K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province

Castaka Agus Sugianto[1)*], Tri Pratiwi Olivia Riska Bokings[2)]
[1)2)]Politeknik TEDC Bandung, Indonesia
[1)]castaka@poltektedc.ac.id, [2)]riskabokings@poltektedc.ac.id

Abstract

The Central Bureau of Statistics is a non-ministerial government institution that reports directly to the President. Based on data from The Central Bureau of Statistics in September 2019, the wealth rate in Indonesia was 9.22% and the number of indigent people in Indonesia reached 24.79 million. The poverty rate in the Bangka Belitung Islands Province was low compared to the national level. This is evidenced by 4.62% of people in Bangka Belitung Island Province were indigent people, which is lower than the national average of 9.22%. The data mining techniques using the K-Means Clustering method are used for this study. The research data was taken from the website of the BPS from 2014-2019 which consisted of 7 districts and/or cities with 3 variables. The variables used are the number of indigent people (in thousands), the average length of school education (years), and adjusted per capita expenditure (thousand rupiahs/year). All data is processed by Rapidminer and 3 clusters are carried out, namely: medium cluster level 0, high cluster level 1, and low cluster level 2. Cluster 0 contains districts/cities whose people have the longest average school time, high per capita expenditure, and a large number of indigent people. Cluster 1 contains districts/cities whose people have a short average school time, low per capita expenditure, a moderate number of indigent people. Cluster 2 contains districts/cities whose people have an average school time, moderate per capita expenditure, a small number of indigent people. Based on the result, the government can prioritize Kabupaten Bangka, Kabupaten Bangka Barat, Kabupaten Bangka Selatan in assisting, especially in the cost of education scholarships and social funds as well as other infrastructure improvements for the welfare of the inhabitants of Bangka Belitung Islands Province.

Keywords: Data Mining; K-Means; Clustering; Bangka Belitung Islands; Poverty

INTRODUCTION

One of the problems faced by Indonesia is the problem of poverty. The problem of poverty is complex and multidimensional so that it becomes a development priority. So far, the government of Indonesia has many programs for poverty alleviation. There are two strategies for poverty alleviation. First, protecting indigent families and groups of people through meeting their needs from various fields. Second, conduct training for them so that they can make efforts to prevent new poverty. In general, the level of education in the province of Bangka Belitung Islands Province has not developed, especially in isolated districts. Based on data from Badan Pusat Statistik (BPS) in 2018, the education success indicator in the form of Angka Partisipasi Sekolah (APS) in the Bangka Belitung Islands Province in the 16-18 years age group only reached 76.41%. The complexity of community problems in the Province of Bangka Belitung Islands has triggered the low APS, especially in socio-economic problems (Sumarno, 2019). Based on BPS's data in 2018 the number of indigent people in the Bangka Belitung Islands reached 4.77 percent and based on research on the Bangka Belitung Island, it is an indigent provincial cluster (Ferezagia, 2018).

The enrollment rate in South Bangka Regency has decreased. This condition occurs because the economy in South Bangka has decreased or the economic capacity of the community is low. This also happened in the other three areas, namely Belitung, Central Bangka, and East Belitung. It is different in two different areas, namely Pangkalpinang City and Bangka Regency, which continues to experience an increase in school enrollment rates that are directly proportional to the decreasing poverty levels of these two areas (Karini, 2018)

To make it easier for agencies to process large amounts of data, a system is needed that can produce a decision to determine the grouping of poverty data in the Bangka Belitung Islands. One of them is using data mining techniques. Data mining is a method used in large-scale data processing. Data processed by data mining techniques then produces new knowledge, the results of data processing can be used in determining decisions. In data mining, some methods

* Corresponding author

can be used such as Classification, Clustering, Estimation, Prediction, Association (Ong, 2013).

Clustering is a method used in data mining that looks for and classifies data that has similar characteristics between one data and another that has been obtained (Ong, 2013). Cluster analysis is one of the techniques in a multivariate analysis which has the main objective of grouping objects so that the objects closest to them are in the same group and those with different characteristics are in other groups (Sajidah, 2016). The clustering techniques most often used in data mining are the K-Means and K-Medoids algorithms (Anggreini & Tresnawati, 2020)(Sugianto, 2015)(Sugianto & Apandi, 2015)(Anggreini, 2019). The characteristic of this clustering method is that it is efficient and fast which can be used in the K-Means algorithm. The way it works is that at first the cluster to be formed is determined, the first element in each cluster can be chosen to be the center point (centroid), then repeating the steps so that no object can be moved again (Alfina & Santosa, 2012). The advantage of the K-Means algorithm is that it can minimize the distance between the data to the cluster. The use of this algorithm in the clustering process depends on the data obtained and what is to be achieved at the end of the process (Agustina et al., 2012). In terms of time, the K-Means algorithm can classify faster when compared to the SFCM algorithm (Widayani & Harliana, 2020). K-means clustering is one of the most efficient partitioning methods and has been used in various applications because of its simplicity and can achieve fast convergence (Fajriani, 2019) The k-means algorithm has been shown to provide more optimal cluster results than the classical algorithm in clustering analysis (Fajriani, 2019) The k-means algorithm is better than the k-medoids algorithm (Sugianto et al., 2020).

LITERATURE REVIEW

Several studies have been conducted by many other researchers previously related to grouping poverty data, as will be explained below:

The research conducted by Suhartini and Ria Yuliani entitled "Application of Data Mining to Cluster Data for the Indigent Using the K-Means Algorithm in Dusun Bagik Endep Sukamulia Timur" classify indigent families to help the village government, so that aid can be distributed appropriately. using the K-Means method. The data used is data in 2019, totaling 200 data with 9 attributes, namely the name of the population, occupation, income/month, the number of children attending elementary school, the number of children attending junior high school, the number of children attending high school, the number of children attending college, the number of children attending elementary school. no school and number of family members. Based on the results of tests carried out by applying the K-Means algorithm, the results obtained with Cluster 1 totaling 18 people with high economic population criteria, Cluster 2 totaling 72 residents with moderate economic population criteria, and Cluster 3 totaling 110 residents with low economic population criteria (Suhartini & Yuliani, 2021).

M. Nanda Variestha Waworuntu and Muhammad Faisal Amin in their research "the application of the k-means method for mapping prospective Jamkesda recipients" for poverty classification can help Kelurahan in making the right decisions to prevent inaccuracy of JAMKESDA recipients. In this study, the application of the K-Means method is implemented in an application made with 2 clusters. This study used a total of 440 data samples. From the results of the calculation of the Davies Bouldin Index, the value of determining the number of clusters is 2 clusters (0.243), 3 clusters (0.256), 4 clusters (0.275). The value used is 2 clusters because the value is close to 0(Waworuntu & Amin, 2018)

Research by Dina Sunia, Kurniabudi Kurniabudi, and Pareza Alam Jusia entitled the application of data mining for clustering data for the indigent using the k-means algorithm, data source from the BPS of the city of Jambi In March 2017 the number of indigent people was 286.55 thousand people (8.19%) Based on these figures, it can be seen that the poverty rate in Jambi City, in general, is still high. Based on these conditions, it is necessary to do clusters to help the Jambi City Social Service in grouping indigent families so that assistance can be distributed appropriately. For clustering using the K-Means algorithm. The results of the calculation show that from 512 samples of indigent population data in South Jambi District, with a total of 5 clusters, with 3 iterations, 13 residents are included as members of cluster 1, 153 residents are in cluster 2, 129 residents are in cluster 3, 138 residents are in cluster 1. cluster 4, and 79 residents are in cluster 5 (Sunia et al., 2019)

Research by Yunita Ratna Sari, Arby Sudewa, Diah Ayu Lestari, and Tri Ika Jaya entitled the application of the k-means algorithm for clustering data on poverty in Banten province using rapid miner. This research utilizes data mining techniques using the K-Means Clustering method. The research data was taken from the website of the Central Statistics Agency (BPS) from 2015-2019 which consisted of 8 districts/cities with 3 variables. The variables used are the number of indigent people (thousand people), the average length of school education (years), and adjusted per

capita expenditure (thousand rupiahs/year). The results of the rapid miner calculation show that Tangerang, Cilegon, and Serang are members of cluster 0, Pandeglang, Lebak, and Serang districts are in cluster 1, Tangerang City, and South Tangerang City are in cluster 2 (Sari et al., 2020).

Meanwhile, the research conducted by Achmad Bahauddin et al entitled "Analysis of Provincial Clustering in Indonesia Based on Poverty Level Using the K-Means Algorithm." In this study, the K-Means algorithm clustering method was used to classify provinces based on their poverty level with the help of Weka software. The results showed that there are 3 provincial clusters in Indonesia based on their poverty level, namely Cluster 0 (provinces with low poverty levels), Cluster 1 (provinces with moderate poverty levels), and Cluster 2 (provinces with high poverty levels). Provinces that are included in the category of provinces with high poverty rates are Maluku, West Papua, and Papua (Bahuddin et al., 2021).

Research by Aniq Noviciatie Ulfah and Shofwatul 'Uyun entitled "Performance Analysis of the Fuzzy C-Means and K-Means Algorithms on Poverty Data". This study uses 15 indicators into 3 groups for the poverty data clustering process. This study aims to analyze the performance between the FCM and K-means algorithms which are implemented in poverty data in Girijati Purwosari Village into 3 clusters. The calculation results are used to compare the FCM algorithm with K-Means. The FCM algorithm has a yield of 50% and the K-Means algorithm is 83.33%. The K-Means algorithm is more appropriate for grouping poverty data based on the three poverty criteria than the FCM algorithm (Ulfah & Uyun, 2015).

Meanwhile, Irmanita Nasution, Agus Perdana Windarto, and M. Fauzan in their research "Application of the K-Means Algorithm in Grouping Data for the Indigent by Province" In this study the data used were sourced from the Central Statistics Agency for 2012-2018. The data mining technique used in data processing is the K-means algorithm. The number of records used was 34 provinces which were divided into 2 clusters, namely the high cluster and the low cluster. The purpose of this study is divided into 2 parts, namely the province group with the highest poverty level and the province group with the lowest poverty level. From the grouping results obtained as many as 8 high clusters and 26 low clusters (Nasution et al., 2020).

Anita Fitria Febrianti et al in her research "k-means clustering with the elbow method for grouping districts and cities in East Java-based on poverty indicators. consists of 5 districts, cluster 3 consists of 19 districts, and cluster 4 consists of 4 districts. The results of the characteristics of the first cluster have a high percentage of the population 15 years and above who work in the agricultural sector which is very high (Febrianti et al., 2018).

The research by Nugroho Irwan Febianto and Nico Dias Palasara entitled K-Means Clustering Analysis on Poverty Information Data in West Java in 2018. 2018. With the characteristics contained in each region, it will certainly be a solid foundation for government administrators to provide the right and fast policies/approaches to alleviate poverty in the region (Febianto & Palasara, 2019)

Meanwhile, Wiwi Widayani and Harliana's research entitled "Comparison of K-Means and SFCM Algorithms for Grouping Indigent Households." This study was conducted to compare the K-Means algorithm and Fuzzy Subtractive Clustering (SFCM) in classifying indigent households. The two algorithms will be compared based on the standard deviation and the validity of the resulting grouping results. Based on 6 tests the K-Means algorithm can classify faster when compared to the SFCM algorithm, but in terms of group standard deviation, standard deviation between groups, and accuracy, the SFCM algorithm has better performance when compared to the K-Means algorithm (Widayani & Harliana, 2020).

Lastly, Disty Wahyuli et al, whose research is entitled "Classifying Poverty Lines by Province Using the K-Medoids Algorithm." In this study, the data used came from the central statistical agency. The research objective was to determine the level of the number of poverty cases by province using k-medoids. In Indonesia, the high poverty rate consists of 23 provinces and the low poverty rate consists of 11 provinces. It is hoped that this research can provide input to the government in increasing employment so that it can improve the economy of the people in Indonesia (Wahyuli et al., 2019).

METHOD

The proposed model is presented in Fig. 1, modeling starting from collecting the dataset to validating the dataset. At this stage the data preparation can be done, data preparation is data cleansing and data integration, then clustering is carried out after the clustering process is validated.
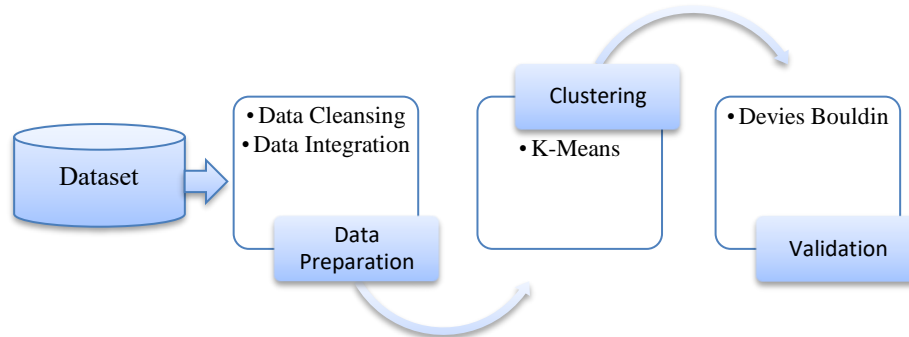
Fig. 1 The Proposed Model

Dataset

The data in this study came from the website of the Central Bureau of Statistics for the Bangka Belitung Islands. Taken from three parts regarding the poverty line, the number of indigent people, and the average length of schooling of the community. The data used are data from 2014 to 2019.

Table. 1
Poverty dataset of the Bangka Belitung Islands from 2014-2019

Region Bangka Belitung	Poverty Standard (Rupiahs/Capita/Year)						Total Population of Indigent						Average Length of School					
	2014	2015	2016	2017	2018	2019	2014	2015	2016	2017	2018	2019	2014	2015	2016	2017	2018	2019
Kepulauan Bangka Belitung	469814	488701	534229	587530	631467	677716	67.2	74.09	72.76	74.09	76.26	68.38	7.35	7.46	7.62	7.78	7.84	7.98
Bangka	375653	385899	427684	475024	497164	541925	16	17.39	17.29	16.45	18.02	16.52	7.92	7.94	7.96	8.19	8.2	8.23
Belitung	563475	580050	610072	652989	704855	766276	12.7	14.58	13.94	14.11	14	11.88	8.07	8.09	8.1	8.11	8.15	8.41
Bangka Barat	388493	399401	440850	497567	520642	549496	6.1	6.01	5.46	6.06	6.35	5.65	6.57	6.68	6.89	7.06	7.18	7.21
Bangka Tengah	512240	528720	586445	653308	680907	693842	9.4	10.19	10.36	11.39	11.12	9.8	6.68	6.7	6.71	6.79	6.8	7.13
Bangka Selatan	364504	374778	415359	462716	483823	523255	7.4	7.2	7.11	7.88	7.58	7.02	5.87	5.88	5.96	6.12	6.36	6.42
Belitung Timur	492652	504668	528572	563309	622396	684277	7.9	8.71	8.48	8.44	8.93	8.51	7.89	7.91	7.95	8	8.14	8.15
Kota Pangkalpinang	509246	527668	604011	654044	700949	748487	7.8	10.02	10.12	9.76	10.27	9	9.63	9.75	9.76	9.77	9.78	9.8

Data Preparation

At this stage the dataset originating from BPS is cleansing and integrated, integration here is combining data from 2014 – 2019, the data can be seen in table 1.

Clustering

For the clustering process, the algorithm used is k-means, K-Means clustering is an algorithm that enables the grouping of similar data based on attributes into a K number of groups. K is a positive integer number. Minimizing the sum of squares of distances between data and then the corresponding cluster centroid is the grouping process in the K-Means algorithm. The weakness of the K-Means algorithm is that the method used to initialize the means is not specified. However, the K-Means method in producing clustering results for many practical applications has proven to be effective and good (Vibha et al., 2007).

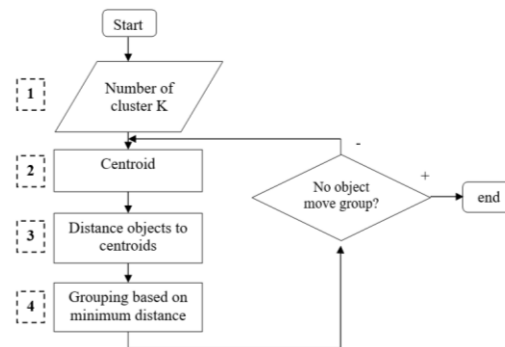Here is a step by step process showing the K-Means clustering algorithm:

Fig. 2 *K-Means Clustering Algorithm.* (Teknomo, 2007)

Concerning Fig. 2 the steps of the K-Means clustering algorithm is further explained as follows:

Step 1. Determine the value of the k = number of clusters

Step 2. Put any initial partition that classifies the data into k clusters. You may assign the training samples randomly, or systematically as follows:

1. Take the first k training sample as single-element clusters
2. Assign each of the remaining (N-k) training samples to the cluster with the nearest centroid. After each assignment, recompute the centroid of the gaining cluster.

Step 3. Take each sample in sequence and calculate its distance from the centroid of each of the clusters. If a sample is not currently in the cluster with the closest centroid, switch this sample to that cluster and update the centroid of the cluster to gain the new sample and the cluster losing the sample.
Step 4. Repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

Validation
To measure clustering performance using the Devices Bouldin Index.

RESULT
The following are the results obtained from research conducted in grouping poverty data in the Province of Bangka Belitung Islands.

Data used
The data is grouped first based on: Average per capita expenditure of the community, Average number of poor people, Average length of schooling for the district/city community in the province of Bangka Belitung Islands from 2014 to 2019, the results can be seen in table 2.

Table 2.
Average Overall Data From 2014-2019

District/City	The average per capita expenditure 2014 - 2019 (thousand rupiahs/year)	The average number of poor people 2014 – 2019 (in thousands)	The average length of school education 2014 - 2019 (years)
Bangka	450558	16.9	8.1
Belitung	646286	13.5	8.16
Bangka Barat	466075	5.9	6.93
Bangka Tengah	609244	10.4	6.8
Bangka Selatan	437389	7.4	6.1

Belitung Timur	565979	8.5	8.01
Kota Pangkalpinang	624068	9.5	9.75

Testing the K-Means Algorithm in Rapid Miner

The process of testing this data is carried out without making any changes to the algorithms contained in Rapidminer. The test is carried out using the k-means algorithm as illustrated in Fig 3.
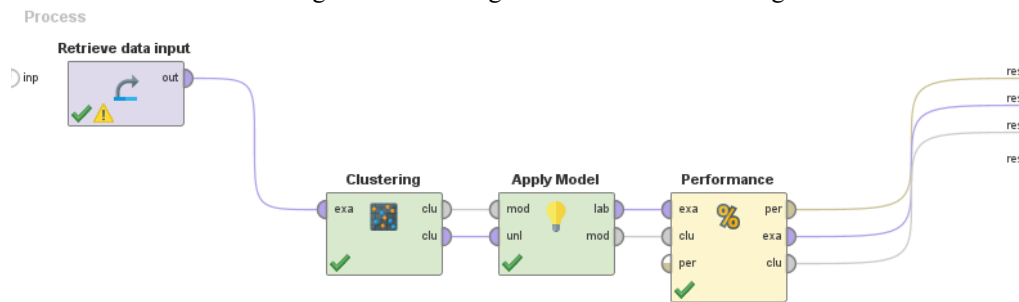

Fig. 3 Clustering Process Using Rapid Miner

By using k-means clustering modeling as shown in Figure 3 above, by initializing the number of clusters as many as 3, the results obtained with the cluster formed are 3, according to the definition of k value with the number of clusters 0 there are 3 items, cluster 1 has 3 and cluster 2 there is 1 item with a total amount of data as much as 7 as for the explanation can be seen in Figure 4.

Cluster Model

```
Cluster 0: 3 items
Cluster 1: 3 items
Cluster 2: 1 items
Total number of items: 7
```

Fig. 4 Cluster Results Using the K-Means Algorithm

Cluster 0: consists of Central Bangka Regency, Belitung Regency, and Pangkalpinang City.
Cluster 1: consists of Bangka Regency, West Bangka Regency, and South Bangka Regency.
Cluster 2: consists of East Belitung Regency.

Where the accuracy based on the average within centroid distance can be seen in Figure 5 with the value of Bouldin Devices of 0.189.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -244242826.033
Avg. within centroid distance_cluster_0: -237796842.680
Avg. within centroid distance_cluster_1: -332103084.730
Avg. within centroid distance_cluster_2: -0.000
Davies Bouldin: -0.189
```

Fig. 5 Results of the K-Means Algorithm Performance

DISCUSSIONS

Results of data clustering related to districts/cities which has the highest number of indigent people with an average length of schooling, and per capita expenditure in the province of Bangka Belitung Islands using Rapid Miner:

1. Bangka Regency, which is included in cluster 1, has an average per capita expenditure of 450,558 thousand rupiahs/year, an average number of indigent people is 16.9 thousand people, an average length of schooling is 8.1 years.
2. Belitung Regency, which is included in cluster 0, has an average per capita expenditure of 646,286 thousand rupiahs/year, an average number of indigent people 13.5 thousand people, an average length of schooling is 8.16 years.
3. West Bangka Regency, which is included in cluster 1, has an average per capita expenditure of 466,075 thousand rupiahs/year, an average number of indigent people is 5.9 thousand people, an average length of schooling is 6.93 years.
4. Kabupaten Bangka Tengah is included in cluster 0 and has an average per capita expenditure of 609,244 thousand rupiahs/year, an average number of indigent people is 10.4 thousand people, an average length of schooling is 6.8 years.
5. South Bangka Regency, which is included in cluster 1, has an average per capita expenditure of 437,389 thousand rupiahs/year, an average number of indigent people is 7.4 thousand people, an average length of schooling is 6.1 years.
6. East Belitung Regency, which is included in cluster 2, has an average per capita expenditure of 565,979 thousand rupiahs/year, an average number of indigent people of 8.5 thousand people, an average length of the school of 8.01 years.
7. The city of Pangkalpinang which is included in cluster 0 has an average per capita expenditure of 624,068 thousand rupiahs/year, an average number of indigent people is 9.5 thousand people, an average length of schooling is 9.75 years.
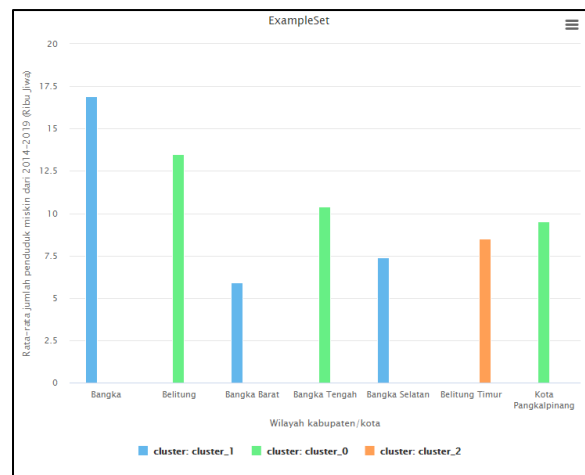

Fig. 6 Visualization of The Average Number of Poor People by Regencies / Cities in The Province of Bangka Belitung Islands From 2014 to 2019
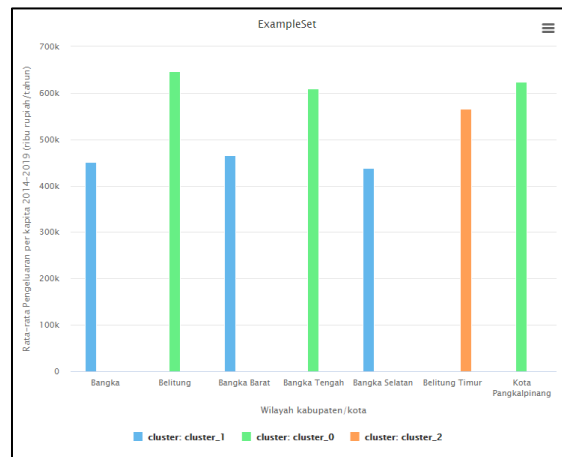
Fig. 7 Visualization of Average per Capita Expenditure of The People of The Districts / Cities of The Bangka Belitung Islands Province From 2014 to 2019
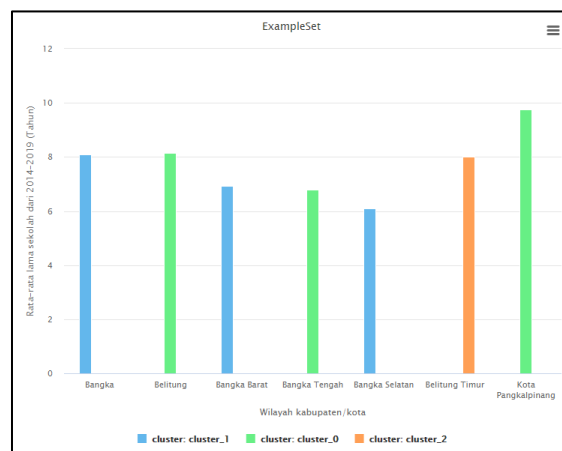

Fig. 8 Visualization of An Average Length of Schooling for The District/City Community in The Province of Bangka Belitung Islands From 2014 to 2019

Cluster 0 (C0): Districts/cities whose people have the longest average school time, high per capita expenditure, and a large number of indigent people.
Cluster 1 (C1): Districts/cities whose people have a short average school time, low per capita expenditure, a moderate number of indigent people.
Cluster 2 (C2): Districts/cities whose people have an average school time, moderate per capita expenditure, a small number of indigent people.

Table 3.
Data Processing Results

District/City	District/city where people have the longest average school time, large per capita expenditures, and a large number of the indigent.	District/city where people have the shortest average school time, low per capita expenditures, and a moderate number of the indigent.	District/city where people have a moderate average school time, a moderate per capita expenditures, and a few numbers of the indigent.
Bangka		C1	
Belitung	C0		
Bangka Barat		C1	
Bangka Tengah	C0		
Bangka Selatan		C1	
Belitung Timur			C2
Pangkalpinang	C0		

CONCLUSION

In this study, the results were obtained from 3 clusters. Cluster (0) is the cluster with the highest poverty rate because it has a large number of indigent people with large per capita expenditures and the longest average length of schooling (education) compared to the other 2 clusters. Cluster (1) is a cluster with a high poverty rate because it has the number of medium indigent people with low/short education levels and the least per capita expenditure compared to other clusters. Furthermore, cluster (2) is the cluster with the lowest poverty rate because it has the least number of indigent people with medium school years and moderates per capita expenditure. In this case, the government can prioritize Bangka Regency, West Bangka Regency, South Bangka Regency in assisting, especially in the cost of education scholarships and social funds as well as other infrastructure improvements for the welfare of the residents of the Bangka Belitung Islands Province.

REFERENCES

Agustina, S., Yhudo, D., Santoso, H., Marnasusanto, N., Tirtana, A., & Khusnu, F. (2012). Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means Algoritma. *Clustering K-Means*, 1–7.

Alfina, T., & Santosa, B. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabugan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS). *Analisa PerbandinganMetode Hierarchical Clustering, K-Means Dan Gabungan Keduanya Dalam Cluster Data*, *1*(1), 1–5.

Anggreini, N. L. (2019). Teknik Clustering dengan Algoritma K-Medoids untuk menangani Strategi Promosi di Politeknik TEDC Bandung. *Junral Teknologi Informasi Dan Pendidikan*, *12*(2), 1–7.

Anggreini, N. L., & Tresnawati, S. (2020). Komparasi Algoritma K-Means dan K-Medoids untuk menangani Strategi Promosi di Politeknik TEDC Bandung. *Jurnal TEDC*, *14*(2), 120–127.

Bahuddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, *4*(1), 1–8.

Fajriani, F. (2019). Persebaran Tingkat Kemiskinan di Sulawesi Selatan. *Seminar Nasional Teknologi Informasi Dan Komputer*, 276–281.

Febianto, N. I., & Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *8*(2), 130–140. https://doi.org/10.32736/sisfokom.v8i2.653

Febrianti, F. A., Antonito, H. C., & Gangga, A. (2018). K-Means Clustering Dengan Metode Elbow Untuk Pengelompokan Kabupaten Dan Kota Di Jawa Timur. *Seminar Nasional Hasil Riset Dan Pengabdian -SNHRP*,

863–870. http://karyailmiah.unipasby.ac.id/wp-content/uploads/2019/04/K-Means-Artikel.pdf

Ferezagia, D. V. (2018). Analisis Tingkat Kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, *1*(1), 1–6. https://doi.org/10.7454/jsht.v1i1.6

Karini, P. (2018). Pengaruh Tingkat Kemiskinan Terhadap Angka Partisipasi Sekolah Usia 16-18 Tahun di Provinsi Kepulauan Bangka Belitung. *Al-Ishlah: Jurnal Pendidikan*, *10*(1), 103–115.

Nasution, I., Windarto, A. P., & Fauzan, M. (2020). Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. *Buliding of Informatics, Technology and Science (BITS)*, *2*(2), 76–83.

Ong, J. O. (2013). Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university. *Jurnal Ilmiah Teknik Industri*, *vol.12, no*(juni), 10–20.

Sajidah, A. (2016). *Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Rakyat dengan Metode c-means dan fuzzy c-means clustering*. http://repository.its.ac.id/51350/

Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, *5*(2), 192–198. https://doi.org/10.24114/cess.v5i2.18519

Sugianto, C. A. (2015). Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan. *Techno.Com*, *14*(4), 336–342. https://doi.org/10.33633/tc.v14i4.992

Sugianto, C. A., & Apandi, T. H. (2015). Algoritma K-Means Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan. *Seminar Nasional Teknologi Informasi Dan Komunikasi Terapan (SEMANTIK)*, 77–81.

Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, *2*(2), 39–44. https://doi.org/10.47292/joint.v2i2.30

Suhartini, & Yuliani, R. (2021). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K- Means di Dusun Bagik Endep Sukamulia Timur. *Infotek:Jurnal Informatika Dan Teknologi*, *4*(1), 39–50.

Sumarno, S. (2019). Angka Partisipasi Sekolah Kasar Sma Rendah Dampak Dari Tingkat Kemiskinan Dan Upaya Mengatasinya Di Provinsi Kepulauan Bangka Belitung. *Jurnal Dinamika Sosial Budaya*, *21*(1), 28–36. https://doi.org/10.26623/jdsb.v21i1.1501

Sunia, D., Kurniabudi, & Jusia, P. A. (2019). Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means. *Jurnal Ilmiah Mahasiswa Teknik Informatika*, *1*(2), 121–134.

Teknomo, K. (2007). *K-Means Clustering Tutorial*. Medicine. http://people.revoledu.com/kardi/Ctutorial/kMean/

Ulfah, A. N., & Uyun, S. (2015). Analisis Kinerja Algoritma Fuzzy C-Means Dak-Mean Pada Data Kemiskinann. *Jatisi*, *1*(2), 139–148.

Vibha, L., Harshavardhan, G. M., Prashanth, S. J., Shenoy, P. D., Venugopal, K. R., & Patnaik, L. M. (2007). A Hybrid Clustering And Classification Technique For Soil Data Mining. *International Conference on Information and Communication Technology in Electrical Sciences*, *Ictes*, 1090–1095.

Wahyuli, D., Handrizal, H., Parlina, I., Windarto, A. P., Suhendro, D., & Wanto, A. (2019). Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, *1*(September), 452–461. https://doi.org/10.30645/senaris.v1i0.51

Waworuntu, N. V., & Amin, M. F. (2018). Penerapan Metode K-Means Pemetaan Calon Penerima JAMKESDA. *Kumpiuan Jurnal Ilmu Komputer (KLIK)*, *05*(02), 190–200. http://klik.ulm.ac.id/index.php/klik/article/view/157 (05 Juni 2020)

Widayani, W., & Harliana, H. (2020). Perbandingan Algoritma K-Means dan SFCM Pada Pengelompokkan Rumah Tangga Miskin. *Jurnal Sains Dan Informatika*, *6*(1), 1–9. https://doi.org/10.34128/jsi.v6i1.200