cluster shapes and sizes also makes it suitable for a wide array of data types and structures, making it a powerful tool for revealing subtle patterns and relationships within the data. In addition, this study also integrates correlation analysis to explore the relationships between population indicators, providing new insights into population dynamics in Aceh.

This study also aims to explore the relationship between population indicators such as In-migrants, out-migrants, fertility and maternal mortality to population growth in the hope of providing recommendations that can be used in population policy-making and regional development. Furthermore, this study aims to identify and analyze clusters of districts/cities in Aceh based on population indicators using the PFCM method, providing a more comprehensive picture of population dynamics in Aceh Province.

2. Materials and Methods

2.1. Study Design

This study adopted an ecological study design, with 23 districts/cities in Aceh Province as the object of study. The ecological approach was used to analyze the clustering of 23 districts/cities in Aceh Province based on population indicators. An ecological study design in population data is a type of research methodology that examines the relationships between various factors at a group or population level rather than focusing on individuals. This approach is useful for understanding broad trends and patterns in public health, epidemiology, and social sciences [10, 11].

2.2. Variables and Data Source

In addition to the total population of each district/city in Aceh Province, the variables In-migrants, Out-migrants, Fertility, and Maternal mortality were used as population indicators. The data on each variable in this study is sourced from the Aceh Population Development Profile Year 2022 published by the Aceh Population Registration Office [12].

In demographic terms, population refers to the total number of people inhabiting a specific area or region at a given time. It encompasses all individuals living in that area, regardless of age, gender, ethnicity, or any other characteristic. Population studies often include analyses of various factors such as age distribution, birth and death rates, migration patterns, and other aspects that influence the size and characteristics of the population. In-migrants are individuals who enter a specific geographical area, region, or country from another place. This migration can be internal (within the same country) or international (from a different country). In-migrants contribute to the population growth of the area they move to and are a key factor in demographic changes. Their movement is often influenced by factors like employment opportunities, education, better living conditions, or escaping unfavorable conditions in their place of origin

Out-migrants are individuals who leave a specific geographical area, region, or country to move to another place. Like in-migrants, this migration can be either within the same country or to a different country. Out-migration results in a decrease in the population of the area being left and is influenced by various factors, including economic, social, political, or environmental reasons.

Fertility in demographic terms refers to the natural capability of producing offspring. More specifically, it often refers to the number of children born or the

childbearing patterns among a population. The Total Fertility Rate (TFR) is a common measure, representing the average number of children a woman would have during her childbearing years (usually considered to be ages 15-49).

Maternal mortality refers to the death of a woman during pregnancy, childbirth, or within a specific period (usually six weeks) after the end of a pregnancy, from causes related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes. The Maternal Mortality Ratio (MMR) is a key measure, indicating the number of maternal deaths per 100,000 live births, which is used to assess the risk associated with pregnancy and childbirth in different regions and populations.

2.3. Statistical Methods

To assess the data, the study computed descriptive statistics, such as minimum, mean, and maximum values. To further assess the data distribution, the study also computed the standard deviation. Descriptive statistics analysis gives a quick overview of the data distribution by assisting in identifying patterns, trends, and significant statistics [13–15].

Using the Shapiro-Wilk test, the study performed a normality test at the inferential statistics step. Particularly in the distribution's tails, the Shapiro-Wilk test has a tendency to be more sensitive to departures from normality [16–18]. Because of this, even with a small sample size, it is an excellent option if you are worried about finding departures from normality. In statistics, testing for data normality aims to determine if the data has a normal distribution [19–21]. The significance of the study lies in the fact that many statistical techniques presume a normal distribution and the accuracy of the analysis results is increased if this assumption is satisfied [22, 23]. Additionally, normality testing enables the researcher to select the most suitable statistical approach and assists in identifying outliers that could influence the study results. Thus, the conclusions are more credible, and the data are easier to explain [24–26].

Depending on how well the data's normal distribution was performed, either the Pearson or Spearman tests were employed to evaluate the relationship between the total population of each district/city and the population indicators included in this investigation. Finding the association between two or more variables is the goal of the statistical correlation test [27–29]. Correlation tests determine how closely changes in one variable are related to changes in another [30–32]. It contributes to our understanding of whether there is no correlation at all, a positive correlation (one variable increases when

the other does, too), or a negative correlation (one variable increases when the other decreases). Using correlation testing, finding potential links between the variables offers valuable insights into data analysis that can be used for planning, decision-making, or additional research [33].

Furthermore, this study uses PFCM for clustering analysis. PFCM is flexible in handling ambiguous or overlapping data. Unlike traditional clustering, which restricts data to a single group, PFCM allows objects to belong to multiple clusters with varying degrees of membership. This adaptability is invaluable when data division is not clear-cut, making PFCM suitable for dynamic or evolving datasets in fields like bioinformatics, image analysis, and social research.

The application of PFCM in this context fills a methodological gap by offering a more nuanced clustering approach suitable for demographic data and contributes to a more informed and region-specific policy-making process. As regional development increasingly seeks precision and personalization in its strategies, techniques like PFCM could become indispensable in the toolkit of demographers and urban planners alike.

In the clustering process, effective optimization plays a crucial role. To assess the quality of clustering, the silhouette coefficient is used, which involves measuring the distance between data points. This measurement typically employs the Euclidean Distance formula. However, it's important to note that the silhouette coefficient and the elbow method serve different purposes. While the silhouette coefficient evaluates the consistency within clusters of data, the elbow method is used to determine the optimal number of clusters by identifying a point where increasing the number of clusters no longer provides significant value in terms of variance.

These values are utilized to determine whether a silhouette coefficient value is good or not can be assessed using the Kaufman and Rousseeuw table, as shown in Table 1 [34]. The visualization output of the PFCM will be displayed in cluster plots and thematic maps.

Table 1. Interpretation of silhouette coefficient.

Silhouette Coefficient	Interpretation
0.71 – 1.00	Strong clusters
0.51 – 0.70	Clusters deemed appropriate
0.26 – 0.50	Weak Cluster
≤ 0.25	Cannot be considered a Cluster

Further, R-4.3.2 software will be used to analyze the data according to the study objectives. Using R software for data analysis offers statistical methods, enhanced performance, and broad community support, making it a robust and efficient choice for modern data analysis needs [35, 36].

3. Results and Discussion

Based on the statistical analysis shown in Table 2, The range of population sizes is quite broad, indicating significant variation in the sizes of the areas or groups being studied. The mean population of 232,517 suggests that most areas fall closer to the smaller end of the scale, as this mean is much closer to the minimum than the maximum. The standard deviation (SD) of 142,17 is quite high, reinforcing the idea of a large variance.

In-migrant numbers are with a minimum of 253 and a maximum of 7,439. The mean of 2,920 indicates a moderate level of in-migration on average. However, the relatively low standard deviation of 1,95 compared to the mean suggests that the in-migration numbers do not vary as widely as the population figures.

For out-migrants, there is a significant range between the minimum (2,237) and maximum (29,906) out-migrants, indicating diverse migration patterns across the areas studied. The mean of 11,188, paired with a standard deviation of 7,780, suggests considerable variability but less extreme than in the population variable.

Further, the fertility rates range from 58 to 79, with a mean of 68 and a low standard deviation of 6. This narrow range and low standard deviation indicate that fertility rates are relatively consistent across the areas

studied. For maternal mortality data, the range from 2 to 28, with a mean of 10 and a standard deviation of 7, indicates variability but not as extreme as seen in some of the other variables, and the data is relatively low numbers.

Based on the results of Shapiro-Wilk normality tests, each variable (Population, In-migrants, Out-migrants, Fertility, Maternal Mortality) is accompanied by a W statistic and a p-value. The W statistic closer to 1 indicates a more normal distribution. The p-value helps decide whether the distribution significantly deviates from normality, with a common threshold being 0.05. Here, in-migrants (W = 0.9166; p-value = 0.0565) and Fertility (W = 0.9595; p-value = 0.4527) show the normal distribution because their values are well above 0.05, suggesting their distribution is not significantly different from normal.

In contrast, other variables, notably Population, Out-migrants, and Maternal Mortality, have lower W values

Table 2. Summary statistics of variables.

No.	Variable	Unit	Central Tendency dan Dispersion	District/City	Value
1	Population	People	Min	Sabang	42,696
			Max	North Aceh	600,163
			Mean	-	232,517
			SD	-	142,17
2	In-migrants	People	Min	Sabang	253
			Max	Aceh Besar	7,439
			Mean	-	2,92
			SD	-	1,95
3	Out-migrants	People	Min	Sabang	2,237
			Max	Banda Aceh	29,906
			Mean	-	11,188
			SD	-	7,780
4	Fertility	Ratio	Min	Nagan Raya	58
			Max	Aceh Jaya	79
			Mean	-	68
			SD	-	6
5	Maternal mortality	Ratio	Min	Central Aceh and Subulussalam	2
			Max	Bireuen	28
			Mean	-	10
			SD	-	7

Table 3. Result of normality test and correlation test.

No.	Variable	Normality test		Correlation test	
		W	p-value	r	p-value
1	Population	0.889	0.015	-	-
2	In-migrants	0.916	0.056*	0.846	1.78 x 10 ⁻⁶ *
3	Out-migrants	0.884	0.012	0.864	2.30 x 10 ⁻⁶ *
4	Fertility	0.959	0.452	0.364	0.09
5	Maternal mortality	0.913	0.048	0.162	0.46

Note:(*) = significant at 5% significance level.

and p-values below 0.05, indicating their distributions significantly deviate from normality. This suggests that the data for these variables might be skewed or have outliers, while in-migrants and fertility data appear to be more normally distributed. The complete information can be seen in Table 3.

The data normality test analysis results direct the study to test the data using the Spearman test. This is because the population variable as the key variable is not normally distributed. The Spearman correlation test results reveal a statistically significant and strong positive correlation between population size and both in-migrants ($r = 0.846$; $p\text{-value} = 1.78 \times 10^{-6}$) and out-migrants ($r = 0.864$; $p\text{-value} = 2.30 \times 10^{-6}$), indicating that areas with larger populations tend to have higher numbers of both incoming and outgoing individuals. However, the correlations between population and fertility ($r = 0.364$; $p\text{-value} = 0.09$) and population and maternal mortality ($r=0.162$; $p\text{-value} =$

0.46) are much weaker and not statistically significant, with p-values exceeding the conventional significance level.

This suggests that while migration patterns are closely linked to population size, fertility, and maternal mortality rates do not show a strong direct relationship with the size of the population. The complete visualization and tabulation of the Spearman test results can be seen in Figure 1 and Table 2.

Further, effective optimization in clusters is essential. The graph in Figure 2 plots the total within the sum of squares (WSS) against the number of clusters k. This type of plot is often called an "elbow graph" because you look for an "elbow" where the rate of decrease sharply changes. This point is likely indicative of the optimal number of clusters. In our graph, the elbow appears to be at the point where the number of clusters k equals 4. After this point, the decrease in WSS begins to level off, suggesting that

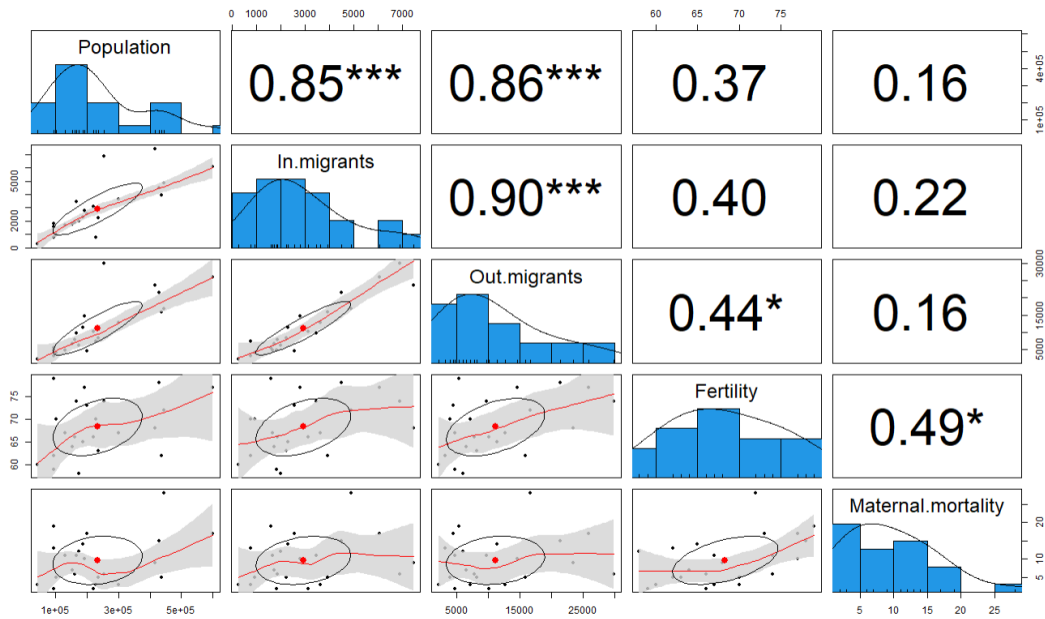


Figure 1. Pairs panel chart of Spearman correlation test.

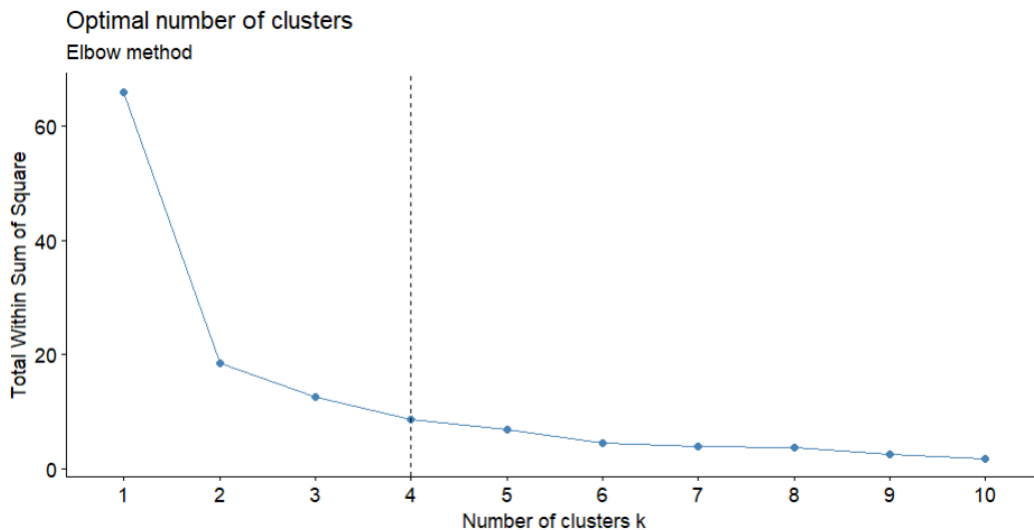


Figure 2. Optimum number cluster using elbow method.

additional clusters do not significantly improve the partitioning. Therefore, according to this method, the optimal number of clusters for the dataset would be 4.

Based on the result of PFCM, there are four clusters identified, as indicated by the four colors and the legend on the right side of the plot (Figure 3). Each cluster appears to contain different districts of Aceh, with the numbers corresponding to the list you provided. The x-axis (Dim1) and y-axis (Dim2) represent the two principal components extracted from the data, which means that the plot is likely a result of a dimensionality reduction technique applied to visualize the clusters in two dimensions. The percentages (e.g., 69.6%) indicate how much of the variance in the data is captured by each principal component.

The spread of the clusters indicates the variance within each cluster. For example, cluster 1 (light blue) contains Aceh Besar (6), North Aceh (8), and Banda Aceh (9) is very tightly clustered, which might suggest that this district is quite distinct from the others in terms of the variables measured.

Cluster 2 (grey) contains somewhat spread-out districts but still forms a distinct group. This includes districts like Aceh Tamiang (16), Bireuen (11), and Pidie (7). Cluster 3 (yellow) has a larger spread and includes districts such as West Aceh (5), Southwest Aceh (12), and Bener Meriah (17). Cluster 4 (pink) is elongated and seems to contain districts that are quite spread out, including Gayo Lues (13), Aceh Jaya (14), and Sabang (20). The positioning of the clusters relative to each other can also give insights

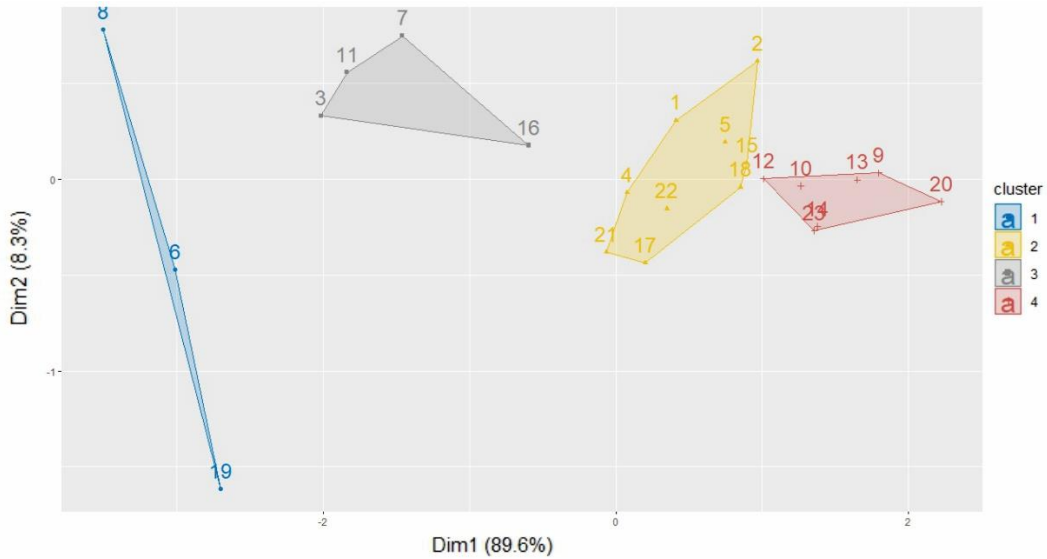


Figure 3. Cluster plot of PFCM result.

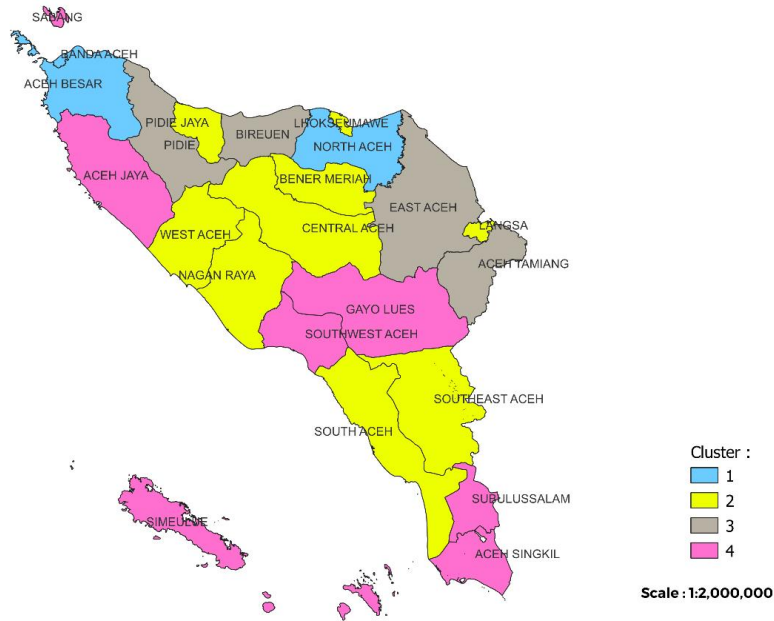


Figure 4. Thematic map of PFCM result.

Table 4. District/city based on the result of PFCM clustering.

Cluster	District/City
1	Aceh Besar, North Aceh, and Banda Aceh
2	South Aceh, Southeast Aceh, Central Aceh, West Aceh, Nagan Raya, Bener Meriah, Pidie Jaya, Lhokseumawe, and Langsa
3	East Aceh, Bireuen, Aceh Tamiang, and Pidie
4	Simeulue, Aceh Singkil, Gayo Lues, Aceh Jaya, Sabang, Southwest Aceh and Subulussalam

into the similarities and differences between the clusters. For example, clusters 3 and 4 are closer to each other than clusters 1 or 2, which might suggest that the districts in clusters 3 and 4 are more similar to each other than those in clusters 1 or 2. A complete explanation of the clustering results can be seen in Table 4 and visualization of the thematic maps can be seen in Figure 4.

This study has introduced PFCM clustering to analyze demographic data in Aceh Province, enhancing the understanding of population dynamics with a novel approach that embraces the inherent uncertainty of demographic data. The PFCM's capacity to handle ambiguity in membership allocation has allowed us to discern the underlying structure in population data that

traditional hard clustering methods could have oversimplified or misrepresented. Our findings echo the sentiments of previous works that highlight the limitations of crisp clustering techniques when applied to complex, real-world datasets [37, 38].

Our results have significant implications for regional development policy and resource allocation. For instance, the tight clustering of districts such as Aceh Besar, Aceh Utara, and Banda Aceh (Cluster 1) suggests a commonality in population for demographic trends that could be attributed to their geographical proximity and socio-economic interactions. These insights corroborate the research conducted by [39], who found geographical factors to play a pivotal role in shaping demographic patterns within regions. Moreover, the distinction between Cluster 4 (including Simeulue, Aceh Singkil, Gayo Lues, Aceh Jaya, Sabang, and Subulussalam) and the others could indicate unique demography, warranting tailored policy interventions [40, 41].

Based on the results, a more impactful approach for this study would be to provide specific recommendations tailored to the unique demographic characteristics of each cluster. For the closely clustered districts of Aceh Besar, Aceh Utara, and Banda Aceh, policy recommendations should focus on leveraging their geographical proximity and socio-economic interactions to maximize regional development. Strategies could include integrated infrastructure projects or shared educational and employment initiatives. Conversely, for Cluster 4, which includes Simeulue, Aceh Singkil, Gayo Lues, Aceh Jaya, Sabang, and Subulussalam, the distinct demography suggests the need for customized policy interventions. This could involve developing localized resource allocation plans or specialized social programs that address the unique needs and challenges of these areas. By providing such targeted recommendations, the study can offer more practical and effective guidance for regional development and resource allocation.

However, this study has its limitations. The choice of variables, while comprehensive, might only capture part of the spectrum of factors influencing population clustering. Future studies could expand the dataset to include socio-demographic variables such as education levels, employment rates, and infrastructure quality, which have been shown to impact demographic population [42–44]. Additionally, while PFCM has provided a robust framework for clustering, the interpretability of fuzzy clusters remains a challenge, requiring decision-makers to navigate the probabilistic nature of the memberships when translating findings into policies [45, 46]. Future directions of this study include expanding the dataset to incorporate socio-

demographic factors like education, employment rates, and infrastructure quality, and improving the interpretability of fuzzy clusters in the Possibilistic Fuzzy C-Means framework to aid in policy-making.

4. Conclusions

The study's findings revealed significant correlations between population size and both in-migrants and out-migrants, indicating that areas with larger populations tend to have higher numbers of both incoming and outgoing individuals. However, the correlations between population size and indicators like fertility and maternal mortality were weaker and insignificant. This suggests that while migration patterns are closely linked to population size, fertility, and maternal mortality rates do not have a strong direct relationship with the size of the population. The PFCM method identified four distinct clusters of districts/cities in Aceh, each showing unique demographic characteristics. This clustering provides a nuanced understanding of demographic patterns and can serve as a basis for more targeted and effective policy-making and regional development in Aceh Province.

Author Contributions: Conceptualization, N.R.S.; methodology, N.R.S., and M.K.; software, M.K. and A.D.; validation, H.Z., R.K., S.M., formal analysis, N.R.S. and M.K.; investigation, F.A., and M.M.; resources A.D.; data curation, N.R.S., and F.A; writing—original draft preparation, N.R.S., D.C.E.S., A.W.M. and T.R.N.; writing—review and editing, N.R.S. and H.Z.; visualization, M.K.; supervision, H.Z.; project administration, A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors

Acknowledgments: We extend our sincere gratitude to the anonymous reviewers whose thorough and constructive feedback greatly strengthened the quality and clarity of this article. Their expertise and invaluable insights have been instrumental in shaping this work.

Conflicts of Interest: All the authors declare that there are no conflicts of interest.

References

1. United Nations. (2023). World Population Prospects 2022, *Department of Economic and Social, Population Division, United Nations*, 1.
2. Statistics of Aceh Province. (2023). Number of Population (People), 2020-2022, *Statistics of Aceh Province*, 1.
3. Syaputri, D., Noprita, P. H., and Romelah, S. (2021). Implementation of K-Means Algorithm for Economic Distribution Clustering Base on Demographics of Population,

- MALCOM: Indonesian Journal of Machine Learning and Computer Science*, Vol. 1, No. 1, 1–6.
4. Arifiyanti, A. A., Darusman, F. S., and Trenggono, B. W. (2022). Population Density Cluster Analysis in DKI Jakarta Province Using K-Means Algorithm, *Journal of Information Systems and Informatics*, Vol. 4, No. 3, 772–783.
 5. Afidah, N. N., and Masrukan. (2023). Penerapan Metode Clustering dengan Algoritma K-Means untuk Pengelompokan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang, *PRISMA, Prosiding Seminar Nasional Matematika*, Jurusan Matematika, Universitas Negeri Semarang, Semarang, 729–738.
 6. Setiawan, D., and Zahra, A. (2023). Pengelompokan Kemiskinan di Indonesia Menggunakan Time Series Based Clustering, *Inferensi*, Vol. 6, No. 1, 83. doi:10.12962/j27213862.v6i1.14969.
 7. Faizah, N., Surohman, Fabrianto, L., Hendra, and Prasetyo, R. (2020). Unbalanced Data Clustering with K-Means and Euclidean Distance Algorithm Approach Case Study Population and Refugee Data, *Journal of Physics: Conference Series*, Vol. 1477, No. 2, 022005. doi:10.1088/1742-6596/1477/2/022005.
 8. Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation, *Electronics*, Vol. 9, No. 8, 1295. doi:10.3390/electronics9081295.
 9. Sasmita, N. R., Sofyan, H., and Subianto, M. (2012). The Comparison of Fuzzy C-Means (FCM) and Fuzzy C-Shell (FCS) Methods By Using Quickbird Satellite Image (Case Study In Peukan Bada), *SEMIRATA 2012*, State University of Medan, Medan, 211–218.
 10. Sasmita, N. R., Phonna, R. A., Fikri, M. K., Khairul, M., Apriliansyah, F., Idroes, G. M., Puspitasari, A., and Saputra, F. E. (2023). Statistical Assessment of Human Development Index Variations and Their Correlates: A Case Study of Aceh Province, Indonesia, *Grimsa Journal of Business and Economics Studies*, Vol. 1, No. 1, 12–24.
 11. Idroes, G. M., Hardi, I., Noviandy, T. R., Sasmita, N. R., and Shabrina, I. (2023). A Deep Dive into Indonesia ' s CO 2 Emissions : The Role of Energy Consumption , Economic Growth and Natural Disasters, 69–81. doi:10.60084/eje.v1i2.115.
 12. Aceh Population Registration Office (DRKA Aceh). (2022). Population Profile Book, *DRKA Aceh*, 1–100.
 13. Sasmita, N. R., Ramadeska, S., Utami, R., Adha, Z., Putri, U., Syarafina, R. H., Reskiaddin, L. O., Kamal, S., Yarmaliza, Y., Muliadi, M., and Saputra, A. (2023). Maternal and Child Healthcare Services in Aceh Province, Indonesia: A Correlation and Clustering Analysis in Statistics, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 22–31. doi:10.60084/ijds.v1i1.88.
 14. Suhendra, R., Suryadi, S., Husdayanti, N., Maulana, A., Noviandy, T. R., Sasmita, N. R., Subianto, M., Earlia, N., Niode, N. J., and Idroes, R. (2023). Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification, *Heca Journal of Applied Sciences*, Vol. 1, No. 2, 54–61. doi:10.60084/hjas.v1i2.85.
 15. Iffaty, A., Salsabila, A., Rafiqhi, A. A., Suhendra, R., Yusuf, M., and Sasmita, N. R. (2023). Enhancing Water Quality Assessment in Indonesia Through Digital Image Processing and Machine Learning, *Grimsa Journal of Science Engineering and Technology*, Vol. 1, No. 1, 1–7.
 16. Azharuddin, A., Sasmita, N. R., Idroes, G. M., Andid, R., Raihan, R., Fadlilah, T., Earlia, N., Ridwan, T., Maya, I., and Farnida, F. (2023). Patient Satisfaction and its Socio-Demographic Correlates in Zainoel Abidin Hospital, Indonesia: A Cross-Sectional Study, *Unnes Journal of Public Health*, Vol. 12, No. 2, 57–67. doi:doi.org/10.15294/ujph.v12i2.69233.
 17. Sofyan, H., Diba, F., Susanti, S. S., Marthoenis, M., Ichsan, I., Sasmita, N. R., Seuring, T., and Vollmer, S. (2023). The state of diabetes care and obstacles to better care in Aceh, Indonesia: a mixed-methods study, *BMC Health Services Research*, Vol. 23, No. 1, 271. doi:10.1186/s12913-023-09288-9.
 18. Maulana, A., Noviandy, T. R., Sasmita, N. R., Paristiowati, M., Suhendra, R., Yandri, E., Satrio, J., and Idroes, R. (2023). Optimizing University Admissions: A Machine Learning Perspective, *Journal of Educational Management and Learning*, Vol. 1, No. 1, 1–7. doi:10.60084/jeml.v1i1.46.
 19. Sasaki, D., Sofyan, H., Sasmita, N. R., Affan, M., and Nizamuddin, N. (2021). Assessing the Intermediate Function of Local Academic Institutions During the Rehabilitation and Reconstruction of Aceh, Indonesia, *Journal of Disaster Research*, Vol. 16, No. 8, 1265–1273. doi:10.20965/jdr.2021.p1265.
 20. Rachmawati, R., Novita, R., Erwandi, E., Fitriyanin, E., Idroes, R., and Reandy Sas, N. (2020). The Effectiveness of Counseling on Complementary Food for Mothers and Supplementary Feeding for Increasing weight of Breastfed babies (12-24 Months of Age) in the Aceh Besar District of Indonesia, *Pakistan Journal of Nutrition*, Vol. 19, No. 6, 279–284. doi:10.3923/pjn.2020.279.284.
 21. Idroes, R., Maulana, A., Noviandy, T. R., Suhendra, R., Sasmita, N. R., Lala, A., and Irvanizam. (2020). A Genetic Algorithm to Determine Research Consultation Schedules in Campus Environment, *IOP Conference Series: Materials Science and Engineering*, Vol. 796, 012033. doi:10.1088/1757-899X/796/1/012033.
 22. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., and Sasmita, N. R. (2023). ANFIS-Based QSRR Modelling for Kovats Retention Index Prediction in Gas Chromatography, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 1–14. doi:10.60084/ijds.v1i1.73.
 23. Rahayu, L., Sasmita, N. R., Adila, W. F., Kesuma, Z. M., and Kruba, R. (2023). Spatial Estimation of Relative Risk for Dengue Fever in Aceh Province using Conditional Autoregressive Method, *Journal of Applied Data Sciences*, Vol. 4, No. 4, 466–479. doi:10.47738/jads.v4i4.141.
 24. Noviandy, T. R., Maulana, A., Sasmita, N. R., Suhendra, R., Irvanizam, I., Muslem, M., Idroes, G. M., Yusuf, M., Sofyan, H., Abidin, T. F., and Idroes, R. (2022). The Prediction of Kovats Retention Indices of Essential Oils at Gas Chromatography Using Genetic Algorithm-Multiple Linear Regression and Support Vector Regression, *Journal of Engineering Science and Technology*, Vol. 17, No. 1, 306–326.
 25. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., Sasmita, N. R., Muslem, M., Idroes, G. M., Kemala, P., and Irvanizam, I. (2021). Application of Genetic Algorithm-Multiple Linear Regression and Artificial Neural Network Determinations for Prediction of Kovats Retention Index, *International Review on Modelling and Simulations (IREMOS)*, Vol. 14, No. 2, 137. doi:10.15866/iremos.v14i2.20460.
 26. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., Sasmita, N. R., Muslem, M., Idroes, G. M., and Irvanizam, I. (2019). Retention Index Prediction of Flavor and Fragrance by Multiple Linear Regression and the Genetic Algorithm, *International Review on Modelling and Simulations (IREMOS)*, Vol. 12, No. 6, 373. doi:10.15866/iremos.v12i6.18353.
 27. Earlia, N., Bulqiah, M., Muslem, M., Karma, T., Suhendra, R., Maulana, A., Amin, M., Sasmita, N. R., Idroes, G. M., and Prakoeswa, C. (2021). Protective Effects of Acehese Traditionally Fermented Coconut Oil (Pliek U Oil) and its Residue (Pliek U) in Ointment against UV Light Exposure: Studies on Male Wistar Rat Skin (*Rattus novergicus*), *Sains Malaysiana*, Vol. 50, No. 5, 1285–1295.
 28. Idroes, R., Husan, I., Muslem, Mahmudi, Rusyana, A., Helwati, Z., Idroes, G. M., Suhendra, R., Yandri, E., Rahimah, S., and Sasmita, N. R. (2019). Analysis of Temperature and Column Variation in Gas Chromatography to Dead Time of Inert Gas and N-Alkane Homologous Series Using Randomized Block Design, *IOP Conference Series: Earth and Environmental Science*, Vol. 364, No. 012020, 1–7. doi:10.1088/1755-1315/364/1/012020.
 29. Nadia, Y., Ramli, M., Muslem, Japnur, A. F., Rusyana, A., Idroes, G. M., Suhendra, R., Muhammad, Sasmita, N. R., Tallei, T. E., Idroes, R., Jafnur, A. F., Rusyana, A., Idroes, G. M., Suhendra, R., Muhammad, Sasmita, N. R., Tallei, T. E., and Idroes, R. (2019).

- Simple Combination Method of FTIR Spectroscopy and Chemometrics for Qualitative Identification of Cattle Bones, *IOP Conference Series: Earth and Environmental Science*, Vol. 364, No. 1, 012040. doi:10.1088/1755-1315/364/1/012040.
30. Maulana, A., Noviandy, T. R., Idroes, R., Sasmita, N. R., Suhendra, R., and Irvanizam, I. (2020). Prediction of Kovats Retention Indices for Fragrance and Flavor using Artificial Neural Network, *2020 International Conference on Electrical Engineering and Informatics (ICELTICs)*, IEEE, 1–5. doi:10.1109/ICELTICs50595.2020.9315391.
31. Noviandy, T. R., Maulana, A., Sasmita, N. R., Suhendra, R., Muslem, Idroes, G. M., Paristiowati, M., Helwani, Z., Yandri, E., Rahimah, S., Muhammad, Irvanizam, and Idroes, R. (2021). The implementation of K-Means clustering in kovats retention index on gas chromatography, *IOP Conference Series: Materials Science and Engineering*, Vol. 1087, No. 1, 012051. doi:10.1088/1757-899X/1087/1/012051.
32. Idroes, G. M., Syahnur, S., Majid, S. A., Sasmita, N. R., and Idroes, R. (2021). Provincial economic level analysis in Indonesia based on the geothermal energy potential and growth regional domestic products using cluster analysis, *IOP Conference Series: Materials Science and Engineering*, Vol. 1087, No. 1, 012079. doi:10.1088/1757-899X/1087/1/012079.
33. Agustia, M., Noviandy, T. R., Maulana, A., Suhendra, R., Muslem, M., Sasmita, N. R., Idroes, G. M., Rahimah, S., Afidh, R. P. F., Subianto, M., Irvanizam, I., and Idroes, R. (2022). Application of Fuzzy Support Vector Regression to Predict the Kovats Retention Indices of Flavors and Fragrances, *2022 International Conference on Electrical Engineering and Informatics (ICELTICs)*, IEEE, 13–18. doi:10.1109/ICELTICs56128.2022.9932124.
34. Kaufman, L., and Rousseeuw, P. J. (1990). *Finding Groups in Data*, Wiley. doi:10.1002/9780470316801.
35. Sasmita, N. R., Ikhwan, M., Suyanto, S., and Chongsuivatwong, V. (2020). Optimal control on a mathematical model to pattern the progression of coronavirus disease 2019 (COVID-19) in Indonesia, *Global Health Research and Policy*, Vol. 5, No. 1, 38. doi:10.1186/s41256-020-00163-2.
36. Sasmita, N. R., Geater, A. F., Kammanee, A., and Chongsuivatwong, V. (2019). Is The Recovery Rate In Latent Tuberculosis Infection Significant In Reducing Tuberculosis Transmission In Indonesia?: A Mathematical Model Study In Epidemiology, *RSU Research Conference 2019*, Rangsit University, Pathum Thani, Thailand, 101–111.
37. Xie, Y., and Shekhar, S. (2019). Significant DBSCAN towards Statistically Robust Clustering, *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, ACM, New York, NY, USA, 31–40. doi:10.1145/3340964.3340968.
38. Aupetit, M., Sedlmair, M., Abbas, M. M., Baggag, A., and Bensmail, H. (2019). Toward Perception-Based Evaluation of Clustering Techniques for Visual Analytics, *2019 IEEE Visualization Conference (VIS)*, IEEE, 141–145. doi:10.1109/VISUAL.2019.8933620.
39. Gabrielli, G., Paterno, A., Salvini, S., and Corazziari, I. (2021). Demographic trends in less and least developed countries: Convergence or divergence?, *Journal of Population Research*, Vol. 38, No. 3, 221–258. doi:10.1007/s12546-021-09264-2.
40. Landis, J., and Reina, V. (2019). Eleven Ways Demographic and Economic Change Is Reframing American Housing Policy, *Housing Policy Debate*, Vol. 29, No. 1, 4–21. doi:10.1080/10511482.2018.1492739.
41. Bijak, J., Courgeau, D., Silverman, E., and Franck, R. (2014). Quantifying paradigm change in demography, *Demographic Research*, Vol. 30, 911–924. doi:10.4054/DemRes.2014.30.32.
42. Fecht, D., Cockings, S., Hodgson, S., Piel, F. B., Martin, D., and Waller, L. A. (2020). Advances in mapping population and demographic characteristics at small-area levels, *International Journal of Epidemiology*, Vol. 49, No. Supplement_1, i15–i25. doi:10.1093/ije/dyz179.
43. Salgado, M., Madureira, J., Mendes, A. S., Torres, A., Teixeira, J. P., and Oliveira, M. D. (2020). Environmental determinants of population health in urban settings. A systematic review, *BMC Public Health*, Vol. 20, No. 1, 853. doi:10.1186/s12889-020-08905-0.
44. Saint Onge, J. M., and Smith, S. (2020). Demographics in Rural Populations, *Surgical Clinics of North America*, Vol. 100, No. 5, 823–833. doi:10.1016/j.suc.2020.06.005.
45. Varshney, A. K., Danish Lohani, Q. M., and Muhuri, P. K. (2020). Improved Probabilistic Intuitionistic Fuzzy c-Means Clustering Algorithm: Improved PIFCM, *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 1–6. doi:10.1109/FUZZ48607.2020.9177574.
46. Pimentel, B. A., and de Souza, R. M. C. R. (2018). A Generalized Multivariate Approach for Possibilistic Fuzzy C-Means Clustering, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 26, No. 06, 893–916. doi:10.1142/S021848851850040X.