Implementation of the K-Means Clustering Method in Clustering Poor Population in Bandar Kumbul Village, Labuhanbatu Regency

Aulia Maharani[1*], Gomal Juni Yanris[2], Fitri Aini Nasution[3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara Indonesia.
[*]Corresponding Author:
Email: aiauliamaharani@gmail.com

Abstract.

Poverty is one of the crucial social problems in rural areas. The problem of poverty in rural areas is increasingly in the spotlight because of its broad impact on the community's economic and social sustainability. The varying levels of poverty require an appropriate analytical approach to design effective intervention programs. In an effort to understand and address this problem, this study uses the K-Means Clustering method to group the poor population. We use K-Means clustering to identify and group hamlets based on their poverty levels. This study aims to categorize the hamlets in Bandar Kumbul Village into multiple clusters according to their poverty levels, thereby identifying which hamlets necessitate more focused attention. The research methods used include collecting data on the number of poor people from 2013 to 2022 in each hamlet, data preprocessing, applying the Elbow method to determine the optimal number of clusters, and applying the K-Means Clustering algorithm to group the hamlets. The results of the study show that there are three main clusters with different characteristics. Cluster 0 includes Hutaimbaru and Mailil Julu hamlets with high poverty levels. Cluster 1 only includes the Pasir Sidimpuan hamlet, which has medium poverty levels. Cluster 2 includes Aek Mardomu, Bandar Kumbul, Mailil Jae, Sidodadi, and Singga Mata hamlets with low poverty levels. Variations in distance from the cluster center indicate significant differences in the distribution of poverty in each hamlet. The K-Means Clustering method is effective in identifying and grouping hamlets based on poverty levels, providing useful insights for the government and stakeholders to design more targeted intervention programs. Clusters with high poverty levels require immediate intervention, while clusters with medium and low poverty levels require maintenance and support to prevent an increase in poverty. This study provides a strong foundation for decision-making and policies to reduce poverty levels in Bandar Kumbul Village more effectively and sustainably.

Keywords: *Bandar Kumbul, Clustering, Elbow Method, K-Means, and Poverty.*

I. INTRODUCTION

Poverty is a complex problem faced by many countries, including Indonesia. Poverty can be defined as a condition where a person or household is unable to meet basic minimum needs, both in terms of economy, social, and education [1]. Poverty has many negative impacts on the lives of individuals and communities, such as difficulty in meeting basic needs, limited access to adequate education and health services, and limited economic opportunities [2]. Factors that influence poverty include income level, unemployment rate, education level, access to health services, and infrastructure conditions [3]. Poverty in rural areas is a complex problem and requires in-depth understanding to find the right solution. The problem of poverty in rural areas is increasingly in the spotlight because of its broad impact on the community's economic and social sustainability. A social structure in rural areas leads to poverty by denying some members of society equal access to economic resources and facilities. Backwardness in communities that lack basic rights such as food, access to health, education, employment, and infrastructure facilities contributes to the problem of poverty in rural areas [4]. The increase in rural poverty is also in the spotlight because of its diverse complexities, as are other social problems such as crime, violence, criminality, and social disabilities [5].It is important to group the poor population in order to identify their characteristics and needs more specifically, so that it can help the government formulate more effective poverty alleviation policies and programs [6].We need to identify and group the population in rural areas appropriately to overcome this problem [7]. The cluster method's importance in identifying patterns in data is a strategic step in evaluating and formulating appropriate policies to combat poverty in rural areas [8]. This approach aims to tailor the resulting solutions to the unique characteristics of each region.

To ensure the success of the implemented assistance programs, in-depth research on the factors causing poverty in each cluster is also necessary [9]. Thus, a more focused approach to eradicating poverty in rural areas can positively impact the local community. In addition, cooperation between the government, non-governmental organizations, and the private sector is also crucial in implementing the policy. We hope

that these programs, working together effectively, can significantly reduce poverty levels in rural areas.Researchers have widely used clustering methods, especially K-means clustering, to group the poor population. Clustering is a technique for grouping data based on similar characteristics without considering class information or data labels [10], [11]. The K-Means Clustering algorithm is one of the popular clustering methods. This algorithm groups data into k clusters based on the closest distance to the centroid [12], [13]. The K-Means Clustering algorithm is able to group data based on similar characteristics, so it can help identify the profile of poor population groups [14]. One cluster will group data with similar characteristics, while different characteristics will place data in different clusters [15], [16]. Several studies have applied the K-means clustering method to group the poor population. In North Sumatra Province, research by [17] has identified three poverty clusters, each consisting of five districts/cities experiencing high levels of poverty. We successfully applied the K-Means method based on the Human Development Index indicator, resulting in the lowest cluster of 13, the medium cluster of 1, and the highest cluster of 19 [18]. Out of 18 sub-districts in Gunungkidul Regency, research by [19] found that 10 sub-districts received a significant amount of PKH assistance, while 8 sub-arrondissements had a high level of social welfare and received only a small amount of PKH assistance.

Research by [20] developed a K-means algorithm based on poverty indicators to classify poverty in Central Java. The results showed that 22 districts or cities in the first cluster were not poor, and 13 districts or cities in the second cluster were poor.Bandar Kumbul is one of the villages in Bilah Barat sub-district, Labuhanbatu Regency, North Sumatra province, Indonesia [21].Of the 10 villages in Bilah Barat sub-district, the one with the largest area is Bandar Kumbul Village, with an area of 36.20 Km^2 and a population of 4,362 people [22]. Bandar Kumbul Village grapples with a significant poverty issue, characterized by varying rates of poverty within each hamlet. According to the collected data, the average level of education only reaches elementary school. Additionally, access to health services and housing conditions remain limited. The analysis results show that the unemployment rate, low levels of education, and limited access to basic services are the main factors influencing poverty in Bandar Kumbul Village. This study aims to apply the K-Means Clustering method to group the poor population in Bandar Kumbul Village based on the number of poor people in each hamlet. The elbow method serves as the foundation for determining the number of clusters. Conducting this research is crucial in identifying the hamlet clusters in Bandar Kumbul Village, taking into account poverty indicators. This will enable the relevant government to devise solutions for addressing the welfare of the people, based on various regional clusters that correspond to the severity of the issue.

II. METHODS

This stage outlines the methodology for applying the K-Means Clustering method to group the poor population in Bandar Kumbul Village. Figure 1 shows the stages of the research, starting from dataset collection, data preprocessing, implementation of the k-means method, and results. We implement all stages of this research using the Python programming language in the Google Colab text editor.

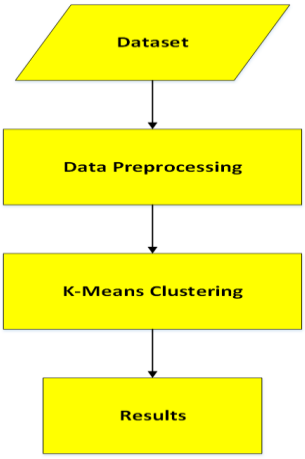

Fig 1. Research Stage

The first stage in this study is to collect the dataset. This dataset contains information about the poor population in Bandar Kumbul Village, Labuhanbatu Regency. This dataset includes eight hamlets with variables indicating the number of poor people from 2013 to 2022. Secondary data is collected directly from the Bandar Kumbul Village office. Table 1 shows the research dataset.

Table 1. Dataset

Dusun	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Aek Mardomu	36	40	46	48	50	50	51	38	40	36
Bandar Kumbul	45	51	59	61	63	64	65	48	50	46
Hutaimbaru	79	89	102	106	110	112	114	84	88	80
Mailil Jae	46	51	59	61	64	64	66	49	51	46
Mailil Julu	72	81	93	96	100	101	103	76	80	73
Pasir Sidimpuan	8	9	10	11	11	11	11	8	9	8
Sidodadi	47	52	60	63	65	66	67	50	52	47
Singga Mata	35	39	45	47	49	49	50	37	39	35

The next step after collecting the dataset is to preprocess the data. Data preprocessing includes several sub-stages, namely: checking missing values, checking duplicate data, labeling data, and determining the average. The third stage is the application of K-Means clustering; this stage is the core of the research where the K-Means clustering method is applied to the preprocessed data. K-Means Clustering is an algorithm that groups data into clusters based on the similarity of existing features. This process involves several steps: Determining the number of clusters and initializing cluster centers (centroids) randomly or using a certain method, calculating the distance of each data point to the cluster center and assigning the data to the nearest cluster, recalculating the cluster center based on the average data in each cluster, repeating the assignment and updating process until the cluster center no longer changes or the change is very small (convergence). The final stage of this research is to present the clustering results. These results can be clustering of poor people, cluster analysis, and visualization of cluster results.

III. RESULT AND DISCUSSION

This section presents the results of applying the K-means clustering method to group the poor population in Bandar Kumbul Village. We describe the study's results sequentially, following the research steps outlined in the previous methodology section.

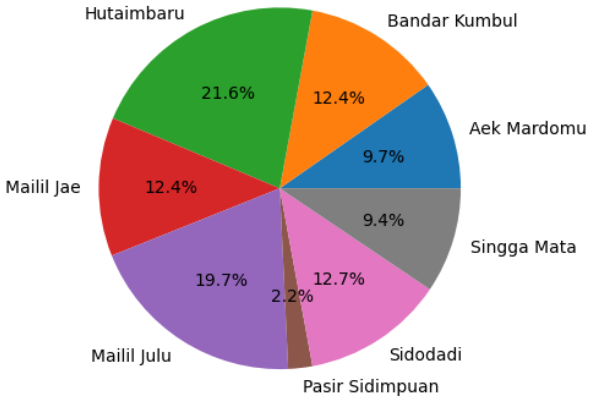

Fig 2. Distribution for Poor Population in 2022

Figure 2 depicts the percentage distribution of poor population data in Bandar Kumbul Village in 2022. This diagram illustrates the distribution of poor people in the village's eight hamlets. In 2022, there were 371 poor people out of a total population of 4362. Hutaimbaru Hamlet has the highest percentage of poor people (21.6%). This shows that around 1/5 of the total poor population in this village lives in Hutaimbaru. Mailil Julu Hamlet also has a significant percentage of poor people, namely 19.7%. This means that nearly 1/5 of the poor population lives in this hamlet. Sidodadi Hamlet has a relatively high percentage of poor people (12.7%). This means that more than 1/10 of the poor people in this village live in Sidodadi. Mailil Jae and Bandar Kumbul Hamlets each have the same percentage, namely 12.4%. This shows that each

has almost the same share of the total poor population in this village. Both hamlets (Aek Mardomu and Singga Mata) have almost the same percentage of poor people, namely 9.7% and 9.4%. This shows that both have a fairly uniform distribution of poor people and are not too different from each other. Pasir Sidimpuan hamlet has the lowest percentage of poor people (2.2%). This shows that this area has a relatively lower poverty problem compared to other hamlets in the village. The distribution of impoverished individuals in Bandar Kumbul Village reveals that Hutaimbaru and Mailil Julu bear the brunt of poverty, accounting for over 40% of the village's total impoverished population. Hamlets with a lower percentage of poor people, such as Pasir Sidimpuan, may have relatively better economic conditions.

Dusun	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Label
Aek Mardomu	36	40	46	48	50	50	51	38	40	36	0
Bandar Kumbul	45	51	59	61	63	64	65	48	50	46	1
Hutaimbaru	79	89	102	106	110	112	114	84	88	80	2
Mailil Jae	46	51	59	61	64	64	66	49	51	46	3
Mailil Julu	72	81	93	96	100	101	103	76	80	73	4
Pasir Sidimpuan	8	9	10	11	11	11	11	8	9	8	5
Sidodadi	47	52	60	63	65	66	67	50	52	47	6
Singga Mata	35	39	45	47	49	49	50	37	39	35	7

Fig 3. Dataset Labeling

Figure 3 shows a table containing data on the number of poor people from various hamlets in Bandar Kumbul Village for the period 2013 to 2022. The Hamlet column lists the names of the hamlets in Bandar Kumbul Village. The year column (2013-2022) represents the number of poor people in each hamlet between 2013 and 2022. The clustering analysis results determine the label column for each hamlet. The number of poor people in Aek Mardomu is relatively stable, from 36 (2013) to 36 (2022), with slight fluctuations between the years. The poor population in Bandar Kumbul Hamlet shows an increase from 45 (2013) to 46 (2022), with a peak in 2018 (63). Hutaimbaru Hamlet has the highest number of poor people overall, increasing from 79 (2013) to 80 (2022), with a peak in 2018 (112). The number of poor people in Mailil Jae Hamlet increased from 46 (2013) to 46 (2022), with a peak in 2018 (66). Mailil Julu Hamlet showed a significant increase from 72 (2013) to 73 (2022), with a peak in 2018 (103). The number of poor people in Pasir Sidimpuan Hamlet was relatively low and stable, from 8 (2013) to 6 (2022). The number of poor people in Sidodadi Hamlet increased from 47 (2013) to 46 (2022), with a peak in 2019 (67). The number of poor people in Singga Mata Hamlet was relatively stable from 35 (2013) to 35 (2022), with slight fluctuations between the years. Most hamlets experienced an increase in the number of poor people between 2013 and 2022, with peaks in certain years, especially 2018. The labels (0 to 7) represent the outcomes of the clustering process, which groups each neighborhood according to similar trends and the proportion of impoverished individuals. These labels help to identify patterns and more effective intervention strategies for each Hamlet group.

Dusun	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Label	Average
Aek Mardomu	36	40	46	48	50	50	51	38	40	36	0	43.5
Bandar Kumbul	45	51	59	61	63	64	65	48	50	46	1	55.2
Hutaimbaru	79	89	102	106	110	112	114	84	88	80	2	96.4
Mailil Jae	46	51	59	61	64	64	66	49	51	46	3	55.7
Mailil Julu	72	81	93	96	100	101	103	76	80	73	4	87.5
Pasir Sidimpuan	8	9	10	11	11	11	11	8	9	8	5	9.6
Sidodadi	47	52	60	63	65	66	67	50	52	47	6	56.9
Singga Mata	35	39	45	47	49	49	50	37	39	35	7	42.5

Fig 4. Determination of average

Figure 4 displays the average values of the data contained in the year columns (2013–2022) for each hamlet. The average values (average column) show significant variations between hamlets. For example, Hutaimbaru has the highest average of 96.4, while Pasir Sidimpuan has the lowest average of 9.6. This suggests that there is a significant difference in the data observed in these hamlets, possibly due to the number of poor people. We use the "average" column to provide an overview of the data in each hamlet from 2013 to 2022. This can help identify which hospitals are more or less severe in terms of the conditions studied.

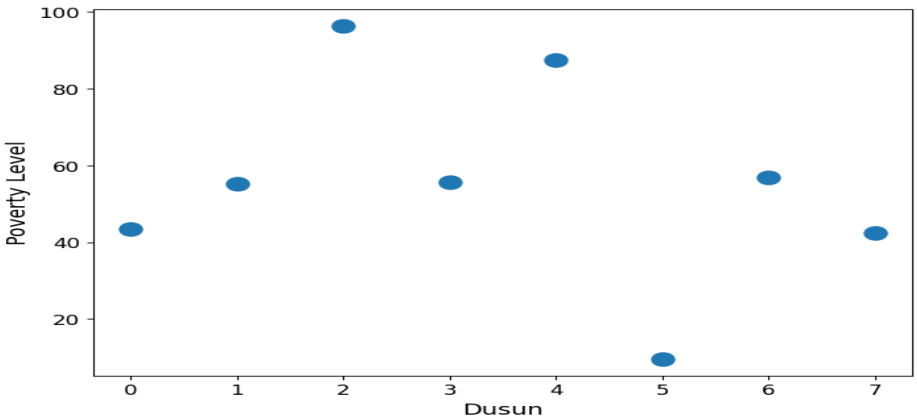

Fig 5. Visualization before clustering

Figure 5 is a visualization of the initial data before the clustering process. The X-axis (Dusun) is a numerical representation of the hamlets in Bandar Kumbul Village, with values ranging from 0 to 7. Each number represents one hamlet, most likely in the same order as the previous figure's table. The Y-axis (Poverty Level) displays the assessed poverty level, which spans from 0 to 100. This is a percentage scale or index that is used to measure poverty. Each point on the graph represents one hamlet and shows the measured poverty level in the hamlet before clustering. This visualization shows the variation in poverty levels between hamlets. Hamlet at point 2 (Hutaimbaru) has a very high poverty level approaching 100, while Hamlet at point 5 (Pasir Sidimpuan) has a very low poverty level approaching 0. There is significant variability in poverty levels between hamlets. The Y-axis displays fairly evenly distributed data points, signifying a significant variance in poverty levels. This visualization demonstrates the formation of potential groups based on the measured poverty levels. There are some hamlets with high poverty rates (around 80-100), some with medium poverty rates (around 40-60), and others with low poverty rates (below 20). This visualization provides an important initial overview before clustering. Looking at the distribution of data points allows us to understand underlying patterns in the data that may not be visible in the raw data table.

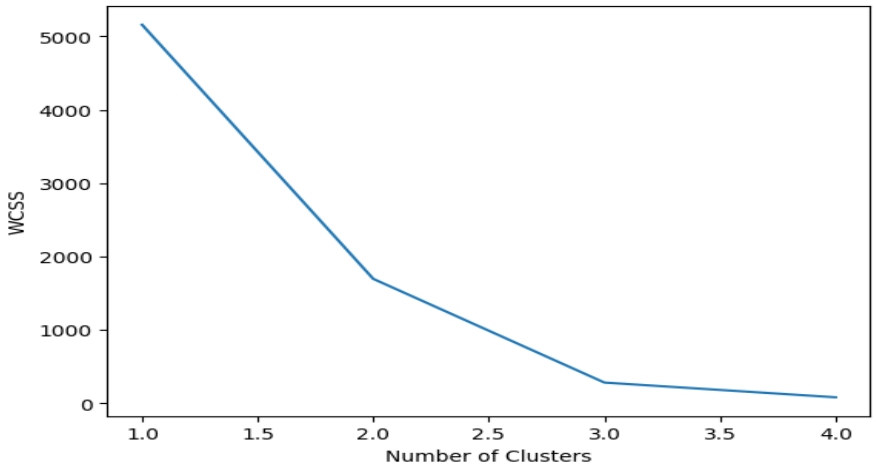

Fig 6. Number of clusters based on Elbow method

Figure 6 illustrates the application of the elbow method to ascertain the ideal number of clusters for the poor in Bandar Kumbul Village. This visualization consists of two parts: the Python code to run the Elbow method and the resulting graph. WCSS (Within-Cluster Sum of Square) is the sum of the squares of

the distances between data points to their respective centroids in a cluster. The elbow graph shows the relationship between the number of clusters (X-axis) and WCSS (Y-axis). We observe a drastic decrease in WCSS from 1 to 2 clusters as the number of clusters increases. This is natural because as the number of clusters increases, the data becomes more segmented, and the distance within the cluster decreases. The elbow method looks for a point where the decrease in WCSS begins to slow down. In the graph, the elbow point appears to be in three clusters; this is where the decrease in WCSS begins to plateau. We chose this point as the optimal number of clusters because, beyond this point, additional clusters only result in a slight decrease in WCSS. According to the graph, the optimal number of clusters for the data is 3.

Table 2. Number of Clusters and Centroid Initiation

Label	Average
1	55.2
5	9.6
0	43.5

Table 2 shows the initial centroids for the K-Means cluster, along with three randomly selected centroids. Random selection of initial centroids is the first step in the K-Means algorithm. The clusters optimized in the next iteration will start from these centroids. The K-Means algorithm will use each row in the table as an initial centroid for the clusters. In the poor population grouping, Label 1 represents the centroid of the population group with an average value of 55.2. Label 5 is the centroid representing the population group with a lower average value of 9.6. Label 0 is the centroid between the other two groups, with an average value of 43.5. Centroid initiation in the K-Means algorithm is a crucial step. The randomly selected centroids will be the starting points of the clusters. Good initiation will help achieve faster convergence and produce more representative clusters. The Number of Clusters (k = 3) determines to divide the dataset into three clusters. The mean values of 55.2, 9.6, and 43.5 indicate significant differences in the data groups. This indicates a distinct variation in the dataset, which K-Means can identify and cluster. After centroid initiation, the next step in the K-Means algorithm is to calculate the distance between each data point and the nearest centroid, cluster the data based on the closest distance, and update the centroid positions based on the average of the data positions in each cluster. This process is repeated until convergence.

Table 3. Distance between Data Point and Centroid

Label	Average	Cluster	Distance
0	43.5	2	0.00
1	55.2	0	0.00
2	96.4	0	2884698.43
3	55.7	0	18.06
4	87.5	0	1107314.24
5	9.6	1	0.00
6	56.9	0	777.85
7	42.5	2	2500.00

Table 3 shows the distance between data points and the centroid in the clustering of poor people in Bandar Kumbul Village. This table details the assignment of each data point to a specific cluster and the distance between the data point and its cluster centroid. Cluster 2 groups data points with Label 0 and Label 7. Cluster 0 groups data points with Label 1, Label 2, Label 3, Label 4, and Label 6. Cluster 1 groups the data point with Label 5. Five data points are distanced by 0.00 from their centroids, indicating that they are the centroids of their own clusters or very close to the centroid. The data point with Label 2 has a very large distance (2884698.43) from its cluster's centroid. That means the data point may be an outlier or far from its cluster center. Cluster 2 groups data points with lower mean values, like 43.5 and 42.5. Cluster 0 groups data points with higher mean values, including 96.4, 55.7, and 87.5. Cluster 1 groups only one data point with a mean value of 9.6, suggesting a potentially very different distribution in this dataset. Cluster 0 has a wide variation in mean values (from 55.2 to 96.4) and different distances from the centroid. This suggests that

Cluster 0 may be a more heterogeneous cluster. Cluster 1 contains one data point with a very low mean value, indicating that this is a small cluster or even an outlier. Cluster 2 has a lower mean value and a smaller distance from its centroid, indicating that it may be more homogeneous.

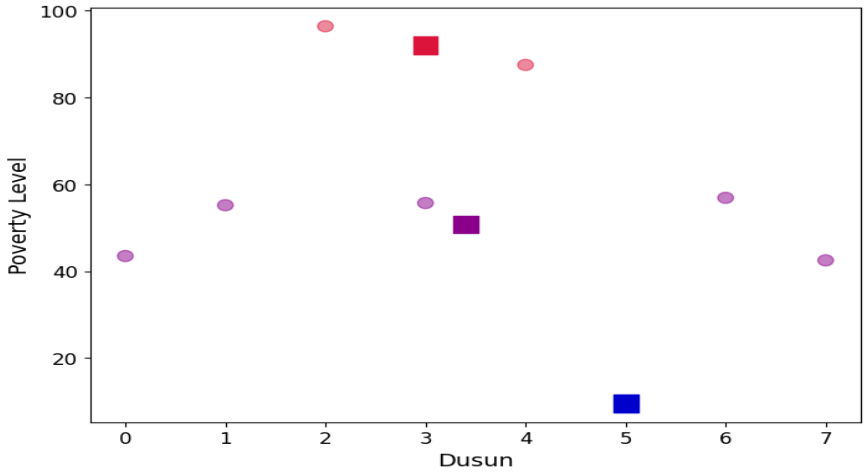

Fig 7. Clustering Results

Figure 7 represents the outcome of clustering using the K-Means method to group impoverished individuals in Bandar Kumbul Village, Labuhanbatu Regency. This section provides an explanation and in-depth analysis of the image. The X-axis (Dusun) is the horizontal axis showing the Hamlet number from 0 to 7. The Y-axis (Poverty Level) is the vertical axis showing the poverty level, with a scale from 0 to 100. The size of the point describes the number of residents or several factors related to the poverty level in the hamlet. The point's color changes, signifying the group or cluster that emerges from the K-Means clustering algorithm. From the image, there are 3 main striking colors, namely red, purple, and blue. This suggests that the clustering results produce three main groups of poor people in the village. Hamlets 2 and 4 have high poverty levels (approaching 100) and are in one cluster. Hamlets 0, 1, 3, 6, and 7 have low poverty levels. Hamlet 5 has a medium poverty level (close to 0) and is in a different cluster itself.

Table 4. Clustering Results

Dusun	Average	Cluster	Distance	Poverty Level
Aek Mardomu	43.5	2	4,130.32	Low
Bandar Kumbul	55.2	2	648.90	Low
Hutaimbaru	96.4	0	432.74	High
Mailil Jae	55.7	2	603.37	Low
Mailil Julu	87.5	0	432.74	High
Pasir Sidimpuan	9.6	1	0.00	Medium
Sidodadi	56.9	2	1,976.66	Low
Singga Mata	42.5	2	6,591.43	Low

Table 4 shows the clustering results of poor people from Bandar Kumbul Village. Cluster 0 with a high poverty rate consists of Hutaimbaru Hamlet (average 96.4, distance 432.74) and Mailil Julu Hamlet (average 87.5, distance 432.74). Both of these hamlets are in a cluster with a very high poverty rate. They are very close to the cluster's center, indicating that they are very representative of it. Cluster 1 with a medium poverty rate consists of Pasir Sidimpuan Hamlet with an average of 9.6 and a distance of 0.00. The only hamlet in this cluster is 0 from the center, indicating that it is the center. Cluster 2 with low poverty levels consists of Aek Mardomu Hamlet (average 43.5, distance 4,130.32), Bandar Kumbul Hamlet (average 55.2, distance 648.90), Mailil Jae Hamlet (average 55.7, distance 603.37), Sidodadi Hamlet (average 56.9, distance 1,976.66), and Singga Mata Hamlet (average 42.5, distance 6,591.43). The hamlets in this cluster show low poverty levels, but the distance from the cluster center varies, with Singga Mata Hamlet having the furthest distance, indicating variation in data distribution within this cluster.This study uses the K-Means Clustering method to group the poor population in Bandar Kumbul Village based on data on the number of poor people in various hamlets from 2013 to 2022. Hutaimbaru Hamlet and Mailil Julu Hamlet demonstrate a significant increase in the number of poor people during the period.

This may be due to various economic, social, and environmental factors that require further analysis. Pasir Sidimpuan and Singga Mata Hamlets show stability in the number of poor people, with relatively small fluctuations, indicating that previous interventions may have been successful or economic conditions are relatively stable.The 2022 pie chart illustrates the concentration of the poor population in several main hamlets: Hutaimbaru (21.6%), Mailil Julu (19.7%), Sidodadi (12.7%), Mailil Jae, and Bandar Kumbul (each 12.4%). Hamlets with a lower percentage of poor people, such as Pasir Sidimpuan (2.2%), may have better economic conditions or better access to resources. In this context, using the elbow method helps determine the optimal number of clusters that are efficient in grouping the poor population in Bandar Kumbul Village. By choosing the right number of clusters, the study can provide more accurate and useful insights into the distribution of poverty in the area. Cluster 0 has data points with varying mean values and distances from the centroid. This indicates that Cluster 0 is the most heterogeneous cluster, with quite a lot of data variation. Cluster 1 contains only one data point with a very low mean value (9.6). This may indicate that Cluster 1 is a small cluster or even contains an outlier. Cluster 2 contains data points with a lower mean value and a relatively smaller distance from the centroid. This suggests that Cluster 2 is a more homogeneous cluster. Some data points have zero distance from the centroid, indicating that they are the centroid itself or are very close to the centroid. This guarantees a well-chosen centroid for data representation.

Data points with a substantial distance from the centroid, such as in Label 2 and Label 4, indicate that they may be outliers or have characteristics that are very different from the rest of the data in their cluster. The wide variation in the mean values suggests that there are other factors that may need to be considered to better cluster this data, such as additional variables or better data normalization techniques. Cluster 1 consisting of only one data point indicates that there is a need to further analyze that data point's characteristics to understand whether it is an outlier or indeed a valid cluster.Further research should consider using more comprehensive data, including other factors such as education, health access, employment, and infrastructure conditions. Extending the time period of data collection can provide a better view of long-term trends in poverty. You can use other methods like Hierarchical Clustering or DBSCAN, in addition to K-Means Clustering, to compare results and gain a deeper perspective on the data. Consider a multidimensional analysis that looks not only at the average poverty but also at the distribution and disparities within different groups.

IV. CONCLUSION

This study successfully used the K-means clustering method to group the poor population in Bandar Kumbul Village, Labuhanbatu Regency. The elbow method yielded three optimal clusters: the high, medium, and low clusters. In the cluster, there are two hamlets with a high poverty rate, one hamlet with a medium poverty rate, and five hamlets with a low poverty rate. The first cluster with a high poverty rate consists of Hutaimbaru and Mailil Julu Hamlets, with a distance from the cluster center of 432.74. The close distance to the cluster center indicates that these hamlets are very representative of the high poverty cluster. This indicates that most of the population in this hamlet lives below the poverty line and needs immediate intervention. Only Pasir Sidimpuan Hamlet contains the second cluster, which has a medium poverty rate and a distance of 0 from the cluster center. Although the poverty rate is medium, it still requires attention to prevent an increase in poverty. The third cluster with a low poverty rate consists of Aek Mardomu, Bandar Kumbul, Mailil Jae, Sidodadi, and Singga Mata Hamlets. In this cluster, the variation in distance from the cluster center shows different distributions. Singga Mata Hamlet has the farthest distance (6,591.43) from the cluster center, indicating that there is variation in poverty data that needs to be considered. This study provides useful insights for the government and stakeholders to design more targeted intervention programs so that they can reduce poverty levels effectively in Bandar Kumbul Village.

REFERENCES

[1] N. K. Aldawiyah *et al.*, "Analisis Faktor Kemiskinan di Provinsi Sumatera Utara berdasarkan Regresi Komponen Utama," ***Var. J. Stat. Its Appl.,*** vol. 6, no. 1, pp. 63–74, 2024, doi: doi.org/10.30598/variancevol6iss1page63-74.

[2] R. H. Sachrrial and A. Iskandar, "Analisa Perbandingan Complate Linkage AHC dan K Medoids Dalam Pengelompokkan Data Kemiskinan di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, 2023.

[3] C. Marito *et al.*, "Analisis Tingkat Pengangguran Terbuka, Human Capital dan Jumlah Penduduk Terhadap Kemiskinan di Sumatera Utara," ***J. Din. Sos. Budaya,*** vol. 25, no. 2, pp. 287–300, 2023.

[4] H. Awalia, S. Hamdi, and A. Nasrullah, "Perangkap Kemiskinan pada Perempuan Pesisir Pantai Cemara Kabupaten Lombok Barat," ***Sosiol. J. Ilm. Kaji. Ilmu Sos. dan Budaya,*** vol. 25, no. 2, pp. 128–151, 2023.

[5] M. A. Lasaiba, "Perkotaan dalam Perspektif Kemiskinan, Permukiman Kumuh dan Urban Heat Island (Suatu Telaah Literatur)," *Geoforum,* vol. 1, no. 2, pp. 63–72, 2022, doi: 10.30598/geoforumvol1iss2pp63-72.

[6] M. Abdul Rahman, N. S. Sani, R. Hamdan, Z. Ali Othman, and A. Abu Bakar, "A clustering approach to identify multidimensional poverty indicators for the bottom 40 percent group," *PLoS One*, vol. 16, no. 8, p. e0255312, Aug. 2021, [Online]. Available: https://doi.org/10.1371/journal.pone.0255312

[7] F. I. Raharja and R. Trivinata, "Klaster Wilayah Penanggulangan Kemiskinan di Kabupaten Malang," ***J. Ilm. Adm. Publisk,*** vol. 6, no. 2, pp. 231–240, 2020, doi: 10.21776/ub.jiap.2020.006.02.10.

[8] N. Afira and A. W. Wijayanto, "Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki," ***Komputika J. Sist. Komput.,*** vol. 10, no. 2, pp. 101–109, 2021.

[9] A. Novianti, I. M. Afnan, R. I. B. Utama, and E. Widodo, "Grouping of Districts Based on Poverty Factors in Papua Province Uses The K-Medoids Algorithm," ***Enthusiastic Int. J. Appl. Stat. Data Sci.,*** vol. 1, no. 2, pp. 94–102, 2020, doi: 10.20885/enthusiastic.vol1.iss2.art6.

[10] A. Azzahra andA.W.Wijayanto,"Perbandingan Agglomerative Hierarchical dan K-Means Dalam Pengelompokkan Provinsi Berdasarkan Pelayanan Kesehatan Maternal,**"*Sist.J.Sist.Inf.,***vol.11,no.2,pp.481–495, 2022.

[11] F. W. Saputri and D. B. Arianto, "Perbandingan Performa Algoritma K-Means, K- Medoids, dan DBSCAN dalam Penggerombolan Provinsi di Indonesia berdasarkan Indikator Kesejahteraan Masyarakat," ***J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.***, vol. 17, no. 2, pp. 138–151, 2023.

[12] W. Andriyani, A. H. Nasyuha, Y. Syahra, and B. Triaji, "Clustering Analysis of Poverty Levels in North Sumatra Province Using the Application of Data Mining with the K-Means Algorithm,**" *J. Media Inform. Budidarma,*** vol. 7, no. 4, pp. 1971–1979, 2023, doi: 10.30865/mib.v7i4.6867.

[13] N. Istiqamah, O. Soesanto, and D. Anggraini, "Application of the K-Means algorithm to determine poverty status in Hulu Sungai Tengah,**" *J. Phys. Conf. Ser.,*** vol. 2106, no. 1, 2021, doi: 10.1088/1742-6596/2106/1/012027.

[14] S. Annas, B. Poerwanto, S. Sapriani, and M. F. S, "Implementation of K-Means Clustering on Poverty Indicators in Indonesia," ***MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.,*** vol. 21, no. 2, pp. 257–266, 2022.

[15] A. A. Arrosyad, A. I. Purnamasari, and I. Ali, "Implementasi Algoritma K-Means Clustering untuk Analisis Persebaran UMKM di Jawa Barat," ***JATI (Jurnal Mhs. Tek. Inform***., vol. 8, no. 3, pp. 3062–3070, 2024.

[16] S. Prakoso, M. Mulyawan, C. Lukman Rohmat, and F. Fathurrohman, "Pengelompokan Wilayah Jawa Barat Berdasarkan Indeks Kedalaman Kemiskinan Dan Jumlah Penduduk Miskin Menggunakan K-Medoids,**" *JATI (Jurnal Mhs. Tek. Inform.*** vol. 8, no. 3, pp. 2929–2935, 2024, doi: 10.36040/jati.v8i3.9609.

[17] M. A. Hanafiah and A. Wanto, "Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra," ***Int. J. Inf. Syst. Technol.,*** vol. 3, no. 2, pp. 315–322, 2020.

[18] H. Sibarani, Solikhun, W. Saputra, I. Gunawan, and Z. M. Nasution, "Penerapan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia,**" *JATI (Jurnal Mhs. Tek. Inform.***, vol. 6, no. 1, pp. 154–161, 2022, doi: 10.36040/jati.v6i1.4590.

[19] F. I. Putri, R. Damayanti, and Kismiantini, "Penerapan Algoritma K-Means Untuk Mengelompokan Kecamatan Di Kabupaten Gunungkidul Berdasarkan Program Keluarga Harapan," *Pros. Semin. Nas. Mat. Stat. dan Apl.*, vol. 2, pp. 408–418, 2022.

[20] D. Istiawan, "Poverty Mapping in Central Java Province Using K-Means Algorithm,**" *J. Intell. Comput. Heal. Informatics*,** vol. 1, no. 1, pp. 1–4, 2020, doi: 10.26714/jichi.v1i1.5380.

[21] wikipedia, "Bandar Kumbul, Bilah Barat, Labuhanbatu," id.wikipedia.org. Accessed: Jun. 20, 2024. [Online]. Available: https://id.wikipedia.org/wiki/Bandar_Kumbul,_Bilah_Barat,_Labuhanbatu

[22] BPS Kabupaten Labuhanbatu, "Kecamatan Bilah Barat Dalam Angka 2023," Labuhanbatu, 2023. [Online]. Available: https://labuhanbatukab.bps.go.id/publication/download.html