Ambidextrous: Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 02, No. 02, 2024, pp. 68-80
E-ISSN: 3031-7002
E-mail:ambidex@takaza.id

Analysis and Application of the K-Means Clustering Algorithm to Identify Dominant Diseases Based on Patient Medical Record Data at Prima Melati Clinic

Elsa Ramadhani

Muhammadiyah University of North Sumatra
Corresponding Author e-mail: elsarmd6@gmail.com

Article History:
Received: 25-06-2024
Revised: 12-08-2024
Accepted: 20-09-2024

Keywords: *Dominant Diseases; K-Means Clustering; Medical Records*

Abstract: *Transformation Digital transformation in the healthcare sector necessitates optimal utilization of medical record data to facilitate more effective decision-making processes. Prima Melati Clinic continues to experience limitations in managing medical record data, which has not undergone systematic analysis to distinguish prevailing disease patterns. The purpose of this study is to analyze and apply the K-Means Clustering algorithm to identify dominant diseases based on patient medical record data at Prima Melati Clinic. The research methodology used is a quantitative approach that utilizes data mining techniques through the Knowledge Discovery in Database (KDD) stages, which include data preprocessing, application of the K-Means algorithm, and interpretation of clustering results. The dataset used consists of approximately 1000 patient medical records covering the period of January 2025 to May 2025. The data preprocessing phase includes data cleaning, missing value management, and data normalization using the StandardScaler technique. Determining the optimal number of clusters is achieved through the Elbow method, using the Sum of Squared Errors (SSE) calculation. The findings indicate that the K-Means algorithm with a cluster size of 3 (k) effectively categorizes patient data into three main clusters based on disease diagnostic characteristics classified by severity (mild, moderate, and severe). Each cluster reveals distinct dominant disease patterns, providing insight into disease distribution in relation to the severity of the patient's condition. The results of this analysis can be utilized by clinics in developing drug procurement strategies, scheduling medical personnel, and designing more targeted disease prevention strategies. Consequently, the implementation of the K-Means Clustering algorithm has demonstrated effectiveness in identifying dominant disease patterns and in strengthening data-driven decision-making at Prima Melati Clinic.*

How to Cite: Elsa Ramadhani. (2024). Analysis and Application of the K-Means Clustering Algorithm to Identify Dominant Diseases Based on Patient Medical Record Data at Prima Melati Clinic. *2*(02). pp https://doi.org/10.61536/ambidextrous.v2i2.499

https://doi.org/10.61536/ambidextrous.v2i2.499

Introduction

Digital transformation has significantly influenced the healthcare sector through the adoption of digital health technologies, health information systems, telemedicine, and electronic medical records (EMRs). These innovations play a crucial role in improving healthcare efficiency, accuracy, accessibility, and quality of service delivery (World Health Organization [WHO], 2021). The implementation of electronic medical records enables healthcare providers to manage patient information more effectively, reduce administrative errors, and support evidence-based clinical decision-making (Kruse et al., 2018). In Indonesia, the Ministry of Health has mandated the implementation of Electronic Medical Records through Regulation of the Minister of Health Number 24 of 2022. However, the adoption of EMRs remains uneven across healthcare facilities due to differences in technological readiness, infrastructure availability, and human resource capabilities (Kementerian Kesehatan Republik Indonesia, 2022). This condition highlights the importance of optimizing health information systems to ensure that healthcare data can be transformed into valuable information for improving healthcare services.

Prima Melati Clinic, a healthcare facility located in Medan, faces challenges similar to those encountered by many healthcare institutions in developing countries. The management of patient medical records is still largely dependent on conventional methods, resulting in inefficiencies in data storage, retrieval, and management. Manual record-keeping systems are often associated with increased risks of data redundancy, data loss, incomplete documentation, and delayed access to patient information (Tsai et al., 2020). Furthermore, inaccurate diagnostic coding and inconsistencies in medical record documentation may negatively affect healthcare planning, clinical decision-making, and resource allocation. These limitations reduce the ability of healthcare organizations to utilize historical patient data strategically for organizational improvement and disease surveillance.

The consequences of manual data processing extend beyond operational inefficiencies and directly affect the quality of healthcare services. Delays in accessing patient records can increase waiting times, hinder clinical workflows, and reduce service effectiveness. Additionally, large volumes of patient data often remain underutilized because healthcare facilities lack analytical tools capable of transforming raw data into meaningful knowledge (Sharma et al., 2021). In this context, data mining techniques offer a promising solution for extracting valuable patterns and insights from large healthcare datasets. Data mining has been widely recognized as an effective approach for identifying hidden trends, supporting clinical decision-making, and improving healthcare management (Han et al., 2022). Among various data mining techniques, clustering methods have gained considerable attention because they enable the grouping of patients with similar characteristics without requiring predefined categories.

Previous studies have demonstrated the effectiveness of the K-Means Clustering algorithm in healthcare data analysis. K-Means has been successfully applied to classify disease severity levels, identify geographical distributions of infectious diseases, analyze diabetes patient profiles, and segment healthcare service users based on demographic and clinical characteristics (Aljohani, 2024; Eken, 2020). The algorithm has also been utilized in clustering

https://doi.org/10.61536/ambidextrous.v2i2.499

patient medical record data based on variables such as age, disease type, and treatment history to support healthcare planning and resource allocation. Its popularity stems from its computational efficiency, scalability, ease of implementation, and interpretability, making it particularly suitable for healthcare datasets that contain large numbers of patient records (Kabir et al., 2024). Moreover, K-Means can provide meaningful insights into disease prevalence patterns that support strategic decision-making in healthcare institutions.

Despite the growing body of literature on healthcare clustering applications, several gaps remain. Many existing studies focus primarily on the technical performance of clustering algorithms without adequately addressing their practical implications for healthcare management. In addition, limited research has examined the implementation of clustering techniques within primary healthcare facilities such as clinics, where medical record systems often face challenges related to incomplete, inconsistent, and unstructured data. Furthermore, previous studies rarely integrate clustering analysis into a web-based decision-support system that can be directly utilized by healthcare administrators for operational planning and disease monitoring. Consequently, there is a need for research that not only applies clustering techniques but also translates analytical findings into actionable insights for healthcare management.

The urgency of this research lies in the increasing volume of patient medical record data generated by healthcare facilities and the need to transform these data into strategic information for improving healthcare services. Without proper analytical approaches, valuable information regarding disease patterns, patient characteristics, and healthcare trends remains hidden within large datasets. The implementation of a clustering-based analytical system can assist healthcare providers in identifying dominant diseases, optimizing drug inventory management, improving healthcare workforce allocation, and designing targeted disease prevention programs. Such capabilities are particularly important for primary healthcare institutions, where efficient resource utilization directly impacts service quality and patient outcomes.

The novelty of this study is reflected in three main aspects. First, this research develops a web-based information system that integrates the complete Knowledge Discovery in Databases (KDD) process, including data preprocessing, cluster determination using the Elbow Method, K-Means clustering, and visualization of analytical results. Second, unlike previous studies that primarily emphasize technical clustering performance, this study focuses on identifying dominant disease patterns and translating clustering results into managerial recommendations for healthcare planning at Prima Melati Clinic. Third, this research explicitly addresses common data quality challenges in medical records, including missing values and data inconsistencies, through systematic preprocessing techniques to improve clustering reliability. These contributions are expected to provide both theoretical and practical value for the implementation of data-driven decision-making in primary healthcare facilities.

Research Methods

Based on observations conducted at Prima Melati Clinic, medical record data management is still carried out using conventional methods, consisting of simple physical archives and passive digital records. As a result, patient data functions primarily as historical documentation of healthcare visits and has not been optimally utilized as a strategic resource

https://doi.org/10.61536/ambidextrous.v2i2.499

for decision-making. In fact, medical record data contain valuable information that can support healthcare planning, disease surveillance, and service improvement when analyzed systematically through data mining approaches (Han et al., 2022). The underutilization of healthcare data remains a common challenge among primary healthcare facilities, particularly in developing countries, where limited technological infrastructure and analytical capabilities often hinder data-driven decision-making (WHO, 2021).

The continued reliance on manual data management systems creates several operational and managerial challenges. These include slow data retrieval processes, reduced service efficiency, difficulties in data integration, and increased risks of redundancy, inconsistency, incompleteness, and data loss. Such issues can compromise the quality of healthcare information and negatively affect the accuracy of clinical and administrative decisions (Kruse et al., 2018; Tsai et al., 2020). Moreover, poor data quality has been identified as one of the major barriers to effective healthcare analytics, reducing the reliability of information extracted from patient records and limiting their usefulness for organizational planning (Sharma et al., 2021). To address these challenges, this study proposes the development of a web-based information system capable of automatically and systematically processing patient medical record data through the Knowledge Discovery in Databases (KDD) framework. The KDD process consists of several stages, including data selection, data preprocessing, transformation, data mining, and knowledge interpretation (Han et al., 2022). In this research, preprocessing activities include data cleaning, handling missing values, and data normalization using the StandardScaler technique to ensure data consistency and improve clustering performance. Data preprocessing is considered a critical stage because the quality of clustering results is highly dependent on the quality of the input data (Aljohani, 2024).

The data mining process is implemented using the K-Means Clustering algorithm, one of the most widely used unsupervised machine learning techniques for grouping data with similar characteristics (Eken, 2020). Patient records are clustered based on variables such as diagnosis, disease severity, and frequency of visits. The determination of the optimal number of clusters is performed using the Elbow Method, which evaluates the relationship between the number of clusters and the Sum of Squared Errors (SSE) value. After identifying the optimal cluster number, the K-Means algorithm iteratively calculates Euclidean distances between data points and cluster centroids, updates centroid positions, and repeats the process until convergence is achieved (Kabir et al., 2024). This approach enables the identification of meaningful patterns within patient data while maintaining computational efficiency.

The resulting clusters are subsequently interpreted into categories representing mild, moderate, and severe disease conditions. To facilitate understanding and decision-making, the clustering outcomes are presented through informative visualizations, including cluster distribution charts, disease pattern graphs, and Elbow Method plots. Data visualization has been shown to enhance the interpretability of analytical results and support healthcare managers in identifying trends and making evidence-based decisions (Fay et al., 2023). Therefore, the proposed system not only improves the efficiency of medical record data processing but also assists Prima Melati Clinic in identifying dominant disease patterns, optimizing healthcare resource allocation, and strengthening data-driven decision-making processes.

https://doi.org/10.61536/ambidextrous.v2i2.499

Results and Discussion

System Requirements

This section describes the specifications of the development and operational environments used to build andoperateThe dominant disease identification system uses the K-Means Clustering algorithm. The environment includes a combination of hardware and software that supports the implementation of the K-Means Clustering algorithm and a web-based interface.

Hardware Requirements

The hardware required for developing and running the system consists of several essential components to ensure optimal performance. The system requires a minimum of an 8th-generation Intel Core i3 processor or its equivalent, such as an AMD Ryzen 3, which is sufficient for data preprocessing activities and the execution of the K-Means clustering algorithm on medium-sized datasets. In addition, a minimum of 8 GB of RAM is needed to support smooth system performance when loading datasets, performing preprocessing tasks, and conducting clustering operations, particularly when handling medical record data that may contain a large number of entries. For storage, a minimum of a 256 GB Solid State Drive (SSD) is recommended, as SSD technology significantly improves operating system boot times, software installation speed, and data read/write performance, thereby enhancing the efficiency of development and testing processes. Furthermore, a stable internet connection is necessary to access online resources during system development and to support the operation of the web-based system once it is implemented.

Software Requirements

The software required for the development and implementation of this system includes several key components that support data processing, machine learning, and web application development. The system can be developed using operating systems such as Windows 10/11, macOS, or Linux distributions like Ubuntu. Python is selected as the primary programming language due to its extensive ecosystem for data science, machine learning, and web development (Kabir et al., 2024). Python is widely recognized as a powerful tool for building predictive and analytical models because it offers a rich collection of libraries for data manipulation, machine learning, and data visualization.

To facilitate coding, testing, and debugging, Integrated Development Environments (IDEs) such as Visual Studio Code, PyCharm, or Jupyter Notebook can be utilized. Several Python libraries are also essential for system development. Scikit-learn serves as the primary machine learning library for implementing the K-Means Clustering algorithm, conducting data preprocessing through tools such as StandardScaler, and evaluating clustering performance using various metrics (Aljohani, 2024; Eken, 2020). The K-Means algorithm is widely recognized for its efficiency, scalability, and effectiveness in clustering numerical datasets. Pandas is employed for reading, managing, and analyzing tabular data from formats such as CSV and Excel files, providing high-performance data structures and user-friendly analytical tools (Eken, 2020). NumPy supports efficient numerical computations, particularly for multidimensional array and matrix operations that form the foundation of many machine learning algorithms. Additionally, Matplotlib and Seaborn are used for data visualization and graphical representation of clustering results, including the Elbow Method used to determine the optimal number of clusters (Eken, 2020; Fay et al., 2023; Kabir et al., 2024).

https://doi.org/10.61536/ambidextrous.v2i2.499

For the development of the web-based application, the Flask framework is utilized. Flask is a lightweight and flexible web framework that simplifies the development of web applications while providing seamless integration with machine learning models. Its simplicity and scalability make it well suited for developing system interfaces, managing backend processes, and presenting clustering results to end users in an accessible manner.

System Implementation

System implementation is the realization stage of the design created in Chapter III. This stage transforms the logical design into a physical, operational form, according to the Prima Melati Clinic's needs. Implementation includes the development of a user interface, data processing modules, and the integration of the K-Means algorithm.

System Interface Implementation

Admin Login Page: On this page, it is the Administrator's responsibility to correctly enter the username and password registered in the system. The authentication process is carried out to ensure that only individuals with the necessary access rights are allowed to log in. The system will assess the validity of the provided credentials. If the submitted information is deemed appropriate, the Administrator will be redirected to the Dashboard page. Conversely, if an error occurs, the system will display a notification message indicating that the username or password is invalid.

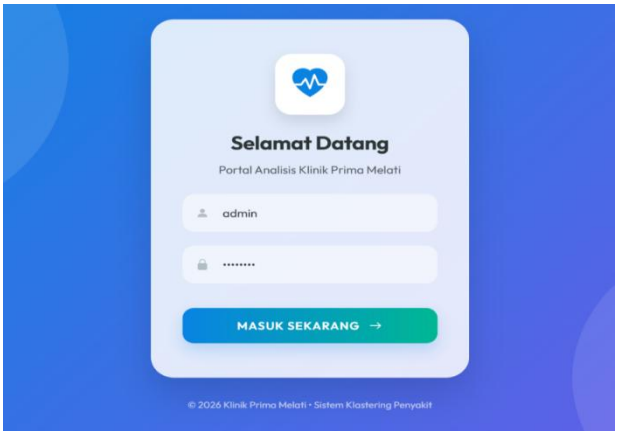

Figure 1. Admin Login Page View

Dashboard Page : The Dashboard page is the main page after successful Administrator authentication. This interface serves as the core of system navigation, facilitating access to all essential functions, such as data upload, clustering, and analysis results. This page also displays a system summary, such as the total number of patient data, the status of processed data, the last analysis time, and a navigation menu to the new data upload feature.

https://doi.org/10.61536/ambidextrous.v2i2.499

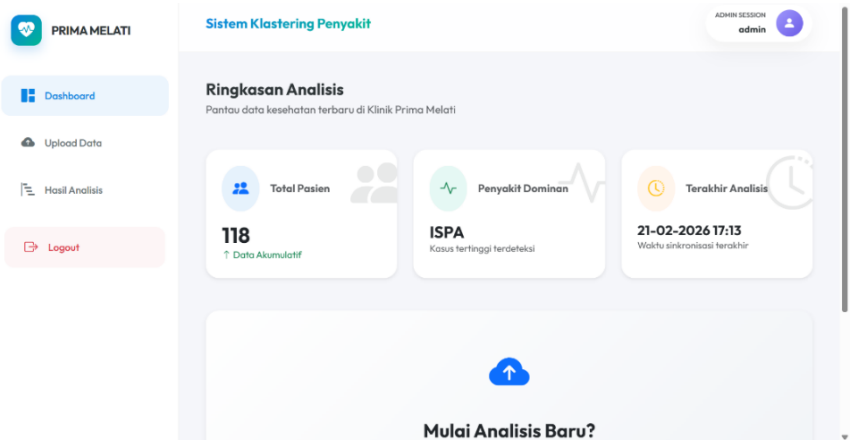

Figure 2. Dashboard View

Data Upload and Preview Page: This page allows the admin to upload patient medical record data files in .xlsx or .csv format. The system will validate the uploaded file format to ensure it complies with the specified data structure. Once the file is successfully uploaded, the data can be displayed first with a preview. The data will then be temporarily stored for processing in the next stage, namely preprocessing and clustering.

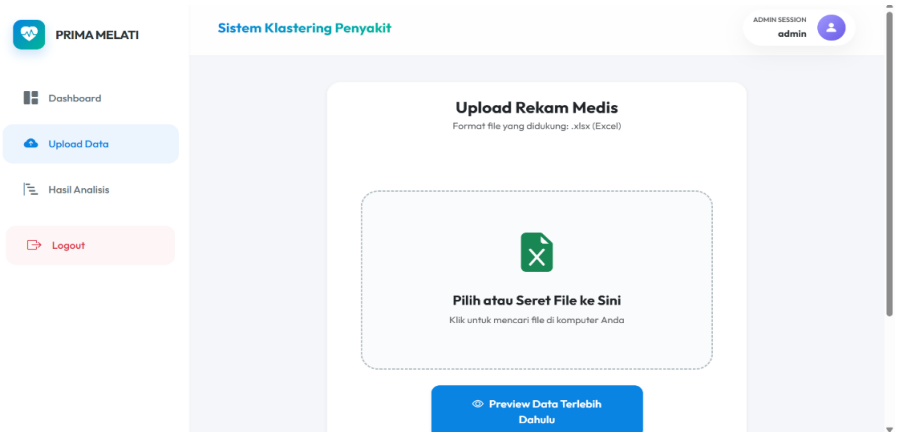

Figure 3. Upload Data Page View

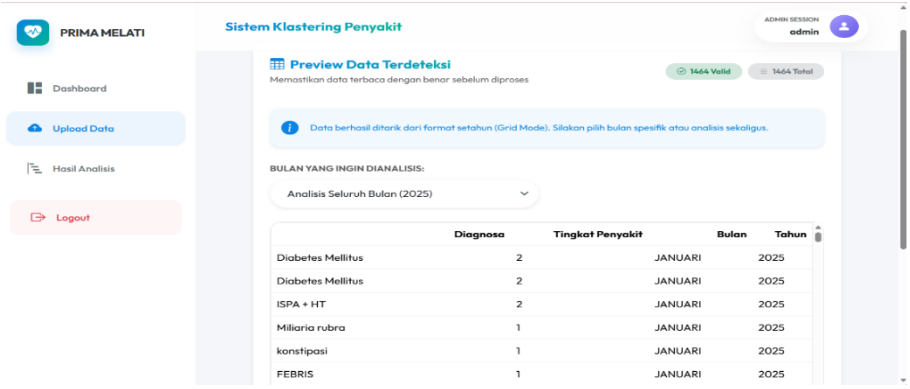

Figure 4. Data Preview Page View

https://doi.org/10.61536/ambidextrous.v2i2.499

Analysis Results Page: Clustering Results Page (K-Means Clustering): This page displays the results of the clustering analysis in the form of an Elbow Method graph, distribution of members per cluster, characteristics of each cluster, and visualization of age distribution.

Figure 5. Analysis Results Page Display

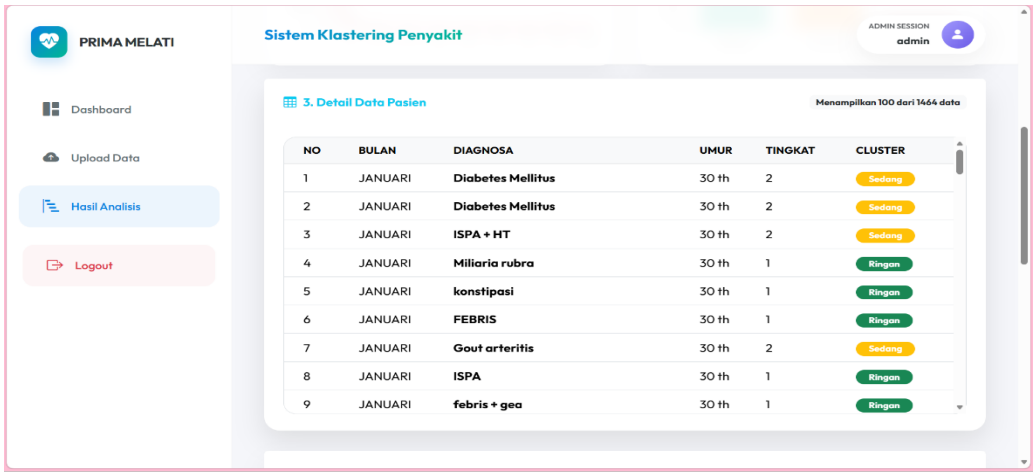

Figure 6. Patient Data Detail View

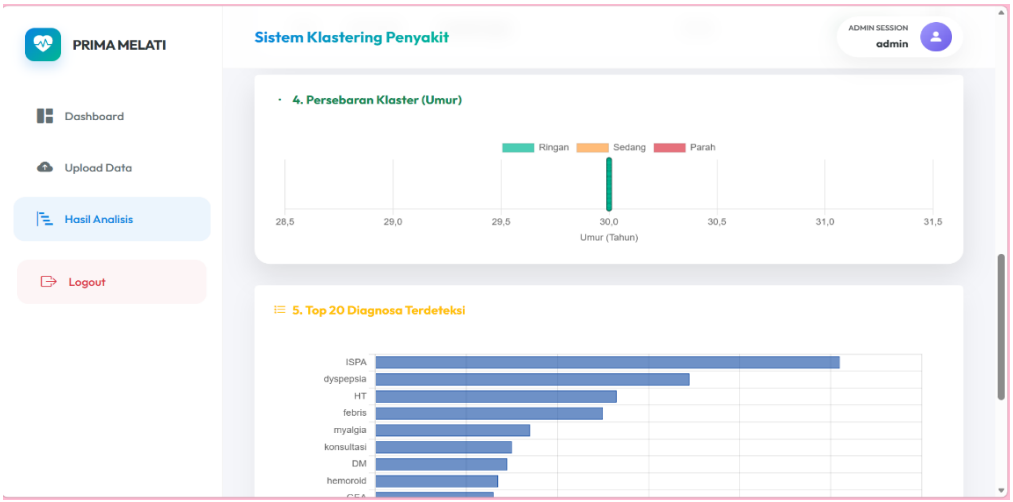

Figure 7. Display of Disease Diagnosis Distribution

Implementation of Data Processing Module and K-Means Algorithm

This module is the core of the system, which implements the algorithm stages described in

https://doi.org/10.61536/ambidextrous.v2i2.499

Chapter III:

a. Data Input Module:Responsible for reading and loading data from uploaded files.
b. Data Preprocessing Module:Implement data cleansing functions (de-duplication,*handling missing values*), variable selection, and data normalization using*StandardScaler*from Scikit-learn.
c. Cluster Number Determination Module (k):Implementing the Elbow Method and Sum of Squared Errors calculation to help Admin in determining the optimal k value.
d. K-Means Clustering Module:Using the K-Means function from Scikit-learn to do this*clustering*This process involves initialization*centroid*(e.g. with K-Means++), Euclidean distance calculation, data clustering, and updating*centroid*until convergence(Aljohani, 2024)The K-Means algorithm is effective for grouping patient medical record data to find disease spread patterns.
e. Output Module Results:Organizing the results*clustering*and present it in an easy-to-interpret format, including identification of the characteristics of each*cluster*and data visualization.

System Testing

System testing seeks to ensure that the system operates in accordance with functional and non-functional requirements, as well as ensuring the accuracy of the results obtained. This testing provides evidence that*input*, process, and*output*the system runs as desired.

Functionality Testing

The system testing in this study was conducted using the Black Box Testing method, a testing approach that emphasizes assessing system functionality without examining the internal architecture or source code. The goal of this assessment is to verify that each feature in the system operates according to the specified requirements.

This testing ensures that each implemented feature works correctly:

a. Admin Login Test:Ensure that only Admins with the correct credentials can access the system.
b. Data Upload Test:Verify that the system can accept and process various file formats (.xlsx, .csv) with varying data.
c. Data Preprocessing Test:Ensure that all preprocessing steps (cleaning, handling missing values, normalization) work according to specifications.
d. Test for Determining the Number of Clusters (k):Verifying that the Elbow Method and SSE calculations provide valid indications for the optimal number of clusters.
e. K-Means Clustering Test:Ensure that the K-Means algorithm runs without errors and produces appropriate clusters.
f. Test Display Results:Verify that clustering results visualizations and reports are displayed correctly and informatively.

Accuracy Testing and Clustering Validation

This test is the most critical for a cloud-based system.*machine learning* (Esnault et al., 2023). The accuracy of the clustering results will be evaluated to ensure that grouping patients based on the dominant disease has clinical meaning.

https://doi.org/10.61536/ambidextrous.v2i2.499

a. Internal Evaluation:Using internal metrics such as Sum of Squared Errors to assess cluster compactness or Silhouette Coefficient serves as a metric to evaluate the degree of similarity of an object to its respective cluster in relation to its similarity to alternative clusters.
b. External Validation:Where possible, clustering results can be validated by comparing them with known clinical data or expert medical judgment to ensure the relevance and validity of the dominant disease patterns found.
c. Statistical Validation:Analyze the statistical characteristics of each cluster to ensure significant and interpretable differences between them. This is important to ensure that the resulting clusters have practical meaning in a medical context.(Aljohani, 2024).
d. Model Accuracy:Accuracy can be measured by comparing clustering results with ground truth (if available) or by using metrics such as Adjusted Mutual Information and Adjusted Rand Index.

Analysis and Interpretation of Results

The K-Means Clustering algorithm has successfully identified dominant disease patterns based on patient disease severity, representing a relevant and effective approach in medical record data analysis. This section discusses the results of clustering patient data using the K-Means Clustering algorithm in more depth. Interpretation is based on the characteristics of each cluster formed, including the number of members, the dominant type of diagnosis, and the disease severity in each group.

Based on the clustering results with the number k = 3, patient data is divided into three main groups that have quite significant differences in the characteristics of disease severity, namely the mild disease group, moderate disease group, and severe disease group.

1. Interpretation of Mild Disease Cluster

The first cluster consisted of 11 patients and was dominated by low-severity illnesses. Based on these characteristics, this cluster can be interpreted as a mild illness group. The most common illnesses in this cluster were fever, dyspepsia, and observational fever. These illnesses are generally mild infections or temporary health problems that do not require intensive treatment.

This indicates that most patients in this cluster experience mild health conditions that can be managed with basic medication and simple monitoring. Therefore, this cluster represents a group of patients with relatively low disease severity. From a managerial perspective, these results suggest that clinics need to ensure the availability of medications for minor illnesses and improve basic health services. Furthermore, clinics can focus efforts on preventive measures and health education for patients to reduce the risk of recurring minor illnesses.

2. Interpretation of Moderate Disease Cluster

The second cluster has the largest number of patients, with 56 patients. Based on its characteristics, this cluster can be interpreted as a group of moderate-severity diseases. The dominant diseases in this cluster are acute respiratory infections (ARI) and dyspepsia. These two types of diseases generally require further medical treatment than milder diseases, but are still not too severe. The high number of patients in this cluster indicates that most visiting patients experience moderate-severity health conditions. This reflects the need for fairly intensive medical services, particularly in the treatment of respiratory infections and digestive disorders. From a managerial perspective, these results indicate that clinics need to prioritize

https://doi.org/10.61536/ambidextrous.v2i2.499

the availability of medical personnel and supporting facilities to treat moderate-severity diseases. Furthermore, preventive efforts such as education on healthy lifestyles and increasing public awareness of the importance of maintaining good health also need to be intensified to reduce the number of cases in this category.

3. Interpretation of Severe Disease Clusters

The third cluster consisted of 51 patients and was dominated by high-severity illnesses. Based on its characteristics, this cluster can be interpreted as a serious illness group. The most common illnesses in this cluster were diabetes mellitus, certain surgical follow-up procedures, and pulpitis. These illnesses are generally associated with chronic conditions or require more complex and ongoing medical treatment. These results indicate that patients in this cluster have more intensive healthcare needs than those in other clusters. The chronic nature of the illnesses and the need for long-term monitoring make this group a priority in healthcare services. From a managerial perspective, these results indicate that clinics need to improve their preparedness for treating serious illnesses, such as providing routine check-ups, regular patient monitoring, and the availability of competent medical personnel. Furthermore, patient recording and monitoring systems need to be optimized to support the ongoing management of chronic illnesses.

System Advantages

The system designed within this research framework offers numerous benefits, as evidenced by the implementation and evaluation results. The system's primary advantage lies in its ability to automatically integrate medical record data processing with the K-Means Clustering algorithm. Processes ranging from data upload, preprocessing, and clustering analysis can be performed within a single web-based system, simplifying data management for users without the need for additional applications. Furthermore, the system displays analysis results in informative visualizations, such as Elbow Method graphs and cluster distributions. These visualizations facilitate user comprehension, particularly in determining the optimal number of clusters and viewing the distribution of patient data based on the resulting clusters. This demonstrates that the system not only generates data but also presents it in an easily understood format.

Another advantage is the data preview feature before the analysis process. This feature allows users to double-check uploaded data to minimize input errors. This step ensures the quality of the data used in the clustering process. The system also features a download feature for analysis results in Excel format. This feature makes it easier for users to document and report analysis results. Processed data can be directly saved and used for further purposes without having to reprocess it. Furthermore, the implemented system interface demonstrates a simple and user-centric design. This is exemplified by the well-defined menu architecture, which includes dashboards, data uploads, and analytical results. With its intuitive design, the system is accessible to users without requiring specialized technical knowledge, thereby improving operational efficiency in clinical settings.

System Deficiencies

Although the system has performed well, several shortcomings were discovered during implementation and testing. One major drawback is that the system uses only one clustering method, K-Means, without any comparison with other methods. This means that the analysis

https://doi.org/10.61536/ambidextrous.v2i2.499

results cannot be validated with other approaches that might provide more optimal results. Furthermore, the system is highly dependent on the quality of the data uploaded by users. If the entered medical record data is incomplete, contains errors, or is inconsistent, the resulting clustering results can be inaccurate. This indicates that the system still requires special attention during the data preprocessing stage. Another limitation is that the system does not yet support real-time data processing. The system can only process data in the form of uploaded files, making it unsuitable for direct or continuous patient data monitoring. This presents a challenge when a system capable of operating dynamically is needed.

In terms of features, the developed system is still limited to use by one type of user, namely the admin. The system does not yet provide user management features or shared access rights, so it does not support simultaneous use by multiple roles, such as doctors or management. Furthermore, the visualizations presented in the system are still basic and not interactive. Users can only view results in static graphs without any advanced data exploration features. Therefore, further development is needed to improve the quality of the visualizations to make them more interactive and informative.

Conclusion and Recommendation

Based on the research results, it can be concluded that the application of the K-Means Clustering algorithm to patient medical record data at the Prima Melati Clinic was successful in being used foridentifydominant disease patterns based on patient age groups. The developed web-based information system is capable of structured data processing, starting from data input, preprocessing, determining the optimal number of clusters using the Elbow method, to presenting the analysis results in an easy-to-understand visualization. The clustering results show the formation of three main patient groups: early age, adulthood, and the elderly, each with different disease characteristics. The early age group is dominated by mild illnesses, the adult group tends to experience infectious diseases and digestive disorders, while the elderly group experiences more chronic illnesses. These results indicate a relationship between patient age and the type of illness experienced. In addition, the use of the K-Means Clustering algorithm is considered effective in assisting the process of segmenting medical record data more systematically, quickly, and accurately compared to manual methods, thereby supporting decision-making in improving health services at the Prima Melati Clinic.

Based on the research findings, it is recommended that Prima Melati Clinic utilize this clustering system sustainably to support patient data management and more targeted healthcare planning. System development can also be done by adding more research variables, such as gender, medical history, patient visit patterns, and environmental factors, to provide more detailed and accurate clustering results. Furthermore, further research is expected to compare the K-Means algorithm with other data mining methods to obtain more optimal clustering results. Improving the quality and completeness of medical record data is also necessary so that the analysis process produces more valid and useful information to support decision-making in the healthcare sector.

https://doi.org/10.61536/ambidextrous.v2i2.499

References

Aljohani, N. R. (2024). *Machine learning clustering techniques in healthcare analytics: A comprehensive review*. *Healthcare Analytics, 6*, 100412.

Eken, S. (2020). A data mining approach for healthcare data classification and clustering. *Journal of Healthcare Engineering, 2020*, 1–12.

Fay, M. P., Smith, J. A., & Brown, T. R. (2023). Data visualization techniques for healthcare analytics and machine learning applications. *Applied Sciences, 13*(5), Article 2874. https://doi.org/10.3390/app13052874

Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Kabir, M. A., Rahman, M. S., & Islam, M. T. (2024). Python-based machine learning applications in healthcare analytics: A systematic review. *Healthcare Analytics, 5*, 100321.

Kementerian Kesehatan Republik Indonesia. (2022). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang rekam medis*. Kementerian Kesehatan Republik Indonesia.

Kruse, C. S., Stein, A., Thomas, H., & Kaur, H. (2018). The use of electronic health records to support population health: A systematic review of the literature. *Journal of Medical Systems, 42*(11), Article 214. https://doi.org/10.1007/s10916-018-1075-6

Sharma, M., Singh, G., & Singh, R. (2021). Big data analytics in healthcare: A systematic literature review. *Journal of Big Data, 8*(1), 1–24.

Tsai, C. H., Eghdam, A., Davoody, N., Wright, G., Flowerday, S., & Koch, S. (2020). Effects of electronic health record implementation and barriers to adoption and use: A scoping review. *JMIR Medical Informatics, 8*(11), e19165. https://doi.org/10.2196/19165

World Health Organization. (2021). *Global strategy on digital health 2020–2025*. World Health Organization.