

Clustering of Community Complaints on Flooding and Waterlogging Using the K-Means Clustering Method at the Public Works and Spatial Planning Agency

Rahmawati Dewi^{1*}, Magdalena Simanjuntak², I Gusti Prahmana³

^{1,2,3} STMIK Kaputama Binjai

rahmazahra202@gmail.com^{1*}, magdalena.simanjuntak84@gmail.com², igustiprahmana4@gmail.com³

Abstract

The Department of Public Works and Spatial Planning (PUTR) plays an important role in infrastructure development, which has a significant impact on regional economic growth. Flooding and waterlogging problems often occur due to high rainfall and suboptimal drainage systems, resulting in losses and disruptions to community activities. This study aims to classify public complaints related to flooding and waterlogging using the K-Means Clustering method. The data used consists of 617 public complaints, which are grouped into three clusters. The results show that Cluster 1 contains 355 data points, Cluster 2 contains 94 data points, and Cluster 3 contains 168 data points. This clustering is expected to assist PUTR in understanding patterns of public complaints and in determining more effective and targeted handling priorities.

Keywords: *Data Mining, Clustering, Public Complaints*

1. Introduction

The rapid development of information technology has had a significant impact on data management, particularly in handling public complaint data related to flooding and waterlogging, which continues to increase over time. Problems such as inadequate drainage systems, unregulated urban development, and slow response in handling issues have resulted in large and complex datasets that require effective analysis. Although these data contain valuable information, analyzing them remains a challenge due to the limitations of conventional methods, making data mining techniques such as clustering essential for grouping data based on similarity. In this study, the K-Means Clustering method is used to group data based on variables such as district, type of complaint, and actions taken, implemented through a MATLAB-based Graphical User Interface (GUI) to facilitate data processing and visualization. The purpose of this research is to identify patterns in public complaints in order to determine the most dominant problem groups, and it is expected to provide useful insights for the Department of Public Works and Spatial Planning (PUTR) in determining more effective and targeted handling priorities.

2. Research Method

2.1 Data Mining Definition

Data mining is the process of analyzing large volumes of data. It is a data analysis process aimed at discovering patterns, relationships, or hidden information from a dataset.

2.2 Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) is the process of discovering knowledge within databases. The complete definition of KDD is the overall process of extracting or identifying patterns, knowledge, and potentially useful information from large datasets. The knowledge and information generated from KDD are valid, novel, understandable, and useful.

2.3 K-Means Algorithm

K-Means is a non-hierarchical (partitioning) clustering method that attempts to divide data into clusters such that data with similar characteristics are grouped into the same cluster, while data with different characteristics are grouped into other clusters.

3. Results and Analysis

Data transformation is the process of converting or integrating data into a suitable format for use in data mining. Often, the data to be used in the data mining process are not in a format that can be directly utilized; therefore, they need to be transformed into an appropriate format

Table 1: Subdistrict

Code	Subdistrict
1	Brandan Barat
2	Gebang
3	Besitang
4	Babalan
5	Sei Lapan
6	Stabat
7	Sawit Seberang
8	Tanjung Pura
9	Batang Serangan
10	Wampu
11	Secanggang
12	Pematang Jaya
13	Hinai
14	Padang Tualang
15	Selesai
16	Bahorok
17	Serapit
18	Kuala

Table 2: Complaint

Code	Complaint
1	Lack of facilities at the evacuation post
2	Flooded road access that is difficult to pass
3	Slow water recession
4	Slow water recession, lack of facilities at the evacuation post, as well as skin diseases and diarrhea

Table 3: Complaint

Code	Action
1	Normalization and improvement of the drainage system to accelerate floodwater recession
2	Elevating or repairing road sections that are prone to flooding
3	Provision of clean water tanks and temporary pipeline installations in evacuation post areas

The existing data are entered into a table to undergo the transformation process. Initially, a total of 10 data points are required. The following are the data to be transformed.

Table 4: Data to be Clustered

Number	Subdistrict	Complaint	Action
1	Brandan Barat Subdistrict	Slow water recession, lack of facilities at evacuation posts, as well as skin diseases and diarrhea	Normalization and improvement of the drainage system to accelerate floodwater recession
2	Gebang Subdistrict	Slow water recession, lack of facilities at evacuation posts, as well as skin diseases and diarrhea	Normalization and improvement of the drainage system to accelerate floodwater recession
3	Besitang Subdistrict	Slow water recession, lack of facilities at evacuation posts, as well as skin diseases and diarrhea	Normalization and improvement of the drainage system to accelerate floodwater recession
4	Babalan Subdistrict	Flooded road access that is difficult to pass	Elevating or repairing road sections that are prone to flooding
5	Sei Lapan Subdistrict	Slow water recession, lack of facilities at evacuation posts, as well as skin diseases and diarrhea	Normalization and improvement of the drainage system to accelerate floodwater recession
6	Stabat Subdistrict	Lack of facilities at evacuation posts	Provision of clean water tanks and temporary pipeline installations in evacuation post areas
7	Sawit Seberang Subdistrict	Lack of facilities at evacuation posts	Provision of clean water tanks and temporary pipeline installations in evacuation post areas
8	Tanjung Pura Subdistrict	Flooded road access that is difficult to pass	Elevating or repairing road sections that are prone to flooding

9	Batang Serangan Subdistrict	Slow water recession	Repair of breached embankments
10	Wampu Subdistrict	Slow water recession	Repair of breached embankments

Table 5: Transformed Data

Initial	Subdistrict	Complaint	Action
A	1	4	1
B	2	4	1
C	3	4	1
D	4	2	2
E	5	4	1
F	6	1	3
G	7	1	3
H	8	2	2
I	9	3	4
J	10	3	4

Table 6: Results of Iteration 1 Calculation

No	X	Y	Z	C ₁	C ₂	C ₃	Group
1	1	4	1	1,00	2,00	7,00	1
2	2	4	1	0,00	1,00	6,16	1
3	3	4	1	1,00	0,00	5,39	2
4	4	2	2	3,00	2,45	3,32	2
5	5	4	1	3,00	2,00	4,12	2
6	6	1	3	5,39	4,69	1,00	3
7	7	1	3	6,16	5,39	0,00	3
8	8	2	2	6,40	5,48	1,73	3
9	9	3	4	7,68	6,78	3,00	3
10	10	3	4	8,60	7,68	3,74	3

After performing the calculations using the existing clustering formula, the grouping based on the minimum distance to the nearest centroid is as follows:

Old Group : {0,0,0,0,0,0,0,0,0}

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

Since there is a change in the group assignment, the process is continued to the next iteration.

Table 7: Results of Iteration 2 Calculation

No	X	Y	Z	C ₁	C ₂	C ₃	Group
1	1	4	1	0,00	3,09	7,61	1
2	2	4	1	0,50	2,13	6,70	1
3	3	4	1	1,50	1,25	5,82	2
4	4	2	2	3,35	1,49	4,18	2
5	5	4	1	3,50	1,25	4,22	2
6	6	1	3	5,77	3,50	2,24	3
7	7	1	3	6,58	4,15	1,43	3
8	8	2	2	6,87	4,27	1,20	3
9	9	3	4	8,14	5,68	1,62	3
10	10	3	4	9,07	6,57	2,37	3

After performing the calculations using the clustering formula, it is found that in Iteration 2 the results are the same as in Iteration 1, and no data points change their group. Therefore, the calculation process can be stopped.

Thus, a cluster graph can be created based on the results obtained. The grouping results from Iteration 1 and Iteration 2 are as follows:

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

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

Since there is no change in the grouping, the clustering process has converged.

A cluster graph is then created based on the calculations that have been performed. The resulting graph is as follows:

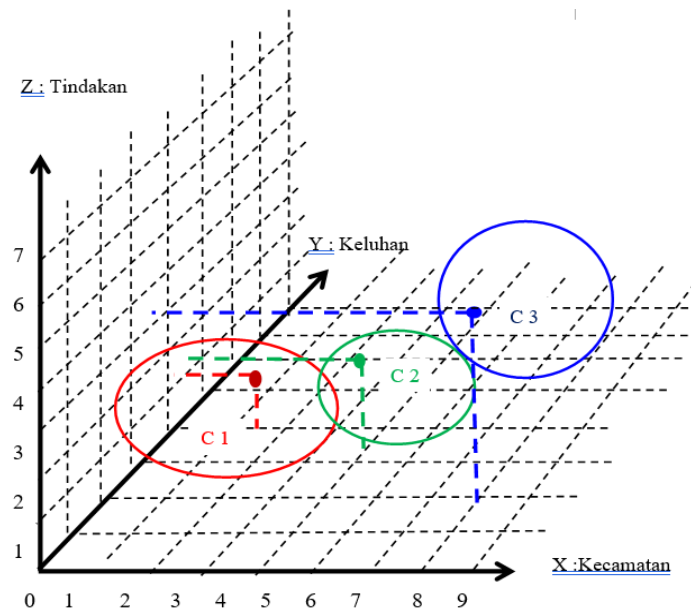


Fig. 1: Cluster Graph Based on the Performed Calculations

- Cluster 1 : 1.50 ; 4.00 ; 1.00
- Cluster 2 : 4.00 ; 3.33 ; 1.33
- Cluster 3 : 8.00 ; 2.00 ; 3.20

Graph Explanation :

From 10 data points, 3 groups (clusters) were obtained. Cluster 1 contains 2 data points, Cluster 2 contains 3 data points, and Cluster 3 contains 5 data points. The largest group is Cluster 3.

1. Cluster 1 (2 Data Points)

1.50 ; 4.00 ; 1.00

It can be seen that Cluster 1 contains 2 data points located in Sei Lapan District. The community complaints include slow water recession, lack of facilities in evacuation posts, as well as skin diseases and diarrhea. The action taken is the normalization and improvement of the drainage system to accelerate floodwater recession.

2. Cluster 2 (3 Data Points)

4.00 ; 3.33 ; 1.33

Cluster 2 contains 3 data points located in Babalan District. The main complaint is slow water recession, and the action taken is the normalization and improvement of the drainage system to accelerate floodwater recession.

3. Cluster 3 (5 Data Points)

8.00 ; 2.00 ; 3.20

Cluster 3 contains 5 data points located in Tanjung Pura District. The complaints include flooded road access that is difficult to pass, and the action taken is the provision of clean water tanks and temporary pipe installations in evacuation areas.

4. Conculation

Based on the analysis of community complaints using the clustering method, the following conclusions can be drawn:

1. From the testing conducted using the clustering method with the K-Means algorithm, it can be identified which district groups, complaint groups, and action groups have the highest frequency and occur most often in the clustering of community complaint data.
2. From 617 community complaint data, 3 groups were obtained, where Group 1 consists of 355 data points, Group 2 consists of 94 data points, and Group 3 consists of 168 data points.
 - a) Cluster 1 (11, 1, 3) is located in Secanggang District, with complaints about the lack of facilities in evacuation posts, and the action taken is the provision of clean water tanks and temporary pipe installations in evacuation areas.
 - b) Cluster 2 (2, 1, 3) is located in Gebang District, with similar complaints and actions.

- c) Cluster 3 (6, 1, 3) is located in Stabat District, also with complaints about the lack of facilities in evacuation posts and the same type of action taken.

References

- [1] Amna, et al. 2023. Data Mining, PT.Global Eksekutif Teknologi, Sumatera Barat.
- [2] Dr. Eng. Ags Naba, 2008. Pengertian GUIDE atau GUI Yogyakarta.
- [3] Eko Prasetyo, 2012. Data Mining Konsep Dan Aplikasi Menggunakan Matlab, PT.Andi, Yogyakarta.
- [4] Farhan Arif Wahyudi, Sudarti, Yushardi. 2023. "Implementasi Kegunaan Sumur Resapan Untuk Mengurangi Banjir Serta Genangan Air." Jurnal Pendidikan Fisika Vol. 7(2), Desember 2023
- [5] Febby Arisca Zurfani, Sawaluddin, Mardiningsih, Muhammad Romi Syahputra. 2024. "Analisis Metode *Clustering K-Means* pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi *Silhouette Coefficient* ." Jurnal Matematika, Ilmu pengetahuan Alam, Kebumihan dan Angkasa Volume.2 No.6 November 2024
- [6] Kusrini, Emha Taufiq Luthfi, 2009. Algoritma Data Mining, Penerbit C.V Andi,Yogyakarta.
- [7] Muhammad Faiz Bamazeza. 2021. "Kinerja Dinas Pekerjaan Umum dan Penataan Ruang dalam Mengoptimalkan Pembangunan Infrastruktur Jalan di Kabupaten Lampung Timur."
- [8] Nova. 2026. *Data Mining: Konsep, Teknik dan Implementasi*. Jakarta: Minhaj Pustaka, hal. 2.
- [9] Sigit Supriyadi, Albert Gamot Malau, 2025. "Clustering dengan Menggunakan K-Means untuk Analisa Dampak Banjir." Jurnal Indonesia Impression Journal (JII) Volume.4 No.9 September.
- [10] Wahyudi, F. A., Sudarti, S., & Yushardi, Y. (2023). *Implementasi kegunaan sumur resapan untuk mengurangi banjir serta genangan air*. Universitas Jember.
- [11] Yatini, B. 2010. *Definisi Flowchart*. Jakarta: PT. Gramedia Widia Sarana Indonesia, hal. 29.
- [12] ([www.http://modul-guideuploder-by-teuinsuska2009-wordpress-com](http://modul-guideuploder-by-teuinsuska2009-wordpress-com)) diakses pada tanggal : 15 April 2026
- [13] ([www. https://dosbing.id/2022/12/03/proses-tahapan-data-mining/](https://dosbing.id/2022/12/03/proses-tahapan-data-mining/)) diakses pada tanggal : 15 April 2026
- [14] ([www. https://dte.telkomuniversity.ac.id/matlab-adalah/](https://dte.telkomuniversity.ac.id/matlab-adalah/)) diakses pada tanggal :29 April 2026