International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 1, Number 2, July 2022, *Page. 34 - 39*
Email : editorijhess@gmail.com

Grouping the Distribution of Diarrhea Based on the Public Health Centers in Binjai City Using the Clustering Method

Rika Hedy Anggraini Prastio [1*)], Akim M.H Pardede [2)], Hermansyah Sembiring [3)]
[1,2,3)] STMIK Kaputama Binjai, Indonesia

*Coresponding Author
Email : rikaprastio2018@gmail.com

Abstract

Diarrhea is one of the endemic diseases that occurs throughout the year and is one of the highest causes of death for everyone, especially children under five in Indonesia. The purpose of this thesis is to analyze the spread of diarrheal disease by forming a grouping of the spread of diarrheal disease and to analyze the characteristics of the spread of diarrheal disease. With the development of technology that is increasingly sophisticated and easy to use, such technological developments motivate us to know the importance of using computers in fast and practical data processing. Clustering is a data analysis method, which is often included as one of the Data Mining methods, whose purpose is to group data with the same characteristics and data with different characteristics into others.

Keywords: Diarrhea, K-Means Algorithm, Clustering, Data Mining.

INTRODUCTION

Diarrhea is one of the endemic diseases that occurs throughout the year and is one of the highest causes of death for everyone, especially children under five in Indonesia. The purpose of this thesis is to analyze the spread of diarrheal disease by forming a grouping of the spread of diarrheal disease and to analyze the characteristics of the spread of diarrheal disease. With the development of technology that is increasingly sophisticated and easy to use, such technological developments motivate us to know the importance of using computers in fast and practical data processing.

To find out the spread of diarrheal disease in every Puskesmas in Binjai City, the Health Office needs to have a data system for classifying the distribution of diarrheal diseases based on the Puskesmas in Binjai City, in order to obtain clear, structured information according to the vision and mission of the strategy. To find out the data on the spread of diarrheal disease based on the patient's age, gender and location, it is necessary to design a data system for the spread of diarrheal disease to determine the identical distribution of diarrheal disease in actual patients. Clustering is a data analysis method, which is often included as one of the Data Mining methods, whose purpose is to group data with the same characteristics and data with different characteristics into others.

One way to find out the spread of diarrheal disease is to group data on diarrheal patients at the puskesmas. Cluster analysis is the work of grouping data (objects) based only on the information found in the data that describes these objects and the relationships between them. Objects that are joined in a group are objects that are similar (or related) to each other and different (or unrelated) to objects in other groups. One of the most commonly used methods for clustering is the K-Means algorithm. Based on the problems above, the writer will influence any criteria (variables) to classify based on the variables of patient age, gender, type of location.

RESEARCH METHODS

(Sembiring et al., 2020) Data mining is a term for pattern recognition which is an algorithm for data processing in order to find data patterns into new knowledge. Data that is processed with data mining techniques will produce knowledge sourced from old data, the results are to determine business decisions.

Data processing techniques with the help of data mining algorithms. processing is done by building a design pattern, then the model forms other data patterns that are not in the databaseIn general, the definition of data mining can be interpreted as follows:

1. The process of finding interesting patterns from large amounts of stored data.
2. Extraction of useful or interesting information (non-trivial, implicit, previously unknown potential use) patterns or knowledge from large amounts of stored data.
3. Exploration of automated or semi-automatic analysis of large amounts of data in search of meaningful patterns and rules.

Knowledge Discovery in Database (KDD) is the whole non-trivial process to find and identify patterns in the data, where the patterns found are valid, new, useful and understandable.
KDD deals with integration techniques and scientific discovery, interpretation and visualization of patterns in a number of data sets.

*Clustering*is a data analysis method, which is often included as a data mining method, whose goal is to group data with the same characteristics.
According to Hermawati (2012, h:123) "Clustering analysis (clustering) is finding objects in one group that are the same (has a relationship) with others and are located (points have a relationship) with objects in other groups".

DestinationThe main purpose of the clustering method is to group a number of data / objects into clusters (groups) so that each cluster will contain data that is as similar as possible. The clustering method tries to place similar objects (closely distanced) in one group and make the distance between groups as far as possible. This means that objects in one group are very similar to each other and different from objects in other groups.
The following is an example of clustering:

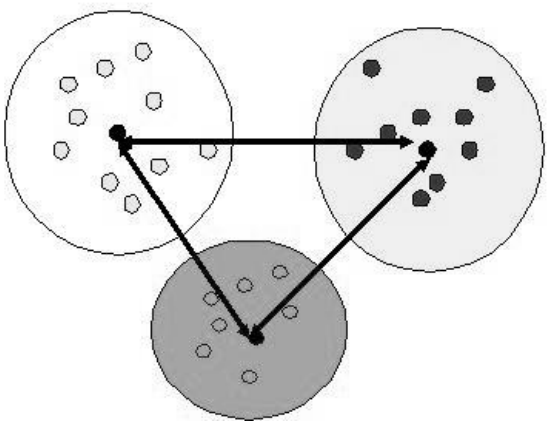

Figure II.2 Examples of Clusters Formed
Source :http://ike.uninet.net.id/clustering(2012)

Broadly speaking, there are 3 ways clusters work, namely:

1. How to measure similarity?
   There are three measures of measuring the similarity between objects, namely the size of the correlation, the size of the distance, and the size of the association.

2. How to form clusters?
   The procedure applied must be able to group objects that have a high similarity into a cluster.
3. How many clusters / groups will be formed?
   In principle, if the number of clusters decreases, the homogeneity within the cluster will automatically decrease.

Algorithm*K-Means*is a relatively simple algorithm for classifying or grouping a large number of objects with certain attributes into K clusters. In the K-Means algorithm, the number of K clusters is predetermined.

(Eko Prasetyo, 2012)states that "K-Means is one of the non-hierarchical grouping methods that tries to partition data into clusters/groups so that data with the same characteristics will be included in the same cluster and data with different characteristics are grouped into clusters. another group."

Data clustering using the K-Means method is generally done by requiring three components, namely:

1. Number of Clusters
   As previously explained, K-Means is part of a non-hierarchical method so that in this method the number of K must be determined first. The number of K clusters can be determined through a hierarchical method approach.
2. *Cluster* Beginning
   Cluster The initial selected is related to the determination of the initial cluster center.
3. Distance Size

In this case, the distance measure is used to place observations into clusters based on the nearest centroid. The distance measure used in the K-Means method is dEuclidean.
The K-Means algorithm in the formation of clusters is as follows:

1. For example, given a matrix X = {Xij) data of size nxp with i = 1,2,….n,j=1,2,…p and assume the number of initial clusters is K.
2. Determine the centroid.
3. Calculate the distance of each object to each centroid using the dEcclidean distance or can be written as follows:
   $J(xici)=\sqrt{(x_i - c_i)^2}$……(1)
4. Each object is arranged to the nearest centroid and the collection of objects will form a cluster.
5. Determine the new centroid of the cluster that will be formed, in which object the new centroid is obtained from the average of each located in the same cluster.
6. Repeat step 3, if the initial and new centroids are not the same.

There isSeveral methods are used to measure the distance of the data to the center of the group, including dEuclidean (Bezdek, 1981), Manhattan/City Block (Miyamoto and Agusta, 1995), and Minkwosky (Miyamoto and Agusta, 1995), each method has advantages and disadvantages. deficiency.

The measurement of distance in the dEuclidean distance space uses the formula:

$$D(x2,x1) = X2 - X1 \| 2 = \sqrt{\sum_{j=1}^{p} |X2j - X1j|^2} \quad \text{.....(2)}$$

D is the distance between the data X2 and X1, and | . | is absolute value.
Measurement of distance in the Munhattan distance space using the formula:

$$D(X2,X1) = X2 - X1 \| 1 = \sqrt{\sum_{j=1}^{p} |X2j - X1j|} \quad \text{.....(3)}$$

Measurement of distance in the Minkowsky distance space using the formula:

$$D(X2,X1) = X2 - X1 \| \lambda = \sqrt{\sum_{j=1}^{p} |X2j - X1j|^{\lambda}} \quad \text{.....(4)}$$

is the Minkowsky distance parameter.

The best waywidely used are dEuclidean and Manhattan. dEuclidean is an option if we want to give the shortest distance between two points (straight distance), as shown in the dEuclidean formula. Meanwhile, Manhattan gives the furthest distance between the two data.

RESULTS AND DISCUSSION

A. Application of the Method

Based on the application of the method, the authors will explore the data to be grouped using the clustering method with the K-Means algorithm, where the variables of the data on the spread of diarrheal disease to be taken for this study are the patient's age, gender and location. Then the data is entered into Matlab so that there are 3 groups of results.

B. Flowchart Design

The flowchart of the hierarchical clustering algorithm and K-means are:

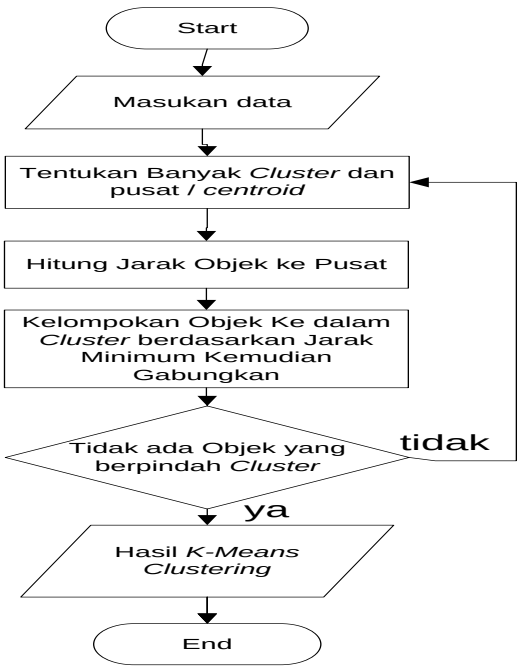

Figure III.2 Flowchart of Hierarchical Clustering Algorithm and K-means

The grouping of data using the K-Means method is generally carried out in the following way:

1. Enter data
2. Determine the number of groups
3. Random allocation of data into groups.
4. Calculate the center of the group (centroid/mean) of the data in each group.
5. Allocate each data to the nearest centroid/average.
6. Return to step 3 if there is still data that moves groups, or if there is a change in the centroid value above the specified threshold value, or if the change in the value of the objective function used is still above the specified threshold value.

C. Measuring Euclidean Distance

In using the clustering method, the initial process for forming clusters is to transform the data into numeric form with predetermined codes, then determine the number of groups (K), calculate the centroid, calculate the distance of the object to the centroid and then group it based on the distance. closest, if there are objects that move the group then the iteration is complete. To determine the

group of an object, the first thing to do is measure the dEuclidean distance between two object points (X and Y) which are defined as follows:

$^{d}Euclidean(X,Y)\ \sqrt{\sum_i (X_i - Y_i)^2}$

D. Clustering Graph

Create a cluster graph based on the calculations that have been done. The graphs obtained are as follows:

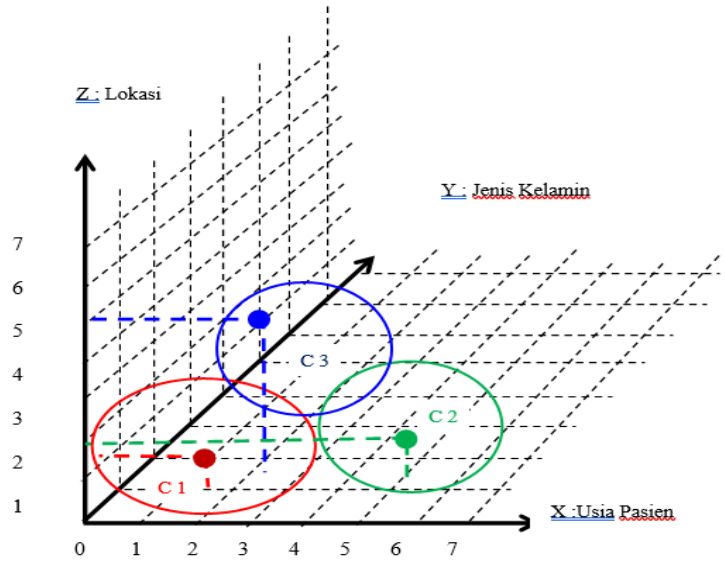

Figure III.3 Cluster Graph based on the calculations that have been done

	*Cluster*1:	1.33;	1.67;	1.50
	*Cluster*2:	5.00;	1.75;	5.75
	*Cluster*3:	2.20;	1.40;	5.00

Graphic Explanation:

From20 data obtained 3 groups, Cluster 1 has 6 data, Cluster 2 has 4 data, and Cluster 3 has 10 data. And obtained the most group is cluster 3.

1. **Cluster 1 There are 6 Data**
   1.33; 1.67; 1.50
   It can be seen that in cluster 1 there are patients aged between <= 5 years, male and female, located at the Binjai Estate and Rambung health centers.

2. **Cluster 2 There are 4 Data**
   5.00; 1.75; 5.75
   It can be seen that in cluster 2 there are patients aged between <=45 years, male and female, who are at the location of the Kebun Lada and Jati Makmur health centers.

3. **Cluster 3 There are 10 Data**
   2.20; 1.40; 5.00
   It can be seen that in cluster 3 there are patients aged between <= 15 years, male sex at the location of the Kebun Lada health center.

CONCLUSION

The distribution of diarrheal diseases based on the Puskesmas in Binjai City was more common at the Kebun Lada Health Center and more in males aged <= 15 years. There is a clustering of diarrhea sufferers in eight sub-districts in Binjai City, namely Binjai Estate Health Center, Rambung, Binjai City, Tanah Tinggi, Pepper Garden, Jati Makmur, Bandar Sinembah, HAH Hasan. By designing the application that will be built, it will be able to obtain information about the highest distribution of diarrheal diseases based on the puskesmas in Binjai City. With the data mining method of the K-means clustering algorithm, it helps to classify the medical record data of diarrhea patients at the puskesmas in Binjai City based on the patient's age, gender, and location.

REFERENCES

Abdurrahman, DD, Agus, F., & Putra, GM (2021). Implementation of Partitioning Around Medoids (PAM) Algorithm to Group Plantation Commodity Production Results (Case Study: Plantation Office of East Kalimantan Province). 16(2).

Hermawati, FA (2012). Data Mining (Andi (ed.)).

Hermawati, FA (2013). Data Mining (Andi (ed.)). https://www.bing.com/search?q=Fajar+Astuti+Hermawati%2C+Data+Mining%2C+Publisher+Andi%2C+Yogyakarta%2C+2013&qs=n&form=QBRE&sp=-1&pq=fajar+astuti+hermawati+process+kdd+data+mining+&sc=8-46&sk=&cvid=CEC2A0E568F04792A5CEC46D88F40351

Noraida, N., Khair, A., Raharja, M., Rohman, H., & Fajarini, N. (2016). Clustering of Diarrhea Cases of Toddler Age in Srandakan Region. Clustering of Diarrhea Cases Age Bali, 11 (Journal of obstetrics), 137–147.

Prasetyo, E. (2012). Data Mining: Concepts And Applications Using MATLAB. Andi.

Sembiring, F., Octaviana, O., & Saepudin, S. (2020). Implementation of the K-Means Method in Clustering Illegal Charges Areas in Sukabumi Regency (Case Study: Population and Civil Registration Office). Journal of Incentive Tech, 14(1), 40–47.

Sy, H., Rismayani, & Syam, A. (2019). Data Mining Using the K-Means Algorithm for Grouping the Distribution of Diarrhea in Makassar City. SISITI : Scientific Seminar on Information Systems and Information Technology, 8(1), 73–82.