

Clustering Disease on Settlements Inhabitant In place seedy With Use Clustering Method

Ruine Buana Br Sitepu^{1*}, Achmad Fauzi², Rusmin Saragih³

^{1,2,3} STMIK Kaputama, Binjai, Indonesia

ruinebuanabrsitepu@gmail.com¹, fauzie.kaputama@gmail.com², evitha12014@gmail.com³

Abstract

Residents living in slum areas often face serious problems related to public health, where the prevalence of disease tends to be high and its spread is difficult to control. The impact of the formation of slums for the community is that safety is threatened, health deteriorates, and social conditions worsen, causing many diseases for people living in slums. Therefore, this study aims to identify patterns and clusters of diseases that exist in residential areas in slums Binjai city using clustering method. The K-Means Algorithm clustering method was chosen because it is able to group data based on similar characteristics, so that it can help identify diseases in a more focused and efficient manner, using the MATLAB application is also very appropriate in this problem so that it can produce output from data mining that can be used in decision making. future decisions. By utilizing the data mining process using the clustering method, clustering can be a problem of grouping diseases in slum settlements. Based on the results of trials with 20 sample data conducted with MATLAB obtained in cluster 1 DHF cases with high slums, Cluster 2 cases of vomiting with moderate slums and cluster 3 cases of diarrhea with moderate slums. The results of this study are expected to provide in-depth insight into disease patterns and clusters in residential areas in slums.

Keywords: Clustering, K-Means, Data_Mining, Disease, Settlements_Slums

1. Introduction

Progress technology information Already the more develop fast in all field life. A lot of data is generated by technology more and more information advanced, start from field industry, economy, science and technology as well as various field life other. Health is same thing valuable for man Because Who just can experience disturbance health, as well as in humans who are very vulnerable to various type disease caused in the environment or settlement slum. Factor reason happening environment seedy that is factor economics, factor social, cultural, factors density amount population, factor quality building, factor availability land. Impact from formation settlement seedy for public is safety threatened, health worsened, and conditions social worsened so that raises Lots disease for living community in settlements seedy, for can grouping disease data in settlements seedy in Binjai city got done with carry out the application of data mining using method clustering on disease data that occurs in settlements inhabitant in place seedy, so it can be identify group frequent illness happen and learn characteristics vulnerable society to disease. With so, parties related can take action prevention or more intervention precise and effective for overcome problem health in the region.

Study this strengthened with study previously by (Rean Ordila, 2020) entitled "Application of data mining for record data grouping medical patient based on type disease with Clustering algorithm (study case PT. Inecda)" The result with use method algorithm Clustering K-means then record data grouping medical patient poly clinic PT. Inecda by region, gender, age and number disease with patient the most is ARI with amount patients (1985 patients) due environment housing area PT. Inecda which is plantation coconut palm oil and PKS (factory coconut oil palm) and also disease other like fall from the motor, check cholesterol, check tension, control pregnancy, and more with amount patients (2142 patients) [1].

2. Research Methods

2.1. Data Mining

Data mining is the process of extracting useful information and patterns from very large data. Data mining includes data collection, data extraction, data analysis, and data statistics. Data mining is also known as Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, and others [2]. Define data mining as a process for get useful information from large database warehouse. Data mining can also interpreted as extraction information just taken from helpful big data chunks in taking decision. The term data mining sometimes also called knowledge discovery [3].

Data Mining is defined as the process of finding patterns in data. The pattern found must be meaningful and the pattern provides an advantage. Characteristics of data mining as follows:

1. Data Mining is related to the discovery of hidden and certain data patterns that were not previously known.
2. Data Mining usually uses very large data. Usually big data is used to create believable results [4] .

2.2. K-Means Algorithm

The K-Means algorithm is one from many algorithms used in grouping Because simplicity and efficiency [5] . The K-Means Algorithm is a popular and widely used clustering algorithm in the industrial world. This algorithm is structured on the basis of a simple idea. Initially it is determined that several clusters will be formed into objects or the first element in the cluster can be selected to serve as the center point (centroid point) of the cluster. The K-Means algorithm will then repeat the following steps until stability occurs (no objects can be moved) The concept of similarity is fundamental in cluster analysis. Similarity between objects is a measure of correspondence between objects. There are three methods that can be applied, namely the correlation measure, distance measure, and association measure. By using a distance measure, the similarity measure that can be used is the dEuclidean and dManhattan City distances. If the first observed object is $X = [X_1, X_2, \dots, X_p]$ and $Y = [Y_1, Y_2, \dots, Y_p]$ between 2 objects of p dimension then:

$$d_{\text{Euclidean}} : X, Y = \sqrt{\sum_i (X_i - Y_i)^2}$$

$$d_{\text{Manhattan}} : X, Y = \sqrt{\sum_i |X_i - Y_i|}$$

Information:

d_e = Euclidean Distance

d_m = Manhattan City

i = Number of objects

(x,y) = Coordinates of the object

The steps using the K-Means Algorithm are as follows:

1. Determine the number of clusters
2. Determining the centroid (coordinate of the midpoint of each cluster) for the first iteration is taken randomly.
3. Calculating the distance of each object to the centroid using the dEuclidean and dManhattan distance formulas.
4. Determine the distance of each object to the midpoint coordinates.
5. Group these objects based on the shortest distance [4] .

2.3. Clustering

Clustering is dividing data into groups that have objects with the same characteristics. Clustering plays an important role in data mining applications, such as data science exploration, information access and text mining, special database applications, and web analytics . Basically there are 2 clustering methods namely Hierarchical clustering method and Non Hierarchical clustering method. The hierarchical method is used if the number of groups is not known beforehand while the non-hierarchical method is used if the number of groups is known from a number of objects. One of the non-hierarchical algorithms is the K-Means Algorithm [4] . Clustering is based grouping of data (objects). only on the information found in describing data object it and the relationship among them [6] .

2.4. Definition of Disease

The word illness and pain is two conditions are different, however its use often swapped. Sick word identical with something it doesn't done or abnormal. Need We distinguish sick people (disturbance physiological / body) with people who have problems. Disease is is term described medical as disturbance in function producing body decrease capacity. Disease happen moment body No balanced as well as abnormal circumstances. Sick is something circumstances Where emotional, physical, social, intellectual, developmental, or somebody disturbed or reduced, no only condition disease process occurs [7] .

2.5. Settlement seedy

Settlement seedy is settlement with level occupancy and density very tall buildings, buildings no regular, quality very low house. Besides that no adequate infrastructure and facilities base such as drinking water, roads, waste water and garbage. Slum area is area where home with condition occupancy the community in the area river banks or in landfills and looks very bad. House nor existing facilities and infrastructure no in accordance with standards apply, fine standard requirement, density building, requirements House healthy, needs clean water, sanitation facilities nor condition completeness infrastructure road, space open, as well completeness facility social other [8] .

3. Results and Discussion

3.1. Research methods

Method study is a process or method scientific for get data that will used for necessity research. In do study this, author follow stages methodology in study this that is as following:

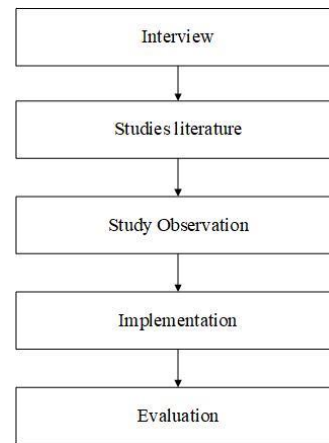


Fig. 1: Research Workflow

Based on the picture above, it can be seen that there are several stages in completing the research, namely:

1. Interview, in method this thing to do is do communication with user later will use system that will built, because in data type clustering disease in residential areas where there are many slums so that need precision to get maximum results, then for the sake of achieving comfortable and suitable with desire user need applied method interview this.
2. Studies literature, research conducted with method understanding to the literature in the form of books mandatory, related with clustering type diseases, settlements, residents and slums.
3. Study field form observation, that is researcher do various observation regarding problem processes type disease in residential areas in slums. after he did observation so writer collect the data that has been got in accordance with need then manage that data into the method clustering so that obtained knowledge new.
4. Implementation Matlab coding for get results calculation clusters and find out connection closest between type of disease, area location, and level of slums. Do revision of the built program so that know experienced coding error.
5. Evaluation is taking conclusions and suggestions that can be done in study this.

3.2. Method Test Analysis

In analysis testing method Clustering in system data grouping used data is needed as input for process and analysis, this research data was taken at two research locations, namely at the Binjai City Health Center data and at the Dinas Permukiman Perumahan (PERKIM) based on the Decree in force at number 8845-1226 IK/YEAR 2021 concerning determination of Slum Housing Locations and Binjai City Slums. the data has made become table in accordance with variable processing to be analyzed with use method Clustering K-Means algorithm, data already transformed the as following:

Table 1: Data Transformation

No.	Type Disease (X)	Area location (Y)	Slum Level (Z)
1	2	17	2
2	4	4	1
3	7	10	2
4	4	3	2
5	5	5	3
6	1	12	2
7	3	19	2
8	2	8	2
9	2	13	3
10	3	14	3
11	4	15	3
12	1	16	3
13	6	20	3
14	8	1	2
15	6	18	2
16	6	11	2
17	4	6	3
18	2	7	3
19	2	2	3
20	5	9	3

Based on the data that has been transformed above, the variables used is Disease type (X), area location (Y), and slum level (Z). The next step is to calculate the data based on the k-means clustering algorithm

Iteration Process 1

Centroid 1 = (2, 17, 2) is taken randomly from data 1

Centroid 2 = (7, 10, 2) is taken randomly from data 3

Centroid 3 = (5, 5, 3) is taken randomly from data 5

Information :

The centroid value is taken randomly from the transformed data

Chapter 1 (2, 17, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(2-2)^2 + (17-17)^2 + (2-2)^2} = 0$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-7)^2 + (17-10)^2 + (2-2)^2} = 8.60$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-5)^2 + (17-5)^2 + (2-3)^2} = 12.41$$

Chapter 2 (4, 4, 1)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(4-2)^2 + (4-17)^2 + (1-2)^2} = 13.19$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-7)^2 + (4-10)^2 + (1-2)^2} = 6.78$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-5)^2 + (4-5)^2 + (1-3)^2} = 2.44$$

Chapter 3 (7, 10, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(7-2)^2 + (10-17)^2 + (2-2)^2} = 8.60$$

$$\text{Distance from C2 (Y)} = \sqrt{(7-7)^2 + (10-10)^2 + (2-2)^2} = 0$$

$$\text{Distance from C3 (Z)} = \sqrt{(7-5)^2 + (10-5)^2 + (2-3)^2} = 5.47$$

Chapter 4 (4, 3, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(4-2)^2 + (3-17)^2 + (2-2)^2} = 14.14$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-7)^2 + (3-10)^2 + (2-2)^2} = 7.61$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-5)^2 + (3-5)^2 + (2-3)^2} = 2.44$$

Chapter 5 (5, 5, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(5-2)^2 + (5-17)^2 + (3-2)^2} = 12.41$$

$$\text{Distance from C2 (Y)} = \sqrt{(5-7)^2 + (5-10)^2 + (3-2)^2} = 5.47$$

$$\text{Distance from C3 (Z)} = \sqrt{(5-5)^2 + (5-5)^2 + (3-3)^2} = 0$$

Chapter 6 (1, 13, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(1-2)^2 + (12-17)^2 + (2-2)^2} = 5.09$$

$$\text{Distance from C2 (Y)} = \sqrt{(1-7)^2 + (12-10)^2 + (2-2)^2} = 6.32$$

$$\text{Distance from C3 (Z)} = \sqrt{(1-5)^2 + (12-5)^2 + (2-3)^2} = 8.12$$

Chapter 7 (3, 19, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(3-2)^2 + (19-17)^2 + (2-2)^2} = 2.23$$

$$\text{Distance from C2 (Y)} = \sqrt{(3-7)^2 + (19-10)^2 + (2-2)^2} = 9.84$$

$$\text{Distance from C3 (Z)} = \sqrt{(3-5)^2 + (19-5)^2 + (2-3)^2} = 14.17$$

Chapter 8 (2, 8, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(2-2)^2 + (8-17)^2 + (2-2)^2} = 9$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-7)^2 + (8-10)^2 + (2-2)^2} = 5.38$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-5)^2 + (8-5)^2 + (2-3)^2} = 4.35$$

Chapter 9 (2, 13, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(2-2)^2 + (13-17)^2 + (3-2)^2} = 4.12$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-7)^2 + (13-10)^2 + (3-2)^2} = 5.91$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-5)^2 + (13-5)^2 + (3-3)^2} = 8.54$$

Chapter 10 (3, 14, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(3-2)^2 + (14-17)^2 + (3-2)^2} = 3.31$$

$$\text{Distance from C2 (Y)} = \sqrt{(3-7)^2 + (14-10)^2 + (3-2)^2} = 5.74$$

$$\text{Distance from C3 (Z)} = \sqrt{(3-5)^2 + (14-5)^2 + (3-3)^2} = 9.22$$

Chapter 11 (4, 15, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(4-2)^2 + (15-17)^2 + (3-2)^2} = 3$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-7)^2 + (15-10)^2 + (3-2)^2} = 5.91$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-5)^2 + (15-5)^2 + (3-3)^2} = 10.05$$

Chapter 12 (1, 16, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(1-2)^2 + (16-17)^2 + (3-2)^2} = 1.73$$

$$\text{Distance from C2 (Y)} = \sqrt{(1-7)^2 + (16-10)^2 + (3-2)^2} = 8.54$$

$$\text{Distance from C3 (Z)} = \sqrt{(1-5)^2 + (16-5)^2 + (3-3)^2} = 11.70$$

Chapter 13 (6, 20, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(6-2)^2 + (20-17)^2 + (3-2)^2} = 5.09$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-7)^2 + (20-10)^2 + (3-2)^2} = 10.10$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-5)^2 + (20-5)^2 + (3-3)^2} = 15.03$$

Chapter 14 (8, 1, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(8-2)^2 + (1-17)^2 + (2-2)^2} = 17.08$$

$$\text{Distance from C2 (Y)} = \sqrt{(8-7)^2 + (1-10)^2 + (2-2)^2} = 9.05$$

$$\text{Distance from C3 (Z)} = \sqrt{(8-5)^2 + (1-5)^2 + (2-3)^2} = 5.09$$

Chapter 15 (6, 18, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(6-2)^2 + (18-17)^2 + (2-2)^2} = 4.12$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-7)^2 + (18-10)^2 + (2-2)^2} = 8.06$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-5)^2 + (18-5)^2 + (2-3)^2} = 13.07$$

Chapter 16 (6, 11, 2)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(6-2)^2 + (11-17)^2 + (2-2)^2} = 7.21$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-7)^2 + (11-10)^2 + (2-2)^2} = 1.41$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-5)^2 + (11-5)^2 + (2-3)^2} = 6.16$$

Chapter 17 (4, 6, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(4-2)^2 + (6-17)^2 + (3-2)^2} = 11.22$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-7)^2 + (6-10)^2 + (3-2)^2} = 5.09$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-5)^2 + (6-5)^2 + (3-3)^2} = 1.41$$

Chapter 18 (2, 7, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(2-2)^2 + (7-17)^2 + (3-2)^2} = 10.05$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-7)^2 + (7-10)^2 + (3-2)^2} = 5.91$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-5)^2 + (7-5)^2 + (3-3)^2} = 3.60$$

Chapter 19 (2, 2, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(2-2)^2 + (2-17)^2 + (3-2)^2} = 15.03$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-7)^2 + (2-10)^2 + (3-2)^2} = 9.48$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-5)^2 + (2-5)^2 + (3-3)^2} = 4.24$$

Chapter 20 (5, 9, 3)

K=3, centroid 1 = (2, 17, 2), centroid 2 = (7, 10, 2), centroid 3 = (5, 5, 3)

$$\text{Distance from C1 (X)} = \sqrt{(5-2)^2 + (9-17)^2 + (3-2)^2} = 8.60$$

$$\text{Distance from C2 (Y)} = \sqrt{(5-7)^2 + (9-10)^2 + (3-2)^2} = 2.44$$

$$\text{Distance from C3 (Z)} = \sqrt{(5-5)^2 + (9-5)^2 + (3-3)^2} = 4$$

For results calculation iteration 1 above can be seen in the table below this:

Table 2: Iteration Results 1

No	Type Disease (X)	Area location (Y)	Slum Level (Z)	Distance From C1	Distance From C2	Distance From C3	Group
1	2	17	2	0	8.60	12.41	1
2	4	4	1	13.19	6.78	2.44	3
3	7	10	2	8.60	0	5.47	2
4	4	3	2	14.14	7.61	2.44	3
5	5	5	3	12.41	5.47	0	3
6	1	12	2	5.09	6.32	8.12	1
7	3	19	2	2.23	9.84	14.17	1
8	2	8	2	9	5.38	4.35	3
9	2	13	3	4.12	5.91	8.54	1
10	3	14	3	3.31	5.74	9.22	1
11	4	15	3	3	5.91	10.04	1
12	1	16	3	1.73	8.54	11.70	1

13	6	20	3	5.09	10.10	15.03	1
14	8	1	2	17.08	9.05	5.09	3
15	6	18	2	4.12	8.06	13.07	1
16	6	11	2	7.21	1.41	6.16	2
17	4	6	3	11.22	5.09	1.41	3
18	2	7	3	10.05	5.91	3.60	3
19	2	2	3	15.03	9.48	4.24	3
20	5	9	3	8.60	2.44	4	2

From the iteration results in the table above, the values in the old group are obtained as follows:

Group old : { 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0 }

New group : { 1,3,2,3,3,1,1,3,1,1,1,1,1,3,1,2,3,3,3,2 }

group change , then proceed to the next iteration.

For group 1 there is 9 data:

$$C_1 X = (2+1+3+2+3+4+1+6+6)/9 = 3$$

$$C_1 Y = (17+13+12+19+13+14+15+16+20+18)/9 = 16$$

$$C_1 Z = (2+2+2+3+3+3+3+3+2)/9 = 3$$

For group 2 there is 3 data:

$$C_2 X = (7+6+5)/3 = 6$$

$$C_2 Y = (10+11+9)/3 = 10$$

$$C_2 Z = (2+2+3)/3 = 2$$

For group 3 there is 8 data:

$$C_3 X = (4+4+5+2+8+4+2+2)/8 = 4$$

$$C_3 Y = (4+3+5+8+1+6+7+2)/8 = 5$$

$$C_3 Z = (1+2+3+2+2+3+3+3)/8 = 2$$

Iteration Process 2

Centroids 1 = (3, 16, 3)

Centroids 2 = (6, 10, 2)

Centroids 3 = (4, 5, 2)

Chapter 1 (2, 17, 2)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(2-3)^2 + (17-16)^2 + (2-3)^2} = 1.73$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-6)^2 + (17-10)^2 + (2-2)^2} = 8.06$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-4)^2 + (17-5)^2 + (2-2)^2} = 12.16$$

Chapter 2 (4, 4, 1)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(4-3)^2 + (1-16)^2 + (1-3)^2} = 12.20$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-6)^2 + (1-10)^2 + (1-2)^2} = 6.40$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-4)^2 + (1-5)^2 + (1-2)^2} = 1.41$$

Chapter 3 (7, 10, 2)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(7-3)^2 + (10-16)^2 + (2-3)^2} = 7.28$$

$$\text{Distance from C2 (Y)} = \sqrt{(7-6)^2 + (10-10)^2 + (2-2)^2} = 1$$

$$\text{Distance from C3 (Z)} = \sqrt{(7-4)^2 + (10-5)^2 + (2-2)^2} = 5.83$$

Chapter 4 (4, 3, 2)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(4-3)^2 + (3-16)^2 + (2-3)^2} = 13.07$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-6)^2 + (3-10)^2 + (2-2)^2} = 7.28$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-4)^2 + (3-5)^2 + (2-2)^2} = 2$$

Chapter 5 (5, 5, 3)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(5-2)^2 + (5-17)^2 + (3-3)^2} = 11.18$$

$$\text{Distance from C2 (Y)} = \sqrt{(5-7)^2 + (5-10)^2 + (3-5)^2} = 5.19$$

$$\text{Distance from C3 (Z)} = \sqrt{(5-5)^2 + (5-5)^2 + (3-3)^2} = 1.41$$

Chapter 6 (1, 13, 2)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

$$\text{Distance from C1 (X)} = \sqrt{(1-3)^2 + (12-16)^2 + (2-3)^2} = 4.58$$

$$\text{Distance from C2 (Y)} = \sqrt{(1-6)^2 + (12-10)^2 + (2-2)^2} = 5.38$$

$$\text{Distance from C3 (Z)} = \sqrt{(1-4)^2 + (12-5)^2 + (2-2)^2} = 7.61$$

Chapter 7 (3, 19, 2)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(3-3)^2 + (19-16)^2 + (2-3)^2} = 3.16$$

$$\text{Distance from C2 (Y)} = \sqrt{(3-6)^2 + (19-10)^2 + (2-2)^2} = 9.48$$

$$\text{Distance from C3 (Z)} = \sqrt{(3-4)^2 + (19-5)^2 + (2-2)^2} = 14.03$$

Chapter 8 (2, 8, 2)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(2-3)^2 + (8-16)^2 + (2-3)^2} = 8.12$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-6)^2 + (8-10)^2 + (2-2)^2} = 4.47$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-4)^2 + (8-5)^2 + (2-2)^2} = 3.60$$

Chapter 9 (2, 13, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(2-3)^2 + (13-16)^2 + (3-3)^2} = 3.16$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-6)^2 + (13-10)^2 + (3-2)^2} = 5.09$$

$$\text{Distance from C3 (Z)} = \sqrt{(2-4)^2 + (13-5)^2 + (3-2)^2} = 8.30$$

Chapter 10 (3, 14, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(3-3)^2 + (14-16)^2 + (3-3)^2} = 2$$

$$\text{Distance from C2 (Y)} = \sqrt{(3-6)^2 + (14-10)^2 + (3-2)^2} = 5.09$$

$$\text{Distance from C3 (Z)} = \sqrt{(3-4)^2 + (14-5)^2 + (3-2)^2} = 9.11$$

Chapter 11 (4, 15, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(4-3)^2 + (15-16)^2 + (3-3)^2} = 1.41$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-6)^2 + (15-10)^2 + (3-2)^2} = 5.47$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-4)^2 + (15-5)^2 + (3-2)^2} = 10.05$$

Chapter 12 (1, 16, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(1-3)^2 + (16-16)^2 + (3-3)^2} = 2$$

$$\text{Distance from C2 (Y)} = \sqrt{(1-6)^2 + (16-10)^2 + (3-2)^2} = 7.87$$

$$\text{Distance from C3 (Z)} = \sqrt{(1-4)^2 + (16-5)^2 + (3-2)^2} = 11.44$$

Chapter 13 (6, 20, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(6-3)^2 + (20-16)^2 + (3-3)^2} = 5$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-6)^2 + (20-10)^2 + (3-2)^2} = 10.05$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-4)^2 + (20-5)^2 + (3-2)^2} = 15.16$$

Chapter 14 (8, 1, 2)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(8-3)^2 + (1-16)^2 + (2-3)^2} = 15.84$$

$$\text{Distance from C2 (Y)} = \sqrt{(8-6)^2 + (1-10)^2 + (2-2)^2} = 9.22$$

$$\text{Distance from C3 (Z)} = \sqrt{(8-4)^2 + (1-5)^2 + (2-2)^2} = 5.65$$

Chapter 15 (6, 18, 2)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(6-3)^2 + (18-16)^2 + (2-3)^2} = 3.74$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-6)^2 + (18-10)^2 + (2-2)^2} = 8$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-4)^2 + (18-5)^2 + (2-2)^2} = 13.15$$

Chapter 16 (6, 11, 2)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(6-3)^2 + (11-16)^2 + (2-3)^2} = 5.91$$

$$\text{Distance from C2 (Y)} = \sqrt{(6-6)^2 + (11-10)^2 + (2-2)^2} = 1$$

$$\text{Distance from C3 (Z)} = \sqrt{(6-4)^2 + (11-5)^2 + (2-2)^2} = 6.32$$

Chapter 17 (4, 6, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(4-3)^2 + (6-16)^2 + (3-3)^2} = 10.05$$

$$\text{Distance from C2 (Y)} = \sqrt{(4-6)^2 + (6-10)^2 + (3-2)^2} = 4.58$$

$$\text{Distance from C3 (Z)} = \sqrt{(4-4)^2 + (6-5)^2 + (3-2)^2} = 1.41$$

Chapter 18 (2, 7, 3)

$$K=3, \text{centroid 1} = (3, 16, 3), \text{centroid 2} = (16, 10, 2), \text{centroid 3} = (4, 5, 2)$$

$$\text{Distance from C1 (X)} = \sqrt{(2-3)^2 + (7-16)^2 + (3-3)^2} = 9.05$$

$$\text{Distance from C2 (Y)} = \sqrt{(2-6)^2 + (7-10)^2 + (3-2)^2} = 5.09$$

Distance from C3 (Z) = $\sqrt{(2 - 4)^2 + (7 - 5)^2 + (3 - 2)^2} = 3$

Chapter 19 (2, 2, 3)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

Distance from C1 (X) = $\sqrt{(2 - 3)^2 + (2 - 16)^2 + (3 - 3)^2} = 14.03$

Distance from C2 (Y) = $\sqrt{(2 - 6)^2 + (2 - 10)^2 + (3 - 2)^2} = 9$

Distance from C3 (Z) = $\sqrt{(2 - 4)^2 + (2 - 5)^2 + (3 - 2)^2} = 3.74$

Chapter 20 (5, 9, 3)

K=3, centroid 1 = (3, 16, 3), centroid 2 = (16, 10, 2), centroid 3 = (4, 5, 2)

Distance from C1 (X) = $\sqrt{(5 - 3)^2 + (9 - 16)^2 + (3 - 3)^2} = 7.28$

Distance from C2 (Y) = $\sqrt{(5 - 6)^2 + (9 - 10)^2 + (3 - 2)^2} = 1.73$

Distance from C3 (Z) = $\sqrt{(5 - 4)^2 + (9 - 5)^2 + (3 - 2)^2} = 4.24$

For results calculation iteration 2 above can be seen in the table below this:

Table 3: Iteration Results 2

No	Type Disease (X)	Area location (Y)	Slum Level (Z)	Distance From C1	Distance From C2	Distance From C3	Group
1	2	17	2	1.73	8.06	12.16	1
2	4	4	1	12.2 0	6.4 0	1.41	3
3	7	10	2	7.28	1	5.83	2
4	4	3	2	13.07	7.28	2	3
5	5	5	3	11.18	5.19	1.41	3
6	1	12	2	4.58	5.38	7.61	1
7	3	19	2	3.16	9.48	14.03	1
8	2	8	2	8.12	4.47	3.6 0	3
9	2	13	3	3.16	5.09	8.3 0	1
10	3	14	3	2	5.09	9.11	1
11	4	15	3	1.41	5.47	10.0 5	1
12	1	16	3	2	7.87	11.44	1
13	6	20	3	5	10.0 5	15.16	1
14	8	1	2	15.84	9.2 2	5.65	3
15	6	18	2	3.74	8	13.15	1
16	6	11	2	5.91	1	6.32	2
17	4	6	3	10.0 5	4.58	1.41	3
18	2	7	3	9.05	5.09	3	3
19	2	2	3	14.03	9	3.74	3
20	5	9	3	7.28	1.73	4.24	2

From the results of iterations I and II in the table above, the values for the old group are obtained as follows:

Old groups: { 1,3,2,3,3,1,1,3,1,1,1,1,3,1,2,3,3,3,2 }

New group: { 1,3,2,3,3,1,1,3,1,1,1,1,3,1,2,3,3,3,2 }

Because on iteration 1st and 2nd position clusters No changed or there is equation, then calculation iteration stopped and get results as following:

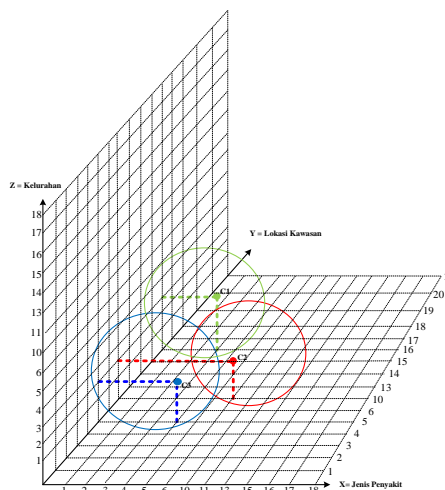


Fig. 2: Clustering Graph

CENTER		
X	Y	Z
(3	16	3)
(6	10	2)
(4	5	2)

Explanation from results on are:

Of the 20 data, there are 3 groups that is group 1 exists 9 data and 2 groups there is 3 data and group 3 exists 8 data. As for the explanation there is from 3 groups the as following:

1. Centroids 1: 3 16 3
Based on calculation on can is known that in cluster 1 group diseases in residential areas in slums group type of disease (X) is Dengue Hemorrhagic Fever (DHF), and for group the location of the area (Y) is the Satria 4 area which is at the slum level (Z), namely High.
2. Centroids 2: 6 10 2
Can is known that in cluster 2 groups disease in residents' settlements in slums in the group type of disease (X) is Muntaber, and for group the location of the area (Y) namely Lorong Sungai which is at the slum level (Z), namely Medium.
3. Centroids 3: 4 5 2
Can is known that in cluster 3 groups disease in residents' settlements in slums in the group type of disease (X) is Diarrhea, and for group the location of the area (Y) is the satria 2 area which is at the slum level (Z), namely Medium.

4. Conclusion

Based on results clustering this give information important about patterns and groups dominant disease occurs in settlements inhabitant in place slums in the city Binjai. Every cluster own characteristic typical disease certain inclined often occurs in settlements seedy on site certain. Cluster 1 includes case fever Dengue Hemorrhagic Fever (DHF) in the Satria 4 area with level slum high. Cluster 2 includes case muntaber in the river alley with level slum medium. Cluster 3 includes case Diarrhea in the Satria 2 area with level slum medium.

References

- [1] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "APPLEMENT OF DATA MINING FOR PATIENT MEDICAL RECORD DATA GROUPING BASED ON TYPE OF DISEASE WITH CLUSTERING ALGORITHM (Case Study: Polyclinic PT. Inecda)," *J , Computing Science.* , vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [2] M. Arhami and M. Nasir, *Data Mining Algorithm and Implementation* . Yogyakarta: ANDI Publisher, 2020.
- [3] H. Siska, S. Aji, and S. Eko, "Implementation of Data Mining to Predict Student Study Period Using the C4.5 Algorithm (Case Study: Dehasen University Bengkulu)," *J. Media Infotama* , vol. 11, no. 2, pp. 130–138, 2020.
- [4] R. Buuon, M. Zarlis, and V. Yasin, *Data Mining Concepts in Implementation* . Jakarta: Publisher Mitra Wacana Media, 2021.
- [5] G. Gustientiedina, MH Adiya, and Y. Desnelita, "Application of the K-Means Algorithm for Drug Data Clustering," *J. Nas. Technol. and Sist. inf.* , vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [6] E. Prasetyo, *Data mining: concepts and applications using MATLAB* , 1st ed. CV Andi Offset, 2012.
- [7] SM Bangun, M. Kes, and A. Saputra, "Basic module of public health science," 2021.
- [8] A. Aguswin, "Structuring Slum Areas in Sukajaya Village, Cibitung District, Bekasi Regency," *J. IKRAITH-TECHNOLOGY Vol 5 No 3 Novemb. 2021* , vols. 5, no. 3, pp. 48–57, 2021.