Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Grouping Data On Infrastructure Development In Langkat District Using The Clustering Method (Case Study: PUPR, Langkat Regency)

Diva Alifya[1], Relita Buaton[2], Suci Ramadani[3]

[1,2]STMIK Kaputama

divaalifya9@gmail.com[1]; bbcbuaton@gmail.com[2]; suci.ramadani23@gmail.com[3]*

Abstract

A building is a man-made structure consisting of walls and a roof permanently erected in a place. Buildings can also be called houses and buildings, namely all facilities, infrastructure or infrastructure in culture as well as human life in building their civilization. Public Works and Public Housing (PUPR) play an important role in increasing the development of national infrastructure in Indonesia so that PUPR can assist in clustering research in infrastructure development in Langkat Regency which is very large every year by grouping the data based on activity names, company names, sub-districts development, and look at the last four years.To classify existing development infrastructure in Langkat Regency with the previous system used by the PUPR Service which is still running by recording in a ledger and hindering reporting performance in grouping PUPR service infrastructure development in road construction, bridge construction and others. So that the existence of grouping using the clustering method helps the PUPR service in clustering infrastructure development data in Langkat Regency to be more effective and efficient.The clustering method is one of the methods that can be applied in classifying infrastructure development data taken from the analysis of Langkat Regency PUPR data regarding developments that have taken place in several sub-districts in Langkat Regency. This clustering method has been widely used by previous studies to group data.

Keywords: *Datamining, infrastructure development, K – Means*

1. Introduction

A building is a man-made structure consisting of walls and a roof permanently erected in a place. Buildings can also be called houses and buildings, namely all facilities, infrastructure or infrastructure in culture as well as human life in building their civilization. Public Works and Public Housing (PUPR) play an important role in increasing the development of national infrastructure in Indonesia so that PUPR can assist in clustering research in infrastructure development in Langkat Regency which is very large every year by grouping the data based on activity names, company names, sub-districts development, and look at the last four years [3]. To classify existing development infrastructure in Langkat Regency with the previous system used by the PUPR Service which is still running by recording in a ledger and hindering reporting performance in grouping PUPR service infrastructure development in road construction, bridge construction and others. So that the existence of grouping using the clustering method helps the PUPR service in clustering infrastructure development data in Langkat Regency to be more effective and efficient.

2. Research Methods

2.1. Data Mining

The process of discovering knowledge in databases is often referred to as data mining or data mining. Data mining is an analysis step of a set of data that is generally large in size to obtain relationships between these data and summarize them in a form that is easy to understand and use. The resulting relationships and summaries are generally in the form of models or patterns [5].

2.2. Clustering

One of the techniques known in data mining is clustering. The definition of scientific clustering in data mining is grouping a number of data or objects into clusters (groups) so that each cluster contains data that is as similar as possible and different from objects in other clusters. The most widely used clustering method is the K-Means clustering method. The main drawback of this method is that the results are sensitive to the selection of the initial cluster center and the calculation of local solutions to achieve optimal conditions. Cluster

analysis is a multivariate technique whose main objective is to group objects based on their characteristics. Cluster analysis classifies objects so that each object that is most closely related to other objects is in the same cluster [8].
The main objective of the clustering method is to group a number of data/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. [1]

2.3. K-Means Algorithm

K-means itself is an algorithm for grouping or clustering. K-means is part of Unsupervised Learning which functions to share data into several groups with a system partition. This algorithm accepts input in the form of label data [2].
The K-Means method attempts to group existing data into several groups, where data in one group have the same characteristics as each other and have different characteristics from data in other groups [4].
The K-Means algorithm in cluster formation is as follows:

1. Suppose given a matrix X = {Xij) data of size n x p with i = 1,2,….n,j=1,2,…p and assume the number of initial clusters is K.
2. Determine the centroids.
3. Calculate the distance of each object to each centroid using the dEcclidean distance or can be written as follows:
   J(xici)=√(*xi* − *ci*) 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 that lies in the same cluster.
6. Repeat step 3, if the initial and new centroids are not the same.

There are several appropriate ways to measure the distance of data to the center of the group, including Euclidean, Manhattan/City Block, and Minkwosky, each method has advantages and disadvantages. Measuring distances in the de-Euclidean distance space uses the formula:

$$D(x2,x1) = \| X2 - X1 \| 2 = \sqrt{\sum} |X2j - X1j| \; 2 \; p \; j=1 \quad (2)$$

Definition 2.1 :
D is the distance between data X2 and X1, and | . | is an absolute value. Distance measurement on the Manhattan distance using the formula:

first_page – end_page

$$D(X2,X1) = \| X2 - X1 \| 1 = \sqrt{\sum} |X2h - X1j| \; p \; j=1 \quad (3)$$

Definition 2.2 :
Measuring distances on the Minkowsky distance using the formula:

$$D(X2,X1) = \| X2 - X1 \| \; \lambda = \sqrt{\sum} |X2h - X1h| \; \lambda \; p \; j=1 \quad (4)$$

Definition 2.3 :
λ is the Minkowsky distance parameter. The most widely used method is Euclidean and Manhattan. Euclidean is the choice if we want to provide the shortest distance between two points (straight distance), as shown in the Euclidean formula. Whereas Manhattan provides the furthest distance in the two data [1].

2.4. Infrastructure development

Development as an effort or a series of growth and change efforts that are planned and carried out consciously by a nation, state and government towards modernity in the framework of nation building. The essence of development is to develop society or the nation as a whole, in order to achieve people's welfare. To be able to build better, people must be better educated and have better morals [6].
The infrastructure system is the main supporter of the functions of the social system and economic system in people's lives. The infrastructure system can be interpreted as basic facilities or structures, equipment, installations that are built and needed for the functioning of the social system and the community's economic system [7].

3. Results And Discussion

3.1. Research methods

In solving a problem in research, of course the researcher must have a way or a method that will be applied in solving the problem so that the research carried out can be resolved properly and in accordance with the expected results. The research method is carried out to look for something systematically by using the scientific method and applicable sources. In this research process it is aimed at the Langkat Regency PUPR office, especially in grouping infrastructure development data based on development by providing results and obtaining the desired information. The results of the conceptualization will be poured into a complete research method with a literature study pattern, collecting the data needed to analyze data mining for infrastructure development based on development using the Clustering method.

3.2. Process Design

The process design flow of the system that will be built using data mining grouping infrastructure development in Langkat Regency using the clustering method can be described in the form of a flowchart as shown below.

3.3. Flow chart

The process or workflow of the infrastructure development grouping system in Langkat Regency using the clustering method can be seen in the form of a flowchart as shown in the figure below.

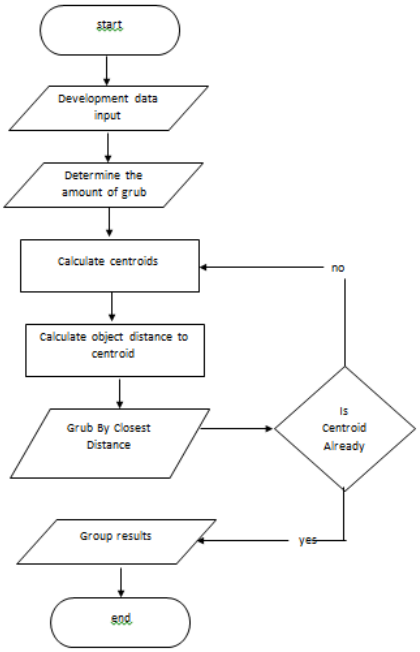

Fig.1: Flowchart of the K-Means Algorithm

3.4. Research Supporting Data

To analyze data in a study, supporting data is needed so that a research can run as expected. Clustering method in obtaining new information about the clustering of infrastructure development data. The data provided is data on infrastructure development based on existing development in Langkat Regency. Based on research that has been carried out at the Langkat Regency PUPR Service, data is obtained which will later be used to analyze infrastructure development data based on development in Langkat Regency. The data is as in the table below.

Table 1: Infrastructure development data

No	Object	company name	Development	Subdistrict
1	A	CV.Samaru Cipta Semesta	DHI office building construction	Stabat
2	B	CV.Kelana Karya Jaya	Trench construction	Stabat
3	C	CV.Berkat Kurnia	road construction	Hinai
4	D	CV.Kelana Karya Jaya	building construction	Stabat
5	E	CV.Kelana Karya Jaya	ditch construction	Stabat
6	F	CV.Indrabudhi Corporation	road construction	Stabat
7	G	CV.Enmo Perkasa	Bridge construction	Gebang
8	H	CV.Kelana Karya Jaya	ditch construction	Kuala
9	I	CV.Enmo Perkasa	Bridge construction	Gebang
10	J	CV.Berkat Kurnia	building construction	stabat
11	K	CV.Anugrah Cipta Bersama	building construction	Stabat
12	L	CV.Hantam Kromo	road construction	Kuala
13	M	CV.Samaru Cipta Semesta	building construction	Stabat
14	N	CV.Kelana Karya Jaya	ditch construction	Gebang
15	O	CV.Hantam Kromo	road construction	Kuala
16	P	CV.Anugrah Cipta Bersama	building construction	Kuala
17	Q	CV.Enmo Perkasa	bridge construction	Kuala
18	R	CV.Kelana Karya Jaya	ditch construction	Hinai

19	S	CV.Enmo Perkasa	bridge construction	Hinai
20	T	CV.Enmo Perkasa	bridge construction	Hinai

Application of the Method; 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 centroids, calculate the object distance to the centroids and then group them based on the closest distance. if no objects are moved or grouped then the iteration is finished.
To determine the group of an object, the first thing to do is measure the Euclidean distance between two object points (X and Y) which is defined as follows.

Table 2: Company name

CODE	Company Name (X)
1	CV.Mitra RIDHO Kontruksi
2	CV.Mitra Selaras
3	CV.Kelana Karya Jaya
4	CV.Berkat Kurnia
5	CV.Samaru Cipta Semesta
6	CV.Enmo Perkasa
7	CV.Hantam Kromo
8	CV.Vinta Makmur
9	CV.Karya Philip Jaya
10	CV.Indrabudhi Corporation
11	CV. Anugrah Cipta Bersama

Tabel 3: Development

CODE	Development (Y)
1	Jalan
2	Jembatan
3	Parit
4	Gedung

Table 4: Subdistrict

CODE	District Name (Z)
1	Kuala
2	Stabat
3	Hinai
4	Gebang
5	Bahorok
6	Serapit
7	Salapian
8	Sei. Bingai
9	Selesai
10	Wampu
11	Batang serangan
12	Sawit seberang
13	Padang Tualang
14	Tanjung pura
15	Secanggang
16	Gebang
17	Babalan
18	Sei. Lepan
19	Brandan Barat
20	Besitang
21	Pangkalan susu
22	Pematang jaya
23	Kutambaru

Then transform the criteria data above to be calculated using the clustering method. The transformation data from the data above can be seen in the table below.

Table 5: Transformation Data

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

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

K = 3 Centroids
C1 = (3,4,2) taken from data D
C2 = (5,4,2) taken from M data
C3 = (6,2,3) taken from S data

So K = 3 Centroid
C1 = (3, 3, 3.5)
C2 = (6,75, 1,5, 2,62)
C3 = (5,4, 3,5, 1,7)

Iteration ke 3
From the above calculation, the results of the 3 iteration calculations are obtained, which are as shown in the table below:

Table 6: Iteration results 3

No	Object	X	Y	Z	C1	C2	C3	Grub
1	A	5	4	2	2,69	3,11	0,70	3
2	B	3	3	2	1,5	4,08	2,46	3
3	C	4	1	3	2,29	2,82	3,14	2
4	D	3	4	2	1,80	4,54	2,46	3
5	E	3	3	2	1,5	4,08	2,46	3
6	F	10	1	2	7,43	3,34	5,24	2
7	G	6	2	4	3,20	1,64	2,81	2
8	H	3	3	1	2,5	4,35	2,54	3
9	I	6	2	4	3,20	1,64	2,81	2
10	J	4	4	2	2,06	3,76	1,51	3
11	K	11	4	2	8,20	4,96	5,63	3
12	L	7	1	1	5,12	1,71	3,04	2
13	M	5	4	2	2,69	3,11	0,70	3

14	N	3	3	4	0,5	4,26	3,36	1
15	O	7	1	1	5,12	1,71	3,04	2
16	P	11	4	1	8,44	5,19	5,66	3
17	Q	6	2	1	4,03	1,85	1,76	3
18	R	3	3	3	0,5	4,05	2,77	1
19	S	6	2	3	3,20	0,97	2,07	2
20	T	6	2	3	3,20	0,97	2,07	2

After calculating using the existing cluster formula in iteration 3, the groups based on the minimum distance to the nearest Centroid are:
Old Group : (3 3 2 3 3 2 2 3 2 3 3 2 3 1 2 3 3 1 2 2)
New Group: (3 3 2 3 3 2 2 3 2 3 3 2 3 1 2 3 3 1 2 2)

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 cluster graphs can be made for grouping data on infrastructure development based on development in Langkat Regency.
Clustering Graph

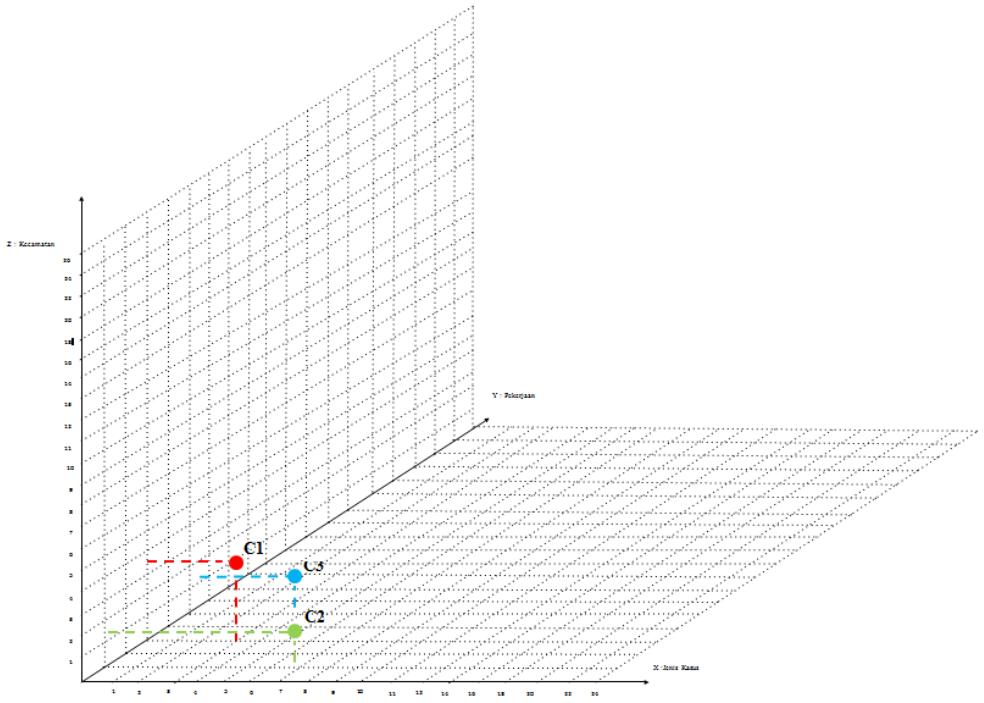

Fig 2: Clustering Result Graph

Cluster I : 3, 3, 3,5
Cluster II : 6.75, 1.5, 2.62
Cluster III : 5.4, 3.5, 1.7

Of the 20 infrastructure development data obtained by groups, cluster 1 has 2 data, cluster 2 has 8 data, cluster 3 has 10 data.

1. Cluster 1 has 2 data.
   3, 3, 3.5 (3 3 3)
   It can be seen that cluster 1 is centered on 3, 3, 3, namely company name 3 = CV.Kelana Karya Jaya, development 3 = Parit and district 3 = Hinai.
2. Cluster 2 contains 8 data
   6.75, 1.5, 2.62 (7 2 3)
   It can be seen that cluster 2 is centered on 7, 2, 3, namely company name 7 = CV. Hit Kromo, Construction 2 = Bridge and District = Hinai.
3. Cluster 3 has 10 data.
   5,4, 3,5, 1,7 (5 4 2)
   It can be seen that cluster 3 is centered on 7, 3, 2, namely company name 5 = CV. Samaru Cipta Semesta, construction 4 = Building and district 2 = Stabat.

4. Conclusions

From the results of the analysis based on the clustering of infrastructure development grouping data, a conclusion can be drawn, from the 933 data there are 3 clusters grouped with infrastructure development data. Can provide new information for PUPR Kab. Langkat regarding data on infrastructure development in sub-districts in Langkat Regency and assisting PUPR Kab. Langkat analyzes and classifies data on existing infrastructure development in Langkat Regency.

References

[1] Sariani, Saripurna, D., & Saragih, R. (2022). *Pengelompokan Data Pengguna Narkoba yang Melakukan Program Rehabilitas Rawat Jalan Menggunakan Metode Clustering (Studi Kasus : BNN Kota Binjai)*, 1-11.

[2] Siahaan, M. (2022). *Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means*, 317-323.

[3] Setiadi Moerwanto Direktorat Jenderal Bina Marga Kementerian Pekerjaan Umum dan Perumahan Rakyat, A., & Junoasmono Direktorat Jenderal Bina Marga Perencanaan Integrasi dan Jaringan Sistem, T. (2017). *STRATEGI PEMBANGUNAN INFRASTRUKTUR WISATA TERINTEGRASI* (Vol. 3, Issue 2).

[4] Tamba, P. S., Kesuma, T. F., & Feryanto. (2019). *Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering*, 67-71.

[5] Khakik, N., Ginting, s., & Novriyenni. (2022). *Data Mining Pengaduhan Hukum Berdasarkan Kasus Dengan menggunakan Metode Clustering* .

[6] Sarah Simbolon dkk, D., Sarah Simbolon, D., Sari, J., Yolanda Purba, Y., Indah Siregar, N., Salsabila, R., & Manulang, Y. (2021). PERANAN PEMERINTAH DESA DALAM PEMBANGUNAN INFRASTRUKTUR. *Jurnal Kewarganegaraan*, *5*(2).

[7] Berkala, J., Efisiensi, I., Posumah, F., Ekonomi, F., Bisnis, D., & Pembangunan, J. E. (n.d.). *PENGARUH PEMBANGUNAN INFRASTRUKTUR TERHADAP INVESTASI DI KABUPATEN MINAHASA TENGGARA*. http://www.

[8] Mochamad, w., & Masitha, R. (2020). yayasan kita menulis . *penerapan algoritma K-means Clustering dan K-medoids Clustering*.