Ambidextrous: Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 5, no. 2, 2026, pp. 113-124
E-ISSN: 3031-7002
E-mail:ambidex@takaza.id

Referral Hospital Recommendation System for Patients Based on Diagnosis Data and Severity Level Using the Random Forest Method at Puskesmas Sigambal

Zahra Amanda

Muhammadiyah University of North Sumatra
Corresponding Author e-mail: zahraamnda254@gmail.com

Article History:
Received: 24-05-2026
Revised: 11-08-2026
Accepted: 11-08-2026

Keywords: *Classification; Diagnosis; Random Forest; Referral Hospital; Severity Level*

Abstract: *The process of determining appropriate referral hospitals at primary healthcare facilities requires accurate and systematic consideration of patient conditions and historical referral patterns. However, referral decisions may still rely on subjective judgment and operational practices, limiting the utilization of available patient data. This study aims to develop a referral hospital recommendation system based on patient diagnosis and severity level using the Random Forest classification method at Puskesmas Sigambal, Labuhanbatu Regency. The study employed a quantitative approach using 120 historical patient referral records containing age, gender, diagnosis, diagnosis category, severity level, and referral hospital as the target variable. The research process included data preprocessing, categorical data transformation, dataset partitioning into training and testing sets using an 80:20 ratio, Random Forest model development, and performance evaluation using a confusion matrix. The evaluation results showed an accuracy of 91.4%, precision of 89.8%, recall of 90.5%, and F1-score of 90.1%. These results indicate that the Random Forest model demonstrated promising performance in identifying referral hospital patterns based on patient characteristics, particularly diagnosis and severity level. The model was subsequently implemented in a web-based recommendation system using PHP and MySQL to support healthcare personnel in obtaining referral recommendations efficiently and systematically. The proposed system is intended as a decision-support tool and does not replace clinical judgment or applicable referral policies. The findings demonstrate the potential of machine learning to support more systematic, data-driven, and consistent referral decision-making at the primary healthcare level.*

How to Cite: Zahra Amanda. (2026). The title.*Ambidextrous Journal of Innovation Efficiency and*

Technology in Organizations. 5(2). Pp 113-124 https://doi.org/10.61536/ambidextrous.v5i02.510
https://doi.org/10.61536/ambidextrous.v5i02.510

Introduction

Advances in information technology have brought about significant changes in the way organizations manage and utilize data. Data, which previously served only as archives, is now viewed as a strategic source of knowledge. Through a data mining approach, data can be processed to discover hidden patterns, relationships, and information that are invisible through conventional analysis. Data mining enables systematic knowledge extraction by utilizing statistical and computational techniques.(Kusrini & Luthfi, 2009);(Syahra, 2023);(Han, Kamber, & Pei, 2012). As machine learning develops, various analytical methods are used to build predictive models, one of which is through classification techniques.(Prasetyo, 2014).

In the context of information management, classification is a technique for grouping data into specific categories based on their attributes. One classification algorithm known for its superior performance is Random Forest. This algorithm works by combining several decision trees through an ensemble learning mechanism to improve prediction accuracy and reduce the risk of overfitting. Random Forest is also considered highly reliable in handling medical datasets with many categorical variables, is tolerant of noise, and produces stable models.(Prasetyo, 2014).

At primary healthcare institutions like the Sigambal Community Health Center (Puskesmas), patient data is routinely collected through administrative processes. This data includes crucial attributes such as disease diagnoses and the severity of the patient's condition. However, observations indicate that data utilization at the Sigambal Community Health Center is currently limited to administrative documentation and reporting. The predictive potential of historical patient data has not been optimally utilized to support clinical and operational decision-making.

The main problem arises in the activity of determining referral hospital recommendations. This process plays a vital role in ensuring patients receive appropriate and prompt follow-up care. However, in practice at the Sigambal Community Health Center, referral decisions are often based on subjective considerations, staff intuition, or operational habits without any historical data-based guidance. This subjective approach has the potential to lead to inconsistent referral decisions, which can ultimately lead to a backlog of patients at certain hospitals (capacity mismatch) or a mismatch between the destination hospital's facilities and the patient's specific needs.

In fact, variables such as diagnosis and severity are strong predictors in classification models. By utilizing machine learning techniques, historical patient data can be analyzed to identify patterns of association between patient characteristics and the most effective referral decisions. Using Random Forest in this case is the right solution because this algorithm can process diverse diagnostic variables and provide decision results based on majority voting from multiple decision trees, resulting in more objective and accurate recommendations.

Although extensive research on Random Forest has been conducted, its specific

application to referral recommendation systems at the community health center level, integrating severity variables, remains relatively limited. This situation indicates a gap between data availability and processing technology in primary healthcare facilities. Based on this description, an analytical approach is needed that can optimize the use of patient data to provide precise referral information.

Research Methods

The research method used in this study is a quantitative research method with a data mining approach using the Random Forest algorithm. The study was conducted at the Sigambal Community Health Center using historical patient referral data as the main data source. The research stage began with a problem analysis to identify the process of determining referral hospitals that is still done manually and not based on data. Next, patient referral data was collected consisting of attributes of age, gender, disease diagnosis, diagnostic category, severity level, and referral destination hospital. The obtained data then went through a preprocessing or data cleaning stage by selecting relevant attributes, deleting unused data, grouping disease diagnoses, and transforming categorical data into numeric form so that it can be processed by the Random Forest algorithm.

After the preprocessing process was completed, the dataset was divided into training data and testing data with a ratio of 80:20 to build and test the classification model. The Random Forest algorithm was used with the stages of bootstrap sampling, decision tree formation, attribute selection using the Gini Index, classification process, and majority voting to generate referral hospital recommendations. The system was then designed and implemented web-based using PHP and MySQL so that it can be used by community health center staff to input patient data and obtain recommendations automatically. Next, the model performance was evaluated using a confusion matrix with parameters of accuracy, precision, recall, and F1-score to determine the level of model performance in classifying referral hospitals based on diagnostic data and patient severity.

Results and Discussion

1. System Requirements

System requirements are the hardware and software specifications required to support the development and implementation process of information systems.(Pressman, 2010)System requirements are fundamental to software development because they determine the operational environment in which the system will run. Clearly identifying these requirements allows the system to run optimally and according to the design. In this study, system requirements are divided into two parts: hardware requirements and software requirements.

a. Hardware Requirements**)**

Hardware is the physical components used to run the system. According to(Stair & Reynolds, 2018), hardware is all the physical equipment in a computer system that functions to process, store, and display information.

The hardware used in this research is as follows:

https://doi.org/10.61536/ambidextrous.v5i02.510

1) Laptop/Computer with minimum specifications:
   a) Intel Core i3 processor or equivalent
   b) Minimum RAM 4 GB
   c) Storage media (Hard disk/SSD) minimum 500 GB
2) Monitor as an output device
3) Keyboard and mouse as input devices

These specifications are sufficient to run local servers and web-based systems.web developed in this research.

b. Software Requirements

Software is a collection of instructions or programs used to operate hardware. According to(Sommerville, 2011), software is a computer program along with the documentation and configuration data needed for the system to function properly.

The software used in this research includes:

1) Operating System: Windows 10
2) Web Server: XAMPP (Apache and MySQL)
3) Programming Language: PHP
4) Database Management System: MySQL
5) Web Browser: Google Chrome
6) Code Editor: Visual Studio Code

This web-based system uses a client-server architecture, where a local server (localhost) serves as the development and testing environment. The Random Forest algorithm is implemented as a classification method to generate referral hospital recommendations based on patient data. With the support of this hardware and software, the system can run optimally and meet user needs.

2. System Implementation

System implementation is the stage of applying the design results described in Chapter III into a workable system. At this stage, the system is developed using the PHP programming language with a MySQL database and run on a local server using XAMPP. This Referral Hospital Recommendation System is built web-based so it can be accessed through a browser at the localhost address. Implementation is carried out based on the interface design, system structure, and the Random Forest algorithm that have been previously determined. The following is the implementation of each system page.

a. Login Page Implementation

The login page is the system's main gateway, serving as a security mechanism to restrict user access. During implementation, this page is built using an HTML form connected to a PHP script for data validation. When a user enters a username and password, the system performs an authentication process by comparing the input data with the data stored in the admin table in the database. If the data matches, the system creates a login session and redirects the user to the dashboard. Conversely, if the data does not match, the system displays an error notification. The use of a session mechanism aims to maintain access security while users are using the system. The implementation of this login page ensures that only authorized users can operate the recommendation system.

https://doi.org/10.61536/ambidextrous.v5i02.510

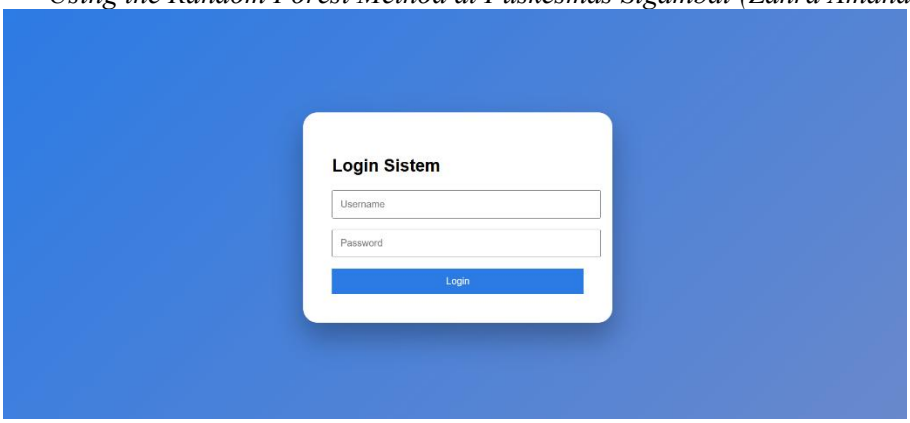


Figure 1. Login Page Implementation

b. Dashboard Page Implementation

The dashboard serves as the system's information and navigation center. During the implementation phase, the dashboard was designed to display a real-time summary of referral data stored in the database. The data displayed on the dashboard is obtained through a MySQL query that calculates the total number of referrals and the distribution of patient severity levels. This information is then visualized in the form of information cards and bar graphs to facilitate data interpretation. Furthermore, the dashboard provides a navigation menu to the data input page, the model evaluation page, and a logout feature. A date filter feature allows users to search for data based on a specific period. The dashboard implementation aims to provide an overview of the referral data status and facilitate user access to system features.

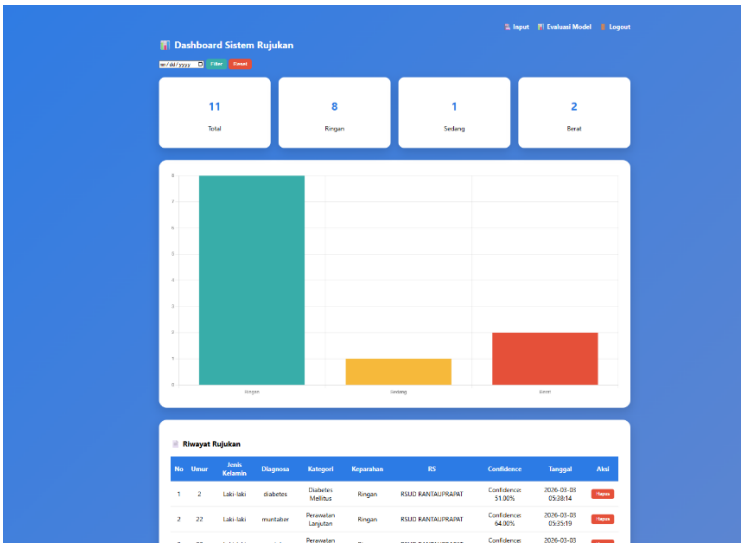


Figure 2. Implementation of the Dashboard Page

c. Implementation of Data Input Page

The data input page is the main part of the system used to enter patient data to be processed by the classification model. During the implementation phase, the input form is built using HTML form elements consisting of fields for age, gender, diagnosis, and severity. Before processing the data, the system performs validation to ensure all fields have been filled in correctly. If the data is complete, the system encodes categorical data such as diagnosis and severity to conform to the format used in the Random Forest model. After the validation and transformation processes are complete, the system sends the data to the classification module

https://doi.org/10.61536/ambidextrous.v5i02.510

for prediction. The implementation of this page serves as the primary interface between the user and the machine learning algorithm used in the system.

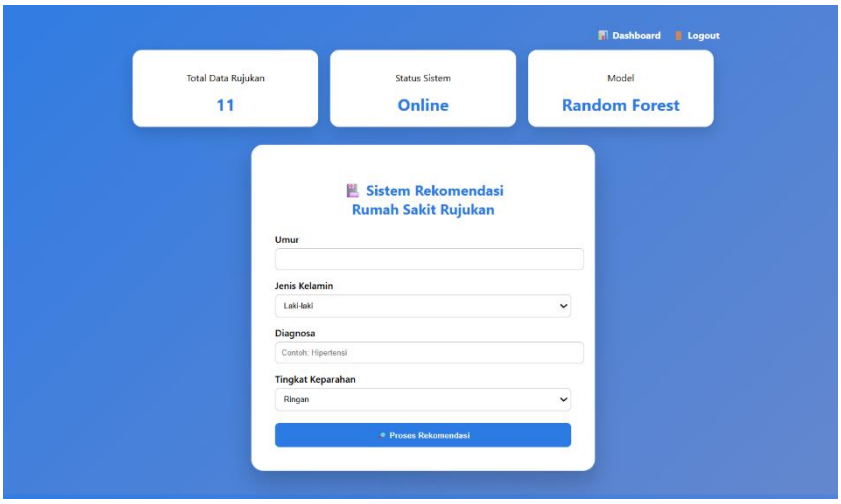


Figure 3. Implementation of the Data Input Page

d. Implementation of the Recommendation Results Page

The recommendation results page displays the output generated by the Random Forest model based on the entered patient data. During the implementation phase, the system receives the predicted results in the form of referral hospital classes and probability values from the model. These results are then displayed as follows:

1) Patient details information
2) Name of recommended hospital
3) Confidence value in percentage form
4) Alternative hospitals and their probabilities

The confidence value is visualized as a progress bar to facilitate user understanding of the model's confidence level. Furthermore, the recommendation results are stored in a database as a reference history for reporting and monitoring purposes. The implementation of this page demonstrates the system's ability to generate reference decisions automatically and based on data.

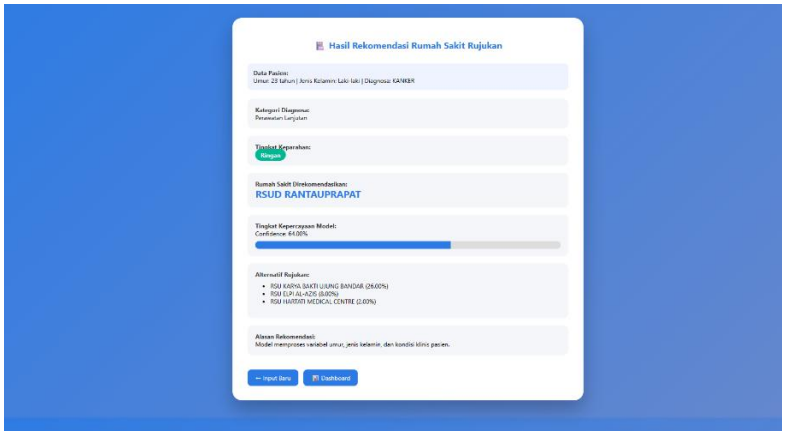


Figure 4. Implementation of the Recommendation Results Page

e. Model Evaluation Page Implementation

The model evaluation page displays the performance of the Random Forest algorithm used in the system. During the implementation phase, the system displays model test results in

https://doi.org/10.61536/ambidextrous.v5i02.510

the form of accuracy, precision, recall, and F1-score values. Evaluation values are obtained from the testing process using previously separated test data. In addition to being displayed numerically, the evaluation metrics are also visualized graphically for ease of interpretation. The implementation of this evaluation page aims to demonstrate that the system not only generates recommendations but is also supported by a measurable machine learning model with a scientifically verifiable level of accuracy.

Figure 5. Implementation of the Model Evaluation Page

3. System Testing

System testing was conducted to ensure that all functions within the Referral Hospital Recommendation System functioned as expected. This testing aimed to demonstrate that the input process, data processing using the Random Forest algorithm, and output in the form of referral hospital recommendations functioned correctly and consistently.(Pressman, 2010)Software testing aims to find errors and ensure that the system meets predetermined functional requirements. In this study, the testing method used was black box testing, which is testing based on system functions without directly viewing the program code.

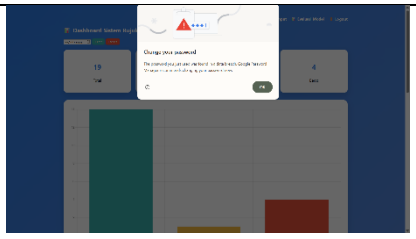
a. System Functional Testing

Functional testing is carried out to ensure that each system feature can run according to design.

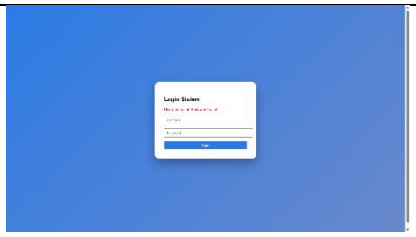
1) Login Testing

Login feature testing is carried out to ensure that the system is able to validate user data before accessing the dashboard and to deny access if the data entered is incorrect.

Table 1. Login Feature Testing

NO	Testing Scenario	Input	Expected Output	Results
1	Login with correct data	Valid Username & Password		Succeed

https://doi.org/10.61536/ambidextrous.v5i02.510

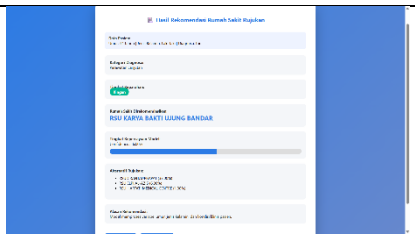
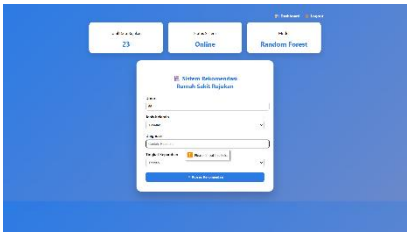
2	Login with incorrect data	Incorrect username/password		Succeed

Based on the test results in Table 1, the system successfully performed the authentication process. The system successfully redirected users to the dashboard when the data was correct and displayed an error message when the data was inaccurate.

2) Patient Data Input Testing

This testing is carried out to ensure that the system can receive patient data correctly and validate the completeness of the data before it is processed by the algorithm.

Table 2. Patient Data Input Testing

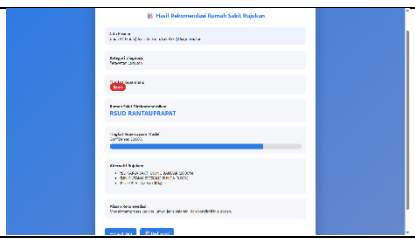
NO	Testing Scenario	Input	Expected Output	Results
1	All data is filled in completely	Age, JK, Diagnosis, Severity		Succeed
2	One of the fields is empty	Diagnosis not filled in		Succeed

Based on the test results in Table 4.2, the system is able to validate the completeness of the input data. The system will not process data if any fields are incomplete.

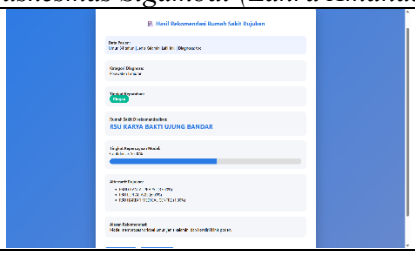
3) Recommendation Process Testing

Recommendation process testing was conducted to ensure the Random Forest algorithm performed effectively and generated referral hospital recommendations based on the patient data entered. Testing was conducted using multiple datasets representing different patient conditions.

Table 3. Recommendation Process Testing

NO	Test Data	Expected Output	System Output	Results
1	Age: 41, Diagnosis: Tuberculosis, Severity: Severe	Rantauprapat Regional Hospital		Succeed

https://doi.org/10.61536/ambidextrous.v5i02.510

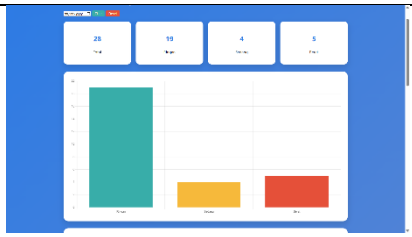
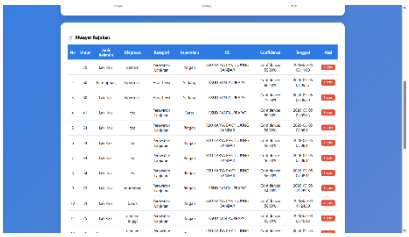
2	Age: 34, Diagnosis: TB, Severity: Mild	KARYA BAKTI Hospital		Succeed

The test results show that the system is able to generate hospital recommendations according to the classification results of the Random Forest model based on the input data provided.

4) Data Storage Testing

This test is carried out to ensure that the recommendations that have been displayed can be saved correctly into the database.

Table 4. Data Storage Testing

NO	Scenario	Expected Output	Results
1	Recommendation process complete		Succeed
2	Open referral history		Succeed

Based on the test results, the system is able to save the recommendation results into the database and display them again on the history page without any data errors.

5) System Testing Conclusion

Based on all the tests that have been carried out, it can be concluded that:

a) The system is able to perform user authentication properly.
b) The system can receive and validate patient data input correctly.
c) The classification process using the Random Forest algorithm ran without error.

The referral hospital recommendation output is displayed and stored securely in the database. Thus, the developed Referral Hospital Recommendation System meets functional requirements and can be used as an aid in the process of determining referral hospitals at the Sigambal Community Health Center in Labuhanbatu Regency.

4. Random Forest Model Evaluation and Analysis

Model evaluation was conducted to measure the performance level of the Random Forest algorithm in classifying referral hospitals based on the dataset used in this study. The dataset used was the patient referral history data of Sigambal Health Center, totaling 120 data with 6 attributes, namely: age, gender, original diagnosis, categorical diagnosis, severity level, and hospital as the target variable (class). Before the model training process was carried out,

https://doi.org/10.61536/ambidextrous.v5i02.510

the dataset went through a preprocessing stage which included data cleaning and transformation of categorical attributes into numeric form. Next, the data was divided into training data and test data with a ratio of 80:20. Thus, a total of 96 data were used as training data and 24 data were used as test data. Model performance evaluation was carried out using accuracy, precision, recall, and F1-score metrics calculated based on the results of the confusion matrix.

a. Confusion Matrix

A confusion matrix is used to compare the model's predictions with the actual test data. Testing 24 test data sets revealed that 22 were correctly classified, while 2 were misclassified. In multi-class classification cases like this study, precision and recall values are calculated for each hospital class and then averaged (macro-averaged) to obtain the overall model value.

b. Accuracy Calculation

Accuracy is the ratio between the number of correct predictions and the total test data.

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data\ Uji}$$

$$Accuracy = \frac{22}{24}$$

$$Accuracy = 0{,}914 = 91{,}4\%$$

The accuracy value of 91.4% shows that the Random Forest model is able to classify most of the test data correctly.

c. Precision Calculation

Precision measures the level of accuracy of the model in providing predictions for each class.

$$Precision = \frac{TP}{TP + FP}$$

Because this study used multi-class classification, the precision value was calculated as the average precision across all hospital classes. Based on the system's calculations, the precision value was 89.8%.

d. Recall

Recall is calculated using the formula:

$$Recall = \frac{TP}{TP + FN}$$

The recall value obtained was 90.5%, which shows that the model is able to recognize most of the actual data correctly.

e. F1-Score

F1-score is calculated using the formula:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Based on the results of the system calculations, the F1-score value obtained was 90.1%.

5. Research Advantages

This research has several advantages in developing a referral hospital recommendation system, including the following:

a. Utilizing diagnostic and severity data as a basis for decision making

https://doi.org/10.61536/ambidextrous.v5i02.510

This study successfully utilized patient diagnosis and severity data as primary variables in the classification process to systematically determine referral hospitals. Utilizing historical patient referral data enabled the system to learn patterns of association between patient conditions and referral decisions, which were previously made manually.

b. Using Random Forest which has good classification performance

The Random Forest method is capable of building a stable classification model through the formation of multiple decision trees and the use of majority voting. Based on the evaluation results using a confusion matrix, the model achieved an accuracy of 91.4%, indicating quite good classification performance.

c. Able to produce web-based recommendation systems

This research not only produces a classification model, but also implements the model into a web-based system that can be used to provide referral hospital recommendations automatically based on patient data.

d. The system is able to display recommendation results in an objective and structured manner.

The system developed is capable of providing automatic referral hospital recommendations based on a classification process using a machine learning model, thereby helping to increase the effectiveness and objectivity of the patient referral determination process.

6. Research Deficiencies

Although this study has succeeded in building a referral hospital recommendation system, this study still has several limitations, including the following:

a. The number of datasets used is still limited

The dataset used in this study consisted of 120 patient data, so the performance of the model built can still be improved if using a larger dataset and a longer data period.

b. The variables used in the model are still limited

The classification model developed only used a few key variables: age, gender, diagnosis, and patient severity. Other factors potentially influencing referral decisions, such as hospital distance, availability of medical facilities, and hospital capacity, were not considered in this study.

c. No comparison has been made with other classification methods.

This study only uses the Random Forest algorithm without comparing its performance with other classification methods, so it is not yet known whether the Random Forest method is the most optimal method in the case of a patient referral system.

d. The system has not been directly integrated with the health information system. The system that was built is still a web-based prototype and has not been directly integrated with the health information system or referral system used by health service facilities.

Conclusion and Recommendation

Based on the research results, it can be concluded that the application of the Random Forest method to the referral hospital recommendation system at the Sigambal Community Health Center can help the patient referral determination process become more effective, systematic, and objective. The use of patient diagnosis and severity data as the main variables

https://doi.org/10.61536/ambidextrous.v5i02.510

has been proven to be able to be used to build a classification model that is able to learn referral decision patterns based on historical patient data. Through the stages of preprocessing, model formation, and performance evaluation using a confusion matrix, the developed system showed good results with a high level of accuracy and consistency in providing referral hospital recommendations. In addition, the developed web-based system is also able to generate recommendations automatically so that it can support first-level health services in the process of making patient referral decisions more quickly and accurately.

Based on the results of this study, recommendations for developing a referral hospital recommendation system can be improved by increasing the number and variety of patient data to make the resulting model more accurate and adaptable to various health conditions. Furthermore, further research is expected to utilize other classification methods as a comparison to achieve more optimal model performance. System feature development is also needed, such as integration with hospital capacity data, location distance, and medical facility availability to ensure recommendations are more relevant to patient needs. With continued development, this system is expected to become a technological solution that supports improving the quality of healthcare services in primary healthcare facilities.

References

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.

Prasetyo, E. (2014). *Konsep dan Aplikasi Machine Learning*. Yogyakarta: Andi.

Pressman, R. S. (2010). *Software Engineering: A Practitioner's Approach* (7th ed.). McGraw-Hill.

Sommerville, I. (2011). *Software Engineering* (9th ed.). Pearson.

Stair, R., & Reynolds, G. (2018). *Principles of Information Systems*. Cengage Learning.

Syahra, Y. (2023). *Penerapan Data Mining dan Sistem Pendukung Keputusan dalam Sistem Informasi*.

https://doi.org/10.61536/ambidextrous.v5i02.510