Published online on the page : https://ejournal.uinbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Implementation of a K-Means-Based Intelligent Patient Complaint Clustering System to Identify Handling Priorities

M. Agung Vafky Ideal[1,], Nurfiah[2] ,Idir Fitriyanto[3,]*

[1,3]*Institut Teknologi Mitra Gama, Bengkalis, Indonesia*
[2]*Universitas Andalas, Padang, Indonesia*

Article Information

Article History:

Submitted:April 29, 2025
Revision: May 15, 2025
Accepted: June 11, 2025
Published: June 30, 2025

Keywords

artificial intelligence
machine learning
software engineering
medical
K-Means

Correspondence

E-mail: mhdagung47@gmail.com

A B S T R A C T

Patient complaints are the body's response to health disturbances, triggered by internal factors such as genetics or external ones like the living environment. Understanding these causes allows community health centers (puskesmas) to take more effective preventive measures and design more targeted services. This study utilizes patient complaint data sourced from medical records, which include biodata and medical history, as well as complaint details that form the research subject. The main goal of this study is to develop an intelligent system that can generate clusters of patient complaints using the K-Means Clustering algorithm. The system is developed using the Research and Development (RnD) method. The clustering process applies a data mining approach, producing clusters based on patient complaints. A total of 600 complaint records, categorized into 72 distinct types, were used. The output consists of three clusters: C1 (high intensity) with 24 categories, C2 (moderate intensity) with 14 categories, and C3 (low intensity) with 34 categories. A practicality test yielded a score of 0.81, indicating the system is highly practical, while an effectiveness test by medical staff scored 0.88, showing the system is highly effective. This system enables health centers to identify trending complaints in the community and develop more focused prevention and treatment strategies. The clustering results also serve as a valuable foundation for strategic decision-making in disease control.

1. Introduction

Puskesmas Balai Makam experiences a significant number of patient visits daily, with each patient presenting various types of complaints. The wide variation in symptoms experienced by patients often makes it difficult for the Puskesmas to determine effective forms of service, such as health education based on the intensity of diseases experienced by the community. Appropriate health education or socialization will help reduce the risk of disease spread in the community.

Patient complaints can be an initial indicator in tracing the source or cause of a disease. By understanding these complaints, the Puskesmas can assess whether the patient's illness stems from internal factors such as abnormal physical or genetic conditions, or from external factors such as the influence of the living environment, lifestyle, or hygiene.

Based on the identified problems, this research aims to solve these issues by creating a web-based system that can perform data clustering. The K-Means clustering data mining algorithm is used in this study. By using this technique, different disease symptoms that the community experiences are grouped into

distinct groups. The developed system can assist the Puskesmas in providing more effective and efficient services, so that appropriate services will reduce the risk of disease spread in the community.

The Puskesmas itself functions as a government-owned public health service institution operating at the sub-district level [1]. Its role is crucial in supporting the functions of higher-level health institutions such as hospitals, especially in preventing and managing public health problems. Therefore, improving the quality of services at the Puskesmas level is an important step in realizing an equitable and quality health service system.

Meanwhile, medical records, known as ICD (International Classification of Diseases), are documents containing medical notes of patients who have undergone examination or treatment at Puskesmas or hospitals. In medical practice, doctors record diagnoses and medical actions in medical language, which are then converted into ICD codes. These codes serve as a standard classification widely used by medical personnel, including general practitioners, to identify diseases based on established guidelines [2].

Knowledge Discovery in Database (KDD) is a model for extracting new knowledge from data collections stored in a database[3]. This KDD is carried out with the aim of discovering new knowledge. Typically, databases are used solely for data storage. However, through the KDD approach, we can extract important information from the data, which can then be used as a guide for analysis and specific actions. In general, the KDD process consists of several stages: data cleaning, data integration, data selection, data transformation (changing the data to fit the k-means analysis format), and data mining, the process of discovering new knowledge from the database.

Data mining itself functions to uncover hidden values and important information from large datasets, which would usually not be detected if analyzed manually[4]. This stage is the core of the entire KDD process. To extract and find valuable information and knowledge from massive datasets, the data mining process employs a variety of strategies, including statistical, artificial intelligence, machine learning, and mathematical computation tools. One technique frequently employed in the knowledge discovery database process is K-Means. One technique commonly employed in machine learning research and data pattern analysis is the partitioning approach in K-Means. The determination of cluster centers, or centroids, is the initial stage of this algorithm, so the final result greatly depends on the selection of these centroids. Because initial centroids are determined randomly, clustering results do not always guarantee an optimal solution[5]. Non-hierarchical clustering methods, including K-Means, produce several data partitions depending on the centroids determined based on characteristic similarity. As a result of this process, data groupings or clusters will be created, with similar data being put into one cluster and diverse data being placed in other clusters.

The K-Means Clustering algorithm has been widely applied in various studies for data grouping. One application was in previous research conducted by the author, where the data used was the same as the current research but no system had been created. In the previous research, clustering still used second-party support software, namely MATLAB. That research produced 3 data clusters ranging from high, medium, and low intensity[6].

The same method was also used in sales clustering by region conducted by Elin Mayoan Fitri, et al. Sales data was utilized in the study. This research produced 3 clusters: 3 high-cluster regions, 11 medium-cluster regions, and 4 low-cluster regions[7].

Another study was conducted by Suryani et al. regarding a geographic information system for mapping road damage using the K-Means clustering algorithm. This research used road damage data in Malang district. This research produced a mapping system for damaged roads in Malang district[8].

Handayani also grouped pupils according to their learning styles using the K-Means method. The data used was learning style data for each student. This research produced an application using K-Means data mining as an algorithm that generates student clusters partitioned by learning styles more efficiently and effectively[9].

Furthermore, K-Means was applied in mapping underprivileged citizens for determining eligibility for aid in the Karang Besuki sub-district by Muhammad Ali Hasymi et al. Data from disadvantaged residents in the Karang Besuki subdistrict area were managed for this study. The research conducted produced a geographic system that generates mapping of underprivileged citizens with the results: 55% of citizens were not eligible for aid, 99% of citizens were less eligible for aid, and 48% of citizens were eligible for aid[10].

Developing an intelligent system that can produce clusters of patient complaints was the goal of this study. The difference between the current research conducted by the author and the previous research by the author is that the previous research focused only on analysis using K-Means without a system, whereas the current one uses a system. The difference with other studies lies in the research object, ranging from road damage, unemployment data, and poor communities, while this research focuses on patient complaint data. The similarity with previous research is the use of K-Means in the data clustering process.

2. Method

This research refers to a framework shown in Figure 1. This framework serves as a reference in directing the research flow, so that each process can run sequentially and systematically. The existence of this framework is expected to help researchers achieve more structured results and facilitate understanding of the entire process undertaken. Figure 1 shows a visualization of this research framework.

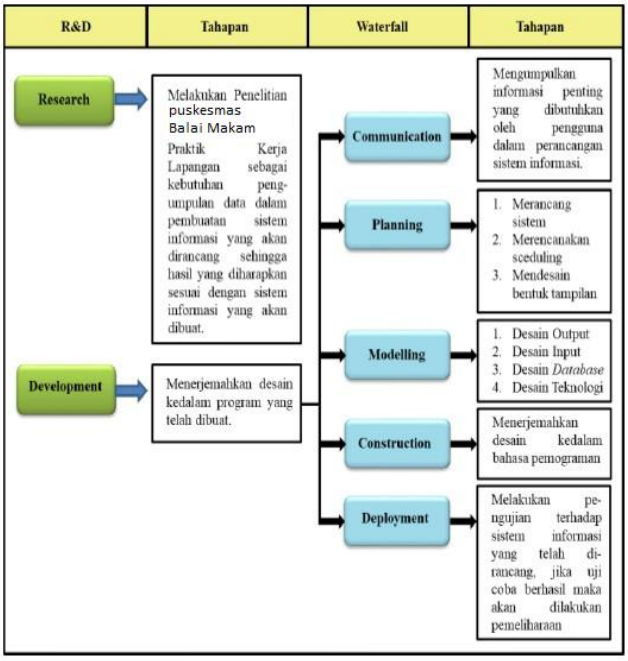


Figure 1. Research framework

This study was carried out utilizing the Research and Development (Rnd) approach, as Figure 1 illustrates. First, in this method, the research process is focused on identifying system needs, both from the system side itself and the data to be used. The R&D methodology is the primary strategy for system development. Its goal is to create a system or product and test it, including determining its degree of efficacy. The RnD method is also utilized in education, such as creating and validating a product that can be used in the learning process. This knowledge leads one to conclude that the purpose of the research and development approach is to create a product and assess its practicality[11].

The System Development Life Cycle (SDLC) model, which is a series of system development procedures beginning with planning, feasibility analysis, and post-implementation assessment, is also used in this study's execution. This model aims to translate organizational needs into application system form [12]. The type of system development used is the waterfall model, which offers a phased and structured approach, resembling a waterfall, in the software creation process. System requirements analysis, system design,

system implementation or creation, system testing, and maintenance of the finished system are the phases in this approach[13]. According to Pressman, the stages in the waterfall model consist of several phases as follows[14]:

Figure 2. Pressman's Waterfall Model

In software creation and development, the Waterfall model is a methodical and sequential process where each step must be completed in the correct sequence and no stage may be bypassed before going on to the next. This approach provides a neat and structured workflow. The following is an explanation of the Waterfall model's steps.

2.1. *Communication*

The initial stage involves interviews to collect data and all requirements for the system to be built. This process aims to ensure the development team thoroughly understands the users' desires and expectations for the system to be built. The information obtained will serve as a reference in designing features that align with user needs.

2.2. *Planning*

This stage includes compiling a software work plan, which covers the division of technical tasks, risk identification, resource allocation, and determination of outputs and completion schedules. In this context, system development time estimates are also designed so that all development activities can be measured, and results can be evaluated periodically according to system needs.

2.3. *Modeling*

In the modeling stage, the identified requirements are transformed into a technical software design. This process includes data structure design, system architecture, interface design, and procedural algorithms. The selection of appropriate design patterns is also carried out to ensure efficient implementation. In this program, data analysis uses the k-means algorithm model with the following stages:

a. Data cleaning, performed to remove inconsistent or irrelevant data.

b. Data selection, performed to obtain suitable data for analysis. In this study, the attributes used are patient complaint data per period (per month).

c. Data transformation, so that the data can be processed using the k-means algorithm.

d. Data mining, after the data has been transformed, it can be processed using the k-means clustering algorithm.

The stages are presented in pseudocode in Algorithm 1.[6]

```
Algorima 1. K-Means Clustering
Input :X,k,n,s
Output: C
Procedure
C←0;
For i←1 to n
Si←RandomSample(x,s)
(ai,ci)←kmeans(Si,k)
C←AppendRows(C,ci)
End for
Return Cfinal←kmeans(C,k)
```

```
End procedure
```

2.4. *Construction*

The construction stage includes the coding process or translating the design into a programming language. Programmers compile program code based on user needs and designed transaction scenarios. Software testing is performed using the black box testing method, which is a testing method based on functional specifications without looking at the code structure. This test is used to verify whether all system features align with user desires in solving problems. This approach may be used for unit, integration, system, and acceptance testing, among other testing levels[15].

2.5. *Deployment*

The completed system is given to users at this point in the development process so they may start using it. After implementation, the software requires periodic maintenance, including bug fixes, addition of new features, or adjustments to changes in user needs and technology. Continuous maintenance ensures the system remains optimal and can be used in the long term.

2.6. *Product Test*

2.6.1. *Validity Test*

To ensure that the instruments used for this research are appropriate, a validity test is conducted on each statement item. This testing is done using Aiken's V statistical formula, It has the following formulation [16]:

$$V = \sum s / [n (c - 1)] \quad (1)$$

Description of the formula :

s: difference between the rating value (r) and the lowest value (lo)

lo: value with the lowest validity

c: highest validity value

r: rater's score

n: number of raters

This formula is used to calculate the validity value based on the opinions of experts or raters. The value s indicates the difference between the given score and the minimum score, while the maximum value and the number of raters are used to determine the instrument's level of conformity with the concept to be measured.

Table 1 Aiken's V Validity Range[17]

Rentang	Keterangan
0,68 < X <=1,00	High Validity
0,34 >= 0,67	Medium Validity
0>x>=0,33	Low Validity

Table 1 presents the validity assessment criteria using Aiken's V formula, classified based on the percentage of scores obtained. This criterion divides validity results into two main categories: if the obtained value is less than 0.6, the item is declared "Not Valid," meaning the instrument does not meet the specified validity standards and is not recommended for use. If the value is more than 0.6, the item can be declared "Valid," indicating that the instrument is feasible for use because it has met the requirements to accurately measure the intended variable. With these criteria, researchers can be more objective in determining the feasibility of using instruments in research or further evaluation.

2.6.2. *Practicality Test*

The purpose of the practicality test is to evaluate how well and efficiently users can apply the fieldwork practice monitoring information system. To conduct the assessment, Puskesmas operators are given questionnaires.

The practicality evaluation of this system is done by summarizing operator responses to a number of statements in the questionnaire. From the responses given, the practicality level is then calculated using the Moment Kappa statistical formula, which is mostly used in measuring the practicality level of a created product. The formula used is as follows[18]:

K = (P - Pe) / (1 - Pe) (2)

The Moment Kappa formula is used to measure the practicality level of a product or system based on evaluations conducted by testers. In the context of this research, The usefulness of the fieldwork practice monitoring information system is evaluated using Moment Kappa.

The following is a description of the Moment Kappa formula:

K: practicality level value

P: obtained value, derived from dividing the sum of testers' scores by the maximum possible score.

Pe: unrealized value, which can be calculated by dividing the maximum score by the amount of the testers' contributions.

Table 2 Criteria for Determining Moment Kappa Practicality[19]

Interval	Keterangan
0,81 – 1,00	Very high
0,61 – 0,80	High
0,41 – 0,60	Medium
0,21 – 0,40	Low
0,01 – 0,20	Very low
≤ 0,00	Not practical

This table shows the criteria for determining practicality based on the obtained value, which is used to measure the practicality level of a product. Each value interval is indicated by a specific category, ranging from "Very high" to "Not practical," describing how well the product meets practicality criteria. Using this table, researchers can easily interpret evaluation results and determine necessary improvement steps to enhance product practicality.

2.6.3. *Effectiveness Test*

The purpose of the effectiveness test is to evaluate how well a learning program or approach can accomplish set objectives. To obtain accurate data regarding this effectiveness, analysis is conducted using appropriate statistical approaches. For this test, the Richard R. Hake (G-Score) formula is used [20]:

$$g = (Sf - Si)/(100 - Si) \quad (3)$$

G-score formula description:
g : G-Score value
Sf: final score
Si: initial score
The g-score's criteria are as follows [20]:

Table 3 Criteria for Determining g-score [21]

Interval	Category
g > 0.7	High effectiveness
0.7 > g > 0.3	Medium effectiveness
g < 0.3	Low effectiveness

The criteria for evaluating a program's or method's efficacy based on the g-value determined from the test sheet are explained in the description above. The program is highly effective in accomplishing its goals when the g-value is more than 0.7, which is indicated by the "High-g" category. The "Medium-g" category indicates medium effectiveness if the g-value is within the interval of 0.3 and 0.7, suggesting that the program is still quite effective but needs some improvements. Meanwhile, the "Low-g" category indicates low effectiveness when the g-value is less than 0.3, meaning that the program has not successfully met the expected expectations and requires further evaluation and revision.

3. Results and Discussion

The research conducted resulted in a clustering system that can cluster patient complaint data based on patient medical records, produce handling priorities, and provide solutions. This research also facilitates the Puskesmas in providing targeted community services.

3.1. *Communication*

The process of determining and evaluating the requirements for the system that has to be created is the main emphasis of this stage. Direct observation techniques and discussions with potential system users are used to collect needs. This stage is carried out to obtain the needs or features required by users, so that the designed system can later meet functionality optimally. The main processes in this stage include project initiation to determine the general scope of the project, and requirements gathering which aims to collect all system needs in a structured manner.

3.1.1. *Project Initiation*

Pre-research interviews were conducted to gather initial information needed before starting further research. In these interviews, various aspects related to system needs and expectations were discussed in depth. Based on the findings from pre-research interviews, it can be concluded: The large variation in symptoms experienced by patients often makes it difficult for the Puskesmas to determine effective forms of service such as health education based on the intensity of diseases experienced by the community. Appropriate health education or socialization will help reduce the risk of disease spread in the community.

3.1.2. *Requirements Gathering*

At this stage, the researcher conducted interviews at the Balai Makam Puskesmas. From the interview results, it was known that for this system, the needs or features to be created in the system were identified; in this system, access rights consist only of an admin.

3.2. *Planning*

The Planning stage is a phase aimed at compiling a system development plan, which includes estimations of various technical tasks to be performed, identification of potential risks that may arise during the development process, calculation of resource needs, and formulation of the product to be produced. Additionally, this stage also includes compiling a work schedule and monitoring the progress of each stage's implementation so that the entire process runs according to the predetermined flow and responsibilities.

3.2.1. *Estimating (Task Estimation)*

In this system, the internal processes are assigned to the admin or Puskesmas operator. The Admin is the individual who manages the entire information system. Admin tasks include logging into the system, inputting patient complaint data, performing iterations, and changing data if necessary.

3.2.2. *Tracking*

The next stage is to design and then implement an offline server. This system can be accessed as long as the Puskesmas server is activated. Thus, this system can only be accessed within the Puskesmas environment.

3.3. *Modeling*

This is the point at which the system's first design is completed. The modeling used includes use case diagrams, and class diagrams are used to define database relations. At this stage, the variables used and the design of the created system are depicted.

3.3.1. *Use Case Diagram*

Use case diagrams are employed to illustrate the connections between the program's actors or users and its cases. A case is a process that can be performed on the system; an actor is a person or user in the program.

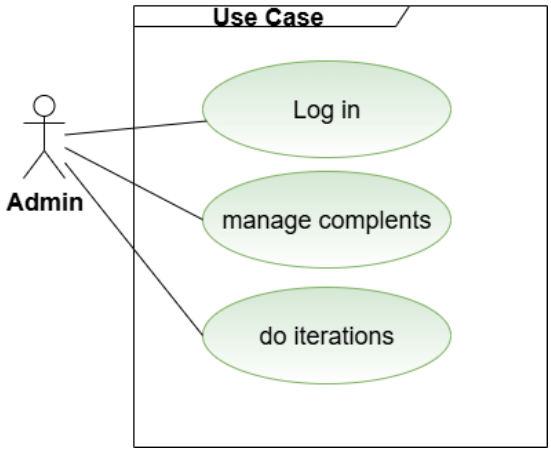


Figure 3. Use Case Diagrams

based on Figure 3's use case diagram there is 1 actor, namely admin, and there are 4 use cases: logging into the system, inputting patient complaint data, performing iterations, and changing data if necessary.

3.3.2. *Class Diagram*

Class diagrams are created with the aim of describing the variables used by each class or entity used within the system. Class diagrams also define the relationships of each class. The following is the class diagram present in the system:

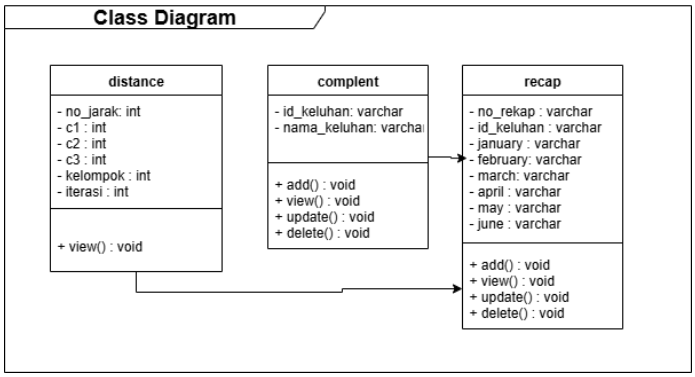


Figure 4. Class Diagrams

As we can see in the figure above, the Class Diagram shows the relationships between entities and the attributes contained within them. The classes in this system include distance (jarak), complaint (keluhan), and recap (rekap).

3.3.3. *Admin Display*

The admin in this system plays an important role in inputting patient complaint data. The admin is also tasked with managing data iteration. The admin dashboard display is as follows:

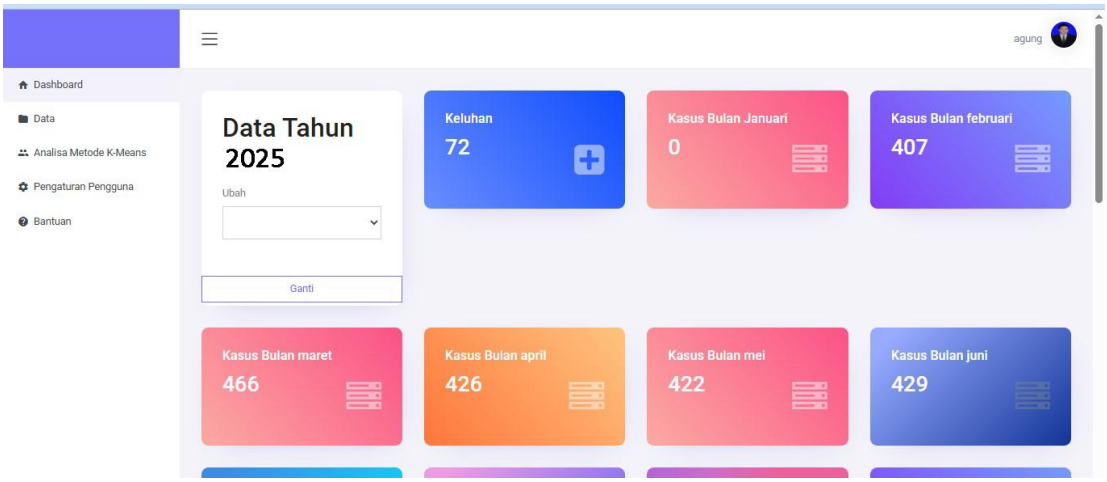
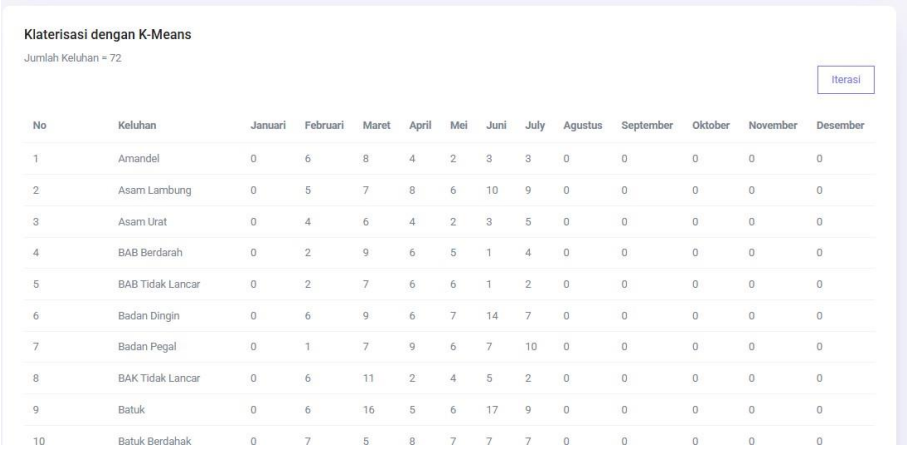


Figure 5. Admin Dashboard

The admin dashboard display contains a recap of patient complaints each month and the system's master data menus. The admin can upload complaint data.

3.3.4. *K Means Display*

In this view the k-means algorithm is applied. At this stage, 600 data in 72 categories will be clustered according to the algorithm that has been applied to the program. After the data input is done, the number of clusters will be determined, in this study there are 3 clusters. The next process is carried out by the iteration process from data input as shown below :



No	Keluhan	Januari	Februari	Maret	April	Mei	Juni	July	Agustus	September	Oktober	November	Desember
1	Amandel	0	6	8	4	2	3	3	0	0	0	0	0
2	Asam Lambung	0	5	7	8	6	10	9	0	0	0	0	0
3	Asam Urat	0	4	6	4	2	3	5	0	0	0	0	0
4	BAB Berdarah	0	2	9	6	5	1	4	0	0	0	0	0
5	BAB Tidak Lancar	0	2	7	6	6	1	2	0	0	0	0	0
6	Badan Dingin	0	6	9	6	7	14	7	0	0	0	0	0
7	Badan Pegal	0	1	7	9	6	7	10	0	0	0	0	0
8	BAK Tidak Lancar	0	6	11	2	4	5	2	0	0	0	0	0
9	Batuk	0	6	16	5	6	17	9	0	0	0	0	0
10	Batuk Berdahak	0	7	5	8	7	7	7	0	0	0	0	0

Figure 6. Iteration Display

The figure above shows the system interface displaying patient complaint data to be clustered. This section contains a data recap for each month over one year. The iteration process will produce clusters of patient complaints.

The clustering display contains clusters for each patient complaint. The clusters consist of 3 categories: high intensity, medium intensity, and low intensity. The cluster display from this system is shown in Figure 7 below:

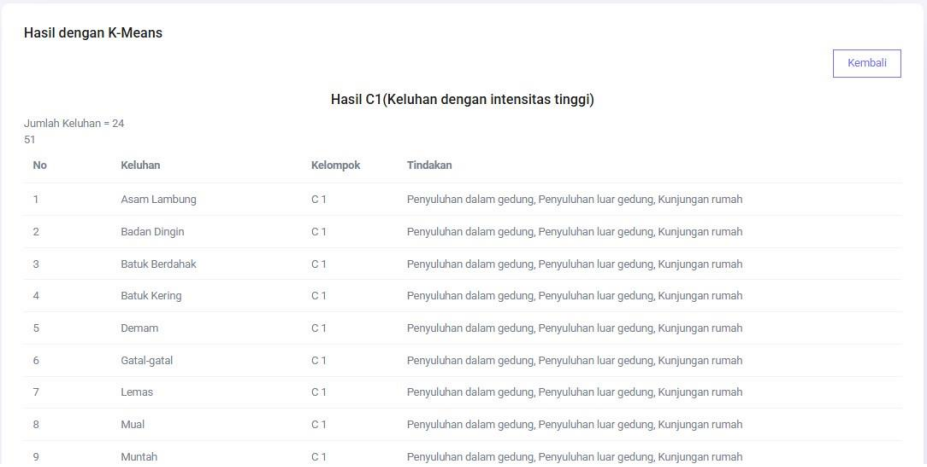


Figure 7. Clustering Display

From the image we can see the solutions provided according to the complaints of patients in an effort to prevent disease in improving health center services. This process produces 3 clusters, namely C1 with high intensity totaling 24 categories, C2 with medium intensity totaling 14 categories, C3 with low intensity totaling 34 categories seperti tabel berikut.

Table 5 Clustering Results

Complaint	Results	Information
Amandel	C2	medium intensity
Asam Lambung	C1	high intensity
Asam Urat	C3	low intensity
BAB berdarah	C3	low intensity
BAB tidak lancar	C3	low intensity
Badan dingin	C3	low intensity
Badan Pegal	C3	low intensity
BAK tidak lancar	C2	medium intensity
Batuk	C2	medium intensity
Batuk berdahak	C3	low intensity

3.4. *Construction*

This stage is the system creation stage. This system is web-based, therefore using HTML, CSS, JavaScript, and PHP. After the system is finished, a black box table will be used for system testing. The black box results will serve as the basis for feedback.

3.4.1. *Testing*

Testing is the testing stage for the information system that has been completed. If errors are discovered after testing, the application will be fixed. If there are no errors, the information system will be tested directly in the field. This testing stage is important to ensure that the system runs according to its functions and meets user needs. Testing also helps identify potential bugs or errors not detected during system creation.

3.4.2. *Testing with Black Box Method*

The following are the results of the black box testing conducted on the system. This black box testing is performed to determine whether each component within the system is functioning properly. Testing was done by opening various pages and input forms in the system, such as the login page, main menu, complaint data, iteration, and clustering display. Each testing step was designed to ensure that the system displays the appropriate page or form when the user accesses a particular menu. The test results show that every tested component functions as desired, in accordance with the initial design objectives. The outcomes of the black box evaluation are as follows.

Table 4 Black Box Testing

No	Rancangan	Hasil	Keterangan
1	login page	admin login page displayed	Success
2	main menu	main menu page displayed	Success
3	complaint data menu	complaint page displayed	Success
4	complaint input form	complaint input form displayed	Success
5	iteration menu	iteration page displayed	Success
6	Iteration process	Iteration results displayed	Success
7	Open clustering result	clustering result page displayed	Success

Based on the black box testing results in Table 3, all components in the system are functioning according to the initially set design. Every tested page and input form was successfully displayed correctly, from the login page to various menus and forms related to complaint data, iteration, and clusters. No failures or display errors were found during the testing process. This indicates that the system is ready for use in its operational context.

3.5. *Deployment*

This stage is the final stage of system creation where the system will be run for users. This system will be trialed with all stakeholders of this system. The processes carried out include delivery, support, and feedback.

3.5.1. *Delivery*

The product is delivered to users by sharing the system link with admin operators. The admin web can be accessed by the admin via a link that directs to the server. This system can be accessed as long as it is within the Puskesmas environment.

3.5.2. *Support*

This system is beneficial for the Balai Makam Puskesmas because the system can be run effectively, is practical in its use, and is inspirational. With features that facilitate management. Additionally, this system also supports fast access, thereby helping to improve services to the community.

3.5.3. *Feedback*

At this stage, corrections or updates have been made to existing deficiencies. This stage is carried out so that the created system operates optimally and precisely according to the initial design previously made. After this stage is carried out, the system is expected to run better, provide better service, and minimize potential errors in the future.

3.6. *Product Test*

In the process, there are 3 categories tested: validation test, product practical use or practicality test, and product effectiveness test. The purpose of this validity test is to determine whether the system's item components are legitimate. The practical or practicality test of the system is carried out to determine the ease of use of the created system. The effectiveness test is carried out to determine the accuracy of the created system in solving problems in this research.

3.6.1. *Validity Test*

At this stage, the test was aimed at several experts in the field of computer systems. The author conducted a validity test, namely construct validity on the program and content validity within the program. With a validity value of 0.82 derived from the test results and computations using Aiken's V formula, this product is deemed legitimate.

3.6.2. *Practicality Test*

This stage was carried out with the aim of knowing the practicality of the created system. This test was conducted on Puskesmas admin operators, and the results from the evaluation aspect were 0.81, which falls into the very practical category.

3.6.3. *Effectiveness Test*

The effectiveness test of this system was conducted through an effectiveness sheet filled out by several staff and patients. It was seen that the results from the evaluation aspect showed an average of 0.88, which falls into the high effectiveness category.

4. Conclusion

From the research that has been done, it can be concluded that the system can be a solution to the problems that have been described where it has succeeded in producing clustering of the data used, where in this study it produces 3 clusters, namely C1 with high intensity totaling 24 categories, C2 with moderate intensity totaling 14 categories, C3 with low intensity totaling 34 categories. In this study, a patient complaint clustering information system has been designed that is valid, practical program usage, and effective in solving problems. This can be seen in the validity test with valid results, practical tests with very practical system results, system effectiveness tests with very effective system results. The clustering results can help the health center to design counseling activities so that they are right on target, more effective, and more efficient, so that they can reduce the potential for the spread of disease based on symptoms that often appear. This research is expected to contribute to improving the quality of health center services in the future, as well as being a reference for further researchers who are interested in similar topics.

References

[1] S. N. Utami and S. Lubis, "Efektivitas Akreditasi Puskesmas Terhadap Kualitas Puskesmas Medan Helvetia," *Publik Reform*, vol. 8, no. 2, pp. 10–21, 2021, doi: 10.46576/jpr.v8i2.1658.

[2] D. Penerapan and M. Fifo, "Tinjauan Literatur Analisis Faktor Penyebab Keterlambatan Penyediaan Rekam Medis Rumah Sakit Di Indonesia," *J. Rekam Medis dan Inf. Kesehat. Indones.*, vol. 01, pp. 17–23, 2023, doi: https://doi.org/10.62951/jurmiki.v1i1.5.

[3] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, doi: https://doi.org/10.51519/journalcisa.v4i2.393.

[4] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, and I. O. Kirana, "Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means," *J. Ilmu Komput. dan Inform.*, vol. 1, no. 2, pp. 109–132, 2021, doi: 10.54082/jiki.13.

[5] F. A. Tanjung, A. P. Windarto, and M. Fauzan, "Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 61, 2021, doi:

10.30645/jurasik.v6i1.271.

[6] M. A. V. Ideal, "Classification of Patient Complaints against Patient Medical Record Data Using the K Means Method," *J. Sistim Inf. dan Teknol.*, vol. 5, pp. 1–6, 2022, doi: 10.37034/jsisfotek.v5i1.151.

[7] E. M. Fitri, R. R. Suryono, and A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz," *J. Komputasi*, vol. 11, no. 2, pp. 157–168, 2023, doi: 10.23960/komputasi.v11i2.12582.

[8] T. Suryani, A. Faisol, and N. Vendyansyah, "Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 380–388, 2021, doi: 10.36040/jati.v5i1.3259.

[9] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.

[10] M. Ali Hasymi, A. Faisol, and F. Ariwibisono, "Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karang Besuki Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 284–290, 2021, doi: 10.36040/jati.v5i1.3269.

[11] M. Waruwu, "Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan," *J. Ilm. Profesi Pendidik.*, vol. 9, no. 2, pp. 1220–1230, 2024, doi: 10.29303/jipp.v9i2.2141.

[12] Ichsan Raksa Gumilang, "Penerapan Metode SDLC (System Devlopment Life Cycle) Pada Website Penjualan Produk Vapor," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 47–56, 2022, doi: 10.55606/jurritek.v1i1.144.

[13] A. Fitri, L. Efriyanti, and R. Silmi, "Pengembangan Modul Ajar Digital Informatika Jaringan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 33–38, 2023.

[14] Z. Rahmah, S. Derta, H. Antoni Musril, and R. Okra, "Perancangan Website Eduji Menggunakan CMS Wordpress," *Intellect Indones. J. Learn. Technol. Innov.*, vol. 1, no. 2, pp. 205–218, 2022, doi: 10.57255/intellect.v1i2.206.

[15] Uminingsih, M. Nur Ichsanudin, M. Yusuf, and S. Suraya, "Pengujian Fungsional Perangkat Lunak Sistem Informasi Perpustakaan Dengan Metode Black Box Testing Bagi Pemula," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 1, no. 2, pp. 1–8, 2022, doi: 10.55123/storage.v1i2.270.

[16] N. R. A. Nabil, I. Wulandari, S. Yamtinah, S. R. D. Ariani, and M. Ulfa, "Analisis Indeks Aiken untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia," *J. Penelit. Pendidik.*, vol. 25, no. 2, pp. 184–191, 2022.

[17] N. Afifah and T. Suhery, "PENGEMBANGAN INSTRUMEN VALIDASI UNTUK EXPERT REVIEW TENTANG 1) Program Studi Pendidikan Kimia , Universitas Sriwijaya 2) Program Studi Pendidikan Kimia , Universitas Sriwijaya Email : tatang_suhery@fkip.unsri.ac.id," *Pros. Semin. Nas. Pendidik. IPA Tahun 2021*, pp. 1–13, 2021.

[18] S. S. Febriani and S. Aini, "Pengembangan Media Pembelajaran Powerpoint Interaktif Berbasis Inkuiri Terbimbing pada Materi Ikatan Kimia Kelas X SMA/MA," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 3, no. 4, pp. 215–222, 2021, doi: 10.38035/rrj.v3i4.343.

[19] D. S Oktavia, S. Zakir, S. Supriadi, and L. Efriyanti, "Perancangan Media Pembelajaran Ipa Kelas Viii Menggunakan Aplikasi Canva Dengan Model Microblogging Di Smpn 1 Lubuk Alung," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1764–1769, 2023, doi: 10.36040/jati.v7i3.7707.

[20] A. W. Alsyabri, "Validitas dan Efektivitas Media Pembelajaran Berbasis Android Mata Pelajaran Komputer dan Jaringan Dasar," *J. Educ. Inform. Technol. Sci.*, vol. 3, no. 1, pp. 1–10, 2021, doi: 10.37859/jeits.v3i1.2602.

[21] U. Yusiana and S. P. Prasetya, "Pengembangan Media E-comic Terhadap Hasil Belajar Peserta Didik Dalam Pembelajaran IPS," *J. Dialekt. Pendidik. IPS*, vol. 2, no. 1, pp. 23–33, 2022, [Online]. Available: https://ejournal.unesa.ac.id/index.php/PENIPS/article/view/44636%0Ahttps://ejournal.unesa.ac.id