Contemporary Mathematics and Applications
Vol. 6, No. 2, 2024, pp. 101-113.

Implementation of K-Means and Single Linkage on Types of Disabilities in East Java Province

Putri Amaningsih[1], Tony Yulianto[2], Faisol[3] & Rica Amalia[4]

[1,2,3,4]Universitas Islam Madura

[1]Corresponding author Email: putriNingsih171@gmail.com

Abstract. The high number of people with disabilities is one of the problems faced by the Indonesian government, especially in Java Province. After West Java Province, East Java Province is in second place as the province with the highest rate of people with disabilities in Indonesia. Disabled people are people with physical disabilities such as not being able to walk, not being able to talk, not being able to see, and so on. The aim of this research is to group districts in East Java Province based on types of disabilities with the hope of facilitating activities in fulfilling the rights of people with disabilities in East Java Province. The grouping was carried out in order to determine the characteristics of each cluster therefore the optimal k-means method was used for clustering using the Euclidean distance method with cluster 1 in 29 districts and cluster 2 in 9 districts. The most optimal single linkage uses the Euclidean distance method with cluster 1 having 8 districts and cluster 2 having 30 districts. From the results of the validity index values, it was found that the single linkage method had the smallest validity value compared to the k-means method.

Keywords: *clustering; east java; k-means; persons with disabilities; single linkage.*

1 Introduction

Every human being wants to be born perfect without any shortcomings and without any defects. In reality, there are some people who are born with defects in their bodies, such as not having hands or feet, or sensory organs that not function. The disability makes the individual have limitations in carrying out daily activities. For normal individuals, this is seen as an unfortunate, troublesome situation and then pity arises. Some people consider people with disabilities as objects worthy of compassion [1].

Disability is any person who have physical, intellectual, mental or sensory limitations for a long period of time, in interacting with the environment, may have obstacles or difficulties in participating fully and effectively with other citizens. There are several types of disabilities, namely, visual impairment, speech impairment, hearing impairment, physical impairment, intellectual disability, hearing impairment and mental disability. Persons with disabilities are also classified as a vulnerable group, namely the group that most often receives discrimination and whose rights are often not fulfilled [2].

Based on current data in 2018 from the Badan Pusat Statistik (BPS), the number of people with disabilities in Indonesia reached 21.8 million. Then, the number of people with disabilities is based on data from the Ministry of Social Affairs' Management Information System for People with Disabilities where East Java Province is in second place with the largest number of people with disabilities [1].

Based on the preamble, it is very clear that the state recognizes the rights of persons with disabilities as citizens which must be fulfilled without discrimination. One of the rights of persons with disabilities specified in UU No.8 Tahun 2016 paragraph (1) is the "right to education". Fulfillment of every disability community by making regulations governing the fulfillment of rights disability [3].

Cluster analysis is often encountered in everyday life, whether related to the social sector, health sector, marketing sector, or agricultural sector. Cluster analysis is a technique that can be used to classify objects or cases into relatively homogeneous groups. Objects or cases in each group tend to be similar to each other and very different (not the same) from objects from other clusters. Judging from the things that are grouped, cluster analysis is divided into two types, namely observation grouping and variable grouping. Meanwhile, in general, cluster analysis is divided into two methods, namely hierarchical methods and non-hierarchical methods. In the hierarchical method itself there are several methods, the methods included in the hierarchical method are the Single Linkage method, the Complete Linkage method, the Centroid Linkage method, the Average Linkage method, and Ward's Method, while the method included in the non-hierarchical method is the K-Means method.

Some of the research on clustering with K-Means method was conducted by [4-6]. Therefore, it is hoped that this research can provide the best cluster solution so that an overview of the types of disabilities can be obtained by comparing district or city data in East Java Province using the K-means and Single Linkage methods, therefore researchers will conduct research with the title "Implementation of K-means and Single Linkage on types of disabilities in East Java Province".

2 Literature Review

2.1 Disability

Disabilities are those who have physical, mental, intellectual, or sensory limitations, for a long period of time where when faced with various obstacles, this can prevent their full and effective participation in society on an equal basis with others, according to John C. Maxwell, a person with a disability is someone who has an abnormality or which can interfere with activities [1]. Persons with disabilities are citizens who have limitations therefore they always have difficulty carrying out daily activities. Apart from that, they often experience obstacles in developing themselves and interacting with

other people. Therefore, they need help from other people when carrying out activities. The types of disabilities are as follows [7].

1. Physical Disorders (Tunadaksa)

A quadriplegic is an individual who has movement disorders caused by congenital neuro-muscular and bone structure disorders, illness or due to accidents (loss of organs) polio or paralysis. The physically disabled are classified into three groups, namely [8]:

a. Mild physical impairment: included in this classification are pure physical impairment and mild combination physical impairment. In general, this type of disabled person experiences few mental disorders and their intelligence tends to be normal. This group is mostly caused by abnormalities in body parts only. Such as: paralysis, reduced limbs (stuns) and other physical disabilities.
b. Moderate level of physical impairment: included in this classification are physical impairment due to congenital defects, mild cerebral palsy and mild polio. This group is often experienced by tuna due to cerebral palsy (mental impairment) and is accompanied by decreased memory, although not far below normal.
c. Severe physical impairment: included in this classification are impairment due to severe cerebral palsy and hereditary impairment due to infection. In general, children who are known to have disabilities have intelligence levels that are classified as debilitated, stubborn and idiotic.

2. Visual Impairment (Blindness)

Blind people are people who have limited sense of sight or even have the inability to see. Based on the level of blindness, the visually impaired are divided into 2 groups, namely: totally blind and low vision.

Meanwhile, based on age, blindness is divided into: blind from birth and blind not from birth (having had the experience of seeing and then experiencing blindness). The limitations or even inability of a blind person to see results in limitations or even inability to receive stimulus or information through the sense of sight (eyes). Therefore, the role of other sensory organs is needed to replace it. In this case, the sense of hearing (ears) and the sense of touch (hands) are the main alternatives for receiving stimulus/information from outside.

3. Hearing Disorders (Deaf)

Deaf people are individuals who have hearing impairments, either permanent or non-permanent. Because they have obstacles in hearing, deaf individuals have obstacles in speaking therefore they are usually called speech impaired. Deaf children are children who experience hearing loss which is classified into deaf and hard of hearing. Deafness has an impact on language and speech development, especially for children who have

been deaf since birth (prelingual). Their language and speech development becomes hampered, resulting in obstacles in developing their potential. The condition of a person's inability to hear something, either totally or partially, is called deafness. A deaf person is someone who is unable to hear therefore they will have difficulty understanding or comprehending other people's speech through their hearing with or without using hearing aids.

4. Speech Disorder (Wise Impairment)

People with speech disorder is someone who has difficulty expressing thoughts through verbal language, making it difficult or even incomprehensible for other people. This speech disorder can be understood by other people. This speech disorder can be functional, which may be caused by deafness, and organic, which is caused by imperfections in the speech organs or disturbances in the motor organs related to speech. A deaf person or speech disability is an individual who experiences difficulties or obstacles in verbal communication and thus experiences difficulty in communicating. This may be caused by the absence or dysfunction of speech organs, such as the oral cavity, tongue, palate and vocal cords, in addition to the absence or dysfunction of hearing organs, resulting in delayed language development.

5. Mental retardation

Mental retardation is a term used to refer to individuals whose intellectual intelligence (IQ) is significantly below average due to developmental, mental, emotional, social and physical obstacles so they are unable to adapt to their environment. Children with mental retardation have mental limitations, who need to be educated and trained to adapt to their surrounding environment. They need more support from their parents and environment so they can live independently. Therefore, mentally retarded children need special services tailored to their abilities. There are several characteristics of mental retardation, namely:

a. Intelligence limitations
b. Social limitations
c. Limitations of other mental functions

6. Tuning in

Tunalaras is an individual who experiences emotional disturbances and behavioral deviations and is unable to adapt to his environment, whether in the family, school or society. Children with developmental disabilities are also often referred to as children with emotional disturbances (emotionally disturbed), children with psychological disorders (psychologically disturbed), or children with emotional handicaps (emotionally handicapped) because they tend not to be able to control their emotions, so they often experience conflict with other people. or with oneself.

7. Disability

Disability Based on UU No. 8 of 2016 Paragraph 1 concerning people with disabilities. Persons with Disabilities are any person who experiences physical, intellectual, mental and/or sensory limitations for a long period of time who, in interacting with the environment, may experience obstacles and difficulties in participating fully and effectively with other citizens based on equal rights.

2.2 Single Linkage Method (Single Linkage)

Hierarchical methods and other agglomerative methods represent attempts to find "good" clusters in data using efficient computational techniques. In general, it is impossible to examine all possible clustering possibilities for a data set, especially a large one. The hierarchical method that will be used in this research is the agglomerative method or known as the single linkage method which consists of single linkage, complete linkage and average linkage. The single linkage method is the simplest hierarchical method, where groups are formed from individual entities by combining the closest individuals, where the term closest individual means the smallest distance or greatest similarity. Complete linkage proceeds in almost the same way as single linkage, with one important exception, namely that at each stage, the distance (similarities) between clusters is determined by the distance (similarities) between two elements, one from each cluster, namely the most distant. Thus, complete linkage ensures that all observations in a cluster are within the maximum distance or (minimum similarity) to each other. Meanwhile, average linkage treats the distance between two clusters as the average Euclidean distance between all pairs of items where one member of the pair belongs to each cluster. In the single linkage method, the distance between the clusters being combined is minimized [6].

Another single link method is a clustering process which is based on the closest distance between objects. If two objects are separated by a short distance, then the two objects will be combined into one cluster. This method starts by determining the objects that have the smallest distance in the proximity matrix **D** = $\{d_{ik}\}$. Objects that have the smallest distance are combined into one cluster, for example U and V are two objects that will be clustered to obtain a cluster (UV). To find the distance between cluster (UV) and cluster W or other clusters, it is obtained using the formula:

$$d_{(UV)W} = \min\{d_{UW,}d_{VW}\} \quad (1)$$

Information:
d_{UW} is the shortest distance from clusters U and W
d_{VW} is the shortest distance from clusters V and W

2.3 K-Means Clustering Method

The k-means method divides data into several groups therefore data with the same characteristics are in the same cluster and data with different characteristics are in different clusters [4]. More specifically, the k-means algorithm is as follows following:

Determine the k value as the number of clusters you want to form.

a. Initialize k as a centroid that can be constructed randomly.
b. Calculate the distance of each data to each centroid using the equation.
c. Group each data based on the closest distance between the data and its centroid.
d. Determine the position of the new centroid (k).
e. Return to step 3 if the position of the new centroid and the old centroid are not the same.

2.4 Internal Cluster Dispersion Rate (Icdrate)

One of the methods used to select the best cluster method is to calculate the average distribution of the Internal Cluster Dispersion Rate (icdrate) over the overall partition [6]. Icdrate is a value used to determine the best cluster criteria. This value describes the level of differences within the cluster, the smaller the icdrate value has low or almost homogeneous differences. Determining the best cluster method in this research is by looking at the icdrate (internal cluster dispersion) value. The smaller the icdrate value, the better the grouping results. The icdrate value is the level of dispersion in the cluster, this value can be written using the equation [4].

The icdrate value can be obtained using the following equation:

$$Icdrate = 1 - \frac{SST-SSW}{SST} = 1 - \frac{SSB}{SST} = 1 - R^2 \quad (2)$$

Where $R^2 = SSW \frac{(SST-SSW)}{SSW}$

R^2 is the sum of the squares of the distance between group centers and the sum of the sample squares to the average of the total.

$$SST = \sum_{j=1}^{c} \sum_{k=1}^{p} \left(x_{ijk} - \bar{\mathrm{x}}_j\right)^2 \quad (3)$$

SST is the total sum of the squares of the distances to the overall mean.

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^{c} \sum_{k=1}^{p} \left(x_{ij}^k - \bar{x}_j^k\right)^2 \quad (4)$$

SSW is the total sum of the squares of the sample distance to the group average.
Where:

C	=	The many variables
P	=	Number of groups
x_{ijk}	=	The i sample in the j variable in the k group

$\bar{x}_j$ = The average of all samples on the *j* variable group

3 Research Methods

3.1 Research Data

In this research, the type of data used includes quantitative data, where the data taken is in the form of numbers. The data in this research comes from BPS for people with disabilities in East Java Province in the form of secondary data, namely data on the number of disabilities based on disability type per district or city in East Java 2018. The data was analyzed using the K-Means method for clustering and single linkage on disability data based on type his disability. The types of disabilities are visually impaired, speech impaired, hearing impaired, physically disabled, mentally retarded, hearing impaired and handicapped and the population numbers are taken from the official website of the East Java Provincial Central Statistics Agency, namely https://jatim.bps.co.id, while data on distances between districts or city can be obtained from Google Maps.

3.2 Research Procedures

This research uses a clustering method consisting of single linkage and k-means which will be used to group types of disabilities in East Java Province. The data in this research was obtained from the 2018 Central Statistics Agency website for 38 cities and regencies in East Java Province. There are seven variables used for this type of disabled person, namely data for the blind (X1), the deaf (X2), the speech impaired (X3), the physically impaired (X4), the intellectually impaired (X5), the hearing impaired (X6) and the handicapped (X7).

The steps of K-means Clustering are described as follows:

1. Determine the k value as the number of clusters you want to form.
2. Initialize k as a centroid that can be constructed randomly.
3. Calculate the distance of each data to each centroid using the equation.
4. Group each data based on the closest distance between the data and its centroid.
5. Determine the position of the new centroid (k).
6. Return to step 3 if the position of the new centroid and the old centroid are not the same.

Steps of the single linkage method:

1. The single linkage method uses the minimum distance principle.
2. Grouping starts by finding the minimum value of D { d_((x,y))}.

3. Combining objects that are close to each other, for example U and V to get a cluster (UV).

In step (3) of the algorithm described previously, the distance between clusters (UV) and another object, for example W.

4 Result and Discussion

The data in this study is data from 2018 in 38 districts or cities in East Java with five variables, namely data on the blind, hearing impaired, speech impaired, physically disabled, mentally disabled and disabled. Before cluster analysis was carried out, descriptive analysis was first carried out.

Table 1 Descriptive analysis of types of disabilities in East Java Province

Variable	Maximum value	Minimum value	Mean	Standard Deviation
$X1$	285	12	157,55263	84,445952
$X2$	247	2	157,55263	68,027293
$X3$	270	9	132,10526	74,848961
$X4$	296	17	160,842105	84,8061466
$X5$	327	17	163,078947	93,8164918
$X6$	240	4	114,36842	73,543303
$X7$	90	0	31,86842	25,74949

Based on Table 1 which consists of 7 disability type variables which are used as observation units to form a Cluster analysis as a reference for grouping types. Disability in East Java Province. In this research, the descriptive statistics used are Maximum Value, Minimum Value, Average Value, Standard Deviation Value.

The districts or cities with the highest scores are the visually impaired (Malang) at 285, the deaf (Lamongan) at 247, the speech impaired (Lamongan) at 270, the physically impaired (Malang) at 296, the mentally retarded (Bojonegoro) at 327, the physically impaired (Lamongan) at 240 and Disability (Lamongan) was 90. The districts/cities with the lowest scores were the visually impaired (Madiun) at 12, the deaf (Blitar City) at 2, the speech impaired (Mojokerto City) at 9, the physically disabled (Blitar and Mojokerto City) at 17, the mentally retarded (Mojokerto City) at 17, Mojokerto) was 17, Tunalaras (Mojokerto) was 4 and Mojokerto City and Batu City had no disabilities. For the average (impaired) 157.55263, (deaf) 157.55263, (impaired) 132.10526, (impaired) 160.842105, (impaired) 163.078947, (impaired) 114.36842 and (defects) amounting to 31.86842. And for the standard deviation (Blind) it is 84.445952, (Hearing Impaired) it is 68.027293, (Hearing Impaired) it is 74.848961, (Hearing Impairment) it is 84.8061466, (Hearing Impairment) it is 93.8164918, (District Hearing) it is 73, 543303

and (Defects) of 25.74949. The highest maximum score is for the deaf in the city of Bojonegoro, namely 327 and the lowest minimum score is for the deaf at 2 in the city of Blitar and there are no disabilities in the city of Mojokerto and Batu.

Comparison of K-means and Single Linkage methods

Based on data on the types of disabilities in East Java in 2018, they can be grouped into 2 clusters. The stage before grouping using the k-means method and single linkage cluster is to determine the number of clusters first. In this research, the method used to determine the number of clusters. The following are the results of determining the number of clusters.

Table 2 Cluster Results

Variable	Cluster 1	Cluster 2
$X1$	20	189,9
$X2$	10,625	144,5333
$X3$	12	161,1667
$X4$	23,5	192,5333
$X5$	23	195,6
$X6$	15	138,9333
$X7$	20	189,9

In Table 2, the results of calculating types of disability using single linkage produce 2 clusters using the Euclidean distance method. The results of single linkage and k-means using the distance method can be seen in Appendix D. The cluster values selected for the Cluster 2 using the Euclidean distance method are as follows.

Table 3 Selected Results from Cluster 2

Kabupaten/Kota	**Cluster 2**
Pacitan	2
Ponorogo	2
Trenggalek	2
Tulungagung	2
⋮	⋮
Madiun	1
Surabaya	2
Batu	1

Table 3 is the result of calculating the type of disability using the Euclidean distance method. The selected one is Cluster 2. Membership of data is determined based on the lowest value. For new membership degree values using other distance methods. And this are the member results and the average on k-means and single linkage is in Table 4.

Table 4 Cluster Member Results on K-means and Single linkage

Cluster	Number of members	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	Average
1 (low)	8	20	10.6250	12	23.5000(	23	15	3.2500	355545,7
2 (tall)	38	189.(	144.533:	161.166'	192.533:	195.600(	138.933:	39.1667	1.516.905

Based on Table 4, the average value for each variable in each cluster can be known so that from these values the cluster interpretation can be carried out as follows:

Cluster 1: A cluster consisting of 8 districts where the first cluster has the lowest average number of speech impaired, physically disabled, disabled, mentally retarded, visually impaired, hard of hearing and hearing impaired.

Cluster 2: A cluster consisting of 30 districts where the third cluster has the highest average number of disabled, hearing impaired, speech impaired, deaf, visually impaired, physically disabled and mentally retarded. Membership of data is determined based on the value of the type of disability. The number of districts and the names of districts for each cluster formed can be seen in the following table.

Table 5 Number of Districts from Clusters

Cluster	Number of districts	District Name
1 (low)	8	Kediri, Blitar, Malang, Probolinggo, Pasuruan, Mojokerto, Madiun, Batu.
2 (tall)	30	Pacitan, Ponorogo, Trenggalek, Tulung Agung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, and Surabaya.

Table 5 Number of districts from clusters for medium and high districts. Comparison of the Best Hierarchical Cluster (Single Linkage) and Non-Hierarchical Cluster (K-Means Cluster). Comparison of hierarchical and non-hierarchical cluster methods will obtain

the best method. To determine the best method from the 2 methods, a comparison test was carried out by calculating the distance method according to the validity index value. The following are the results of the values for each index for Single Linkage and K-means Cluster using the distance method in Appendix B.

To see which method is the best, the validity index value can be calculated using the SST, SSW, R Squared and Icdrate index calculations for each method with a total of 2 clusters. The validity index values will be compared to obtain the optimal number of clusters based on the criteria for each validity index which can be seen in Table 6.

Table 6 Value of Each Method

Method	$\boldsymbol{SST}$	$\boldsymbol{SSW}$	$\boldsymbol{R^2}$	$\boldsymbol{Icdrate}$
Single Linkage	37,65306	36,18129	0,040678	0,959322
K-means	37,65306	36,97138	0,018438	0,981562

Table 6 The value of each validity index from single linkage and k-means shows that the lowest value shown by the Single Linkage method is the Icdrate value, R Squared value, SSW and the largest is SST. Furthermore, the lowest value is shown by the K-means method in Icdrate, R Squared, SSW and the largest is in SST. Thus, it can be concluded that the cluster with the best value is the cluster formed using the Single Linkage method, namely using the Icdrate Value. The Single Linkage method dendogram showing the clustering of data using the distance method that is formed is in Appendix C.

Table 7 Icdrate Value Results

Method	**Icdrate value**
Single Linkage	0,959322
K- Means	0,981562

Based on Table 7, it is found that the Icdrate value in the Single Linkage method has the lowest validity value compared to the Icdrate value in K-means. Therefore, the best method in this research using data on types of disabilities in East Java Province in 2018 is Single Linkage using Icdrate values.

5 Conclusion

For comparison results of the K-means and Single Linkage methods on types of disabilities in East Java Province based on the validity index value using SSW, SST, R Squared and Icdrate calculations, there are:

1. In clustering using the optimal k-means method using the euclidean distance method with cluster 1 there are 29 districts and cluster 2 there are 9 districts. The most optimal single linkage uses the Euclidean distance method with cluster 1 having 8 districts and cluster 2 having 30 districts.
2. From the results of the validity index value, it is found that the single linkage method has the smallest validity value of the icdrate method compared to the k-means method. So it can be concluded that the best method in this research uses the Icdrate method using Euclidean distance from data on types of disabilities in East Java Province in 2018.

6 Suggestions

The suggestions for further research are as follows:

1. Suggestions for further research include using different cluster methods with the same case study.
2. Can apply the hierarchical cluster method to all types of disabilities, not only those categorized as People with Social Welfare Problems (PMKS).

7 References

[1] Adioetomo, S.M., Mont, D. & Irwanto, 2014, *Persons with Disabilities in Indonesia: Empirical Facts and Implications for Social Protection Policies*, Jakarta: Lembaga Demografi Fakultas Ekonomi Universitas Indonesia.

[2] Nurhayati, D.A. & Ambari, 2020, *Implementation of Fulfillment of Rights with Disabilities in Obtaining Work in Indonesia*, International Journal of Business, Economics and Law, **23**(1), 49-54.

[3] Tamami & Suryawati, D., 2021, *Implementation of Protection Policy and Fulfillment of The Rights of People with Disabilities Perspective In G. Edward III (1980): Study on Fulfillment of Facilities and Infrastructure Needs for Persons with Disabilities in Jember Regency Green Space*, Regional Dynamic: Journal of Policy and Business Science, **2**(1), 29-38.

[4] Nirmala, I.D. & Atika, P.D., 2020, *Implementation of K-Means Algorithm as a Clustering Method for Selecting Achievement Students Based on Academic Grade*, Pilar Nusa Mandiri : Journal of Computing and Information System, **16**(2), 199-204.

[5] Nasution, S.A.Y., Poningsih & Okprana, H., 2021, *Penerapan Algoritma K-Means Pada Penjualan Frozen Food Pada UD Soise Sosis Pematangsiantar*, Jurnal Sistem Komputer dan Informatika, **2**(2), 171-177.

[6] Zainudin, M.N.S., Sulaiman, M.N., Mustapha, N. & Mohamed, R., 2018,

Comparison of Expectation Maximization and K-means Clustering Algorithms with Ensemble Classifier Model, WSEAS Transaction on Computers, **17**, 253-259.

[7] Putra, M.K.S., Hijri, Y.S., & Kamil, M., 2022, *The Government's Efforts in Fulfilling the Accessibility of the Rights of Persons with Disabilities through Social Rehabilitation in East Java*, Jurnal Public Policy (JPP), **8**(3), 177-186.

[8] Anlianna, Sunanto, Nursalim, M. & Rahmasari, D., 2023, *Problems of Children with Intellectual and Mental Disabilities at School*, Sentra Cendekia, **4**(2), 80-92.