J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

J-INTECH
Journal of Information and Technology

Implementation of K-Means Clustering in a Web-Based Public Complaint System to Improve Efficiency and Transparency

Lalu Yazid Adnan[1*], Ahmad Heru Mujianto[2]

[1,2]*Universitas Hasyim Asy'ari (UNHASY), Jl. Irian Jaya No.55, Tebuireng, Cukir, Diwek, Jombang, Jawa Timur 61471, Indonesia*

Keywords
Community Complaint System;K-Means Clustering;Text Clusterting;TF-IDF;Village Digital Governance

***Corresponding Author:**
laluadnan@mhs.unhasy.ac.id

Abstract
The community complaint system in Sumengko Village still relies on conventional methods such as direct visits to the village office and deliberation forums, making it difficult for village officials to systematically categorize complaints and determine the appropriate handling priorities. This research aims to design and develop a web-based community complaint system integrated with the K-Means Clustering algorithm to automatically group complaints based on the similarity of text characteristics. The system is built using the Laravel framework based on PHP with a MySQL database, and applies text preprocessing stages including case folding, tokenizing, stopword removal, and stemming, followed by TF-IDF weighting to convert the text into a numerical representation before the clustering process is performed. The dataset used consists of 15 complaint documents representing three dominant topics: infrastructure, environment, and health. Validation of the number of clusters using the Elbow Method shows k=3 as the optimal point with the lowest significant SSE value. The clustering results show that the K-Means algorithm converges at the 2nd iteration with the following cluster distribution: C_1 Infrastructure (60%), C_2 Environment (33.3%), and C_3 Health (6.7%), indicating that infrastructure-related complaints are the dominant issue reported by residents. Usability evaluation using the System Usability Scale (SUS) yielded an overall mean score of 79.25, categorized as "Good" with an acceptability level of "Acceptable," confirming that the system is well-received by both village administrators and community members. This system not only serves as a medium for submitting complaints digitally but also as an automatic analysis tool that helps village officials prioritize complaint handling more efficiently and transparently.

1. Introduction

Advances in information technology have opened opportunities for communities to communicate with various parties more quickly and efficiently. The absence of communication media channels that bridge the community with the village government is one of the factors causing dependence on social media as a communication channel for conveying aspirations, information, and complaints [1]. A community complaint or aspiration is an important omponent of a government institution, especially at the village level. Through complaints from the community, a village government institution can easily improve the quality of its services to the community [2].

In Sumengko Village, reports and unfulfilled citizen wishes indicate problems with the complaint system. To submit their complaints, the community can visit the village office directly, discuss them at Deliberation forums, or use inter-community communication media. Conversely, village administration work is very extensive and generates a large amount of data. Because there is no systematic complaint grouping system, village apparatus faces particular challenges with this condition. As a result, village apparatus faces challenges in determining the appropriate priority of complaint handling and problem patterns [3].

One approach that can be used to address these problems is the application of data mining technology with the K-Means Clustering method. This method is a data grouping algorithm capable of identifying patterns and similarities of characteristics between data, thereby producing groups (clusters) that have certain similarities [4]. The use of this method has been widely applied in various fields, showing that the K-Means algorithm is effective in grouping community complaint assessment data [5].

Previous research on complaint data clustering and the development of community complaint systems has been conducted using various approaches. Text clustering of community complaint data has been performed using the K-Means algorithm in a case study at the DKI Jakarta Social Service for the 2021–2022 period [6]. A web-based community complaint information system has also been developed to improve complaint management at the village level [7]. In addition, the K-Means algorithm has been applied for clustering customer complaint data in a regional water utility company, demonstrating its effectiveness in organizing complaint data into meaningful groups [8]. Building upon these previous studies, the present research integrates the K-Means algorithm into a web-based community complaint system for Sumengko Village to enable automatic and efficient complaint grouping.

Unlike previous studies that applied K-Means clustering as a standalone offline analysis tool, this study integrates the K-Means algorithm directly into a web-based complaint management system. This integration enables automatic and real-time complaint grouping at the point of submission, eliminating the need for separate data processing steps. To the best of the authors' knowledge, no previous study has implemented such direct integration of K-Means clustering within a village-level web-based complaint system in Indonesia, making this study a novel contribution to the field of digital village governance.

This research aims to: (1) design and develop a web-based community complaint system that can be accessed by the community of Sumengko Village to submit complaints digitally; and (2) integrate the K-Means algorithm into the web-based community complaint system to perform automatic complaint grouping based on the similarity of text characteristics.

2. Research Method

This research falls within the field of software engineering, aiming to build a web-based community complaint system integrated with the K-Means Clustering data mining method. The research methodology consists of several interconnected stages as described in the following subsections.

2.1 Text Preprocessing

Text preprocessing is the stage of cleaning and normalizing text data before computational processing. The preprocessing process is carried out in four sequential stages: (1) Case folding: converting all text characters to lowercase while removing punctuation, numbers, and special characters; (2) Tokenizing: breaking sentences into separate word units based on spaces; (3) Stopword Removal: removing common words that do not contribute significant meaning such as conjunctions and prepositions; (4) Stemming: changing affixed words into base words using Indonesian language rules. The stemming process in this study was implemented using native PHP string functions without employing a dedicated Indonesian stemming library such as Sastrawi. This approach applies basic affix-stripping rules manually coded to handle common Indonesian prefixes and suffixes. While this method is functional for the limited dataset used in this study, it is acknowledged as a limitation, as a dedicated library such as Sastrawi would provide more comprehensive and linguistically accurate stemming results.

2.2 TF-IDF Weighting

After preprocessing, the TF-IDF (Term Frequency-Inverse Document Frequency) method is used to convert text documents into numerical representation. TF-IDF gives higher weight to words that frequently appear in a particular document but rarely appear in other documents, so these words are considered important for that document [8]. The formulas used are:

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \quad (1)$$
$$TF(t,d) = f(t,d) / \Sigma f(w,d) \quad (2)$$
$$IDF(t) = log_{10}(N / df(t)) \quad (3)$$

Where: f(t,d) = number of occurrences of term t in document d; Σf(w,d) = total words in document d; N = total number of documents; df(t) = number of documents containing term t.

2.3 K-Means Clustering Algorithm

K-Means Clustering is the most popular and widely used partitioning algorithm for grouping data. K-Means works iteratively to minimize variation within each cluster [9]. The algorithm steps are: (1) Determine the number of clusters (k); (2) Initialize initial centroids randomly; (3) Calculate the distance from each data point to the centroid using Euclidean Distance; (4) Assign each data point to the nearest cluster; (5) Update centroids as the average of all member data; (6) Repeat steps 3-5 until convergence.

$$d(x,y) = \sqrt{\Sigma}(x_i - y_i)^2 \quad (4)$$

Where: x and y are two data points; x_i and y_i are the i-th attribute values; n is the number of attributes/dimensions.

2.4 Use Case Diagram

The system has two main actors: User (Community) and Administrator (Village Apparatus). The Use Case Diagram describes the interaction between actors and the system built. Table 1 describes the use case specifications for the User actor, while Table 2 describes the use case specifications for the Admin actor.

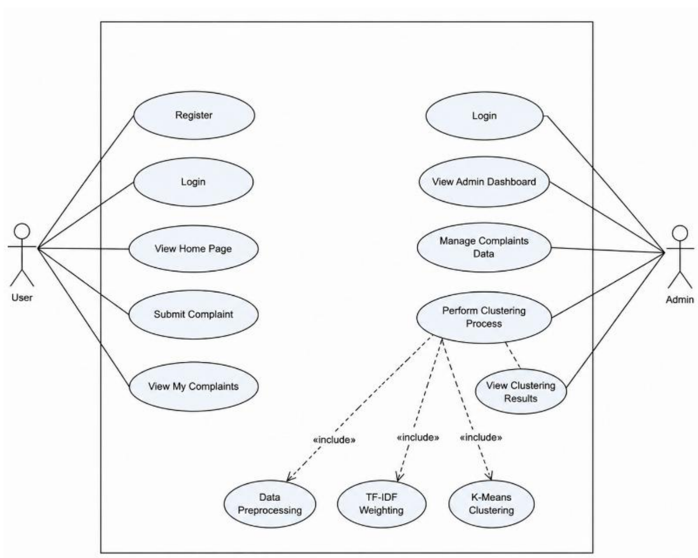

Figure 1 Use Case Diagram

Figure 1 shows the integrated Use Case Diagram for both the User and Administrator actors. The User actor only has access to complaint reporting and monitoring features, while Admin has additional access to manage data and perform clustering analysis.

2.5 System Architecture and Database Design

The system is built using PHP with the Laravel (MVC) framework as the backend, and HTML, CSS, and JavaScript for the frontend interface. The database uses MySQL with four main tables: users, pengaduan (complaints), clustering, and hasil_clustering (clustering results). The entity relationships follow a one-to-many pattern between users and complaints, and a junction entity for clustering results connecting complaints and clusters.

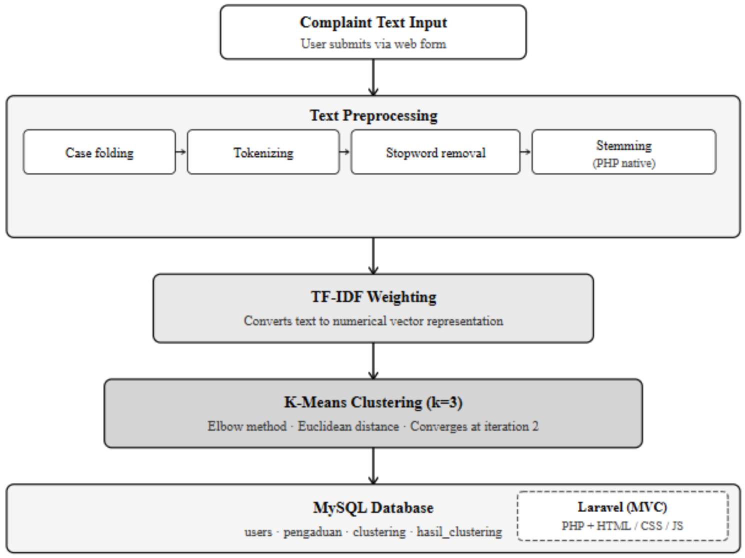

Figure 2 System Architecture Diagram Interaction between Text Preprocessing, TF-IDF Weighting, K-Means Clustering, and MySQL Database

Figure 2 illustrates the overall system architecture, showing the interaction between the text preprocessing pipeline, TF-IDF weighting, K-Means clustering process, and the MySQL database managed through the Laravel MVC framework.

3. Result and Discussions

3.1 Dataset

The dataset used consists of 15 community complaint documents from Sumengko Village, selected representatively from three dominant complaint topics: Infrastructure (D1-D5), Environment (D6-D10), and Health (D11-D15). Table 3 shows a sample of the dataset used.

Tabel 1. Sample Complaint Dataset

ID	Original Complaint Text	Category
D1	s very rotted, it feels like it will collapse when crossed.	Infrastructure
D3	to RT 08 is starting to crack and there are parts that are damaged during heavy rain.	Infrastructure
D7	has become a thicket, a mosquito breeding ground, and a dumping site for illegal waste..	Environment
D9	plastic waste, and when it rains, it overflows and the water enters the yard.	Environment
D12	ɜxperiencing rashes allegedly due to the overflow of wastewater into their yards.	Health
D15	ı-brown and has a strange smell, its quality needs to be checked.	Health

3.2 Determination of Number of Clusters (k=3)

The number of clusters was determined using the Elbow Method and domain knowledge related to the classification of public complaints in Sumengko Village. Based on the SSE analysis, the optimal value was obtained at k = 3, representing three dominant complaint categories: Infrastructure, Environment, and Health. Therefore, the K-Means clustering process in the implemented system uses three clusters.

Tabel 2. SSE Values for Elbow Method Validation

k Value	SSE	Description
1	0.1030	Very sharp decrease
2	0.0666	Sharp decrease
3	0.0467	Optimal Elbow Point
4	0.0352	Decrease slowing
5	0.0249	Very small decrease

Figure 3 presents the SSE graph generated from the Elbow Method validation. As shown in the graph, a significant decrease in SSE occurs from k=1 to k=3, after which the rate of decrease slows considerably. This inflection point at k=3 confirms it as the optimal number of clusters for this dataset.

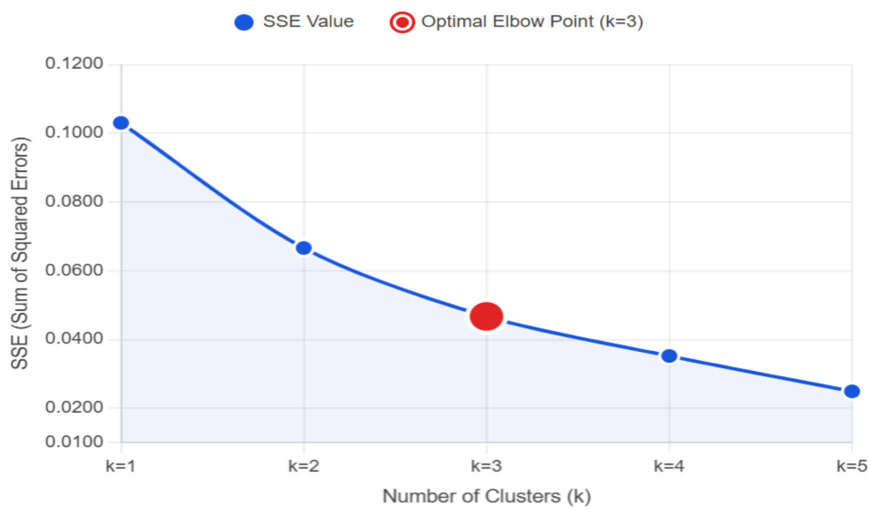

Figure 3 SSE Value Graph Using the Elbow Method

3.3 Text Preprocessing Results

The preprocessing process is carried out in four sequential stages. Table 5 shows sample preprocessing results for several complaint documents**.**

Tabel 3. Preprocessing Results per Document (Sample)

ID	Original Text (Summary)	Preprocessing Result
D3	e of the road on RT 08 started to crack and get damaged when it rains	ıg to crack, small damage, heavy rain has passed
D5	ıe public well in RT 06 has had a broken pump for a month, causing difficulties	well's pump is broken, and it has been difficult to fix it for a month
D9	ditch is full of plastic waste, rain overflow, and garbage.	own the stairs, full of plastic trash, rain floods the yard
D14	Residents of RT 01 are feeling dizzy and nauseous in the morning due to burning smoke	complaining about headaches and nausea in the morning from the smoke of burning trash behind

3.4 TF-IDF Weighting Results

After preprocessing, the TF-IDF method is used to convert complaint text documents into numerical vectors. Eight main terms were selected based on their relevance: rusak, hujan, diperbaiki, sampah, limbah, pembakaran, sumur, and asap. Table 6 shows the IDF values for each selected term.

Tabel 4. IDF Calculation for Each Term

Term	df(t)	IDF = $\log_{10}(15/df)$	Documents Containing Term
broken	2	0.8751	D3, D5
rain	4	0.5740	D2, D3, D8, D9
corrected	2	0.8751	D4, D5
garbage	4	0.5740	D7, D9, D11, D14
waste	2	0.8751	D9, D12
burning	2	0.8751	D10, D14
well	2	0.8751	D5, D15
smoke	2	0.8751	D10, D14

3.5 K-Means Clustering Process

The K-Means process uses k=3 with initial centroids determined through random initialization by the system. In this simulation, the random initialization produced C_1 = D3 (Infrastructure), C_2 = D9 (Environment), and C_3 = D12 (Health) as the initial centroids. The algorithm converged at iteration 2 with no document migration between clusters. Table 5 shows the final clustering results.

Tabel 5. Final K-Means Clustering Results

Cluster	Label	Count	ominant Terms	Priority	Recipient Unit
C_1	Infrastructure	9 (60%)	residents, rain, road, minor, severe	High	Village Infrastructure Section
C_2	Environment	5 (33,3%)	residents, waste, trees, village, caused	Medium	Environmental and Sanitation Section
C_3	Environmental & Health Disturbance	1 (6,7%)	children, waste, medical, used, syringes	High	Community Health Center (Puskesmas) / Health Department

3.6 Cluster Interpretation

Cluster C1 contains 9 complaint documents (60%) and is dominated by the terms warga, hujan, jalan, kecil, and banget. Most complaints in this cluster are related to damaged roads, drainage issues, and village infrastructure conditions requiring repair and maintenance. This cluster represents the largest complaint category submitted by residents.

Cluster C2 contains 5 complaint documents (33.3%) dominated by the terms warga, sampah, pohon, desa, and gara. The complaints mainly concern environmental cleanliness, waste management, and public environmental conditions. These issues require attention from the environmental management unit of the village administration.

Cluster C3 contains 1 complaint document (6.7%) dominated by the terms anak, sampah, medis, bekas, and suntikan. This cluster is associated with public health concerns, particularly medical waste and sanitation-related issues that may pose health risks to the community.

The cluster distribution obtained in this study — Infrastructure (60%), Environment (33.3%), and Health (6.7%) — demonstrates a highly skewed pattern in which infrastructure-related complaints overwhelmingly dominate. This finding is directly comparable to Nugraha [6], who applied K-Means text clustering to community complaint data at the DKI Jakarta Social Service for the period 2021–2022 and similarly reported that infrastructure and public facility complaints constituted the largest complaint category. The alignment between the present study and Nugraha [6] suggests that infrastructure dissatisfaction represents a structurally persistent concern in Indonesian community complaint datasets, regardless of the administrative level or geographic context. However, a meaningful contrast emerges when comparing with Kholidah and Indriyanti [10], who applied both K-Means Clustering and Fuzzy C-Means Clustering to complaint data from the Wargaku Surabaya application directed at the Surabaya City Civil Service Police Unit (Satpol PP). Their results revealed a relatively balanced distribution across complaint categories, with no single cluster exceeding 50% dominance. This divergence from the present study's findings can be attributed to the difference in complaint scope and administrative context: the Wargaku Surabaya application covers a broader urban population with more diverse complaint types, whereas Sumengko Village represents a small, rural community where physical infrastructure — roads, drainage, and public wells — constitutes the most immediate and unmet public need. The concentration of 60% of complaints within a single infrastructure cluster therefore reflects the specific socioeconomic and developmental conditions of Sumengko Village rather than a methodological artifact.

The determination of k=3 as the optimal number of clusters in this study was validated using the Elbow Method, based on the observation of a significant SSE inflection point between k=1 and k=3, after which the rate of SSE decrease slowed considerably (SSE values: k=1: 0.1030; k=2: 0.0666; k=3: 0.0467; k=4: 0.0352; k=5: 0.0249). This pattern — a sharp initial decline followed by a pronounced deceleration — is consistent with the theoretical and empirical framework established by Marutho et al. [11], who specifically investigated the determination of optimal cluster numbers using the Elbow Method applied to headline news data. Marutho et al. [11] demonstrated that the elbow point, defined as the cluster value at which additional clusters yield diminishing SSE reduction, provides a reliable and interpretable criterion for cluster selection. In the present study, the SSE drop from k=2 to k=3 (0.0199, approximately 29.9% reduction) is substantially larger than the drop from k=3 to k=4 (0.0115, approximately 24.6% reduction) and from k=4 to k=5 (0.0103, approximately 29.3% reduction), confirming k=3 as the point of significant inflection. This pattern closely mirrors the SSE behavior reported by Marutho et al. [11], where the elbow point was similarly characterized by a clear transition between steep and gradual decline phases. Furthermore, the alignment of k=3 with domain knowledge — three natural complaint categories (Infrastructure, Environment, Health) — provides additional external validity beyond purely mathematical criteria, reinforcing the methodological soundness of this study's cluster selection.

The overall SUS score of 79.25 obtained in this study falls within the "Good" category and achieves an "Acceptable" acceptability level according to the adjective rating scale proposed by Bangor et al. [21]. To contextualize this result within the broader usability literature, it is important to compare it against SUS scores reported in comparable web-based information systems. Sucipto et al. [12] evaluated a web-based village letter administration system (e-Surat) deployed in Menturus Village, Jombang, and reported usability results consistent with the "Good" to "Acceptable" range, indicating that village-level web applications generally achieve moderate-to-good usability when designed for non-expert users. Similarly, Sushanty et al. [13] reported usability outcomes in a comparable range for a web-based village administrative service system in Central Kalimantan. The SUS score of 79.25 obtained in the present study is comparable to or slightly above the scores reported in these studies, which is notable given that the current system incorporates additional functional complexity in the form of an integrated K-Means clustering module. This indicates that the integration of machine learning-based clustering functionality into a web-based complaint management system does not necessarily compromise usability, provided that the clustering output is presented through a well-designed dashboard interface. Furthermore, the score of 79.25 approaches the threshold of 80.3, which Bangor et al. [21] associate with the "Excellent" category, suggesting that with targeted UI improvements — particularly for respondents who rated the system as "OK" (R5 and R8, both scoring 70.00) — the system has the potential to achieve an "Excellent" usability rating in future iterations.

The use of Euclidean Distance as the distance metric in the K-Means process has proven effective for measuring proximity between data points in multidimensional space, supporting the accuracy of the clustering results [14]. The effectiveness of K-Means clustering has been demonstrated in various domains, including satisfaction measurement, social media analysis, and information system development [15][16][17]. These studies confirm that K-Means is a reliable and widely applicable clustering algorithm. Furthermore, the Elbow Method has been recognized as a standard technique for determining the optimal number of clusters, as it provides a clear visual justification based on SSE reduction [18][11].

These results further confirm that the K-Means algorithm is simple, efficient, and easy to implement, making it widely applicable across various domains such as student grouping, sales analysis, customer segmentation, and complaint data grouping [19].

However, it is important to acknowledge that the dataset used in this study consists of only 15 complaint documents, which is extremely limited for a comprehensive clustering evaluation. This limitation affects the generalizability of the results. Therefore, the current implementation should be regarded as a simulation or prototype stage, intended to demonstrate the feasibility of integrating K-Means clustering into a web-based complaint system rather than as a fully validated production system.

Despite the promising results, a critical evaluation of the methods used is necessary. TF-IDF weighting, while effective for general text representation, has known limitations when applied to short complaint texts such as those in this study. Short texts produce sparse vector representations, where most term weights are zero, which may reduce the discriminative power of the features and affect clustering accuracy. Furthermore, K-Means clustering assumes that clusters are spherical and equally sized, which may not always reflect the natural distribution of complaint data. The convergence at iteration 2 observed in this study, while efficient, may also indicate that the dataset is too small and homogeneous to fully challenge the algorithm. These limitations suggest that the current results, while valid as a proof of concept, should be interpreted with caution and validated with a larger and more diverse dataset in future work.

3.7 System Interface

The system provides several main pages: (1) Login and Registration Page for user and admin authentication; (2) Home Page displaying general village information and navigation menus; (3) Complaint Input Page with digital forms for submitting complaints; (4) View Complaint Page showing submitted complaint lists with current status; (5) Admin Dashboard showing total complaints, clustering button, and complaint management; and (6) Clustering Results Page showing grouping reports with filter by cluster, dominant keywords, and export to PDF/Excel features. The development of this web-based system follows the approach demonstrated in previous village digital service implementations, which proved effective in improving efficiency, speed, and accuracy of village administrative services, as well as producing better structured and well-documented data management[20]

.

3.7.1 Clustering Results Page

The Clustering Results Page, shown in Figure 4, presents the output of the K-Means clustering process in dashboard form. The system processed 15 complaint records and automatically grouped them into three clusters.

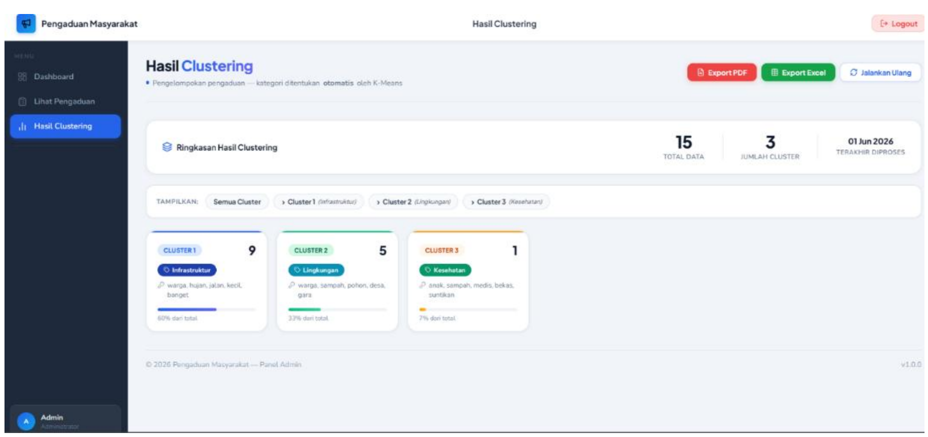

Figure 4. Clustering Results Page

The summary section displays 15 total records, 3 clusters, and the latest clustering execution date. The clustering results consist of Cluster 1 (Infrastructure) with 9 complaints (60%), Cluster 2 (Environment) with

5 complaints (33.3%), and Cluster 3 (Health) with 1 complaint (6.7%). Each cluster card displays the cluster label, number of complaint records, dominant keywords, and percentage distribution. The system also provides filtering functionality and export features in PDF and Excel formats to support reporting and decision-making activities by village administrators [13].

3.7.2 System Usability Evaluation

To evaluate the usability of the developed system, a System Usability Scale (SUS) evaluation was conducted involving 10 respondents consisting of 5 village officials and 5 community members. The SUS instrument consists of 10 Likert-scale questions (1–5) with alternating positive and negative items. The SUS score for each respondent is calculated using the formula: sum of adjusted scores for odd-numbered items (score – 1) and even-numbered items (5 – score), multiplied by 2.5, yielding a final score on a 0–100 scale. The evaluation results are presented in Table 5. The overall mean SUS score obtained was 79.25, which falls within the "Good" category on the adjective rating scale, with an acceptability level of "Acceptable" and a grade equivalent of B+ [21]. Individual scores ranged from 70.00 to 95.00, with four respondents achieving "Excellent" ratings (≥85), four respondents receiving "Good" ratings (71–84), and two respondents receiving "OK" ratings (51–70). These results indicate that the system is usable and well-received by both village administrators and community members, confirming its readiness as a functional prototype for complaint management in a village-level governance context.

Table 5. Usability Evaluation Results – System Usability Scale (SUS)

Resp.	Role	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score	Category
R1	Village Official	4	2	4	2	4	1	4	2	4	2	82.50	Excellent
R2	Village Official	4	1	4	2	3	2	4	2	3	2	77.50	Good
R3	Village Official	5	1	4	1	4	1	5	1	5	1	95.00	Excellent
R4	Village Official	4	2	3	2	4	2	3	2	4	2	75.00	Good
R5	Village Official	3	2	4	2	3	2	4	3	3	2	70.00	OK
R6	Resident	4	1	4	2	4	2	4	2	4	1	85.00	Excellent
R7	Resident	4	2	3	2	4	2	3	2	4	2	75.00	Good
R8	Resident	3	2	4	3	3	2	4	2	3	2	70.00	OK
R9	Resident	4	2	4	1	4	1	4	2	4	2	85.00	Excellent
R10	Resident	4	2	3	2	4	2	3	2	3	2	72.50	Good
Mean	Total	-	-	-	-	-	-	-	-	-	-	79.25	Good

4. Conclusions and Future Works

Based on the research results, it can be concluded that the design and development of a web-based community complaint system integrated with the K-Means Clustering algorithm has been successfully implemented with two main actors, namely User (Community) and Admin (Village Apparatus). The K-Means algorithm combined with TF-IDF weighting is capable of automatically grouping complaint data into three clusters based on textual similarity without requiring predefined category labels. The clustering process conducted on 15 complaint documents produced three clusters consisting of Infrastructure with 9 complaints (60.0%), Environment with 5 complaints (33.3%), and Health with 1 complaint (6.7%). The Infrastructure cluster represents the largest proportion of complaints submitted by residents, indicating that infrastructure-related issues are the dominant concern in Sumengko Village. The clustering process converged efficiently and demonstrated the capability of the proposed approach to identify thematic patterns within community complaint data. The Elbow Method validated k=3 as the optimal number of clusters for the dataset used in this study. This system not only functions as a digital complaint submission medium but also serves as an automatic analysis tool that can assist village administrators in determining complaint-handling priorities more efficiently and transparently.

For future research, it is recommended to: (1) conduct testing using larger and more representative complaint datasets from Sumengko Village; (2) compare the K-Means algorithm with other clustering methods such as DBSCAN or Fuzzy C-Means to evaluate performance differences; (3) incorporate Indonesian-language Natural Language Processing (NLP) techniques and advanced stemming approaches to improve text preprocessing quality; (4) develop complaint category prediction features using supervised classification algorithms to complement the clustering process; and (5) implement and evaluate the system using larger real-world complaint datasets from multiple villages or regions to improve the generalizability and scalability of the proposed approach.

In addition, a System Usability Scale (SUS) evaluation involving 10 respondents, consisting of 5 village officials and 5 community members, produced an average score of 79.25. This result falls within the "Good" category and indicates that the developed system is acceptable and easy to use in supporting community complaint management and analysis activities.

5. References

[1] M. Rizky and Muhammad Arifin, "Implementasi Aplikasi Pengaduan Masyarakat Berbasis Web di Desa Suwawal Kecamatan Mlonggo Kabupaten Jepara : Peningkatan Responsivitas Pemerintah Desa," *PROFICIO*, vol. 6, no. 1, pp. 564–571, Nov. 2024, doi: 10.36728/jpf.v6i1.4304.

[2] I. P. Rakhayu, Purwanto, and L. Nur Azizah, "Analisis pola pengaduan warga menggunakan data mining K-Means di Desa Larangan Kabupaten Purbalingga," Global Research and Innovation Journal, vol. 1, no. 3, pp. 2120–2125, 2025.

[3] S. Sasmita and S. Muntari, "Penerapan algoritma K-Means clustering pada data keluhan pelanggan PT. PLN Persero Kota Pagar Alam," *Jurnal Ilmiah Teknosains*, vol. 9, no. 1, pp. 9–12, 2023, doi: 10.26877/jitek.v9i1/Mei.15366.

[4] D. Afida, E. D. Udayanti, and E. Kartikadarma, "Aplikasi Text Mining untuk Klasterisasi Aduan Masyarakat Kota Semarang Menggunakan Algoritma K-means," *Jurnal Transformatika*, vol. 18, no. 2, pp. 215–224, Jan. 2021, doi: 10.26623/transformatika.v18i2.2362.

[5] A. Nugraha, "Analisis Pengelompokan Teks Pada Data Pengaduan Masyarakat Menggunakan Algoritma K-Means Periode 2021-2022," Doctoral Dissertation, 2023.

[6] K. Kurniasih and H. Mulyono, "Sistem Informasi Pengaduan Masyarakat Berbasis Web Pada Kantor Desa Ladang Peris Kecamatan Bajubang," *Jurnal Manajemen Sistem Informasi*, vol. 7, no. 4, pp. 678–688, Dec. 2022, doi: 10.33998/jurnalmsi.2022.7.4.692.

[7] Dea Syafitri, Yani Maulita, and Lina Arlianan Nur Kadim, "Pengelompokan Data Keluhan Masyarakat Terhadap Fasilitas Umum diKota Binjai Menggunakan Metode Clustering," *Jurnal Penelitian Teknologi Informasi dan Sains*, vol. 2, no. 3, pp. 158–170, Aug. 2024, doi: 10.54066/jptis.v2i3.2381.

[8] A. K. R. N. F. S. R. N. L. R. F. M. Dewi, "Deteksi Berita Hoax Pada Platform X Menggunakan Pendekatan Text Mining dan Algoritma Machine Learning," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 33–46, 2025.

[9] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.

[10] I. A. Kholidah and Aries Dwi Indriyanti, "Perbandingan Penerapan Text Mining Pada Aplikasi Wargaku Surabaya Untuk Aduan Masyarakat Terhadap Satpol PP Kota Surabaya Dengan Algoritma K-Means Clustering Dan Fuzzy C-Means Clustering," *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 04, pp. 1180–1189, Jul. 2025, doi: 10.26740/jinacs.v6n04.p1180-1189.

[11] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, “The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News,” in *2018 International Seminar on Application for Technology of Information and Communication*, IEEE, Sep. 2018, pp. 533–538. doi: 10.1109/ISEMANTIC.2018.8549751.

[12] Hadi Sucipto, Ahmad Heru Mujianto, Chamdan Mashuri, Mahrus Ali, and Mahfudiyanto, “Implementasi Aplikasi E-Surat pada Perangkat Desa di Desa Menturus Kudu Jombang,” *ABIDUMASY Jurnal Pengabdian Kepada Masyarakat*, vol. 1, no. 1, pp. 34–41, Mar. 2020, doi: 10.33752/abidumasy.v1i1.654.

[13] P. Sushanty, Ariesta Lestari, Devi Karolita, Jadiaman Parhusip, and Licantik, “Sistem Informasi Pelayanan Administrasi Surat Desa Bumi Subur, Katingan, Kalimantan Tengah Berbasis Web,” *Prosiding Seminar Nasional KONSTELASI*, vol. 3, no. 1, pp. 517–528, Jan. 2026, doi: 10.24002/prosidingkonstelasi.v3i1.13879.

[14] F. M. Sarimole and L. Hakim, “Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang,” *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 846–854, Feb. 2024, doi: 10.55338/saintek.v5i3.2709.

[15] D. I. Masruri, S. Widodo, and F. E. Purwiantono, “Implementasi k-Means Untuk Sistem Informasi Penyebaran Penyakit Epidemi Di Kota Malang,” *J-INTECH*, vol. 9, no. 02, pp. 101–107, Dec. 2021, doi: 10.32664/j-intech.v9i02.638.

[16] D. P. Dewi, I. H. Santi, and W. D. Puspitasari, “Perhitungan Penilaian Tingkat Kepuasan Pelanggan Dengan Menerapkan Algoritma K-Means,” *J-INTECH*, vol. 11, no. 2, pp. 257–265, Dec. 2023, doi: 10.32664/j-intech.v11i2.981.

[17] V. P. Ramadhan and G. M. Namung, “Klasterisasi Komentar Cyberbullying Masyarakat di Instagram berdasarkan K-Means Clustering,” *J-INTECH*, vol. 11, no. 1, pp. 32–39, Jul. 2023, doi: 10.32664/j-intech.v11i1.846.

[18] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, “Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means,” *IEEE Access*, vol. 10, pp. 115025–115032, 2022, doi: 10.1109/ACCESS.2022.3215568.

[19] F. Handayani, “Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar,” *Jurnal Teknologi dan Informasi*, vol. 12, no. 1, pp. 46–63, Mar. 2022, doi: 10.34010/jati.v12i1.6733.

[20] M. Aidin, “Transformasi Digital Administrasi Desa Melalui Sistem Informasi Desa: Kajian Pustaka Tentang Faktor Pendukung,” *Nusantara Journal of Multidisciplinary Science*, vol. 3, no. 8, pp. 1661–1674, 2025.

[21] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.