KnE Social Sciences
ICIGR 2021
ICIGR Conference Proceedings
Volume 2022

Knowledge E
enriching | engaging | empowering

Research article

Cluster Analysis for Grouping Districts in Sidoarjo Regency Based on Education Indicators

Cindy Cahyaning Astuti* and Vanda Rezania

Universitas Muhammadiyah Sidoarjo, Jalan Mojopahit 666B Sidoarjo, Indonesia

ORCID
Cindy Cahyaning Astuti: https://orcid.org/0000-0003-3128-0458

Abstract.
The purpose of education is to develop self-ability, knowledge, skills, and habits that are passed down from one generation to the next and from educators to students through a teaching process to shape a personality physically and spiritually. It also serves as a benchmark of success for a region. Education indicators can be used as a measuring tool to analyze the quality of education in an area. The current study aimed to determine the level of education in the districts of the Sidoarjo Regency, Indonesia. In this study, the sub-districts of the Sidoarjo Regency were grouped based on the education indicators using cluster analysis. Cluster analysis is a multivariate analysis that groups objects into different categories. Based on the results, two clusters were formed. Of the 18 districts in Sidoarjo Regency, the first cluster comprised of 14 districts (Prambon, Tulangan, Krembung, Tarik, Wonoayu, Gedangan, Porong, Buduran, Candi, Sukodono, Tanggulangin, Sedati, Jabon, Balongbendo), while the second included 4 (Sidoarjo, Waru, Taman, Krian). The results showed higher education indicators in the second cluster. Therefore, the researchers recommend using the results of this study as a reference for developing an equal distribution of education in the Sidoarjo Regency.

Keywords: education indicators, cluster analysis, Sidoarjo districts

Corresponding Author: Cindy Cahyaning Astuti; email: cindy.cahyaning@umsida.ac.id

Published 20 June 2022

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICIGR 2021 Conference Committee.

1. Introduction

Education indicators in area can be used as measuring tool to see how well the quality of education in that area. Based on Law Number 2 of 1989 concerning the National Education System, it stipulates that basic education is held to develop attitudes and abilities as well as to provide basic knowledge and skills needed to live in society and to prepare students who meet the requirements for secondary education. Therefore, Indonesian government has launched a 9 year compulsory education program which means that it is mandatory to receive a minimum of basic education, namely from elementary to junior high school which is stated in PP No. 47 of 2008 to maintain quality and quality, education in Indonesia uses National Education Standards [1]. Government of Sidoarjo Regency in collaboration with the Sidoarjo Regency Central Statistics Agency

published the book ”Sidoarjo Dalam Angka 2021” [2] which is an annual publication in the form of a comprehensive collection of data from various sectors. The main purpose of this publication is provide an overview of the global situation in Sidoarjo Regency every year. One sector that plays an important role development of region is education sector. Data is an important supporting sector in regional development planning. Based on the existing education indicator data, it can be seen that the distribution of basic education in Sidoarjo Regency can be seen. One method that can be applied to analyze Education indicators is to use cluster analysis method. In the context of equal distribution of education in order to improve welfare of the community, research was carried out on Education indicators in Sidoarjo Regency. Education indicators can be used as a measuring tool to see how good the quality of education in an area. To find out the distribution of basic education in this study, sub-districts in Sidoarjo Regency were grouped based on education indicators using cluster analysis. Sidoarjo Regency consists of 18 sub-districts, with a total of 12 variables used in this study. The results of the cluster analysis on the Education indicator can be used as a reference for the distribution of basic education in Sidoarjo Regency. Several relevant previous studies have been conducted by Oktavianty et al [3] about the grouping of Education indicators in districts or cities in Sulawesi. The results showed that there were 4 clusters formed based on Education indicator variable in the Districts in Sulawesi. Another study was conducted by Sidik and Firmansyah [4] on the evaluation of development performance in the education sector in Ngawi Regency. The results of study indicate that it is necessary to increase the bureaucracy in the field of education to improve the quality of education in Ngawi Regency. Based on the description above, in this study an analysis of education indicators in Sidoarjo Regency was carried out using cluster analysis.

2. Research Methods

Analysis Data analysis used in this research is cluster analysis using orange statistical software. Cluster analysis used is Hierarchical Cluster Analysis with Complete Linkage algorithm. Cluster analysis is a technique used to classify objects into relatively homogeneous groups, called clusters. Objects in each group tend to be similar to each other and very different from objects from other clusters. Cluster analysis is also called classification analysis or numerical taxonomy, because it relates to the clustering procedure where each object is only included in one cluster, there is no overlapping or overlapping [5]. Several stages in cluster analysis are measuring the similarity between objects and creating clusters. According to principle of cluster analysis, namely grouping

objects that have similarities, the first process is measure how far there are similarities between objects. The method used to measure correlation between a pair of objects on several variables and measure the distance between two objects. There are several kinds of measurement methods, but the most popular and most frequently used method is Euclidian Distance method [6].

Next is stage of creating clusters. There are two types of methods in making clusters, namely hierarchical cluster analysis and non-hierarchical cluster analysis. The hierarchical cluster analysis method begins with grouping two or more objects that have the closest similar characteristics [7]. Then the process is passed to another object that has the second proximity and so on. While the non-hierarchical cluster analysis method begins with the desired number of clusters (two clusters, three clusters and so on). After the number of clusters is known, then the cluster process is carried out without following the hierarchical process. This method is also called the k-means cluster method [8]. In the cluster hierarchy method, there are two basic types, namely agglomerative and divisive. In agglomerative method, each object or observation is considered as a separate cluster. In the next stage, two clusters that have similarities are combined into a new cluster and so on. In contrast, in the divisive method, from a large cluster consisting of all objects or observations. Furthermore, the objects or observations with the highest dissimilarity values are separated and so on [9].

There are four approaches in hierarchical cluster analysis namely, Single Linkange, Complete Linked, Average Linked and War'd Linked. Single Linkage is a method that groups two objects that have closest distance first. Complete Linkage is a complete linking method based on maximum distance. The distance between one cluster and another is measured based on object that has the furthest distance. Average Linkage, is a method that groups objects based on variations of the single linkage and complete linkage algorithms, namely calculating the distance between two clusters which is called the average distance where the distance is calculated in each cluster by minimizing the average distance between the combined cluster pairs. Ward's Linked is a method that groups objects with the aim of obtaining clusters that have smallest possible internal cluster variance [10].

There are 12 variables used in this study. This variable is an indicator of education in Sidoarjo district. These variables include the following:

X1 = Number of State Elementary Schools (SD) in Sidoarjo

X2 = Number of Private Elementary Schools (SD) in Sidoarjo

X3 = Number of State Elementary School (SD) Teachers in Sidoarjo

X4 = Number of Private Elementary School (SD) Teachers in Sidoarjo

X5 = Number of State Junior High Schools (SMP) in Sidoarjo

X6 = Number of Private Junior High Schools (SMP) in Sidoarjo

X7 = Number of State Junior High School (SMP) Teachers in Sidoarjo

X8 = Number of Private Junior High School (SMP) Teachers in Sidoarjo

X9 = Number of State Senior High Schools (SMA) in Sidoarjo

X10 = Number of Private Senior High Schools (SMA) in Sidoarjo

X11 = Number of Public High School (SMA)Teachers in Sidoarjo

X12 = Number of Private High School (SMA) Teachers in Sidoarjo

Using cluster analysis, it can be seen the number of clusters formed based on the Education indicators in Sidoarjo which are represented by 12 variables. Furthermore, based on the results of the cluster formed, it can be concluded and interpreted using descriptive statistical analysis of the characteristics of education in Sidoarjo.

3. Result and Discussion

The first stage in cluster analysis is to determine distance. The distance used is Euclidean distance. After determining distance, grouping process can be carried out. The method used is hierarchical grouping method with the War'd Linkage algorithm. The results of cluster analysis in 18 sub-districts in Sidoarjo Regency using 12 variables that reflect education indicators can be presented in Figure 1 as follows:

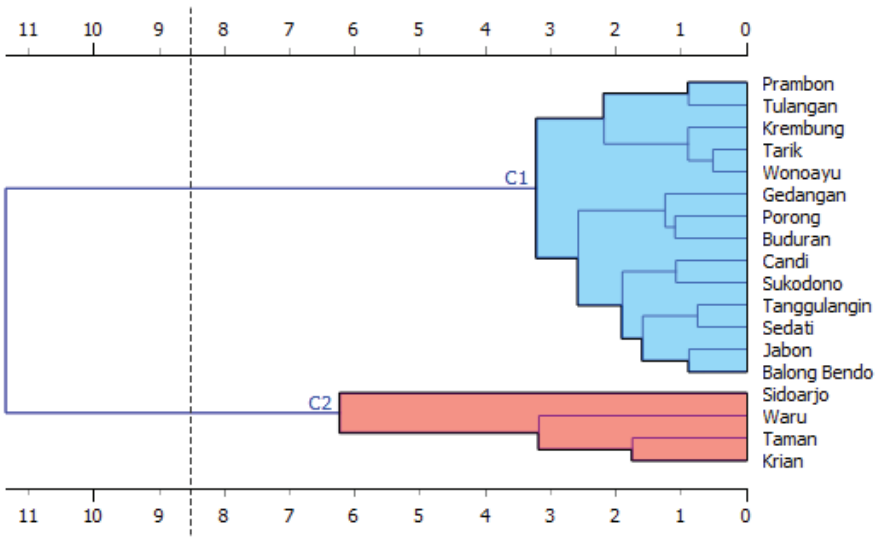

Figure 1: Dendogram of Cluster Analysis.

Based on the results of cluster analysis on education indicator data in each District of Sidoarjo Regency, it can be concluded that 2 clusters are formed. Of the total 18 sub-districts in Sidoarjo Regency, the first cluster consists of 14 sub-districts and cluster 2 consists of 4 sub-districts. Cluster 1 consists Districts of Prambon, Tulangan, Krembung, Tarik, Wonoayu, Gedangan, Porong, Buduran, Candi, Sukodono, Tanggulangin, Sedati, Jabon and Balong Bendo. Meanwhile, Cluster 2 consist Districts of Sidoarjo, Waru, Taman and Krian. To find which characteristics are the most dominant in each cluster, then the highest average of the variables for each cluster is sought. A summary of the average values in each cluster can be seen in Table 1. as follows:

Table 1: Average Value in Each Cluster.

Variabel	Cluster 1	Cluster 2
X1	44.9	31.8
X2	6.8	15.8
X3	580.7	534.3
X4	95.7	318.3
X5	4.0	4.0
X6	8.5	17.0
X7	164.9	187.0
X8	128.9	335.5
X9	0.8	1.8
X10	3.6	7.5
X11	37.3	103.8
X12	57.6	132.8

In Cluster 1 dominant variables are X1 (Number of State Elementary Schools (SD) in Sidoarjo) and X3 (Number of State Elementary School (SD) Teachers in Sidoarjo). This shows that the average number of public elementary schools and the highest number of public elementary school teachers are in cluster 1. In cluster 2 dominant variable is X2 (Number of Private Elementary Schools (SD) in Sidoarjo), X4 (Number of Private Elementary School (SD) Teachers in Sidoarjo), X5 (Number of State Junior High Schools (SMP) in Sidoarjo), X6 (Number of Private Junior High Schools (SMP) in Sidoarjo), X7 (Number of State Junior High School (SMP) Teachers in Sidoarjo), X8 (Number of Private Junior High School (SMP) Teachers in Sidoarjo), X9 (Number of State Senior High Schools (SMA) in Sidoarjo), X10 (Number of Private Senior High Schools (SMA) in Sidoarjo), X11 (Number of Public High School (SMA) Teachers in Sidoarjo) and X12 (Number of Private High School (SMA) Teachers in Sidoarjo). This shows that the average private elementary school, the average number of private elementary school teachers, the average junior and senior high school both public and private and the average

junior and senior high school teacher both public and mostly are in cluster 2. Based on the results of the analysis it is known that cluster the second has higher education indicators than cluster 1, so the results of this study can be used as a reference for equal distribution of education in Sidoarjo Regency by increasing the Education indicator in the lower cluster and maintaining the Education indicator in the cluster with good Education indicators.

4. Conclusion

Based on cluster analysis in education indicator data in 18 sub-districts of Sidoarjo Regency using 12 variables that reflect the education indicators that 2 clusters have been formed. From a total of 18 districts in Sidoarjo Regency, the first cluster consists of 14 districts and cluster 2 consists of 4 districts. The method used in this cluster analysis is a hierarchical grouping method with the War'd Linkage algorithm. Cluster 1 consists of the Districts of Prambon, Tulangan, Krembung, Tarik, Wonoayu, Gedangan, Porong, Buduran, Candi, Sukodono, Tanggulangin, Sedati, Jabon and Balong Bendo. Meanwhile, Cluster 2 consists of the Districts of Sidoarjo, Waru, Taman and Krian. Based on the analysis, it is also known that the second cluster has higher education indicators than the first cluster, so the results of this study can be used as a reference for equal distribution of education in Sidoarjo Regency by increasing the Education indicator in the lower cluster and maintaining the Education indicator in the cluster with good Education indicators.

Acknowledgment

Thanks to Muhammadiyah University Sidoarjo for support this article. Hopefully the results of this study can be used as a reference for equal distribution of education in Sidoarjo Regency.

References

[1] Kemendikbud RI. Indikator pendidikan di Indonesia. Jakarta: Kemendikbud RI; 2007.

[2] BPS Kabupaten Sidoarjo. Kabupaten sidoarjo dalam angka 2021. Sidoarjo: CV. Insert Coin; 2021.

[3] Oktavianty E, et al. District/city clustering in sulawesi based on education indicators using average linkage cluster and median linkage analysis. Natural Science.

2019;8(3):191–197.

[4] Sidik FF. Evaluasi Kinerja pembangunan bidang pendidikan di kabupaten ngawi tahun 2020. FOUNDASIA. 11(1) 2020..24-34 Available from: http://journal.uny.ac.id/index.php/foundasia/index

[5] Yim O, Ramdeen KT. Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data. The Quantitative Methods for Psychology. 2015;11(1):8–21. https://doi.org/10.20982/tqmp.11.1.p008

[6] Gülağız FK, Şahin S. Comparison of hierarchical and non-hierarchical clustering algorithms. International Journal of Computer Engineering and Information Technology. 9 (1)2017, 6–14 2017. Available from: www.ijceit.org

[7] Dini SK, Fauzan A. Clustering provinces in Indonesia based on community welfare indicators. EKSAKTA: Journal of Sciences and Data Analysis.1(1) 2020:56–63. https://doi.org/10.20885/eksakta.vol1.iss1.art9

[8] Adha R, Nurhaliza N, Soleha U. Perbandingan algoritma DBSCAN dan k-means clustering untuk pengelompokan kasus Covid-19 di dunia. Jurnal Sains, Teknologi dan Industri. 2021;18(2):206–211.

[9] Sari IP, Batubara IH. Cluster analysis using k-means algorithm and fuzzy c-means clustering for grouping students' abilities in online learning process. Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE). 2021;2(1):139–144. https://doi.org/10.30596/jcositte.v2i1.6504

[10] Nagari SS, Inayati L. Implementation of clustering using k-means method to determine nutritional status. Jurnal Biometrika dan Kependudukan. 2020;9(1):62-68 https://doi.org/10.20473/jbk.v9i1.2020.62-68