Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2023. Vol. 3. No. 1; e-ISSN : 2808-4519

Grouping Patient Data Based On Work And Place Of Residence On Perceived Complaints

Jhody Alkhalis Sembiring[1], Yani Maulita[2], Suci Ramadani[3]

[1,2,3]Informatics Enginering, STMIK Kaputama
Jl. Veterans No. 4A-9A, Binjai, North Sumatra, Indonesia
jhodysembiring26@gmail.com[1]*

Abstract

Every day the Sawit Seberang Health Center serves many patients with various kinds of disease complaints from various areas in Sawit Seberang District. The number of patients can even reach tens of people in one day resulting in a large number of patient visit data. Limited information regarding the spread of diseases that are often suffered by patients in several areas at the Sawit Seberang Health Center has resulted in less optimal policy action, anticipation of treatment and prevention of disease in the community. To find information about grouping patient data based on work and place of residence for perceived complaints, a large or large data mining technique is needed, namely data mining techniques using the clustering method. The purpose of this study is to process and cluster patient data based on work, place of residence and complaints that are felt using the Clustering method, to analyze the results of applying data mining using K-Means Clustering in grouping patient data based on work, place of residence and complaints that are felt and find out the results of the settlement grouping patient data based on work and place of residence on perceived complaints using clustering and data mining methods.

Keywords: *Data Mining, Work occupational patient data, residence, complaint*

1. Introduction

The Sawit Seberang Health Center serves many patients with various kinds of disease complaints from various areas in Sawit Seberang District. The number of patients can even reach tens of people in one day resulting in a large number of patient visit data. Limited information regarding the spread of diseases that are often suffered by patients in several areas at the Sawit Seberang Health Center has resulted in less optimal policy action, anticipation of treatment and prevention of disease in the community. To find information about grouping patient data based on work and place of residence for perceived complaints, a large or large data mining technique is needed, namely data mining techniques using the clustering method. In addition, the different patient complaints at the Sawit Seberang Health Center resulted in more complex data to be processed. Applying data mining to grouping patient data based on work, place of residence, and perceived complaints using the K-Means algorithm. So a data mining technology design is needed to maximize the performance of the Sawit Seberang Health Center in knowing which types of diseases are most commonly suffered by patients at the Sawit Seberang Health Center who tend to experience or attack a patient with the most victims or sufferers and can provide benefits regarding counseling, treatment anticipation and find out what is the cause by paying attention to all kinds of factors.

2. Metodologi

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. Data mining bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan. According to Erwin (2017, h.1) data mining is one of the main parts or processes of Knowledge Discovery in Database (KDD) whose form of activity is collecting and using past data to find regularities, patterns or relationships in a larger data set. Broadly speaking, KDD includes three stages, namely pre-processing, process (data mining) and post-processing .Data mining has now become a powerful new technology with great potential to help companies focus on the most important information in the data they have collected about the behavior of their customers and potential customers. Through data mining, companies can find information in large amounts of data through precise and effective

processing with various methods available in data mining so that in simple terms data mining can be described as a pattern or model or rule or knowledge generated from data mining.

2.2 Algoritma K-Means

The K-Means algorithm is a relatively simple algorithm for classifying or grouping a large number of objects with certain attributes into groups (clusters) of K. In the K-Means algorithm, the number of K clusters is predetermined. According to Baginda Harahap (2021) states that K-Means is a non-hierarchical grouping method that tries to partition data into clusters or groups so that data with the same characteristics will be included in the same cluster and data with different characteristics are grouped into other groups.

2.3 Clustering

Clustering is a data analysis method, which is often included as a data mining method, the purpose of which is to group data with the same characteristics. According to (Zulfadhillah et al., 2016) cluster analysis (clustering) is finding objects in one group that are the same (have a relationship) with others and are located (points have a relationship) with objects in other groups. The main goal of the Clustering method is to group a number of data or objects into clusters (groups) so that each cluster contains data that is as similar as possible. The clustering method seeks to place similar (closely spaced) objects 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 as can be seen in Figure 1.

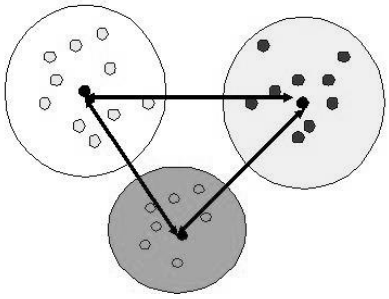

Figure 1: Examples of Clusters Formed

2.4 Patient

A patient is a person who has physical or mental weaknesses submits supervision and care, receives and follows the treatment prescribed by a health worker. According to the Ministry of Health, a patient is anyone who consults about their health problems to obtain the necessary health services, either directly or indirectly at the hospital.

2.5 Work

In human life always hold various activities. One of these activities is manifested in movements called work. Work implies carrying out a task that ends with the fruit of work that can be enjoyed by the human being concerned. An important driving factor that causes humans to work is the existence of needs that must be met.

2.6 Residence

Residence is the location where a person or group lives and resides. A residence can be a house, apartment, condominium or other place used for the purpose of living. Residence can also refer to the area or region in which a person or group lives, such as a city or village. In a broader context, place of residence can also refer to a country or a larger geographic area.

2.7 Complaint

A complaint is a statement expressing dissatisfaction or discomfort with something, such as a product, service, situation or person. Complaints can be expressions of dissatisfaction with the quality or quantity of a product or service, discomfort with certain situations or circumstances, or dissatisfaction with someone's actions or behavior. Complaints can be stated orally or in writing and can be made directly to the party concerned or through the complaint channels provided by the company or organization. Complaints are often used as a form of feedback to help companies or organizations improve their products or services and increase customer satisfaction.

3. Results and Discussion

3.1. Calculation Clustering

In using the clustering method, the initial process carried out to form clusters is to transform data into numeric form with predetermined codes, then determine the number of groups (K), calculate the centroids, calculate the use of objects to the centroids and then group them based on work and place of residence against complaints that are felt, if no objects move or groups then the iteration is complete. Then transform the criteria data above to be calculated using the clustering method. The data transformation from the data above can be seen in the table below.

Table 1: Transformation data

No	Object	X	Y	Z
1	A	1	2	1
2	B	4	2	1
3	C	1	2	2
4	D	1	3	3
5	E	3	4	5
6	F	4	1	4
7	G	4	4	4
8	H	1	2	3
9	I	3	1	2
10	J	1	2	1
11	K	3	4	5
12	L	3	1	3
13	M	4	1	4
14	N	1	3	1
15	O	2	2	2
16	P	3	4	5
17	Q	2	2	3
18	R	4	1	2
19	S	1	2	3
20	T	2	2	1

Then form a cluster into 3 groups (K = 3) and determine the centroid center point. The clustering calculation process is as follows.

K=3 Centroid

C1= (1, 2, 3) taken from the data H

C2= (3, 1, 2) taken from the data I

C3= (2, 2, 3) taken from the data Q

Then do the calculations like the calculation process below: Iteration 1 :

1. A (1,2,1)

   C_1= (1, 2, 3) = $\sqrt{(1-1)^2+(2-2)^2+(1-3)^2} = 2$

   C_2= (3, 1, 2)= $\sqrt{(1-3)^2+(2-1)^2+(1-2)^2} = 2{,}45$

   C_3= (2, 2, 3) = $\sqrt{(1-2)^2+(2-2)^2+(1-3)^2} = 2{,}24$

2. B (4,2,1)

   C_1= (1, 2, 3) = $\sqrt{(4-1)^2+(2-2)^2+(1-3)^2} = 3{,}61$

   C_2= (3, 1, 2)= $\sqrt{(4-3)^2+(2-1)^2+(1-2)^2} = 1{,}73$

   C_3= (2, 2, 3) = $\sqrt{(4-2)^2+(2-2)^2+(1-3)^2} = 2{,}83$

3. C (1,2,2)

   C_1= (1, 2, 3) = $\sqrt{(1-1)^2+(2-2)^2+(2-3)^2} = 1$

   C_2= (3, 1, 2)= $\sqrt{(1-3)^2+(2-1)^2+(2-2)^2} = 2{,}24$

   C_3= (2, 2, 3) = $\sqrt{(1-2)^2+(2+(2-3)^2} = 1{,}41$

4. D (1,3,3)

   C_1= (1, 2, 3) = $\sqrt{(1-1)^2+(3-2)^2+(3-3)^2} = 1$

   C_2= (3, 1, 2)= $\sqrt{(1-3)^2+(3-1)^2+(3-2)^2} = 3$

   C_3= (2, 2, 3) = $\sqrt{(1-2)^2+(3-2)^2+(3-3)^2} = 1{,}41$

5. E (3,4,5)

C_1= (1, 2, 3) = $\sqrt{(3-1)^2+(4-2)^2+(5-3)^2} = 3{,}46$

C_2= (3, 1, 2)= $\sqrt{(3-3)^2+(4-1)^2+(5-2)^2} = 4{,}24$

C_3= (2, 2, 3) = $\sqrt{(3-2)^2+(4-2)^2+(5-3)^2} = 3$

6. F (4,1,4)

C_1= (1, 2, 3) = $\sqrt{(4-1)^2+(1-2)^2+(4-3)^2} = 3{,}32$

C_2= (3, 1, 2)= $\sqrt{(4-3)^2+(1-1)^2+(4-2)^2} = 2{,}24$

C_3= (2, 2, 3) = $\sqrt{(4-2)^2+(1-2)^2+(4-3)^2} = 2{,}45$

7. G (4,4,4)

C_1= (1, 2, 3) = $\sqrt{(4-1)^2+(4-2)^2+(4-3)^2} = 3{,}74$

C_2= (3, 1, 2)= $\sqrt{(4-3)^2+(4-1)^2+(4-2)^2} = 3{,}74$

C_3= (2, 2, 3) = $\sqrt{(4-2)^2+(4-2)^2+(4-3)^2} = 3$

8. H (1,2,3)

C_1= (1, 2, 3) = $\sqrt{(1-1)^2+(2-2)^2+(3-3)^2} = 0$

C_2= (3, 1, 2)= $\sqrt{(1-3)^2+(2-1)^2+(3-2)^2} = 2{,}45$

C_3= (2, 2, 3) = $\sqrt{(1-2)^2+(2-2)^2+(3-3)^2} = 1$

9. I (3,1,2)

C_1= (1, 2, 3) = $\sqrt{(3-1)^2+(1-2)^2+(2-3)^2} = 2{,}45$

C_2= (3, 1, 2)= $\sqrt{(3-3)^2+(1-1)^2+(2-2)^2} = 0$

C_3= (2, 2, 3) = $\sqrt{(3-2)^2+(1-2)^2+(2-3)^2} = 1{,}73$

10. J (1,2,1)

C_1= (1, 2, 3) = $\sqrt{(1-1)^2+(2-2)^2+(1-3)^2} = 2$

C_2= (3, 1, 2)= $\sqrt{(1-3)^2+(2-1)^2+(1-2)^2} = 2{,}45$

C_3= (2, 2, 3) = $\sqrt{(1-2)^2+(2-2)^2+(1-3)^2} = 2{,}24$

11. K (3,4,5)

C_1= (1, 2, 3) = $\sqrt{(3-1)^2+(4-2)^2+(5-3)^2} = 3{,}46$

C_2= (3, 1, 2)= $\sqrt{(3-3)^2+(4-1)^2+(5-2)^2} = 4{,}24$

C_3= (2, 2, 3) = $\sqrt{(3-2)^2+(4-2)^2+(5-3)^2} = 3$

12. L (3,1,3)

C_1= (1, 2, 3) = $\sqrt{(3-1)^2+(1-2)^2+(3-3)^2} = 2{,}24$

C_2= (3, 1, 2)= $\sqrt{(3-3)^2+(1-1)^2+(3-2)^2} = 1$

C_3= (2, 2, 3) = $\sqrt{(3-2)^2+(1-2)^2+(3-3)^2} = 1{,}41$

13. M (4,1,4)

C_1= (1, 2, 3) $= \sqrt{(4-1)^2 + (1-2)^2 + (4-3)^2} = 3{,}32$

C_2= (3, 1, 2)= $\sqrt{(4-3)^2 + (1-1)^2 + (4-2)^2} = 2{,}24$

C_3= (2, 2, 3) = $\sqrt{(4-2)^2 + (1-2)^2 + (4-3)^2} = 2{,}45$

14. N (1,3,1)

C_1= (1, 2, 3) $= \sqrt{(1-1)^2 + (3-2)^2 + (1-3)^2} = 2{,}24$

C_2= (3, 1, 2)= $\sqrt{(1-3)^2 + (3-1)^2 + (1-2)^2} = 3$

C_3= (2, 2, 3) = $\sqrt{(1-2)^2 + (3-2)^2 + (1-3)^2} = 2{,}45$

15. O (2,2,2)

C_1= (1, 2, 3) $= \sqrt{(2-1)^2 + (2-2)^2 + (2-3)^2} = 1{,}41$

C_2= (3, 1, 2)= $\sqrt{(2-3)^2 + (2-1)^2 + (2-2)^2} = 1{,}41$

C_3= (2, 2, 3) = $\sqrt{(2-2)^2 + (2-2)^2 + (2-3)^2} = 1$

16. P (3,4,5)

C_1= (1, 2, 3) $= \sqrt{(3-1)^2 + (4-2)^2 + (5-3)^2} = 3{,}46$

C_2= (3, 1, 2)= $\sqrt{(3-3)^2 + (4-1)^2 + (5-2)^2} = 4{,}24$

C_3= (2, 2, 3) = $\sqrt{(3-2)^2 + (4-2)^2 + (5-3)^2} = 3$

17. Q (2,2,3)

C_1= (1, 2, 3) $= \sqrt{(2-1)^2 + (2-2)^2 + (3-3)^2} = 1{,}00$

C_2= (3, 1, 2)= $\sqrt{(2-3)^2 + (2-1)^2 + (3-2)^2} = 1{,}73$

C_3= (2, 2, 3) = $\sqrt{(2-2)^2 + (2-2)^2 + (3-3)^2} = 0$

18. R (4,1,2)

C_1= (1, 2, 3) $= \sqrt{(4-1)^2 + (1-2)^2 + (2-3)^2} = 3{,}32$

C_2= (3, 1, 2)= $\sqrt{(4-3)^2 + (1-1)^2 + (2-2)^2} = 1$

C_3= (2, 2, 3) = $\sqrt{(4-2)^2 + (1-2)^2 + (2-3)^2} = 2{,}45$

19. S (1,2,3)

C_1= (1, 2, 3) $= \sqrt{(1-1)^2 + (2-2)^2 + (3-3)^2} = 0$

C_2= (3, 1, 2)= $\sqrt{(1-3)^2 + (2-1)^2 + (3-2)^2} = 2{,}45$

C_3= (2, 2, 3) = $\sqrt{(1-2)^2 + (2-2)^2 + (3-3)^2} = 1$

20. T (2,2,1)

C_1= (1, 2, 3) $= \sqrt{(2-1)^2 + (2-2)^2 + (1-3)^2} = 2{,}24$

C_2= (3, 1, 2)= $\sqrt{(2-3)^2 + (2-1)^2 + (1-2)^2} = 1{,}73$

C_3= (2, 2, 3) = $\sqrt{(2-2)^2 + (2-2)^2 + (1-3)^2} = 2$

From the calculation above, the results of the iteration 1 calculation are obtained, which are as shown in the table below.

Table 2: Iteration Result Data 1

No	X	Y	Z	C1	C2	C3	Group
1	1	2	1	2	2,45	2,24	1
2	4	2	1	3,61	1,73	2,83	2
3	1	2	2	1	2,24	1,41	1
4	1	3	3	1	3	1,41	1
5	3	4	5	3,46	4,24	3	3
6	4	1	4	3,32	2,24	2,45	2
7	4	4	4	3,74	3,74	3	3
8	1	2	3	0	2,45	1	1
9	3	1	2	2,45	0	1,73	2
10	1	2	1	2	2,45	2,24	1
11	3	4	5	3,46	4,24	3	3
12	3	1	3	2,24	1	1,41	2
13	4	1	4	3,32	2,24	2,45	2
14	1	3	1	2,24	3	2,45	1
15	2	2	2	1,41	1,41	1	3
16	3	4	5	3,46	4,24	3	3
17	2	2	3	1	1,73	0	3
18	4	1	2	3,32	1	2,45	2
19	1	2	3	0	2,45	1	1
20	2	2	1	2,24	1,73	2	2

After calculating using the existing cluster formula, the groups based on the minimum distance to the nearest centroid are:

Old Group: (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)

New Group: (1 2 1 1 3 2 3 1 2 1 3 2 2 1 3 3 3 2 1 2)

There is a group change, followed by the following iteration::

Table 3: Iteration Result Data 2

No	X	Y	Z	C1	C2	C3	Group
1	1	2	1	1,04	2,91	3,76	1
2	4	2	1	3,18	1,70	3,48	2
3	1	2	2	0,29	2,57	3,02	1
4	1	3	3	1,23	3,03	2,11	1
5	3	4	5	3,99	3,76	1,21	3
6	4	1	4	3,83	1,70	2,61	2
7	4	4	4	3,99	3,19	1,34	3
8	1	2	3	1,04	2,60	2,48	1
9	3	1	2	2,38	0,67	3,08	2
10	1	2	1	1,04	2,91	3,76	1
11	3	4	5	3,99	3,76	1,21	3
12	3	1	3	2,58	0,77	2,54	2
13	4	1	4	3,83	1,70	2,61	2
14	1	3	1	1,23	3,30	3,53	1
15	2	2	2	1,04	1,65	2,54	1
16	3	4	5	3,99	3,76	1,21	3
17	2	2	3	1,44	1,70	1,86	1
18	4	1	2	3,26	0,77	3,29	2
19	1	2	3	1,04	2,60	2,48	1
20	2	2	1	1,44	2,14	3,39	1

After calculating using the existing cluster formula, in iteration 3 it is the same as in iteration 2 and there is no data that moves groups again so the calculation can be stopped. So that a cluster graph can be made for grouping patient data based on work and place of residence for the complaints that are felt.

Old Group: (1 2 1 1 3 2 3 1 2 1 3 2 2 1 3 3 3 2 1 2)

New Group: (1 2 1 1 3 2 3 1 2 1 3 2 2 1 1 3 1 2 1 1)

Table 4: Iteration Result Data 3

No	X	Y	Z	C1	C2	C3	Group
1	1	2	1	1,06	3,25	4,81	1
2	4	2	1	2,89	1,89	4,32	2
3	1	2	2	0,36	2,87	4,08	1
4	1	3	3	1,32	3,25	3,02	1
5	3	4	5	3,89	3,73	0,35	3
6	4	1	4	3,57	1,38	3,18	2

7	4	4	4	3,81	3,15	1,06	3
8	1	2	3	1,06	2,81	3,48	1
9	3	1	2	2,08	0,96	4,08	2
10	1	2	1	1,06	3,25	4,81	1
11	3	4	5	3,89	3,73	0,35	3
12	3	1	3	2,31	0,76	3,48	2
13	4	1	4	3,57	1,38	3,18	2
14	1	3	1	1,32	3,64	4,49	1
15	2	2	2	0,73	1,98	3,62	1
16	3	4	5	3,89	3,73	0,35	3
17	2	2	3	1,24	1,89	2,94	1
18	4	1	2	2,95	0,76	4,14	2
19	1	2	3	1,06	2,81	3,48	1
20	2	2	1	1,24	2,50	4,43	1

After calculating using the cluster formula in iteration 3, the groups based on the minimum distance to the nearest centroid are:

Old Group (1 2 1 1 3 2 3 1 2 1 3 2 2 1 1 3 1 2 1 1)

New Group (1 2 1 1 3 2 3 1 2 1 3 2 2 1 1 3 1 2 1 1)

3.2. Clustering Chart

The following is a cluster graph based on the calculation of the results of data mining iterations of grouping patient data based on work and place of residence for the complaints that are felt. The graphs obtained are as follows:

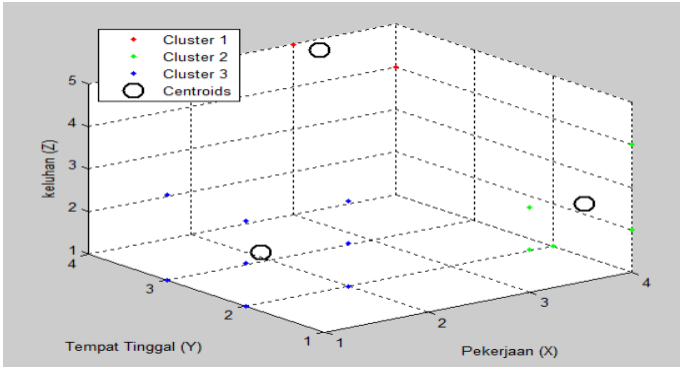

Figure 2: Cluster Chart

Graphic Explanation :

Of the 20 patient data groupings based on work and place of residence for complaints that were felt, 3 groups were obtained, cluster 1 contained 10 data, cluster 2 contained 6 data, and cluster 3 contained 4 data.

1. Cluster 1 There are 10 Data
   It can be seen that cluster 1 is centered on 1.30, 2.20, 2, namely patient data for employee work, residence of Sawit Hulu with complaints of Coughing.
2. Cluster 2 There are 6 Data
   It can be seen that cluster 2 is centered on 3.67, 1.17, 2.67, namely patient data for Housewife work, where Sawit Seberang lives with complaints of Dizziness.
3. Cluster 3 There are 4 Data
   It can be seen in cluster 3 centered on 3.65, 4, 4.75, namely patient data for Housewife work, where Alur Gadung lives with complaints of Itching.

3.3 Interface Design

1. Home

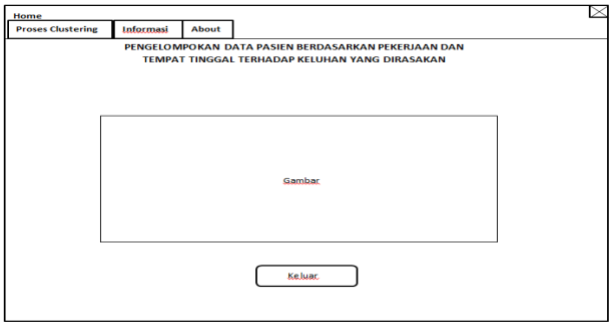

Figure 3: Home Menu

2. Clustering Process

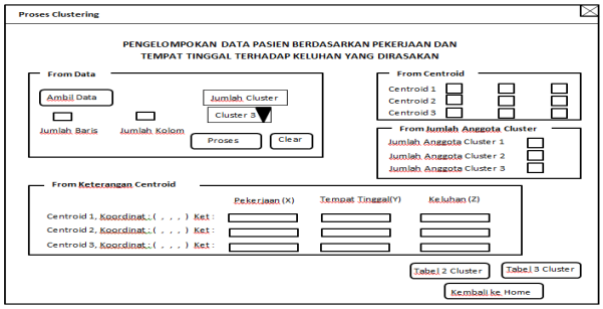

Figure 4: Clustering Process

3. Data Information

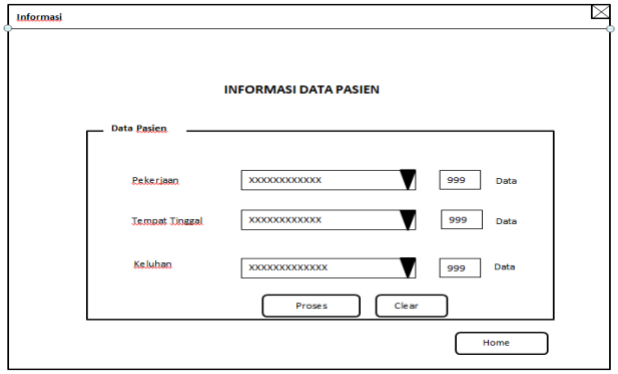

Figure 5: Data Information

4. About

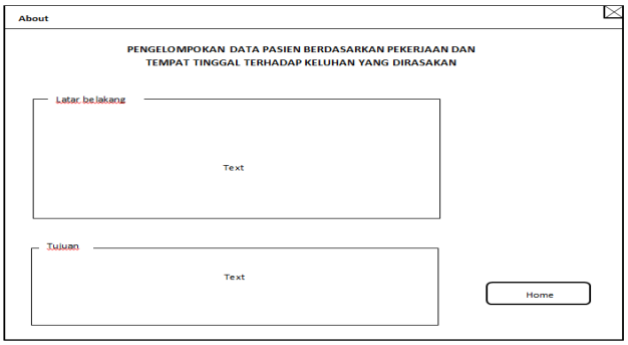

Figure 6: About

References

[1] Amir Tjolleng, 2017, Pengantar Pemrograman MATLAB, Jakarta: PT. Alex Media.
[2] Anindya Khrisna Wardhani, 2016, Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan, Jurnal: Politeknik Rukun Abdi Luhur.
[3] Arhami, Muhammad dan Muhammad Nasir. 2020. Data Mining Algoritma dan Implementasi. Yogyakarta: CV Andi Offset.
[4] Lamhot Sitorus, 2015, Algoritma dan Pemrograman, Yogyakara : Andi Offset
[5] Muhammad Arhami dan Muhammad Natsir, 2020, Data Mining Algoritma dan Implementasi, Yogyakarta: CV. Andi Ofset.
[6] Prasetyowati, Erwin. 2017. Data Mining Pengelompokan Data untuk Informasi dan Evaluasi. Pamekasan: Duta Media Publishing.
[7] Teguh Widiarsono, 2015, Tutorial Praktis Belajar Matlab, Jakarta : Republika.