Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of the K-Means Method for Clustering Regency/City in North Sumatra based on Poverty Indicators

Syafira Eka Wardani[1)*], Syaiful Zuhri Harahap[2)], Rahma Muti'ah[3)]
[1,2,3)]Universitas Labuhanbatu, Indonesia
[1)]syafiraekawardani0@gmail.com, [2)]syaifulzuhriharahap@gmail.com, [3)]rmuthea5@gmail.com

Submitted : Jun 3, 2024 | **Accepted** : Jun 19, 2024 | **Published** : Jul 1, 2024

Abstract: Poverty has many negative effects on people's lives, such as difficulty meeting basic needs, limited access to adequate health and education services, and limited economic opportunities. North Sumatra faces significant poverty problems as one of the largest provinces in Indonesia. This requires special attention and a thorough investigation. Reducing poverty is a very important issue for the government of North Sumatra Province. Poverty-alleviation strategies can no longer be applied uniformly. Instead, it is necessary to consider all the factors that cause poverty in each region. This means that the approach that must be given to each regency or city based on its poverty level must be adjusted. To overcome this problem, clustering must be carried out to identify areas with different levels of welfare. The aim of this research is to cluster regencies and cities in North Sumatra Province using the K-means method based on poverty indicator variables. This research only uses three poverty indicators: gross regional domestic product, human development index, and unemployment rate. The optimal number of clusters is determined based on the results of the silhouette coefficient. The research method begins with dataset collection, exploratory data analysis, data preprocessing, and k-means clustering. The value k = 6 produces a silhouette coefficient of 0.4135. This research produced six regency/city clusters. Cluster 1 consists of 11 regencies and 1 city; cluster 2 consists of 1 regency and 2 cities; cluster 3 consists of 4 regencies; cluster 4 consists of 7 regencies; cluster 5 consists of 4 cities; and cluster 6 consists of 2 regencies and 1 city. The variables gross regional domestic product, human development index, and unemployment rate have a big influence on the cluster results. This will enable the government to adopt policies to tackle poverty quickly and effectively.

Keywords: Clustering; K-Means; North Sumatra; Poverty; Silhouette.

INTRODUCTION

Poverty is a problem faced by many countries, including Indonesia. Poverty is a social and economic problem where a person or group of people is unable to fulfill their basic rights to live and develop a dignified life (Novianti, Afnan, Utama, & Widodo, 2020). Because people cannot meet basic needs such as clothing, food, and shelter, the poverty rate in Indonesia continues to increase. On the other hand, this is one of the main problems in conditions of poverty because the average Indonesian population continues to experience hunger and food shortages, which causes the death rate to increase every year (Aprilia & Sembiring, 2021). Poverty has many negative effects on a person's life and society, such as

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

difficulty meeting basic needs, limited access to adequate health and education services, and limited economic opportunities (Sachrrial & Iskandar, 2023).

North Sumatra is the Indonesian province with the fourth largest population, which has an area of 72,981.23 km^2 (Sujarwo, 2022). This province consists of 25 regencies and 8 cities, with the capital being Medan (Badan Pusat Statistik Sumatera Utara, 2024). In 2023, the population in North Sumatra will be 15,386,640 people, with the number of poor people being 1,239,710 people (Utara, 2024b). As one of the largest provinces in Indonesia, North Sumatra faces significant poverty problems that require special attention and thorough analysis (Andriyani, Nasyuha, Syahra, & Triaji, 2023). The North Sumatra Provincial Government is very concerned about reducing poverty as an important issue (Suyani, Arnita, Nabila, & Fitria, 2023). The Medium-Term Regional Development Plan for North Sumatra Province sets the goal of creating an advanced and prosperous society. The government needs to determine priority development areas to improve the welfare of the people of North Sumatra so that the level of welfare can increase and be evenly distributed. To do this, the people of North Sumatra are grouped into groups that have the same characteristics (Sihombing & Sihombing, 2021).

Currently, poverty alleviation strategies can no longer be implemented in a uniform way. On the other hand, it is necessary to consider every aspect of the causes of poverty in each region, which means that it needs to be handled differently for each region according to its poverty level (Novianti et al., 2020; Raharja & Trivinata, 2020). It is difficult to identify which areas have high levels of poverty, which in turn causes the government to be unable to take sufficient action. To solve this problem, clustering must be carried out to identify which areas are included in groups with high poverty levels and groups with low poverty levels. This will enable the government to implement policies to tackle poverty quickly and precisely (Afira & Wijayanto, 2021; Nasution, Windarto, & Fauzan, 2020). The cluster method can group observation objects into various groups or clusters based on similar characteristics (Indah, Sari, & Dar, 2023; Kasim, Bahri, & Amir, 2021; Lubis, Harahap, & Ritonga, 2024; Putri, Damayanti, & Kismiantini, 2022; Sitorus, Masrizal, & Muti'ah, 2023). This grouping is based on existing information about the objects studied (Ni'matuzzahroh, Andrea Tri Rian, & Adrianingsih, 2022).

The cluster method has been widely applied by previous researchers to classify poverty. By using a time series-based clustering approach and a complete linkage algorithm, (Setiawan & Zahra, 2023) grouped poverty in Indonesia with optimal clustering results in 3 groups with low, medium, and high poverty categories. Using the same approach as the poverty case study in East Java, optimal results were obtained using the average linkage method with a silhouette coefficient value of 0.8161 (Riani & Sofro, 2023). The best clustering results were also obtained using the average linkage algorithm (Azzahra & Wijayanto, 2022). In research conducted by (Abdul Rahman, Sani, Hamdan, Ali Othman, & Abu Bakar, 2021; Annas, Poerwanto, Sapriani, & S, 2022; Istiqamah, Soesanto, & Anggraini, 2021), the K-Means algorithm has succeeded in grouping poverty. In research conducted by (Saputri & Arianto, 2023), the best grouping methods have been produced, namely K-Means and DBSCAN, based on the highest Silhouette coefficient and the lowest Davies-Bouldin index.

The application of the clustering method in grouping poverty in North Sumatra has also been carried out by previous research. Using the k-means algorithm, three poverty line clusters were obtained in North Sumatra, with 5 regencies or cities having a high level of poverty, 18 regencies or cities having a medium level, and 10 regencies or cities having a low category (Hanafiah & Wanto, 2020). Based on the human development index indicator, (Sibarani, Solikhun, Saputra, Gunawan, & Nasution, 2022) have succeeded in applying the K-Means method to the lowest cluster of 13, the medium cluster of 1, and the highest cluster of 19. Results of research conducted by (Andriyani et al., 2023) show that the K-means method can help in making more efficient decisions in analyzing poverty level groupings in North Sumatra Province using data from 2013 to 2022. By applying the average linkage method, two clusters have been produced. In grouping regencies and cities in North Sumatra based on poverty indicators, cluster one consists of 25 regencies and 7 cities, and cluster two consists of 1 city (Novitasari & Arofah, 2023). From these studies, it was found that the application of the K-means method has been successful in grouping poverty at the regency/city, provincial, and national levels.

*Syafira Eka Wardani

SinkrOn
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This research aims to apply the K-means method to cluster regencies and cities in North Sumatra Province based on poverty indicator variables. This research only focuses on the use of poverty indicator variables, namely, gross regional domestic product, human development index, and unemployment rate for North Sumatra Province in 2023, and analyzes the influence of these three variables in forming clusters. In this research, the number of clusters (k) is determined based on the results of the silhouette coefficient. It is important to carry out this research to determine the regency and city clusters in North Sumatra province based on poverty indicators so that the relevant government can implement treatment solutions for people's welfare based on different regional clusters according to the level of the problem.

LITERATURE REVIEW

Research related to regency and city clusters in North Sumatra has been carried out by several previous researchers. Research conducted by (Hanafiah & Wanto, 2020) only used one poverty indicator variable, namely, the poverty line per regency or city in North Sumatra, based on data from the North Sumatra Central Statistics Agency from 2017–2019. In research conducted by (Sihombing & Sihombing, 2021), the number of clusters was determined using the elbow method. Meanwhile, the number of indicator variables used are population density, unemployment, gross regional domestic product per capita based on current prices, percentage of poor people, gross enrollment rate at the primary school level, percentage of households that do not have defecation facilities, and hope rate life. Research conducted by (Sibarani et al., 2022) focuses on the use of human development index variables in North Sumatra, which consist of life expectancy, literacy rate, average years of schooling, and per capita expenditure. Research conducted by (Andriyani et al., 2023) used data from 2013 to 2022, which shows the number of poor people by regency or city.

In contrast to previous research, research conducted by (Novitasari & Arovah, 2023) applied a hierarchical method, namely average linkage. In that research, the poverty indicator variables used were malnutrition among children under five, the number of health workers, the population who use water that is not suitable for drinking, the population using PLN, and the number of unemployed. In research conducted by (Suyani et al., 2023), a combined k-means and generalized regression neural network (GRNN) method was applied. The poverty level is mapped using k-means, while the GRNN method is applied for poverty modeling and prediction. In this research, the indicator variables used include population, health, education, unemployment, and asset ownership. The data source used is time series data from 2010–2020, which comes from the Central Statistics Agency. Research conducted by (Hasibuan, Cipta, & Dur, 2024) used a model-based clustering method by applying five poverty indicator variables, namely, human development index, poor population, unemployment rate, gross domestic product, and health.

The gap with the research previously explained is that this study uses three poverty indicator variables, namely, gross regional domestic product, human development index, and unemployment rate. These three variables have a very significant influence on poverty in North Sumatra (N. I. Sari, 2022). Gross regional domestic product has a negative and significant influence in North Sumatra Province (Damanik & Sidauruk, 2020; J. Sari, 2019), meaning that the poverty rate will decrease if the growth rate of gross regional domestic product increases (Aldawiyah et al., 2024). The human development index has a significant (J. Sari, 2019) and negative (Masni, 2023) effect on poverty levels in the regencies and cities of North Sumatra. This is a very useful variable in reducing poverty rates (Marito et al., 2023). The unemployment rate has a positive and significant effect (Masni, 2023) on poverty in North Sumatra, meaning that if the unemployment rate increases, the poverty rate will also increase (Aldawiyah et al., 2024; Marito et al., 2023; Rozaini, Maharani, Azhari, & Maisyaroh, 2024; Syuhada, Yafiz, & Irham, 2024).

METHOD

This stage explains the methodology used to apply the k-means method to clustering regency and city poverty levels in North Sumatra Province. This research method is a combination of methods that have been applied by (Sibarani et al., 2022) and (Andriyani et al., 2023). Figure 1 shows the research stages starting from dataset collection, exploratory data analysis, data preprocessing, implementation of

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the k-mean method, and results. The exploratory data analysis process up to the results stage was carried out using the Python programming language.

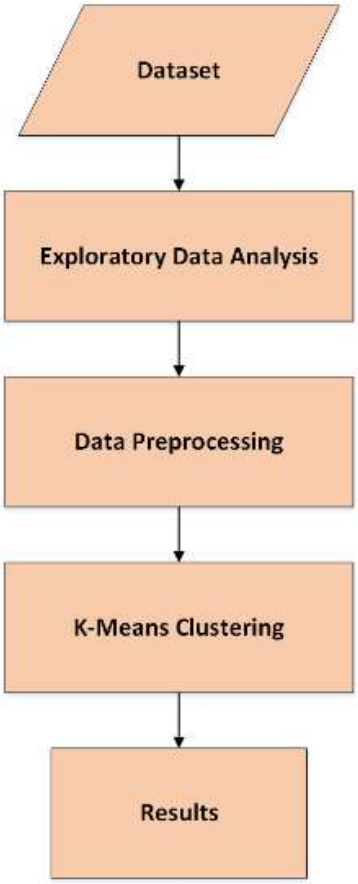


Fig 1. The Research Stage

Dataset

The dataset used in this research is data obtained from the North Sumatra Province Central Statistics Agency (BPS) regarding poverty and its indicators. The data taken is data released in 2023 regarding Gross Regional Domestic Product (Utara, 2024c), Human Development Index (Utara, 2024a), and Unemployment Rate (UR) (Utara, 2024d).

Exploratory Data Analysis

Exploratory data analysis is very useful for studying hidden structures that include important data attributes sequentially by extracting patterns to obtain useful information (Muliani, Sihombing, & Munthe, 2024). At this stage, null and duplicate analysis is carried out to calculate missing values in the dataset and see data redundancy. Next, univariate analysis was carried out on the indicator variables gross regional domestic product, human development index, and unemployment rate. At this stage, the correlation between each variable will also be seen using a correlation heatmap. This stage will also show the distribution of data for each variable by carrying out a graphical plot.

Data Preprocessing

In this phase, the value of each variable that measures poverty is measured through feature scaling. Input values are treated by machine learning as simply numeric numbers, and it cannot understand the units of feature values. Therefore, data normalization is necessary for features whose data does not show a normal distribution. In this study, the dataset was converted into a scale ranging from 0 (min) to 1 (max). This is done by using the MinMaxScaler() function to perform min-max normalization.

*Syafira Eka Wardani

SinkrOn
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

K-Means Clustering

At this stage, the k-means method is implemented. Then, the number of clusters (k value) is determined using the silhouette coefficient. K-Means is implemented in 33 regencies and cities with three poverty indicator variables. This stage begins by determining the k value randomly. The silhouette value function is useful for determining the optimal K value. In this research, researchers used the Scikit-Learn library to determine silhouette values. Silhouette coefficient values can range between -1 and 1. Higher values indicate the best number of clusters.

RESULT

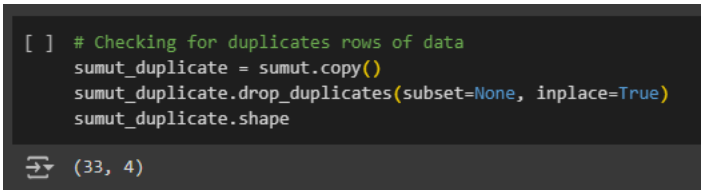
The entire dataset used comes from the North Sumatra Central Statistics Agency. This dataset consists of three poverty indicator variables, namely, gross regional domestic product, human development index, and unemployment rate. The amount of data taken came from 33 regencies and cities in North Sumatra Province. This research conducted clustering based on regency and city poverty levels, which were determined based on cluster results from the K-means method. The data is presented in the form of a.csv file format, as shown in Table 1.

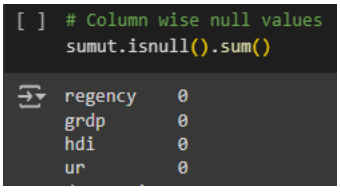
Table 1. The part of Dataset

No	regency	grdp (%)	hdi (%)	ur (%)
1	Nias	3.82	64.56	2.31
2	Mandailing Natal	4.93	68.56	7.45
3	Tapanuli Selatan	5.11	71.55	3.49
4	Tapanuli Tengah	4.23	70.91	7.81
5	Tapanuli Utara	4.75	74.65	1.03
6	Toba	4.93	76.38	1.30
7	Labuhanbatu	5.03	73.69	5.99
8	Asahan	4.87	71.56	6.12
9	Simalungun	5.07	74.29	5.35
10	Dairi	5.04	73.27	1.23

Table 1 shows a sample research dataset from the entire dataset, which consists of 33 rows and 4 columns. In the dataset, there is 1 categorical feature and 3 numerical features with the following description:

1. **regency**: the name of the 33 regencies or cities in North Sumatra Province; this variable includes category features of the object data type.
2. **grdp**: growth rate of gross regional domestic product by business field based on constant prices in 33 regencies and cities of North Sumatra Province in 2023. This variable includes numerical features in the type of float data.
3. **hdi**: human development index in 33 regencies and cities of North Sumatra Province in 2023. This variable includes numerical features with a float data type.
4. **ur**: open unemployment rate for the population aged 15 years according to 33 regencies or cities in North Sumatra Province in 2023. This variable includes numerical features in the type of float data.





(a) Duplicate check *(b) Null check*

Fig 2. Null and Duplicate Analysis

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 2 shows the results of duplication checks and null checks on the dataset, which is the first stage of the exploratory data analysis process. From Figure 2(a), it can be seen that the size of the dataset after checking for duplication is the same as the size of the previous dataset, namely, 33 rows and 4 columns. The result shows that there are no duplicates in the dataset. Figure 2(b) shows a value of 0 for each feature; this proves that there is no missing data value.

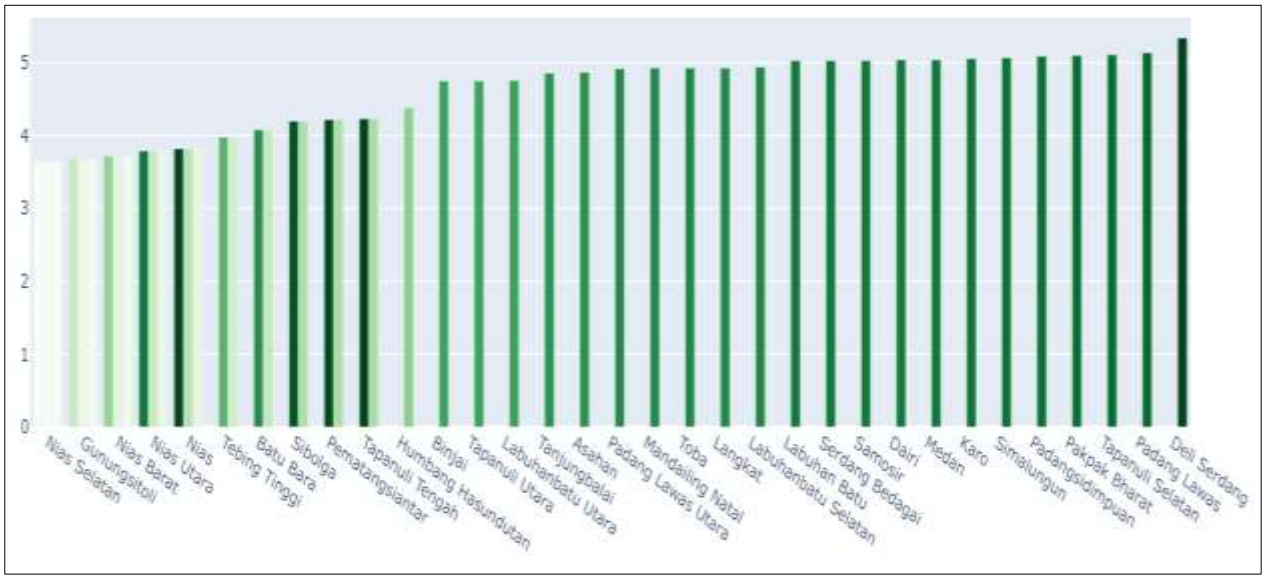


Fig 3. Gross Regional Domestic Product

Figure 3 is a univariate analysis graph of the gross regional domestic product indicator variable. The bottom five regencies and cities based on gross regional domestic product value are South Nias, Gunungsitoli, West Nias, North Nias, and Nias. Meanwhile, Deli Serdang is the regency that has the highest gross regional domestic product in North Sumatra.

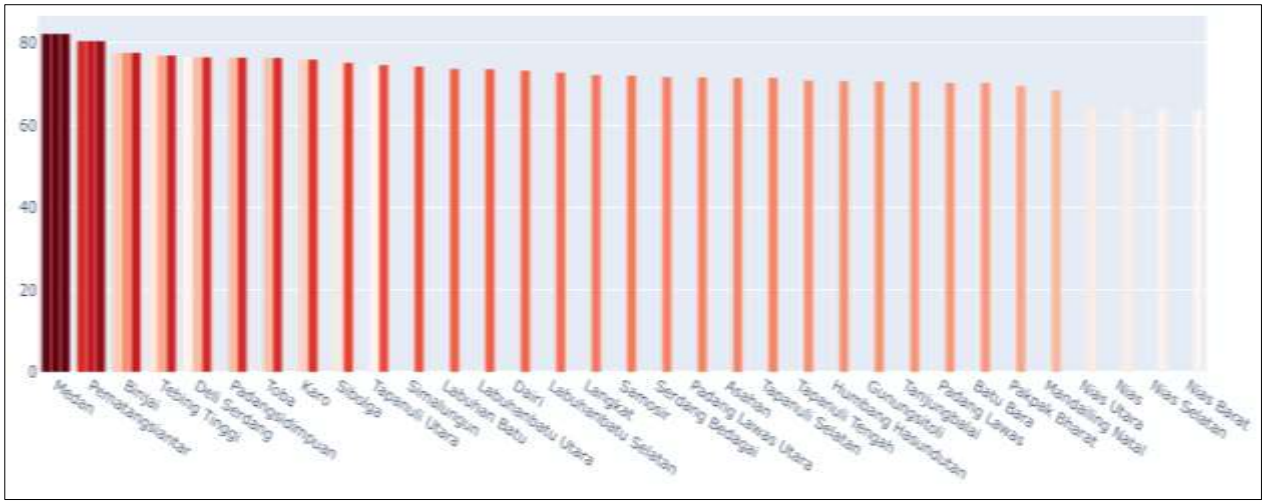


Fig 4. Human Development Index

Figure 4 is a univariate analysis graph of the human development index indicator variable. The bottom four regencies or cities based on the human development index value are West Nias, South Nias, Nias, and North Nias. Meanwhile, the cities of Medan and Pematangsiantar are the ones that have the highest human development index in North Sumatra.

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

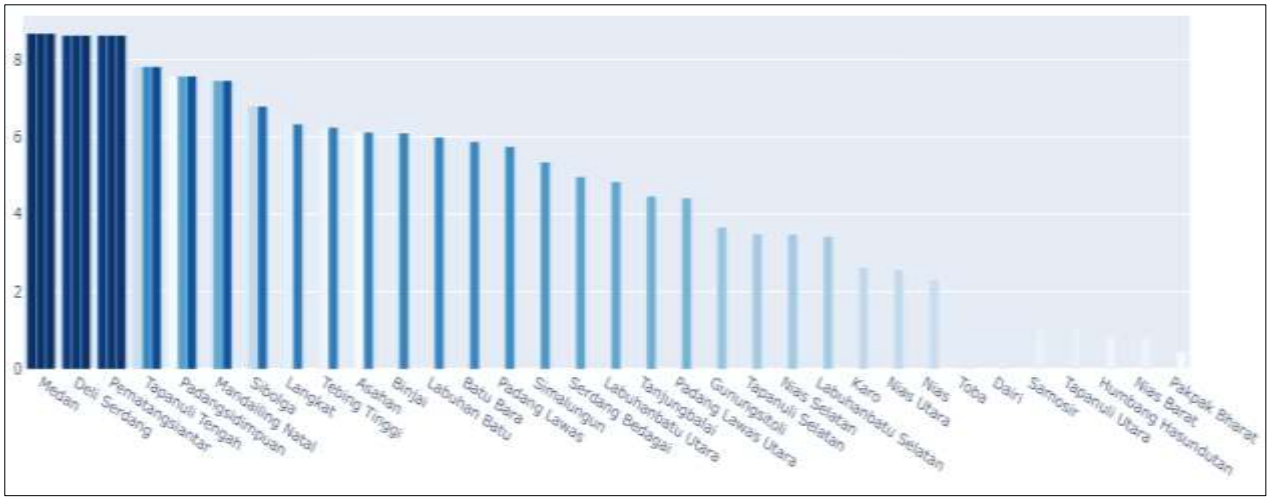


Fig 5. Unemployment Rate

Figure 5 is a univariate analysis graph of the unemployment rate indicator variable. The three regencies or cities that have the highest unemployment rates are Medan, Deli Serdang, and Pematang Siantar. Meanwhile, the regencies that have the lowest unemployment are Pakpak Bharat, West Nias, Humbang Hasundutan, North Tapanuli, Samosir, Dairi, and Toba.

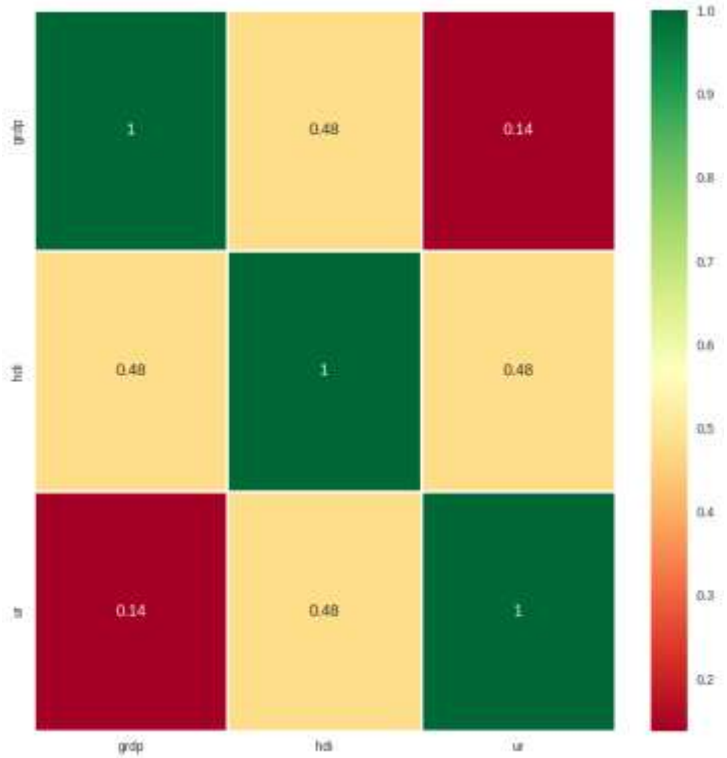


Fig 6. Correlation Matrix

Figure 6 shows the correlation between each feature in the dataset, which is formed using the correlation heatmap function. The grdp and hdi features have a positive correlation of 0.48. This shows that regencies and cities that have a high gdp tend to have higher hdi values. The gdp and your features have a small positive correlation, namely 0.14. Meanwhile, the hdi and ur features also have a positive correlation of 0.48. This also shows that regencies and cities that have high hdi values tend to have higher ur values.

*Syafira Eka Wardani

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

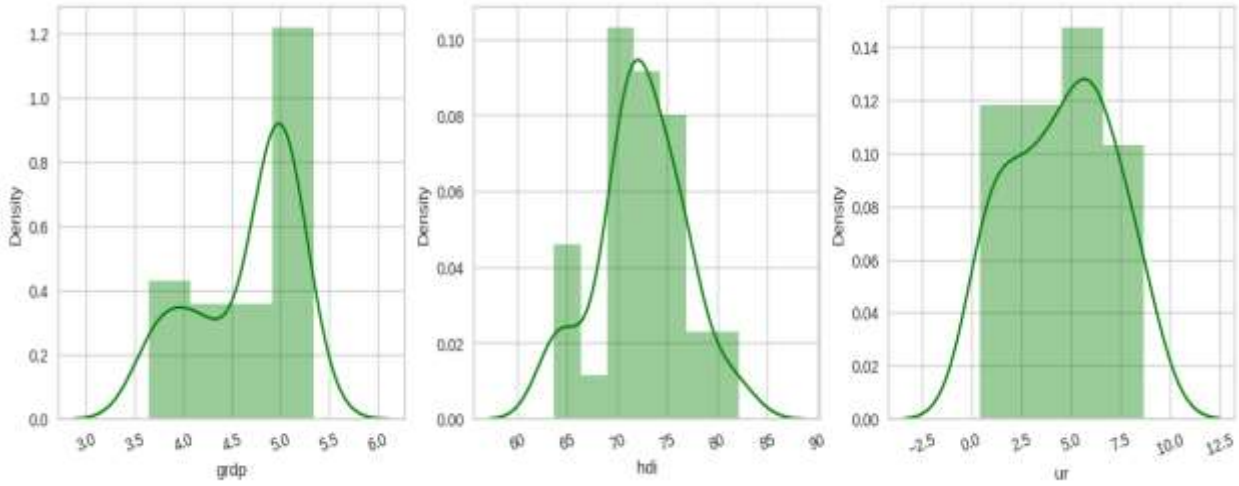


Fig 7. Data Distribution

Figure 7 shows the data distribution of each variable in the dataset. Hdi values tend to follow a normal distribution. Meanwhile, the ur and grdp values do not follow a normal distribution. This shows that the variance and range of data are different. For this reason, it is necessary to standardize the data before building the model. It is important to ensure that features with a larger value range do not exceed the weight of features with a smaller value range. Therefore, scaling down all features to the same normal scale is important.

	regency	grdp	hdi	ur
0	Nias	3.82	64.56	2.31
1	Mandailing Natal	4.93	68.56	7.45
2	Tapanuli Selatan	5.11	71.55	3.49
3	Tapanuli Tengah	4.23	70.91	7.81
4	Tapanuli Utara	4.75	74.65	1.03

Before normalization

	grdp	hdi	ur
0	0.822485	1.000000	1.000000
1	1.000000	0.693348	0.993917
2	0.337278	0.906436	0.993917
3	0.343195	0.389941	0.895377
4	0.852071	0.689021	0.866180

After normalization

Fig 8. Scaling Feature

The preprocessing stage of scaling features is shown in Figure 8. From the data normalization results, it can be seen that the data range has changed between 0 and 1. There is no longer any data that exceeds the 0–1 range.

```
For n_clusters=2, the silhouette score is 0.381197002768855593
For n_clusters=3, the silhouette score is 0.39801614517192635
For n_clusters=4, the silhouette score is 0.3744507081042942
For n_clusters=5, the silhouette score is 0.40062621988 11673
For n_clusters=6, the silhouette score is 0.4134521999071438
For n_clusters=7, the silhouette score is 0.4031704357784583
For n_clusters=8, the silhouette score is 0.3666100177278859
```

Fig 9. Silhouette Coefficient

Figure 9 shows the silhouette coefficient value. This stage begins by determining the k value randomly from the range of 2 to 8. From this process, the optimal k value is obtained in the 6th cluster with a silhouette value of 0.4135. Next, the model will be trained by applying the obtained value k = 6 according to the silhouette coefficient results.

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2. Cluster Results

Cluster	Regency/City	Number	GRDP	HDI	UR
Cluster 1	Mandailing Natal, Tapanuli Selatan, Labuhanbatu, Asahan, Simalungun, Langkat, Serdang Bedagai, Padang Lawas Utara, Padang Lawas, Labuhanbatu Selatan, Labuhanbatu Utara, Tanjungbalai	12	4.97	71.89	5.22
Cluster 2	Deli Serdang, Medan, Padangsidimpuan	3	5.16	78.38	8.29
Cluster 3	Nias, Nias Selatan, Nias Utara, Nias Barat	4	3.75	64.26	2.29
Cluster 4	Tapanuli Utara, Toba, Dairi, Karo, Humbang Hasundutan, Pakpak Bharat, Samosir	7	4.90	73.24	1.22
Cluster 5	Sibolga, Pematangsiantar, Tebing Tinggi, Binjai	4	4.29	77.56	6.94
Cluster 6	Tapanuli Tengah, Batu Bara, Gunungsitoli	3	4.00	70.63	5.79

Table 2 shows the six clusters produced by the k-means method from 33 regencies and cities in North Sumatra based on three poverty indicators. Cities and regencies that are in the same cluster have similar characteristics for each poverty indicator variable used. Cluster 1 consists of 11 regencies and 1 city. The average gross regional domestic product growth was 4.97%, the average human development index reached 71.89%, and the average unemployment rate was 5.22%. Cluster 2 consists of 1 regency and 2 cities. The average gross regional domestic product growth was 5.16%, the average human development index reached 78.38%, and the average unemployment rate was 8.29%. Cluster 3 consists of 4 regencies. The average gross regional domestic product growth was 3.75%, the average human development index reached 64.26%, and the average unemployment rate was 2.29%. Cluster 4 consists of 7 regencies. The average gross regional domestic product growth was 4.90%, the average human development index reached 73.24%, and the average unemployment rate was 1.22%. Cluster 5 consists of 4 cities. The average gross regional domestic product growth was 4.29%, the average human development index reached 77.56%, and the average unemployment rate was 6.94%. Cluster 6 consists of 2 regencies and 1 city. The average gross regional domestic product growth was 4.00%, the average human development index reached 70.63%, and the average unemployment rate was 5.79%.

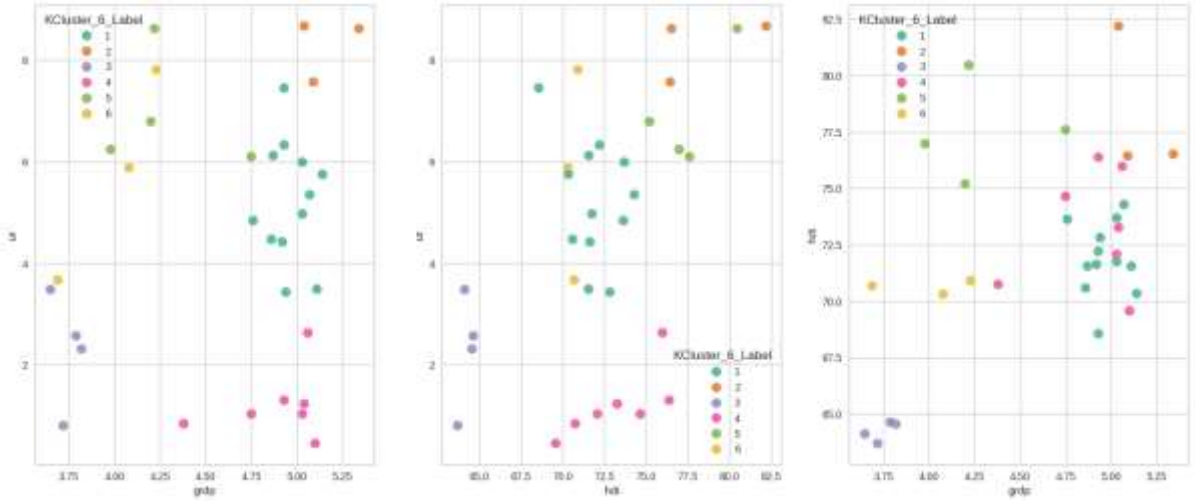


Fig 10. Cluster by scatter plots

Figure 10 is a graph of the scatter plot results that form six regency/city clusters in North Sumatra based on the correlation of poverty indicator variables. It can be seen that cluster 1 is at the middle level for the human development index and unemployment rate variables. Cluster 2 is at the highest level for all variables. Cluster 3 is at the lowest level for all variables. Cluster 4 is at the second-lowest level for the unemployment rate variable. Cluster 5 has a low tendency for the gross regional development product variable. Meanwhile, cluster 6 has a high tendency for the unemployment rate variable.

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

The application of the k-means method has produced six clusters from 33 regencies and cities in North Sumatra province based on three poverty indicators. The largest proportion of regency/city clusters is in cluster 1, with a percentage of 36.4%, consisting of 11 regencies, namely: Mandailing Natal, South Tapanuli, Labuhanbatu, Asahan, Simalungun, Langkat, Serdang Bedagai, North Padang Lawas, Padang Lawas, South Labuhanbatu, North Labuhanbatu, and 1 city, namely Tanjungbalai. The average growth of the gross regional development product is in second place, the achievement of the human development index is in fourth place, and the unemployment rate is in fourth place. The second largest proportion is in cluster 4, with a percentage of 21.2%, which consists of 7 regencies, namely: North Tapanuli, Toba, Dairi, Karo, Humbang Hasundutan, Pakpak Bharat, and Samosir. The average growth of the gross regional development product is in third place, the achievement of the human development index is in third place, and the unemployment rate is in sixth place. The third largest proportion is in clusters 3 and 5, with the same percentage, namely 12.1%. Cluster 3 consists of 4 regencies, namely: Nias, South Nias, North Nias, and West Nias. The average growth of the gross regional development product is in sixth place, the achievement of the human development index is in sixth place, and the unemployment rate is in fifth place. Cluster 5 consists of 4 cities, namely: Sibolga, Pematangsiantar, Tebing Tinggi, and Binjai. The average growth of the gross regional development product is in fourth place, the achievement of the human development index is in second place, and the unemployment rate is in second place. The fourth largest proportion is in clusters 2 and 6, with the same percentage, namely 9.09%. Cluster 2 consists of 1 regency, namely: Deli Serdang, and 2 cities, namely: Medan and Padangsidimpuan. The average growth of the gross regional development product is in first place, the achievement of the human development index is in first place, and the unemployment rate is in first place. Cluster 6 consists of 2 regencies, namely: Central Tapanuli and Batu Bara, as well as 1 city, namely: Gunungsitoli. The average growth of the gross regional development product is in fifth place, the achievement of the human development index is in fifth place, and the unemployment rate is in third place. Based on geographical location, cluster 1 is predominantly occupied by regencies and cities on the east coast of North Sumatra. Cluster 3 is entirely filled by regencies on Nias Island. Cluster 4 is entirely filled with regencies in the northern Tapanuli region.

This research does not categorize which regencies or cities have the highest, middle, or low levels of poverty, as produced by research (Hanafiah & Wanto, 2020), and research (Sihombing & Sihombing, 2021), which concludes that regencies or cities have the highest levels of welfare. lower. However, this research only looks at the influence of the three poverty indicator variables in forming existing clusters. For further research, a comparative test of the results of the silhouette coefficient and elbow method can be carried out to determine the optimal number of k. Apart from that, it is also necessary to conduct more in-depth data exploration to see outliers in the dataset.

CONCLUSION

This research has successfully applied the k-means method to cluster regencies and cities in North Sumatra province based on poverty indicator variables, namely, gross regional domestic product, human development index, and unemployment rate. Based on these three variables, six regency/city clusters have been produced. Cluster 1 consists of 11 regencies and 1 city; cluster 2 consists of 1 regency and 2 cities; cluster 3 consists of 4 regencies; cluster 4 consists of 7 regencies; cluster 5 consists of 4 cities; and cluster 6 consists of 2 regencies and 1 city. The grouping of regencies and cities in cluster 1 is influenced by the average GRDP with the second largest value, along with the average HDI and Unemployment Rate with the fourth largest value. The grouping of regencies and cities in cluster 2 is influenced by the average GRDP, HDI, and Unemployment Rate, which have the largest values compared to other clusters. The grouping of regencies in cluster 3 is influenced by the average GRDP and HDI with the smallest values compared to other clusters, along with the average Unemployment Rate with the second smallest values. The grouping of regencies in cluster 4 is influenced by the average GRDP and HDI with the third largest value, along with average Unemployment Rate with the smallest value. The grouping of cities in cluster 5 is influenced by the lowest average HDI and Unemployment Rate compared to other clusters, along with the average grdp with the fourth largest value. Meanwhile,

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the grouping of regencies and cities in cluster 6 is influenced by the average GRDP and HDI with the second lowest value, along with the average Unemployment Rate with the third largest value.

REFERENCES

Abdul Rahman, M., Sani, N. S., Hamdan, R., Ali Othman, Z., & Abu Bakar, A. (2021). A clustering approach to identify multidimensional poverty indicators for the bottom 40 percent group. *PLOS ONE*, *16*(8), e0255312. Retrieved from https://doi.org/10.1371/journal.pone.0255312

Afira, N., & Wijayanto, A. W. (2021). Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki. *Komputika: Jurnal Sistem Komputer*, *10*(2), 101–109. https://doi.org/10.34010/komputika.v10i2.4317

Aldawiyah, N. K., Astuti, A., Kurnia, R. D., Amalia, N. K., Amelia, D., Mardianto, M. F. F., & Ana, E. (2024). Analisis Faktor Kemiskinan di Provinsi Sumatera Utara berdasarkan Regresi Komponen Utama. *Variance: Journal of Statistics and Its Applications*, *6*(1), 63–74. https://doi.org/doi.org/10.30598/variancevol6iss1page63-74

Andriyani, W., Nasyuha, A. H., Syahra, Y., & Triaji, B. (2023). Clustering Analysis of Poverty Levels in North Sumatra Province Using the Application of Data Mining with the K-Means Algorithm. *Jurnal Media Informatika Budidarma*, *7*(4), 1971–1979. https://doi.org/10.30865/mib.v7i4.6867

Annas, S., Poerwanto, B., Sapriani, S., & S, M. F. (2022). Implementation of K-Means Clustering on Poverty Indicators in Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(2), 257–266. https://doi.org/10.30812/matrik.v21i2.1289

Aprilia, K., & Sembiring, F. (2021). Analisis Garis Kemiskinan Makanan Menggunakan Metode Algoritma K-Means Clustering. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 1–10.

Azzahra, A., & Wijayanto, A. W. (2022). Perbandingan Agglomerative Hierarchical dan K-Means Dalam Pengelompokkan Provinsi Berdasarkan Pelayanan Kesehatan Maternal. *SISTEMASI: Jurnal Sistem Informasi*, *11*(2), 481–495. Retrieved from http://download.garuda.kemdikbud.go.id/article.php?article=2929100%5C&val=10006%5C&title=Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services

Badan Pusat Statistik Sumatera Utara. (2024). Provinsi Sumatera Utara Dalam Angka 2024. In *Badan Pusat Statistik Sumatera Utara* (Vol. 52). Medan.

Damanik, R. K., & Sidauruk, S. A. (2020). Pengaruh Jumlah Penduduk Dan PDRB Terhadap Kemiskinan Di Provinsi Sumatera Utara. *Jurnal Darma Agung*, *28*(3), 358. https://doi.org/10.46930/ojsuda.v28i3.800

Hanafiah, M. A., & Wanto, A. (2020). Implementation of Data Mining Algorithms for Grouping Poverty Lines by Regency/City in North Sumatra. *International Journal of Information System & Technology*, *3*(2), 315–322. https://doi.org/10.30645/ijistech.v3i2.66

Hasibuan, M. A. F., Cipta, H., & Dur, S. (2024). Model Based Clustering for Regency/City Grouping Based on Community Welfare Indicators in North Sumatra. *Jurnal Pijar MIPA*, *19*(3), 479–487. https://doi.org/10.29303/jpm.v19i3.6595

Indah, I. C., Sari, M. N., & Dar, M. H. (2023). Application of the K-Means Clustering Agorithm to Group Train Passengers in Labuhanbatu. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *7*(2), 825–837. https://doi.org/10.33395/sinkron.v8i2.12260 e-ISSN

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Istiqamah, N., Soesanto, O., & Anggraini, D. (2021). Application of the K-Means algorithm to determine poverty status in Hulu Sungai Tengah. *Journal of Physics: Conference Series*, *2106*(1). https://doi.org/10.1088/1742-6596/2106/1/012027

Kasim, R. J., Bahri, S., & Amir, S. (2021). Implementasi Metode K-Means Untuk Clustering Data Penduduk Miskin Dengan Systematic Random Sampling. *Prosiding SISFOTEK*, *5*(1), 95–101.

Lubis, N. J. A., Harahap, S. Z., & Ritonga, I. (2024). Analisis K-Means dan Naive Bayes untuk Pengelompokan Rawan Bencana di Kabupaten Labuhanbatu. *Informatika*, *12*(1), 81–92. https://doi.org/10.36987/informatika.v12i1.5492

Marito, C., Nisa, F., Pardede, N. N., Erza, D. S. D. Y., Sinaga, D., Hidayat, N., & Suharianto, J. (2023). Analisis Tingkat Pengangguran Terbuka, Human Capital dan Jumlah Penduduk Terhadap Kemiskinan di Sumatera Utara. *Jurnal Dinamika Sosial Budaya*, *25*(2), 287–300. https://doi.org/10.26623/jdsb.v25i4.7943

Masni, N. U. (2023). *Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan di Provinsi Sumatera Utara*. Universitas Muhammadiyah Sumatera Utara.

Muliani, S. S., Sihombing, V., & Munthe, I. R. (2024). Implementation of Exploratory Data Analysis and Artificial Neural Networks to Predict Student Graduation on-Time. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *8*(2), 1188–1199. https://doi.org/10.33395/sinkron.v8i2.13658

Nasution, I., Windarto, A. P., & Fauzan, M. (2020). Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. *Building of Informatics, Technology and Science (BITS)*, *2*(2), 76–83. https://doi.org/10.47065/bits.v2i2.492

Ni'matuzzahroh, L., Andrea Tri Rian, D., & Adrianingsih, N. Y. (2022). Clustering Regencies/Cities in Kalimantan Island Based on Poverty Indicators using Agglomerative Hierarchical Clustering (AHC). *Jurnal Matematika, Statistika, Dan Komputasi*, *19*(1), 79–89. https://doi.org/10.20956/j.v19i1.20882

Novianti, A., Afnan, I. M., Utama, R. I. B., & Widodo, E. (2020). Grouping of Regencies Based on Poverty Factors in Papua Province Uses The K-Medoids Algorithm. *Enthusiastic : International Journal of Applied Statistics and Data Science*, *1*(2), 94–102. https://doi.org/10.20885/enthusiastic.vol1.iss2.art6

Novitasari, P., & Arovah, I. (2023). Penerapan Metode Clustering Average Linkage Untuk Mengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Kemiskinan. *MathVvsion: Jurnal Matematika*, *5*(1), 22–27. https://doi.org/10.55719/mv.v5i1.604

Putri, F. I., Damayanti, R., & Kismiantini. (2022). Penerapan Algoritma K-Means Untuk Mengelompokan Kecamatan Di Kabupaten Gunungkidul Berdasarkan Program Keluarga Harapan. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, *2*, 408–418.

Raharja, F. I., & Trivinata, R. (2020). Klaster Wilayah Penanggulangan Kemiskinan di Kabupaten Malang. *Jurnal Ilmiah Administrasi Publisk (JIAP)*, *6*(2), 231–240. https://doi.org/10.21776/ub.jiap.2020.006.02.10

Riani, R. A., & Sofro, A. (2023). Pengelompokan Berdasarkan Garis Kemiskinan Pendekatan Time Series Based Clustering di Provinsi Jawa Timue. *MATHUNESA: Jurnal Ilmiah Matematika*, *11*(03), 478–488.

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Rozaini, N., Maharani, S., Azhari, D., & Maisyaroh. (2024). Pengaruh Pertumbuhan Ekonomi dan Tingkat Pengangguran terhadap Kemiskinan di Provinsi Sumatera Utara. *JESYA: Jurnal Ekonomi & Ekonomi Syariah*, *7*(1), 396–402. https://doi.org/10.36778/jesya.v7i1.1405

Sachrrial, R. H., & Iskandar, A. (2023). Analisa Perbandingan Complate Linkage AHC dan K Medoids Dalam Pengelompokkan Data Kemiskinan di Indonesia. *Building of Informatics, Technology and Science (BITS)*, *5*(2). https://doi.org/10.47065/bits.v5i2.4310

Saputri, F. W., & Arianto, D. B. (2023). Perbandingan Performa Algoritma K-Means, K- Medoids, dan DBSCAN dalam Penggerombolan Provinsi di Indonesia berdasarkan Indikator Kesejahteraan Masyarakat. *Jurnal Teknologi Informasi : Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, *17*(2), 138–151.

Sari, J. (2019). Analisis Tingkat Kemiskinan Masyarakat Di Provinsi Sumatera Utara. *Al-Masharif: Jurnal Ilmu Ekonomi Dan Keislaman*, *7*(2), 290–304. https://doi.org/10.24952/masharif.v7i2.2190

Sari, N. I. (2022). *Identifikasi Faktor-Faktor yang Mempengaruhi Kemiskinan Akibat Pandemi COVID-19 di Sumatera Utara dengan Metode Least Absolute Shrinkage and Selection Operator (LASSO)* (Universitas Islam Negeri Sumatera Utara). Universitas Islam Negeri Sumatera Utara. Retrieved from http://repository.uinsu.ac.id/19116/

Setiawan, D., & Zahra, A. (2023). Pengelompokan Kemiskinan di Indonesia Menggunakan Time Series Based Clustering. *Inferensi*, *6*(1), 83. https://doi.org/10.12962/j27213862.v6i1.14969

Sibarani, H., Solikhun, Saputra, W., Gunawan, I., & Nasution, Z. M. (2022). Penerapan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(1), 154–161. https://doi.org/10.36040/jati.v6i1.4590

Sihombing, S. C., & Sihombing, D. A. (2021). Pengelompokan Tingkat Kesejahteraan Masyarakat di Sumatera Utara dengan Metode K-Means Clustering. *Jurnal Matematika Integratif*, *17*(2), 127. https://doi.org/10.24198/jmi.v17.n2.35025.127-135

Sitorus, M., Masrizal, M., & Muti'ah, R. (2023). Decision Making System for Visitor Satisfaction of Labuhanbatu Library Using K-Means Method. *JINAV: Journal of Information and Visualization*, *4*(1), 36–44. https://doi.org/10.35877/454ri.jinav1591

Sujarwo. (2022). K-Means Clustering Gross Participation Rate Regency/City Area In North Sumatra. *Jurnal Mantik*, *5*(36), 2392–2399. Retrieved from http://www.iocscience.org/ejournal/index.php/mantik/article/view/1844%0Ahttp://www.iocscience.org/ejournal/index.php/mantik/article/download/1844/1567

Suyani, N., Arnita, A., Nabila, R. C., & Fitria, A. (2023). Prediction of the Poor Rate K-Means and Generalized Regression Neural Network Algorithms (Case Study: North Sumatra Province). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *17*(1), 0467–0474. https://doi.org/10.30598/barekengvol17iss1pp0467-0474

Syuhada, M. R., Yafiz, M., & Irham, M. (2024). Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Provinsi Sumatera Utara. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, *9*(1), 513–526. https://doi.org/10.30651/jms.v9i1.21800 Volume

*Syafira Eka Wardani

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13720
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Utara, B. S. (2024a). Indeks Pembangunan Manusia (Metode Baru) 2021-2023. Retrieved from sumut.bps.go.id website: https://sumut.bps.go.id/indicator/26/59/1/indeks-pembangunan-manusia-metode-baru-.html

Utara, B. S. (2024b). Jumlah Penduduk Miskin Menurut Kabupaten/Kota (Jiwa), 2021-2023. Retrieved from sumut bps website: https://sumut.bps.go.id/indicator/12/65/1/jumlah-penduduk-menurut-jenis-kelamin-dan-kabupaten-kota.html

Utara, B. S. (2024c). Laju Pertumbuhan PDRB Menurut Lapangan Usaha Atas Dasar Harga Konstan 2022-2023. Retrieved from sumut.bps.go.id website: https://sumut.bps.go.id/indicator/156/99/1/-seri-2010-laju-pertumbuhan-pdrb-menurut-lapangan-usaha-atas-dasar-harga-konstan-subkategori-.html

Utara, B. S. (2024d). Tingkat Pengangguran Terbuka (TPT) Penduduk Umur 15 Tahun Keatas Manurut Kab/Kota (Persen), 2021-2023. Retrieved from sumut.bps.go.id website: https://sumut.bps.go.id/indicator/6/44/1/tingkat-pengangguran-terbuka-tpt-penduduk-umur-15-tahun-keatas-manurut-kab-kota.html

*Syafira Eka Wardani