

Analysis of Diabetes Mellitus Patient Medical Records through Data Mining Implementation at Hospital Aloe Saboe Gorontalo

Gladis A. Ismail^{1*}, Mohamad Reza Moontuno² Merlin Abd. Rahman³

^{1,3} STIKES Bakti Nusantara Gorontalo, Gorontalo, Indonesia

² BPJS Kesehatan Gorontalo, Gorontalo, Indonesia

ORCID/Email:

First AUTHOR : - / gladisismail.gi@gmail.com

Second AUTHOR : - / rmohamad.moontuno@gmail.com

Third AUTHOR : - / merlind.rahman@gmail.com

Corresponding author email: gladisismail.gi@gmail.com

Article Info

Received: 10/02/2026

Revised: 23/04/2026

Accepted: 16/05/2026

Online: 30/06/2026

Abstract

This study aims to analyze the medical record data of Diabetes Mellitus patients through the application of data mining techniques to uncover patterns and information that can support decision-making in hospitals. The data used are secondary data derived from patient medical records at Hospital Aloe Saboe Gorontalo, encompassing variables such as age, sex, blood glucose level, disease history, and type of treatment. This research employs a quantitative approach, with the analysis following the Knowledge Discovery in Databases (KDD) process, which includes data selection, data cleaning, data transformation, the data mining process, and evaluation of the results. The data mining techniques applied consist of a classification method using Decision Tree and a clustering method using K-Means. The results show that the clustering technique was able to group patients according to similar characteristics, such as condition severity and risk of complications, while the classification method was able to predict patients' risk levels with a fairly good level of accuracy. Variables such as age, blood glucose level, and disease history were found to have a significant influence on patients' condition. These findings indicate a consistent pattern between risk factors and the severity of Diabetes Mellitus. In conclusion, applying data mining techniques to Diabetes Mellitus patients' medical records can generate useful information and support improved quality of healthcare services. The methods used proved effective in processing large volumes of data and in supporting more accurate and efficient decision-making.

Keywords: data mining, medical records, diabetes mellitus, classification, clustering



© 2026 by the author(s)

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Diabetes Mellitus is one of the chronic diseases that has become a major concern in global health (World Health Organization, 2021; International Diabetes Federation, 2021). This disease is characterized by elevated blood glucose levels resulting from impaired insulin secretion, impaired insulin action, or both. The rising prevalence of Diabetes Mellitus from year to year shows that this disease is not only an individual health problem, but has also become a burden on the healthcare service system (Ogurtsova et al., 2022; Wahidin et al., 2024). In Indonesia, the number of Diabetes Mellitus patients continues to increase in line with changes in lifestyle, dietary patterns, and a lack of physical activity in the population (Kementerian Kesehatan Republik Indonesia, 2020; Wahidin et al., 2024). This condition demands more effective efforts in the management and prevention of the disease.

Hospitals, as healthcare institutions, play an important role in the treatment of Diabetes Mellitus patients. Every healthcare service activity provided to a patient generates data, particularly in the form of medical records. Medical records contain complete information regarding a patient's health history, diagnosis, medical procedures, laboratory examination results, and the therapy given (Kementerian Kesehatan Republik Indonesia, 2022). Along with the development of information technology, the volume of medical record data stored in hospitals continues to increase and has become a highly valuable data source.

Nevertheless, the large volume of stored medical record data is often not optimally utilized. Such data is generally only used as an administrative archive and service documentation, without further analysis to generate valuable information. In fact, if medical record data can be processed properly, it can provide important insights related to disease patterns, risk factors, treatment effectiveness, and predictions of a patient's future condition.

One approach that can be used to process large volumes of data is the data mining technique. Data mining is the process of extracting hidden information or knowledge from a large collection of data through statistical, mathematical, and artificial intelligence methods (Han et al., 2012; Larose & Larose, 2014). This technique is able to uncover patterns, relationships, and tendencies that are not directly visible (Witten et al., 2016; Tan et al., 2019). In the context of healthcare, data mining can be utilized to support clinical decision-making, improve service quality, and support health policy planning (Turban et al., 2011).

The application of data mining in the analysis of Diabetes Mellitus patients' medical records holds significant potential. Using this technique, patients can be grouped based on specific characteristics, factors influencing disease severity can be identified, and the risk of complications can be predicted (Novaliendry et al., 2023; Sapitri et al., 2025). In addition, the results of the analysis can help medical personnel determine more appropriate and efficient treatment strategies. Thus, the utilization of data mining not only increases the usability value of medical record data, but also contributes to improving the overall quality of healthcare services.

Nevertheless, the implementation of data mining techniques in the healthcare field still faces several challenges. One of them is data quality that is not always complete or consistent. In addition, the limited number of human resources with competence in data processing is also an obstacle. On the other hand, aspects of patient data privacy and security must also be a primary concern in every data processing step (Oturugbum, 2023). Therefore, a systematic and structured approach is needed in applying data mining to medical record data.

Based on this background, this study focuses on the analysis of Diabetes Mellitus patients' medical records through the implementation of data mining techniques at Hospital Aloe Saboe Gorontalo. This research was conducted to explore the information contained in medical record data, so as to provide useful insight for the hospital in improving the quality of service and patient management. By utilizing appropriate data mining techniques, it is hoped that important patterns can be found that can be used as a basis for decision-making.

The problems examined in this study include the characteristics of Diabetes Mellitus patients' medical record data at Hospital Aloe Saboe Gorontalo, how data mining techniques are applied in processing such data, and what information can be generated from the analysis process. These issues are important to study given the substantial potential of medical record data that has not been optimally utilized.

The purpose of this study is to analyze Diabetes Mellitus patients' medical record data using data mining techniques in order to discover useful patterns and information. In addition, this study also aims to provide recommendations for the utilization of medical record data as a basis for decision-making within the hospital environment. Thus, the results of this study are expected to contribute both theoretically and practically to the development of data mining applications in the field of healthcare.

RESEARCH METHOD

This study uses a quantitative approach with an experimental method and secondary data analysis sourced from the medical records of Diabetes Mellitus patients at Hospital Aloe Saboe Gorontalo. The research design used is descriptive-analytic, utilizing data mining techniques to explore patterns and information from the available data. This study aims to identify patient characteristics and discover patterns that can be used as a basis for decision-making in healthcare services.

Research Object and Data

The object of this study is the medical record data of patients who have been diagnosed with Diabetes Mellitus within a specific period at Hospital Aloe Saboe Gorontalo. The data used includes variables such as age, sex, blood glucose level, disease history, laboratory examination results, and the type of treatment given. The data used are secondary data that have undergone an anonymization process to protect patient identity confidentiality.

Data Collection Procedure

The data collection process was carried out through documentation, namely by retrieving medical record data from the hospital information system with permission obtained from the relevant parties. The data collection stages include selecting data relevant to the research, followed by data cleaning to address incomplete, duplicate, or inconsistent data. After that, the data was transformed into a format suitable for further analysis.

Data Mining Techniques

This study employs data mining techniques comprising classification and clustering methods. The classification method is used to predict a specific category, such as the level of risk or likelihood of complications in patients, while the clustering method is used to group patients based on the similarity of their characteristics. The algorithms used include Decision Tree for classification (Perveen et al., 2016; Shojaee-Mend et al., 2024) and K-Means for clustering (Novaliendry et al., 2023; Sapitri et al., 2025). These methods were selected based on their ability to process large volumes of data and produce information that is easy to interpret.

Data Analysis Technique

The data analysis stages refer to the Knowledge Discovery in Databases (KDD) process, which consists of several steps, namely: (1) data selection, (2) preprocessing or data cleaning, (3) data transformation, (4) the data mining process, and (5) evaluation and interpretation of results (Han et al., 2012; Larose & Larose, 2014). Model evaluation was carried out using parameters such as accuracy, precision, and recall for the classification method, as well as distance measurement and group homogeneity for the clustering method.

Tools

The equipment used in this study includes hardware in the form of a computer or laptop with specifications that support data processing, as well as software such as data processing applications and data mining tools, for example RapidMiner or Python with supporting libraries such as pandas, scikit-learn, and matplotlib. The use of this software is intended to facilitate the analysis and data visualization process.

Data Quality Control

Quality control in this study was carried out by ensuring the quality of the data used, such as performing data validation and avoiding bias in sample selection. In addition, repeated model testing was conducted to ensure the consistency of the results obtained. Research ethics were also considered by maintaining the confidentiality of patient data and not including personally identifiable information. With this systematic and structured research method, it is hoped that the results of the study can provide a clear picture of the medical record data patterns of Diabetes Mellitus patients and generate information that can be utilized to improve the quality of healthcare services at Hospital Aloe Saboe Gorontalo.

RESULTS AND DISCUSSION

This section presents the results of the analysis of Diabetes Mellitus patients' medical record data at Hospital Aloe Saboe Gorontalo using data mining techniques, along with a discussion linking the research results to relevant theory and prior research. This study uses classification and clustering methods to explore patterns in the available data.

Description of Research Data

The data used in this study are the medical records of Diabetes Mellitus patients obtained from Hospital Aloe Saboe Gorontalo within a specific period. After going through the selection and preprocessing stages, a clean dataset ready for analysis was obtained. The variables used in this study include age, sex, blood glucose level, disease history, and the type of treatment given. The results of the data description show that most patients fall within the adult to elderly age range. This is consistent with the theory stating that the risk of Diabetes Mellitus, particularly type 2, increases with age (Sadiq et al., 2025; Wahidin et al., 2024). In addition, it was found that patients with certain disease histories, such as hypertension, have a higher tendency toward complications.

Results of the Clustering Method Application

In the clustering stage, the K-Means algorithm was used to group patients based on the similarity of their characteristics. The grouping results show that the data can be divided into several main clusters, for example: Cluster 1: Patients with high blood glucose levels and a history of complications; Cluster 2: Patients with moderate blood glucose levels without significant complications; and Cluster 3: Patients with a relatively stable and controlled condition

These results indicate a clear pattern in the patient data. This grouping makes it easier for medical personnel to identify patient groups based on disease severity level. Theoretically, clustering is indeed used to discover hidden structures in unlabeled data, and the results of this study support that concept.

Compared with previous research, these results share similarities in that clustering techniques are able to group patients based on clinical characteristics (Novaliendry et al., 2023; Sapitri et al., 2025; Aulia et al., 2024). However, the difference lies in the number of clusters and the variables used, which were adjusted to the condition of the data at Hospital Aloe Saboe Gorontalo

Results of the Classification Method Application

The classification method uses the Decision Tree algorithm to predict the level of complication risk in Diabetes Mellitus patients. Based on the analysis results, the model built is able to classify patients into specific risk categories, namely low, moderate, and high risk.

The variables that most influence the classification process include blood glucose level, age, and disease history. This is consistent with medical theory, which states that these factors are important indicators in determining the condition of a Diabetes Mellitus patient (Sadiq et al., 2025; Lai et al., 2025). The model evaluation results show a fairly good level of accuracy, so the model can be used as a decision-support tool. Compared with previous research using similar algorithms, the results of this study show consistency in that Decision Tree is effective in classifying health data because of its ability to produce rules that are easy to understand (Perveen et al., 2016; Shojaee-Mend et al., 2024; Olusanya et al., 2022). However, there are differences in the level of accuracy obtained, which may be caused by differences in data quality, data volume, and the variables used in the study.

Pattern Analysis and Research Findings

Based on the clustering and classification results, several important patterns were found, including: Patients with high blood glucose levels tend to have a greater risk of complications; Older age is correlated with a higher level of disease severity; and A history of comorbidities such as hypertension worsens a patient's condition

These findings are consistent with existing theory in the field of health, which states that Diabetes Mellitus is influenced by various risk factors, both internal and external (Wahidin et al., 2024; Lai et al., 2025). Thus, the results of this study reinforce existing theory while also providing empirical evidence based on medical record data at Hospital Aloe Saboe Gorontalo. Compared with previous research, there are similarities in terms of the factors that influence a patient's condition. However, this study provides additional value by directly integrating data mining techniques into medical record data, resulting in a more data-driven analysis.

Implications of the Research Results

The results of this study have several important implications, both practical and theoretical. Practically, the results of the analysis can be used by the hospital to: Identify high-risk patients more quickly; Develop more appropriate treatment strategies based on patient groups; and Improve the efficiency of healthcare services. Theoretically, this study shows that the application of data mining techniques in the field of healthcare, particularly on medical record data, can make a significant contribution to the development of scientific knowledge.

Limitations of the Study

Although this study provides fairly good results, there are several limitations that need to be noted, including: Limitations in the volume and quality of available data; The possibility of incomplete or inconsistent data; and The use of algorithms that are still limited to a few methods

These limitations may affect the results of the analysis and should be taken into consideration for future research.

Comparison with Previous Research

In general, the results of this study are consistent with previous research stating that data mining is effective in analyzing health data (Perveen et al., 2016; Olusanya et al., 2022; Shojaee-Mend et al., 2024). The similarity found lies in the ability of classification and clustering techniques to discover patterns and make predictions. The differences lie in the research context, the type of data used, and the methods applied. This study focuses more specifically on medical record data at Hospital Aloe Saboe Gorontalo, thereby providing a more specific picture of local conditions. Based on the results and discussion presented, it can be concluded that the application of data mining techniques to the medical record data of Diabetes Mellitus patients is able to generate useful information. These results not only support existing theory, but also provide a real contribution to improving the quality of healthcare services.

CONCLUSION

Based on the results of the research conducted on the analysis of Diabetes Mellitus patients' medical records through the implementation of data mining techniques at Hospital Aloe Saboe Gorontalo, it can be concluded that the application of data mining methods, specifically the classification algorithm (Decision Tree) and clustering (K-Means), is able to process medical record data into more meaningful information. The research results show a significant pattern between variables such as age, blood glucose level, and disease history in relation to a patient's level of complication risk.

The hypothesis of this study, which states that data mining techniques can be used to discover patterns and assist in decision-making using Diabetes Mellitus patients' medical record data, was confirmed. The classification method was able to predict patients' risk levels with a fairly good level of accuracy, while the clustering method successfully grouped patients based on similar characteristics. Thus, data mining techniques proved effective in supporting the analysis of health data, particularly patient medical records (Olusanya et al., 2022; Sadiq et al., 2025).

The advantages of the method used in this study lie in its ability to process large volumes of data, discover hidden patterns, and produce information that is easy to understand and can be used as a basis for decision-making. In addition, the algorithms used are relatively simple and efficient to implement.

Nevertheless, this study also has several shortcomings, including its dependence on the quality of the data used, limitations in the number of variables analyzed, and the use of algorithms that are still limited. In addition, the analysis results may also be influenced by incomplete or inconsistent data in the medical records.

Based on the results of this study, it is recommended that future research use a larger and more diverse dataset, as well as combine several data mining algorithms to improve the accuracy of the analysis results. In addition, an integrated data-mining-based system could be developed in conjunction with the hospital information system to support real-time decision-making.

The potential for developing data mining methods in the field of healthcare remains very broad, particularly in the analysis of chronic diseases such as Diabetes Mellitus. With the continued advancement of technology, it is hoped that this method can make a greater contribution to improving the quality of healthcare services and help medical personnel provide more precise and effective treatment.

ACKNOWLEDGMENTS

The authors would like to thank the management of Hospital Aloe Saboe Gorontalo for granting permission and access to the patient medical record data used in this study.

AUTHOR CONTRIBUTIONS

Conceptualization, G.A.I. and M.R.M.; Methodology, G.A.I.; Software, G.A.I.; Validation, G.A.I. and M.A.R.; Formal Analysis, G.A.I.; Investigation, G.A.I. and M.R.M.; Resources, M.R.M.; Data Curation, G.A.I.; Writing – Original Draft Preparation, G.A.I.; Writing – Review & Editing, G.A.I., M.R.M. and M.A.R.; Supervision, M.R.M.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Aulia, W., Siahaan, A. P. U., Marlina, L., Khairul, K., & Iqbal, M. (2024). K-Means clustering algorithm analysis for grouping patient medical record data based on disease type. *Jurnal Info Sains: Informatika dan Sains*, 14(4), 832–843.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- International Diabetes Federation. (2021). *IDF Diabetes Atlas* (10th ed.). International Diabetes Federation.
- Kementerian Kesehatan Republik Indonesia. (2020). *Profil Kesehatan Indonesia 2019*. Kemenkes RI.
- Kementerian Kesehatan Republik Indonesia. (2022). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis*. Kementerian Kesehatan RI.
- Lai, J., Hao, M., Huang, X., Chen, S., Yan, D., & Li, H. (2025). Novel model predicts type 2 diabetes mellitus patients complicated with metabolic syndrome using retrospective dataset from First Affiliated Hospital of Shenzhen University, China. *International Journal of Endocrinology*, 2025, Article 9558141.
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (2nd ed.). Wiley.
- Novaliendry, D., Wibowo, T., Ardi, N., Evi, T., & Admojo, D. (2023). Optimizing patient medical records grouping through data mining and K-Means clustering algorithm: A case study at

- RSUD Mohammad Natsir Solok. *International Journal of Online and Biomedical Engineering*, 19(12), 144–155.
- Ogurtsova, K., Guariguata, L., Barengo, N. C., Ruiz, P. L.-D., Sacre, J. W., Karuranga, S., Sun, H., Boyko, E. J., & Magliano, D. J. (2022). IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. *Diabetes Research and Clinical Practice*, 183, 109118.
- Olusanya, M. O., Ogunsakin, R. E., Ghai, M., & Adeleke, M. A. (2022). Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: A systematic survey and meta-analysis approach. *International Journal of Environmental Research and Public Health*, 19(21), 14280.
- Oturugbum, R. O. (2023). Preserving the safety and confidentiality of data mining information in health care: A literature review. *arXiv preprint arXiv:2312.00016*.
- Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, 82, 115–121.
- Sadiq, I. Z., Katsayal, B. S., Ibrahim, B., Ibrahim, M., Hassan, H. A., Ghali, U. M., Usman, A. G., Usman, A., & Abba, S. I. (2025). Data-driven diabetes mellitus prediction and management: A comparative evaluation of decision tree classifier and artificial neural network models along with statistical analysis. *Scientific Reports*, 15, 19339.
- Sapitri, A., Nurdin, N., & Afrilia, Y. (2025). Implementation of clustering method using K-Means algorithm for grouping BPJS Health patient medical record data. *Journal of Applied Informatics and Computing*, 9(5), 2391–2398.
- Shojaee-Mend, H., Velayati, F., Tayefi, B., & Babae, E. (2024). Prediction of diabetes using data mining and machine learning algorithms: A cross-sectional study. *Healthcare Informatics Research*, 30(1), 73–82.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). Pearson.
- Turban, E., Sharda, R., Delen, D., & King, D. (2011). *Business Intelligence: A Managerial Approach* (2nd ed.). Pearson.
- Wahidin, M., Achadi, A., Besral, B., Kosen, S., Nadjib, M., Nurwahyuni, A., Ronoatmodjo, S., Rahajeng, E., Pane, M., & Kusuma, D. (2024). Projection of diabetes morbidity and mortality till 2045 in Indonesia based on risk factors and NCD prevention and control programs. *Scientific Reports*, 14(1), 5424.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.
- World Health Organization. (2021). *Diabetes*. Geneva: WHO. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/diabetes>