Research Article

Clustering Analysis Using K-Medoids on Poverty Level Problems in Central Java by District/City

Fitriani Dwi Ratna Sari, Sotya Partiwi Ediwijojo

Politeknik Dharma Patria, Kebumen, Indonesia

Abstract.
Poverty describes the absence of property and poor income or the circumstance that the food, shelter, and clothing needs cannot be met. The performance of Central Java in poverty reduction has risen and declined during the last decade. This research aims to perform mapping analyses utilizing artificial intelligence techniques as clusters on the number of poverty levels in Central Java districts or cities. Since Central Java is after West Java and East Java is the third most populous province, this was necessary to achieve in the last few years through regional mapping of a macro-picture of the poverty level. The dataset used is from the statistics agency website on the number of poor people (millennia) in 2017–2019. The data used are from the Central Java Statistical Agency. The way to map the clusters is using the k-medoids method which is part of data clustering. The number of clusters utilized for mapping poverty levels is high and low. The results showed six provinces (17%) in the high and 29 (83%) in the low. In the high cluster (cluster 1) and in the low cluster (cluster 1) and {18.6, 19.4, 20.1} the final centroid values for each cluster were {293.2, 309.2, 343.5}. The results of mapping can help address the poor in places in which the high cluster (cluster 1), Cilacap District, Banyumas District, Kebumen District, Grobogan District, Pemalang District, and Brebes District are a priority of the government in Central Java province.

Keywords: K-Medoids, clustering, poverty

Corresponding Author: Fitriani Dwi Ratna Sari; email: fitrianidwiratnasaridp@gmail.com

Published 26 May 2023

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICASI Conference Committee.

1. Introduction

Poverty is a complex and phenomenal problem facing developing countries [1], especially the Indonesian people. Poverty is a condition where there is a shortage of everyday necessities like food, clothing, drinking water and shelter. Poverty can also entail a lack of access to education and jobs that can help people solve their problems and gain respect as citizens. The main problem in efforts to reduce poverty in Indonesia today is related to the fact that economic growth is not evenly distributed throughout the region. Thus, one indication of poverty alleviation is increased Economic growth is a concept that incorporates both economic progress and national revenue [2]. The poverty that will be discussed in this paper is precisely in the districts/cities in the province of Central

How to cite this article: Fitriani Dwi Ratna Sari, Sotya Partiwi Ediwijojo, (2023), "Clustering Analysis Using K-Medoids on Poverty Level Problems in Central Java by District/City" in *International Conference on Advance & Scientific Innovation*, KnE Life Sciences, pages 78–87. DOI 10.18502/kss.v8i9.13321

Java. Central Java, after West Java and East Java, is the province with the third highest population. Based on data from the Indonesian Central Statistics Agency in Semester 1 of 2020, Central Java is the second province in Indonesia with the highest number of poor people with 3,980.90 thousand people after East Java with 4,419.10 thousand people. The region of Central Java should be able to become an example for other provinces in Indonesia. However, it is unfortunate because, in reality, the poverty rate in this province is still very high. Although the number of poor individuals has decreased according to the Central Java Provincial Statistics Agency, in previous decades, the unpredictable economic situation in Indonesia, along with the protracted Covid-19 crisis, has the ability to re-increase the poverty rate.

Therefore, this paper proposes an analysis of the poverty level in each district/city in the province of Central Java in the form of clustering which areas have the potential to increase. The purpose of this research proposal is to classify the number of poor people by District/City in Central Java employing regional mapping to generate over the last five years, a macro picture of poverty has emerged (high, medium, low). The research dataset is in the form of data on the number of poor people by regencies/cities in Central Java Province from 2017 to 2019 consisting of 29 regencies and 6 cities obtained from the Central Java Statistics Agency [3]. This research proposal uses the K-Medoids data mining algorithm. Because data mining is an algorithm that is widely used to deal with data classification problems [4]–[8], as well as data clustering [9]–[13].

Several studies that become references in solving this problem, such as research to measure the performance of lecturers in implementing the Tri Dharma of higher education, use the K-Means and K-Medoids method approaches. In this study, the grouping is divided into three categories; satisfying, reasonable, and wrong. The evaluation results show that K-Medoids have a better performance than K-Means. The DBI value for the K-Means technique is -0.417, while the K-Medoids is 0.652. The significant difference indicates that K-Medoids work better in determining lecturer performance [14]. The subsequent research analyses and implements data mining techniques for clustering crude oil exports to destination countries using the K-Medoids method. The result is that K-Medoids can be applied with high cluster results (C1) consisting of 3 countries (Japan, Thailand and the United States) and low cluster (C2) composed of 6 countries (South Korea, Taiwan, China, Singapore, Malaysia and Singapore) [15]. Further research is on taking representative samples for rock connection surfaces of series sizes. According to the roughness statistical frequency distribution, to screen a representative combined sample for each sampling size, a new sampling approach combining the coating principle with the K-Medoids clustering algorithm is suggested. The possibility of achieving an intelligent sampling approach is indicated by sample results that meet the test accuracy standards. The findings of the comparison with the classic stratified sampling method show that the proposed method is more stable [16]. Based on this, this is required in order to use regional mapping to construct a macro picture of poverty levels during the last few years (high, medium, low). So that further action can be taken.

2. Methods

2.1. Flowchart K-Medoids

The flowchart of the K-Medoids Clustering method can be seen in Figure 1.

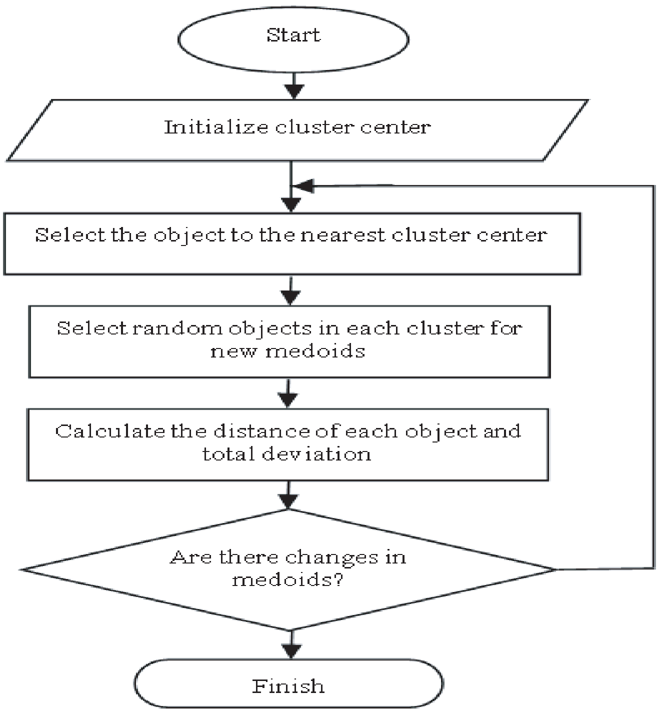

Figure 1: Flowchart K-Medoids.

The steps of the K-Medoids method include: Initialization of cluster centers k (amount of clusters). Using the Euclidian Distance Measure equation, assign each data (object) to the closest Cluster. Objects in each Cluster should be chosen at random as potential newcomers Medoids. Determine the distance between every object in every Cluster and the new Medoids hopeful. Take the difference between the new and old total distances

DOI 10.18502/kss.v8i9.13321

to calculate the total deviation (S). If S is less than zero, to build a new collection of k objects as Medoid, swap items with Cluster data. Steps 3–5 should be repeated until there is no change in the number of Medoids, so that Clusters and their respective Cluster members are obtained.

3. Results and Discussion

3.1. Research Material

The research material used is a data on the number of poor people in Central Java Province based on regencies/cities from 2017 to 2019 consisting of 6 cities and 29 regencies obtained from the Central Java Statistics Agency. The analysis was carried out using the K-medoids technique.

3.2. K-Medoids method analysis calculation

The attribute used as a class is the number of poor people.

TABLE 1: Initial Medoid Value.

No	2019	2018	2017
3	140.10	144.20	171.90
7	131.30	138.30	159.20

Iteration 1

c(1,1) = $\sqrt{(185.20-140.10)^2+(193.20-144.20)^2+(283.30-171.90)^2}$ = 6855.06

and so on until:

c(35,2) = $\sqrt{(18.60-131.30)^2+(19.40-138.30)^2+(20.10-159.20)^2}$ = 33598.72

Iteration 2

c(1,1) = $\sqrt{(185.20-293.20)^2+(193.20-309.20)^2+(238.30-343.50)^2}$ = 24631.04

and so on until:

c(35,2) = $\sqrt{(18.60-9.20)^2+(19.40-9.20)^2+(20.10-9.60)^2}$ = 223.69

S = new total cost – initial total cost

= 128145.25 – 98668.35

= 29476.90

DOI 10.18502/kss.v8i9.13321

Table 2: Calculation Result of Initial Medoid Distance Iteration 1.

No	Central Java Region	C1	C2	Nearest Distance	Data Clustering
1	Cilacap District	6855.06	9324.72	6855.06	C1
2	Banyumas District	19183.19	23182.71	19183.19	C1
3	Purbalingga District	0.00	204.90	0.00	C1
4	Banjarnegara District	238.26	22.12	22.12	C2
5	Kebumen District	8003.70	10531.80	8003.70	C1
6	Purworejo District	9115.28	6724.50	6724.50	C2
7	Wonosobo District	204.90	0.00	0.00	C2
8	Magelang District	219.43	36.11	36.11	C2
9	Boyolali District	5242.65	3477.45	3477.45	C2
10	Klaten District	107.86	226.00	107.86	C1
11	Sukoharjo District	15348.98	12188.36	12188.36	C2
12	Wonogiri District	4146.97	2603.69	2603.69	C2
13	Karanganyar District	7474.57	5342.81	5342.81	C2
14	Sragen District	3093.55	1736.15	1736.15	C2
15	Grobogan District	704.86	1430.00	704.86	C1
16	Blora District	5381.09	3552.33	3552.33	C2
17	Rembang District	5449.93	3644.81	3644.81	C2
18	Pati District	1345.23	525.91	525.91	C2
19	Kudus District	18727.99	15191.23	15191.23	C2
20	Jepara District	8700.30	6355.08	6355.08	C2
21	Demak District	375.00	83.50	83.50	C2
22	Semarang District	13259.29	10296.41	10296.41	C2
23	Temanggung District	12042.95	9256.87	9256.87	C2
24	Kendal District	6828.79	4760.67	4760.67	C2
25	Batang District	14347.77	11317.33	11317.33	C2
26	Pekalongan District	6681.28	4691.50	4691.50	C2
27	Pemalang District	6989.02	9299.04	6989.02	C1
28	Tegal District	1842.22	909.80	909.80	C2
29	Brebes District	56824.66	63335.20	56824.66	C1
30	Magelang City	44265.85	38767.85	38767.85	C2
31	Surakarta City	23231.74	19300.28	19300.28	C2
32	Salatiga City	44697.19	39169.07	39169.07	C2
33	Semarang City	13333.46	10376.28	10376.28	C2
34	Pekalongan City	37741.95	32674.83	32674.83	C2
35	Tegal City	38739.78	33598.72	33598.72	C2

Because the Difference value is > 0, then the Cluster process is terminated. So that the members of each cluster contained in Iteration 1.

Table 3: New Medoid Grades.

No	2019	2018	2017
29	293.20	309.20	343.50
32	9.20	9.20	9.60

According to the findings, six provinces (17%) fell into the high cluster, whereas 29 provinces (83%) fell into the low cluster. Cilacap, Banyumas, Kebumen, Grobogan District, Pemalang, and Brebes District are among the high clusters. While the low clusters are Purbalingga District, Banjarnegara District, Purworejo District, Wonosobo District, Magelang District, Boyolali District, Klaten District, Sukoharjo District, Wonogiri District, Karanganyar District, Sragen District, Blora District, Rembang District, Pati District, Kudus District, Jepara District, Demak District, Semarang District, Temanggung District, Kendal District, Batang District, Pekalongan District, Tegal District, Magelang City, Surakarta City, Salatiga City, Semarang City, Pekalongan City, Tegal City.

4. Conclusions

Based on the results of the study, it can be concluded that the application of the k-medoids method can be carried out in the form of mapping the area of the number of poor people in the province of Central Java. The mapping results can be important information for tackling the poor in Central Java province, namely Cilacap District, Banyumas District, Kebumen District, Grobogan District, Pemalang District, and Brebes District, where the high cluster (Cluster 1) is a priority for the government.

Acknowledgements

Our thanks go to Direktorat Riset dan Pengabdian Masyarakat, Deputi Bidang Penguatan Riset and Pengembangan, Kementerian Riset and Teknologi/ Badan Riset and Inovasi Nasional (Ristek/BRIN) who has provided novice lecturer research grants (PDP) in 2020 for 2021 funding. Furthermore, thank you to the chairman and staff of the Dharma Patria Polytechnic LPPM who have facilitated PDP activities starting from the preparation of proposals to research reports.

Table 4: New Medoid Distance Calculation Results Iteration 2.

No	Central Java Region	C1	C2	Nearest Distance	Data Clustering
1	Cilacap District	24631.04	86335.69	24631.04	C1
2	Banyumas District	10606.69	122148.36	10606.69	C1
3	Purbalingga District	56824.66	44697.19	44697.19	C2
4	Banjarnegara District	63070.24	39350.99	39350.99	C2
5	Kebumen District	22314.16	90078.79	22314.16	C1
6	Purworejo District	111127.50	13514.49	13514.49	C2
7	Wonosobo District	63335.20	39169.07	39169.07	C2
8	Magelang District	62353.13	39923.60	39923.60	C2
9	Boyolali District	96294.91	19411.74	19411.74	C2
10	Klaten District	56817.60	44590.31	44590.31	C2
11	Sukoharjo District	130850.28	7715.25	7715.25	C2
12	Wonogiri District	91416.11	21709.62	21709.62	C2
13	Karanganyar District	105253.55	15701.10	15701.10	C2
14	Sragen District	85531.49	24683.80	24683.80	C2
15	Grobogan District	46277.80	54970.91	46277.80	C1
16	Blora District	96558.75	19258.88	19258.88	C2
17	Rembang District	97178.03	19016.70	19016.70	C2
18	Pati District	75233.53	30716.30	30716.30	C2
19	Kudus District	140232.65	5632.48	5632.48	C2
20	Jepara District	109585.24	14041.95	14041.95	C2
21	Demak District	63856.42	38775.41	38775.41	C2
22	Semarang District	124331.99	9400.96	9400.96	C2
23	Temanggung District	120777.93	10405.68	10405.68	C2
24	Kendal District	102571.01	16704.50	16704.50	C2
25	Batang District	127970.71	8462.12	8462.12	C2
26	Pekalongan District	102251.90	16929.89	16929.89	C2
27	Pemalang District	24315.56	86229.47	24315.56	C1
28	Tegal District	78930.20	28581.55	28581.55	C2
29	Brebes District	0.00	201773.21	0.00	C1
30	Magelang City	200866.67	1.26	1.26	C2
31	Surakarta City	152286.80	3516.93	3516.93	C2
32	Salatiga City	201773.21	0.00	0.00	C2
33	Semarang City	124687.32	9293.85	9293.85	C2
34	Pekalongan City	186661.69	305.10	305.10	C2
35	Tegal City	188846.20	223.69	223.69	C2

References

[1] Bednar DJ, Reames TG. Recognition of and response to energy poverty in the United States. Nat Energy. 2020;5(6):432–439.

Table 5: Cluster Model.

Cluster	Amount of cluster
Cluster 1	6
Cluster 2	29

Table 6: Cluster Results.

Cluster 1	Cluster 2
Cilacap District Banyumas District Kebumen District Grobogan District Pemalang District Brebes District	Banjarnegara District Purworejo District Wonosobo District Magelang District Boyolali District Sukoharjo District Wonogiri District Karanganyar District Sragen District Blora District Rembang District Pati District Kudus District Jepara District Demak District Semarang District Temanggung District Kendal District Batang District Pekalongan District Tegal District Klaten District Purbalingga District Magelang District Surakarta City Salatiga City Semarang City Pekalongan City Tegal City

[2] Wanto A, Hardinata JT. Estimations of Indonesian poor people as poverty reduction efforts facing industrial revolution 4.0. *IOP Conf Series Mater Sci Eng*. 2020;725(1):1–8.

[3] BPS. "Kemiskinan 2017-2019," Badan Pusat Statistik Provinsi Jawa Tengah, 2020. [Online]. Available: https://jateng.bps.go.id/indicator/23/34/1/kemiskinan.html. [Accessed: 25-Oct-2020].

[4] Damanik IS, Windarto AP, Wanto A, Poningsih SR, Andani SR, Andani SW, et al. "Decision tree optimization in C4.5 algorithm using genetic algorithm". *J Phys Conf Ser*. 2019 Aug;1255(1):1–6.

[5] Katrina W, Damanik HJ, Parhusip F, Hartama D, Windarto AP, Wanto A. C.45 Classification rules model for determining students level of understanding of the subject. *J Phys Conf Ser*. 2019;1255(1):1–7.

[6] Siahaan H, Mawengkang H, Efendi S, Wanto A, Perdana Windarto A. Application of classification method C4.5 on selection of exemplary teachers. *J Phys Conf Ser*. 2019;1235(1):012005.

[7] Parlina I, Yusuf Arnol M, Febriati NA, Dewi R, Wanto A, Lubis MR, et al. Naive Bayes algorithm analysis to determine the percentage level of visitors the most dominant zoo visit by age category. *J Phys Conf Ser*. 2019;1255(1):1–5.

[8] Hartama D, Perdana Windarto A, Wanto A. The application of data mining in determining patterns of interest of high school graduates. *J Phys Conf Ser*. 2019;1339(1):1–6.

[9] Hanafiah MA, Wanto A, Indonesia PB. Implementation of data mining algorithms for grouping poverty lines by District/City in North Sumatra. *Int J Inf Sys Technol*. 2020;3(2):315–322.

[10] Febriyati NA, GS AD, Wanto A. "GRDP growth rate clustering in Surabaya City uses the K-Means algorithm". Int J Inf Sys Technol. 2020;3(2):276–283.

[11] Sudirman S, Windarto AP, Wanto A. Data mining tools | RapidMiner : K-Means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia. IOP Conf Series Mater Sci Eng. 2018;420(012089):1–8.

[12] Abbas SA, Aslam A, Rehman AU, Abbasi WA, Arif S, Kazmi SZ. K-Means and K-Medoids: Cluster analysis on birth data collected in City Muzaffarabad, Kashmir. IEEE Access. 2020;8:151847–151855.

[13] Hutagalung J, Ginantra NL, Bhawika GW, Parwita WG, Wanto A, Panjaitan PD. COVID-19 cases and deaths in Southeast Asia clustering using K-Means algorithm. *J Phys Conf Ser*. 2021;1783(1):012027.

[14] Milkhatun M, Rizal AA, Asthiningsih NW, Latipah AJ. "Performance assessment of university lecturers: A data mining approach." *Khazanah Informatika: Jurnal Ilmu*

Komputer dan Informatika. 2020;6(2):73–81.

[15] Rahman F, Ridho II, Muflih M, Pratama S, Raharjo MR, Windarto AP. Application of data mining technique using K-Medoids in the case of export of crude petroleum materials to the destination country. *IOP Conf Series Mater Sci Eng*. 2020;835(1):012058.

[16] Huang M, Hong C, Ma C, Luo Z, Du S, Yang F. A new representative sampling method for series size Rock joint surfaces. *Sci Rep*. 2020 Jun;10(1):9129.