INTERNATIONAL JOURNAL OF APPLIED ARTIFICIAL INTELLIGENCE AND ROBOTICS
E-ISSN: 0000-0000, DOI: 00.000/XXXX.YYYY.II.NNN
Vol. 1 No. 1 2025 pp. 01-10

Application of K-Means Clustering in the Prevention and Complaint System of Sexual Violence at Politeknik Negeri Lhokseumawe

Sepkha Ayunda Sari[1], Muhammad Rizka[2], Hendrawaty[3]

[1,2,3] Department of Information and Computer Technology, Lhokseumawe State Polytechnic, Indonesia

*Corresponding Author: rizka@pnl.ac.id
Article info: Received 09/09/2025, Revised 4/10/2025, Accepted 09/10/2025

Abstract

Sexual violence on campus, particularly at Politeknik Negeri Lhokseumawe, is an issue that requires serious attention and effective handling. To improve the current manual complaint management system, the application of information technology through the K-Means Clustering method is necessary for the prevention and complaint system of sexual violence. This approach is expected to assist in grouping complaint data from victims, making it easier to identify levels of sexual violence and to design more targeted responses. This research was conducted by developing and managing a sexual violence complaint system, where the complaints were grouped into three clusters based on the calculation of the data. The clusters were validated using the silhouette method, resulting in a score of 0.50 for cluster 0 (moderate level), 0.20 for cluster 1 (moderate level), and 0.58 for cluster 2 (low level). These results indicate that the K-Means Clustering method is effective in categorizing complaint data and can support more effective handling of sexual violence cases on campus.

Keywords: *Sexual Violence, Prevention, K-Means, Clustering*

1. Introduction (14 pt)

Sexual violence within the campus environment is a critical issue that continues to be actively addressed to raise awareness, understanding, and concern not only among the victims who experience such acts but also among individuals who witness them. Sexual violence itself can be defined as any act involving sexual exploitation, harassment, or coercion toward an individual without explicit and informed consent. It encompasses various forms of behavior, ranging from verbal harassment to physical actions that violate individual rights and compromise the well-being of victims [1]. This issue demands serious attention and comprehensive handling. Like other higher education institutions, Lhokseumawe State Polytechnic bears a significant responsibility to create a safe, supportive, and violence-free environment for all its members, especially students and staff [2].

The process of prevention and complaint management involves a complex series of efforts, including the identification of potential risk factors, raising public awareness of this issue, and developing strategies that are not only comprehensive but also effective. Key measures to achieve success in these preventive efforts include establishing a responsive complaint system that allows victims to report incidents safely and without fear, as well as providing adequate support for victims. These are integral initial steps to address psychological and emotional consequences and ensure justice for those affected—thereby guaranteeing safety, fairness, and the protection of human rights, which should be fundamental for every individual [3].

To address the problem of sexual violence on campus, it is essential to develop an effective system for managing complaints and preventing such incidents. Currently, the complaint handling and prevention processes at Lhokseumawe State Polytechnic still rely on manual methods, which tend to be inefficient and prone to errors. Therefore, it is necessary to integrate information technology that can assist in managing sexual violence complaints more efficiently and accurately, minimizing errors in reported data. This can be achieved by developing a Sexual Violence Prevention and Complaint System that implements the K-Means Clustering method.

The use of information technology in the form of a complaint and prevention system applying the K-Means Clustering algorithm can help classify sexual violence complaints based on existing patterns into different severity levels. This enables the identification of the degree of sexual violence experienced by victims based on their submitted reports. Consequently, institutions can respond to sexual violence incidents more effectively and design more targeted prevention programs. Furthermore, implementing the K-Means Clustering method in the sexual violence prevention and complaint system at Lhokseumawe State Polytechnic is expected to improve responsiveness in handling complaints and assist in identifying the severity of cases that may occur on campus. This contributes to creating a safer and more supportive campus environment for all members of the academic community and provides a broader contribution to reducing sexual violence in society.

2. Methods

2.1. System Architecture

The system architecture, as illustrated in Figure 1, begins with the Data Transformation stage, where data from the Data Source are processed and prepared for the subsequent steps. After the transformation process, the workflow continues to the Clustering Using the K-Means Method, which is a data grouping technique based on specific similarities among data points. Once the clustering process is completed, the results proceed to the Interpretation and Evaluation phase, in which the clustering outcomes are analyzed to ensure accuracy and relevance. Following this evaluation stage, the next step is the Distribution of Complaint Results, where the clustering outcomes are utilized to determine appropriate follow-up actions for each group of data. System Architecture shown in Fig. 1.

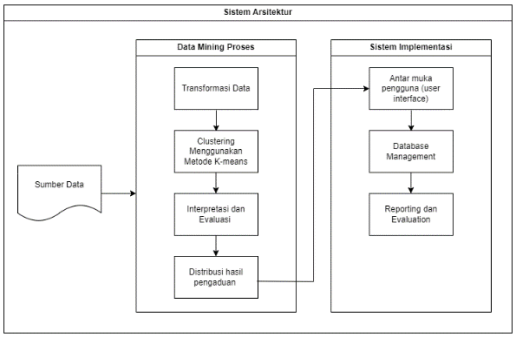

Figure 1. System Architecture

On the system implementation side, the results from the Data Mining Process are then displayed through the User Interface, where users can view and manage the clustering data results. In addition, this system also has Database Management, which functions to manage and store complaint data, as well as Reporting and Evaluation, which are used to provide reports and evaluations based on the processed data. The interaction between these two components produces an integrated system for managing sexual violence complaints using the K-Means method.

2.2. K-Means Clustering Method

K-Means is one of the algorithms that belongs to the unsupervised learning category. The K-Means algorithm functions to group data into several clusters. The advantage of this algorithm is its ability to accept data without any predefined category labels [4]. By using this method, data can be grouped based on certain similarities in features or attributes without requiring any assistance from labels or prior supervision.

Simply put, K-Means is an algorithm that is commonly used for document clustering. The main principle of K-Means is to arrange k prototypes or centroids from a set of multidimensional data [5]. Before applying the K-Means algorithm process, the documents will first go through a preprocessing stage. Then, the documents are represented as vectors that contain terms with specific values [6].

The K-Means algorithm requires an input parameter k and divides a set of n objects into k clusters, where the similarity level among members within the same cluster is high, while the similarity level with members in other clusters is very low. The similarity between members and the cluster is measured by the closeness of the object to the mean value in the cluster, or what is called the cluster centroid [7].

The basic concept of K-Means is the iterative search for the cluster center. The cluster center is determined based on the distance of each data point to the cluster center, as follows:

- The clustering process begins by identifying the data to be clustered, xij (i=1,...,n; j=1,...,m) where n is the number of data to be clustered and *m* is the number of variables.
- At the beginning of the iteration, the center of each cluster is determined randomly, ckj (k=1,...,K; j=1,...,m).
- Then, the distance between each data point and each cluster center is calculated [4].

- To calculate the distance of the i data (Xi) to the k cluster center (Ck), named (dik), the Euclidean distance formula can be used:

$$dij = \sqrt{\sum_{j=1}^{m} (xij - ckj)^2}$$

- A data point will become a member of cluster *J* if the distance of that data point to the cluster center *J* has the smallest value compared to the distance to other cluster centers.
- Next, the data that belong to each cluster are grouped.
- The new cluster center value can be calculated by finding the average value of the data that belong to the cluster, using the following formula [5]:

$$ckj = \frac{\sum_{h}^{p} 1yhj}{p}; yhj = xhj \epsilon clusterke - k$$

2.3. System Design

The system design of this research includes several stages, namely the design of the use case diagram, flowchart, sequence diagram, and class diagram, which aim to describe comprehensively the working process of the system in performing clustering on sexual violence complaint data within the campus environment.
This design stage is very important because it provides a clear overview of how the system will operate and interact with users as well as other components.

1) Use Case Diagram

A use case is a system description used to illustrate the interaction between users (or actors) and the system, aiming to achieve a specific goal. In the context of software engineering, a use case is used to define the functional requirements of a system, namely how the system should function from the user's point of view [8]. This diagram shows how two main actors, namely User and Admin, interact with the system. The User, who may consist of students, staff, or other parties, has access to four main functions in this system: Login, Complaint, Article, and Logout.The user begins their interaction with the system through the Login process to access the system. After logging in, the user can submit Complaints related to sexual violence cases they have experienced or witnessed. In addition, the user can also read Articles available in the system to gain information related to the prevention and handling of sexual violence. After finishing using the system, the user can exit using the Logout function. Use case diagram in for this application shown in Fig. 2.

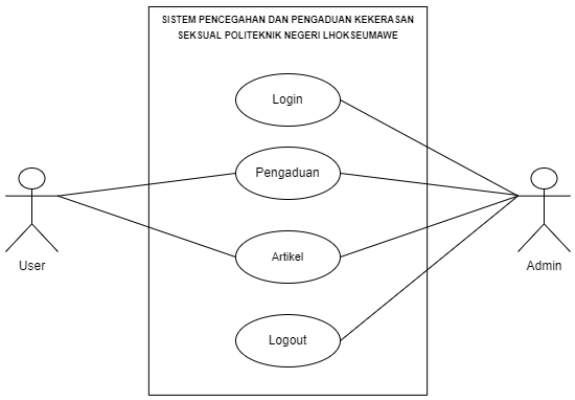

Figure 2. Use Case Diagram

2) Flowchart

The process begins by entering the data to be analyzed, followed by determining the desired number of clusters. Next, the user inputs the initial central points (centroids) for each cluster. The algorithm then calculates the distance of each data point to the corresponding cluster center. Based on the shortest distance, the data are grouped into the most appropriate cluster. This process of distance calculation and grouping is repeated until there are no significant changes, resulting in the final division of data into optimal clusters. Once the clustering results are obtained, the process is completed. Fig. 3 shown the flowchart use in this study.

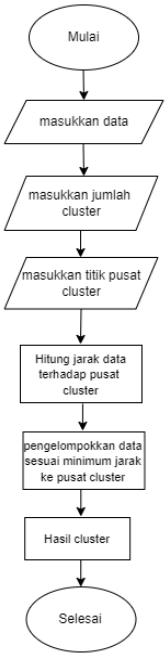

Figure 3. Flowchart

3) Sequence Diagram of Complaint Input

A sequence diagram illustrates the chronological order of actions that occur. Each message has a marker indicating the sequence of communication between objects [9]. In Fig. 4, The process begins when the user clicks the Complaint button, which triggers the system to display the complaint page. After that, the user inputs the complaint data, which is then sent to the database for storage. The system then performs a selection process on the complaint data stored in the database and displays the resulting complaint data back to the user. This diagram shows the communication flow between the user, the complaint module, and the database in managing complaint data in a structured manner.

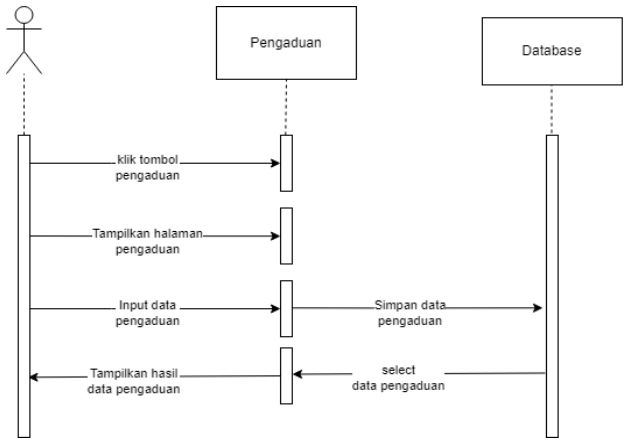

Figure 4. Sequence Diagram of Complaint Input

4) Sequence Diagram of Admin Article

This diagram illustrates the interaction between the user, the article input module, the article deletion module, and the database. The process begins when the user clicks the Article menu to open the article page. After the article page is displayed, the user can add a new article by filling in the required data. The system then receives this input and stores the article data into the database. After the storage process is successful, the system retrieves the article data from the database and displays it to the user as the input result.

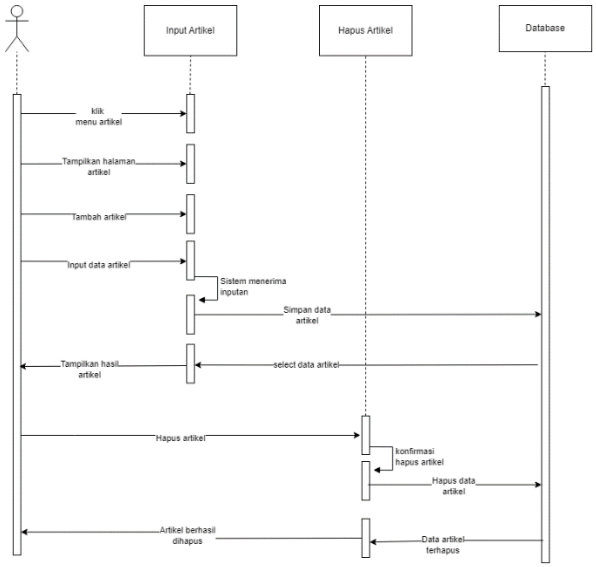

Figure 5 Sequence Diagram of Admin Article

In addition to the input process, this diagram also shows the workflow for deleting an article. After the article results are displayed, the user can choose to delete the article. The article deletion module processes this request by sending a delete command to the database. The system then asks the user for deletion confirmation. If confirmed, the article data will be deleted from the database, and the system will notify that the article has been successfully deleted. Sequence Diagram of Admin Article in detail shown in Fig. 5.

5) Class Diagram

Fig. 6 represents the class diagram, which shows two main classes, namely GUI and Result, along with their functionalities. The GUI class is responsible for receiving input and managing the clustering process, with methods such as Input_data(), determine_cluster(), process_cluster(), and Kmeans_clustering(), which are used to group data using the K-Means algorithm. After the clustering process is completed, the results are passed to the Result class, which has three methods: cluster_low(), cluster_medium(), and cluster_high(). Each method displays the results of data grouping into three categories or clusters (low, medium, and high), which are visualized in the form of tables and graphs.

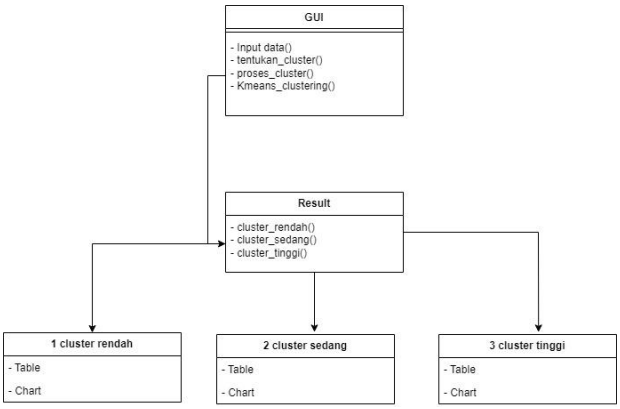

Figure 6 Class Diagram

3. Result and Discussions

3.1. User Interface

The explanation regarding the implementation of the user interface includes all design elements that have been applied, such as layout, navigation, visual aesthetics, and interactive functions designed to facilitate users in reporting and preventing incidents of sexual violence. In addition, this subsection also describes the process and results of system testing that have been implemented. The testing covers various aspects, ranging from functionality testing, to ensure that each feature works according to the planned specifications, to usability testing, which evaluates the comfort and ease of use of the interface by users.

1. Home Page

The homepage is the main entry point of a website — it's the first page visitors see and serves as the digital front door to the entire site. Its primary goal is to introduce the site's purpose, guide users to key sections, and create a strong first impression. Fig.7 shown the Homepage for this application.

Figure 7 Home Page

2. Admin Dashboard Page

The dashboard page consists of several main elements displayed in two sections. On the left side, there is a vertical navigation menu that includes several options; Dashboard, Complaint, Data, Article, and Settings; which allow users to navigate between different features in the system. The main section of the dashboard displays several pieces of information related to Total Complaints, Total Question Data, Total Articles, and Total Clustering Data. Below this section, there are two charts that provide data visualization. The first chart, titled "Clustering", shows a line graph indicating the levels of sexual violence categorized as Low, Moderate, and High. The second chart, titled "Total Clustering", is a pie chart that presents the percentage of each category (High, Moderate, and Low) in a visual form that is easy to understand.

Figure 8 Admin Dashboard Page

3. Complaint Input Page

The complaint form page of the sexual violence prevention and complaint system features the title "Sexual Violence Prevention and Complaint" located at the top, followed by several navigation elements such as "Home", "Complaint", "Article", and a "Login" button. The form is divided into two main sections: Identity and Questions. The Identity section includes input fields for Reporter's Name, Identification Number, Email, and Reporter's Status. Each field is accompanied by a text box or a dropdown menu for entering the corresponding information. The Questions section contains three main questions that request information about: the physical and emotional symptoms experienced by the victim after the sexual violence incident, the extent of social support and mental assistance received by the victim after the incident, and the long-term impact of sexual violence on the victim's

life. Each question is provided with a large text box to allow users to enter their detailed responses.

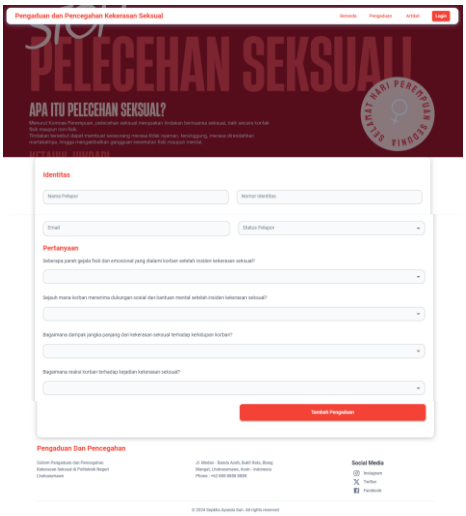

Figure 9 Complaint Input Page

4. Article Page

The article page of the sexual violence prevention and complaint system features a header at the top with the title “Sexual Violence Prevention and Complaint”, along with a navigation menu consisting of “Home”, “Complaint”, “Article”, and a “Login” button. Below the header, there is an information section that explains what sexual harassment is, complete with a red background and descriptive text highlighting the importance of understanding sexual harassment.

Figure 10 Article Page

5. Complaint Data Page

Figure 11 shows the design result of the complaint data page, which displays the “Complaint” option currently selected, indicating that the user is on the complaint page. In the main section of the page, there is a table containing the complaint data that has been submitted. The table displays several columns, including Number (No), Complaint Code, Reporter’s Name, Identification Number, Contact Information, Reporter’s Status, Cluster Level, and Action. These columns provide detailed information about each complaint recorded in the system, including the reporter’s identity, contact details, and the severity level of the complaint, which is categorized into clusters (Low, Moderate, and High). Above the table, there is a search box that allows users to easily search for complaint data based on the reporter’s name.

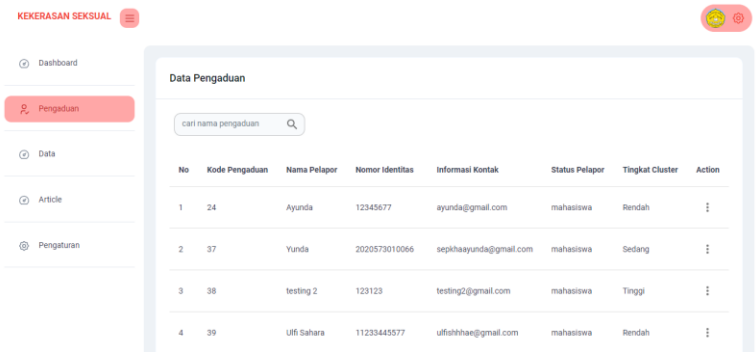

Figure 11 Complaint Data Page

3.2. Silhouette Testing

The silhouette test is a method used to evaluate the quality of clustering or data grouping results. The silhouette score provides an overview of how well the data have been grouped by considering two main aspects: how close the data are within the same cluster (intra-cluster similarity) and how far they are from other clusters (inter-cluster dissimilarity) [10]. The results of the K-Means Clustering were evaluated using the silhouette analysis method, which is a technique used to measure how well the data have been grouped. The silhouette value indicates how close each complaint is to the center point of its cluster and how far it is from the other clusters. This value ranges from -1 to 1, where values closer to 1 indicate that the complaint fits very well within its cluster and is far from other clusters, while values closer to -1 indicate that the complaint would be more appropriately placed in another cluster.Figure 11 shows the results of the silhouette score evaluation for the three generated clusters. In this figure, there are three clusters labeled 0, 1, and 2. Cluster 0 has a silhouette score of 0.5369, cluster 1 has a score of 0.2256, and cluster 2 has a score of 0.5837. The average silhouette score for all clusters is 0.4487.
These results indicate that clusters 0 and 2 have better separation compared to cluster 1, which has a lower silhouette score, showing that the objects in cluster 1 are less well-separated from the objects in other clusters.

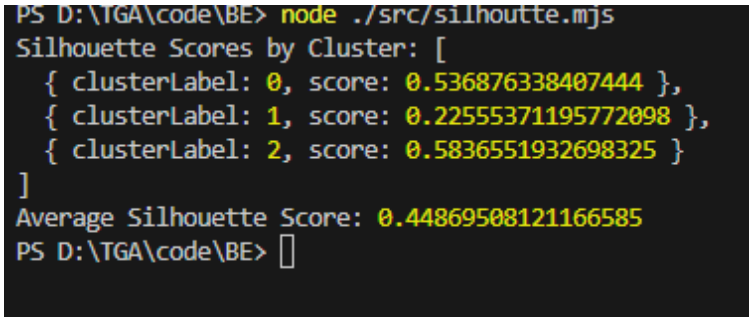

Figure 12 Silhouette Testing

4. Conclusions

After conducting an evaluation of the main objective of this research, which is to develop a system aimed at supporting prevention and complaint handling within the campus area of Politeknik Negeri Lhokseumawe by applying the K-Means Clustering method, several important conclusions can be drawn. The Prevention and Complaint System designed in this study has been successfully developed and is capable of classifying incoming complaints based on the severity level inputted by the reporter. The use of the K-Means Clustering method allows for a more structured classification of complaints, assisting relevant parties in addressing issues based on priority. In addition, the system design process for the sexual violence prevention and complaint handling system in this research utilizes an activity diagram to provide a clear and structured overview of the workflow for each feature within the system. The test results for this system show that the K-Means Clustering method used is capable of performing data grouping effectively. This is demonstrated by the average silhouette score obtained from all clusters, which is 0.4487. Although this value indicates that the clustering process is functioning properly, there is still room for further improvement to achieve more optimal results.Upon deeper examination, the silhouette scores for each cluster vary: Cluster 0 has a score of 0.5369, Cluster 1 has a score of 0.2256, and Cluster 2 records a score of 0.5837. Based on these results, the cluster with the highest silhouette score indicates that the data within that cluster have clearer separation compared to the others. The highest score reflects that, in certain clusters, the method performs clustering more optimally, while for clusters with lower scores, improvements are still needed to achieve more accurate and consistent results.

REFERENCES

[1] U. Marfu'ah, S. Rofi'ah, and Maksun, "Sistem Pencegahan Dan Penanganan Kekerasan Seksual Di Kampus," Kafa'ah J., vol. 11, no. 1, pp. 96–106, 2021, [Online]. Available: http://kafaah.org/index.php/kafaah/index

[2] N. Sudjud and Mohamad Ali Akbar, "Rancang Bangun Aplikasi Pelaporan Anti Kekerasan Seksual (AKAS) Berbasis Android," Inf. (Jurnal Inform. dan Sist. Informasi), vol. 14, no. 2, pp. 162–169, 2022, doi: 10.37424/informasi.v14i2.182.

[3] I. Firdiyansyah, R. Riki, and ..., "Rancang Bangun Sistem Pengaduan Penanganan Kekerasan Seksual Di Universitas Maritim Raja Ali Haji," Jursima …, vol. 10, no. 3, 2022, [Online]. Available: https://ejournal.indobarunasional.ac.id/index.php/jursima/article/view/499%0Ahttps://ejournal.indobarunasional.ac.id/index.php/jursima/article/download/499/308

[4] R. A. Faujia, E. S. Setianingsih, and H. Pratiwi, "Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia," Semin. Nas. Off. Stat., vol. 2022, no. 1, pp. 1249–1258, 2022, doi: 10.34123/semnasoffstat.v2022i1.1511.

[5] A. K-means, A. Fadilah, M. N. P, S. Lumbanbatu, and S. Defiyanti, "Pengelompokan kabupaten/kota di indonesia berdasarkan faktor penyebab stunting pada balita menggunakan algoritma k-means," vol. 6, no. 2, pp. 223–230, 2022.

[6] J. P. Tanjung, B. A. Wijaya, and M. Ridho, "Implementasi Algoritma K-Means Dalam Mengukur Tingkat Kepuasan Siswa Terhadap Pelayanan Pada SMA Swasta Bani Adam AS," Data Sci. Indones., vol. 3, no. 1, pp. 1–11, 2023, doi: 10.47709/dsi.v3i1.2775.

[7] M. Y. Matdoan, F. Y. Risdiana, and G. Haumahu, "Application of the K-Means Cluster for the Classification of Disadvantaged Districts/Cities in Maluku Province," JRST (Jurnal Ris. Sains dan Teknol., vol. 6, no. 1, p. 61, 2022, doi: 10.30595/jrst.v6i1.11637.

[8] M. Naomi, H. Noprisson, F. I. Komputer, U. Mercu, and B. Jakarta, "Analisa Dan Perancangan Sistem Pengaduan Mahasiswa Berbasis Web (Studi Kasus : Universitas Mercu Buana Kranggan)," 2, vol. 1, no. 5, pp. 185–193, 2020.

[9] Haviluddin, "Memahami Penggunaan UML (Unified Modelling Language)," Memahami Pengguna. UML (Unified Model. Lang., vol. 6, no. 1, pp. 1–15, 2011, [Online]. Available: https://informatikamulawarman.files.wordpress.com/2011/10/01-jurnal-informatika-mulawarman-feb-2011.pdf

[10] A. Septianingsih, "Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni," J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat., vol. 3, no. 1, pp. 224–241, 2022, doi: 10.46306/lb.v3i1.116.