Poverty Mapping in Central Java Province using K-means Algorithm

Research Article

Deden Istiawan [iD]

Department of Statistics, Akademi Ilmu Statistika Muhammadiyah Semarang, Semarang 50185, Indonesia

deden.istiawan@aismuh.ac.id

ABSTRACT

Prosperity has a relative, dynamic, and quantitative meaning. Until now, the formula is not finished because it will continue to grow along with the times. Public welfare is a condition where all citizens are always in a condition that is completely adequate in all their needs. Poverty in Central Java Province is still above national poverty. Poverty grouping is one way to focus on the people's budget in each region so that they can take development policies and strategies that are right on target and effective. In this study, the proposed K-means algorithm for classifying poverty in Central Java is based on poverty indicators. The results of the first cluster study consisted of 22 districts/cities with the category of not poor, the second cluster consisted of 13 districts / cities that were categorized as poor.

Keywords : Property, central java, clustering, k-means.

1. INTRODUCTION

This Welfare has a relative, dynamic and quantitative meaning. The formulation is never final because it will continue to develop along with the development of the needs of human life. Poverty is a problem that is thought to be as old as human age itself. In reality poverty is a manifestation of the results of human interaction invo-lving almost all aspects that humans have in their lives. Poverty alleviation and minimizing inequality are at the core of all development policies. In Indonesia the problem of poverty from year to year continues to enter the national strategic research list 1. (Widiastuti and Yusuf A. G 2012). In 2011, the National Team for the Acceleration of Poverty Reduction (TNP2K) made priority areas of people's welfare in an area that prevailed in any province in Indonesia. These priority areas are made as a benchmark for overall people's welfare in addition to other fields 2 (Ramdhani, Hoyyi, and Mukid 2015).

According to BPS data from 35 regencies / cities in Central Java, there are 15 regencies whose percentage of poor population is above the provincial figure and above the national figure there are 24 regencies. As the program progresses, the problem of poverty cannot be solved. One reason for the failure of the program carried out by the government to overcome the problems of poverty that occur in the community is the occurrence of equalization or uniformity of policies in each region without first paying attention to the problems that cause poverty in these areas, whereas the characteristics and issues of poverty vary in each region. As a result of poverty alleviation programs that are not on target, the economy of the poor is very vulnerable and easily returns to the poverty line. Therefore, poverty alleviation programs that are more targeted are needed in order to truly help the poor improve their living standards. To support the successful implementation of development programs, especially those related to poverty alleviation (Widiastuti and Yusuf A. G 2012).

Mapping and grouping regions based on poverty characteristics and welfare indicators can be done using cluster analysis. Cluster analysis is a technique in data mining that is unsupervised learning. Klastering aims to group unlabeled data into several groups based on their respective characteristics (Mashfuufah and Istiawan 2018).

2. THE MATERIAL AND METHOD

The Classification of poverty is one way to identify the characteristics of the level of welfare of the people in each region so as to take development policies and strategies right on target and effective. Cluster analysis is useful in several exploratory patterns-analysis, grouping, and decision making (Hafiludien and Istiawan 2018). In the last few years, there have been a number of studies that have proposed clustering algorithms for grouping districts/cities based on welfare indicators in Central Java Province. In the last few years, there have been a number of studies that have proposed clustering algorithms for grouping districts / cities based on welfare indicators in Central Java Province. Yulianto and Hidayatullah in (Hafiludien and Istiawan 2018) proposed a hierarchical method of producing three groups of regencies/cities in Central Java based on indicators of people's welfare with variables used including GRDP per capita, population density, poor population, number of labor force, adjusted per capita real expenditure, life expectancy and average length of school.

Another study on welfare grouping was conducted by Putriana in (Hafiludien and Istiawan 2018) of Central Java Province. In this study a comparison of the K-means algorithm and the Hierarchy Method were made. The results showed that the K-means algorithm is superior to the Hierarchy Method by producing three clusters. Research on the classification of poverty in Central Java using the Fuzzy C-Means method was conducted by (Nidyashofa and Istiawan 2017). The study grouped districts / cities in Central Java into three groups based on poverty indicator variables, namely the number of residents, female household heads (KRT), children not in school, disabled individuals, individuals who have chronic diseases, unemployment, unprotected drinking water sources, non-electric lighting sources, cooking fuel using firewood/ charcoal/ kerosene and defecation facilities (BAB) are not available. (Hidayat, Wasono, and Darsyah 2017) also made a comparison between K-means and Fuzzy C- Means algorithms for poverty grouping in Central Java Province. The results also mentioned that the K-means algorithm is better than the Fuzzy C-Means algorithm (Hidayat et al., 2017). Based on the results of the study above, the K-Means algorithm is superior to other algorithms for grouping poverty in Central Java. In this study the K-Means algorithm will be used to classify poverty in Central Java based on poverty indicators.

K-means algorithm grouping objects into several groups or clusters so that objects in one cluster have a high similarity, while between clusters have a very low similarity. The K-means algorithm starts by determining the number of clusters as, then generates random cluster centers. Furthermore, each object will be grouped according to the distance closest to the center of the cluster, the cluster center is updated based on the data points in each cluster. This process is repeated until the convergence criteria are met. The following are the stages of the K-means algorithm:

1. Select cluster center randomly to be the initial cluster center;
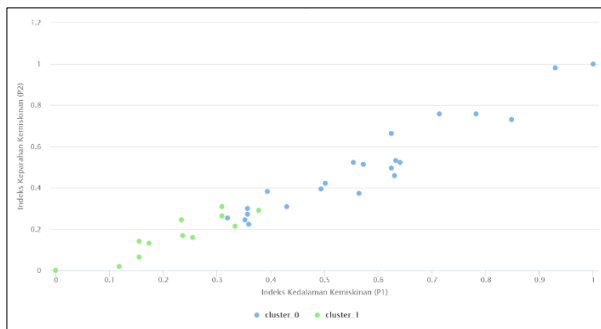2. Allocate all data to the nearest cluster center with distance matrix;

Fig. 1. The k-means grouping results of central java proferty agency.

3. Recalculate the new cluster center based on data that follows each cluster;

4. Repeat steps 3 and 4 until the convergence condition is reached or no data is moved from one cluster to another.

The similarity between the data can be known by calculating the distance between the data types with the cluster center. For the similarity used is the euclidean distance formulated by the following equation:

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(x_i - c_k)^2} \qquad (1)$$

Each cluster representation is relocated to the cluster center by the arithmetic mean of each cluster. This is also what causes this algorithm is often called the cluster mean or centroid cluster like the name that is owned.

3. RESULT AND DISCUSSION

The data used in this study are secondary data published by the Central Java Poverty Agency (BPS) of Central Java Province. The variables used in this study are the percentage of poor population (X1), index into poverty (X2), poverty severity index (X3), poverty line (X4), and Raskin recipient households (X5). In this study, the number of clusters to be formed is two clusters. The results of stabbing can be seen in Table 1 and Fig. 1.

Table 1. The k-means clustering results of central java proverty agency.

No	Districts/Cities	Clusters
1	Cilacap	cluster_0
2	Banyumas	cluster_0
3	Purbalingga	cluster_0
4	Banjarnegara	cluster_0
5	Kebumen	cluster_0
6	Purworejo	cluster_0
7	Wonosobo	cluster_0
8	Magelang	cluster_0
9	Boyolali	cluster_1
10	Klaten	cluster_0
11	Sukoharjo	cluster_1
13	Wonogiri	cluster_0
14	Karanganyar	cluster_1
15	Sragen	cluster_0
16	Grobogan	cluster_0
17	Blora	cluster_0
18	Rembang	cluster_0
19	Pati	cluster_0

No	Districts/Cities	Clusters
20	Kudus	cluster_1
21	Jepara	cluster_1
22	Demak	cluster_0
23	Semarang	cluster_1
24	Temanggung	cluster_0
25	Kendal	cluster_0
26	Batang	cluster_0
27	Pekalongan	cluster_0
28	Pemalang	cluster_0
29	Tegal	cluster_1
30	Magelang	cluster_1
31	Surakarta	cluster_1
32	Salatiga	cluster_1
33	Semarang	cluster_1
34	Pekalongan	cluster_1
35	Tegal	cluster_1

The results of the K-Means algorithm clustering can be seen in Table 1, cluster 0 consists of 22 districts/cities in Central Java. Cluster 0 is a group of regencies/cities that are not poor. Characteristics in cluster 0 are the percentage of poor population, poverty depth index, poverty severity index, poverty line and low percentage of Raskin recipients. Cluster 1 consists of 13 districts/cities, in this cluster are poor clusters due to the percentage of poor population, poverty depth index, poverty severity index, poverty line and high percentage of Raskin recipients.

4. CONCLUSION

Application of K-Means algorithm in clustering poverty in Central Java province with 2 clusters. The first cluster consists of 22 districts / cities with the category of not poor, the second cluster consists of 13 districts / cities that are categorized as poor. The results of the grouping are very important to get a picture of the problem of poverty so that they can take policies in terms of setting targets and providing recommendations for interventions for people with social welfare at the level of Central Java Province. Interventions are emphasized on poor districts / cities so they can be suppressed as low as possible and seriously handled to realize equitable development to all levels of society.

REFERENCES

Hafiludien, Amrul, and Deden Istiawan. 2018. "Penerapan Algoritma Self Organizing Maps Untuk Pemetaan Penyandang Kesejahteraan Sosial (PMKS) Di Provinsi Jawa Tengah Tahun 2016." Pp. 84–92 in *The 7th University Research Colloqium 2018 : URECOL*. Surakarta: STIKES PKU Muhammadiyah Surakarta.

Hidayat, Rahman, Rochdi Wasono, and Moh Yamin Darsyah. 2017. "Pengelompokan Kabupaten/Kota Di Jawa Tengah Menggunakan Metode K-Means Dan Fuzzy C-Means." Pp. 240–50 in *Seminar Nasional Pendidikan, Sains dan Teknologi*. Semarang: Universitas Muhammadiyah Semarang.

Mashfuufah, Syayidati, and Deden Istiawan. 2018. "Penerapan Partition Entropy Index , Partition Coefficient Index Dan Xie BeniIndex Untuk Penentuan Jumlah Klaster Optimal Pada Algoritma Fuzzy C-Means Dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah." Pp. 51–60 in *The 7th University Research Colloqium 2018 : URECOL*. Surakarta: STIKES PKU Muhammadiyah Surakarta.

Nidyashofa, Nurika, and Deden Istiawan. 2017. "Penerapan Algoritma Fuzzy C-Means Untuk Pengelompokan Kabupaten / Kota Di Jawa Tengah Berdasarkan Status Kesejahteraan Tahun 2015." Pp. 23–30 in *The 6th University Research Colloquium 2017 : URECOL*. Magelang: Universitas Muhammadiyah Magelang.

Ramdhani, Fitra, Abdul Hoyyi, and Moch. Abdul Mukid. 2015. "JURNAL GAUSSIAN Vol . 4 , No . 4 , Tahun 2015 Halaman." *Jurnal Gaussian* 4(2006):875–84.

Widiastuti, Minawati, and Edy Yusuf A. G. 2012. "Pemetaan Kemiskinan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2002 Dan 2010 Menggunakan Analisis Klaster." *Diponegoro Journal of Economics* 1(1):1–14.