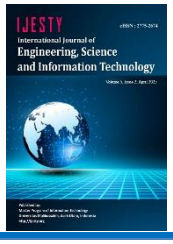
International Journal of Engineering, Science and Information Technology
Volume 5 No. 3 (2025) pp. 1-11
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.865
Research Paper, Short Communication, Review, Technical Paper



Classification Of Outpatient Visit Status Walking at Dr. Zubir Mahmud Hospital Using Algoritma C4.5

Putri Fikria*, Rozzi Kesuma Dinata, Yesy afrillia

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: putri.200170183@mhs.unimal.ac.id*

The manuscript was received on 18 November 2024, revised on 28 December 2024, and accepted on 28 March 2025, date of publication 9 May 2025

Abstract

This study aims to classify the status of outpatient visits at RSUD Dr. Zubir Mahmud into three main categories, namely "Very Urgent", "Urgent", and "Not Urgent", using the C4.5 algorithm. The web-based system uses the PHP programming language and MySQL database to ensure ease of implementation and efficient data management. The classification process is done by setting threshold parameters, calculating entropy, and the gain ratio to form an accurate and reliable decision tree. The results show that the C4.5 algorithm can classify patient visit data with a reasonably high accuracy rate, which is 93.75% for 2022 data and reaches 100% for 2023 data. In 2022 the "Very Urgent" category had 9 True Positives (TP); in 2023, the number remained consistent. However, in both years, there were also False Negatives in the same category, with 4 cases in 2022 and 5 cases in 2023. The "Urgent" and "Not Urgent" categories show suboptimal classification performance due to uneven data distribution, which causes the precision and recall values in these categories low. Model evaluation was conducted using evaluation metrics such as precision, recall, and F1 score. The evaluation results show that the model works very well in identifying high-priority categories, but further development is needed to improve classification in other categories. This system is expected to be a reliable tool in decision-making in health services, especially in determining the priority of patient services appropriately and efficiently. With further development, this system has the potential to be widely applied in various other hospitals.

Keywords: *C4.5 Algorithm, Classification, Confusion Matrix, Decision Tree, Outpatient Visits*.

1. Introduction

In the world of health, technological advances are needed to support the performance of medical personnel. Many hospitals and health centers have started using information technology to process patient data. This, of course, makes it easier for medical personnel to process patient data and group the status of visits to one of the health institutions, with the status of visits to one of the health agencies, by comparing data grouping from previous data and comparing data grouping of earlier data. Regional public hospitals are very much needed by people who experience health problems to carry out treatment.

Public awareness of health can increase the number of visitors to the hospital. Patient handling can be improved, and patients can be well served with enhanced quality. One of the units affected by the increasing number of visits to the hospital is the outpatient unit. If the status of outpatient visits can be adequately grouped, it will help the organization in planning actions that must be taken by medical personnel. Due to the increasing number of outpatient visits and data requests every year, it is necessary to have an information system as a case reporting tool so that hospital staff can group and handle the cases. In addition, the regional general hospital Dr. Zubir Mahmud requires an information system as efficient data processing so that health workers can access it at the regional general hospital dr. Zubir Mahmud [1]. What must be done by the hospital is to group outpatient visitor disease data so that it can be classified well and accurately. Classified accurately and adequately, so that the hospital can provide better services, patients can receive better care, and patients are happier to visit for treatment.

The C4.5 algorithm is one of the algorithms used to form decision trees. The C4.5 algorithm has input in training and test samples, where training samples are in the form of sample data that will be used. Samples and samples, training samples in the form of sample data that will be used to build a tree that has been tested, while samples to build a tree that has been tested, while samples are data fields that will be used as parameters. are data fields that will later be used as parameters in classifying data[2]. This algorithm will determine which outpatient

visitor status is urgent, very urgent, and non-urgent. Urgent, very urgent, and not urgent. It is expected to help hospital staff classify the status of outpatient visitors. In looking for the classification of the status of outpatient visitors, the author takes 2022 and 2023 data [3].
Like the research conducted by Siska Febriani and Heni Sulistiani, with the research title "Analysis of Diagnosis Results Data for Classification of Personality Disorders Using the C4.5 Algorithm"[4]. Personality Disorders Using the C4.5 Algorithm" (2021). Based on the research results, the C4.5 algorithm was successfully applied to classify psychological disorders accurately with a relatively good level of accuracy. On training data testing, this algorithm achieved an accuracy of 57.50%, with the same precision and recall. According to the testing data, the accuracy increased to 72.67%, with a recall of 100%. Reaching 100%. The prediction showed that 73% of the people experienced major depressive disorder, 25% had generalized anxiety disorder (GAD), and 2% experienced conversion disorder. These results demonstrate the potential of the C4.5 algorithm in supporting the process of diagnosing psychological disorders quickly and precisely, so that it can provide helpful information for people to understand their psychological condition.
Similar research was also conducted by Siti Riska Julianti Ihnur Alham, Efy Yosrita, Rizqia Cahyaningtyas, with the research title "Diagnosis System for Coronary Heart Disease Using Website-Based C4.5 Algorithm". Coronary Heart Disease By Using Website-Based C4.5 Algorithm (Case Study: Dr. Soedarso Pontianak Hospital)" [5]. Based on research, C4.5 algorithm was successfully applied for the diagnosis of coronary heart disease with a very high accuracy, which reached with very high accuracy, reaching 94.4%. This algorithm processes various risk factors such as age, gender, systolic and diastolic blood pressure, total cholesterol levels, HDL, LDL, triglycerides, and heredity. By building a decision tree based on the classified data, this study demonstrates the ability of the C4.5 algorithm to support the diagnosis process quickly, accurately, and in a structured manner. Diagnosis process in a fast, accurate, and structured manner, making it useful for clinical decision-making in the health sector. Clinical decision-making in the health sector.

2. Literature Review

2.1 Data Mining

According to David Hand, Heikki Manilla, and Padhraic Smyth from MIT, data mining is the analysis of data (usually large data) to analyze it. Data mining is the analysis of data (usually large data) to find clear relationships and infer them that were not known before in a currently understood and useful way to the data owner, before in a way that is currently understood and useful for the owner of the data[6]. Data Mining is the process of collecting and processing data that aims to extract important information from the data. Data mining applies two-step methods, namely unsupervised learning and supervised learning. Learning. Unsupervised learning does not use guidance or instructors, but uses data names as teachers, while supervised learning involves guidance and instructors (Dayera, 2011). involves guidance and instructors[7]. Data mining can be divided into several stages, a series of processes.
Data mining is also a series of processes that extract additional value in the form of knowledge that is not known manually. Additional knowledge that has not been known manually from a data set[8]. Data mining is often also called knowledge discovery in databases (KDD). KDD is an activity that historically collects and uses data to find regularities, patterns, or relationships in large data sets[9]. Data mining is related to other fields of science, such as database systems, data warehousing, statistics, and data science: warehousing, statistics, machine learning, information retrieval, and high-level computing[10]. In addition, data mining is supported by other sciences such as neural networks, pattern recognition, spatial and analysis, image databases, and signal processing—database, signal procesing. Data mining is the process of finding patterns in data[11].

2.2. Outpatient Care

Outpatient service (ambulatory service) is one form of existing medical services. Existing medical services, which are defined as outpatient services, are medical services provided to patients, not in the form of hospitalization. Hospitalization. Outpatient services are for observation, diagnosis, treatment, medical rehabilitation, and other health services without staying at the hospital [12].

2.3. Classification

Classification is finding models (functions) that describe and distinguish between data classes. Models (functions) that can describe and differentiate classes of data or concepts can be used to describe and differentiate classes of data or concepts. Concepts, to use the model to predict the class of an object. Class of an object [13]. The process of finding a model or function that describes or distinguishes concepts or classes of data to estimate the class of an object whose label is unknown. Classification can also be interpreted as assigning a new data record to one of the following categories (or classes) previously defined[14].
Classification processes are typically divided into two phases: learning and testing. In the learning phase, a portion of the data with known class labels is used to build an estimation model. Then, in the testing phase, the formed model is tested with another portion of the data to determine the model's accuracy[15]. Classification itself is divided into two stages: classification and learning. A classification algorithm constructs a model in the learning stage by analyzing the training data. This learning stage can also be viewed as forming a function or mapping y = f(x), where *y* is the predicted class and *x* is the tuple whose class is to be[16].

2.4. C4.5 Algorithm

The C4.5 algorithm is one of the classification algorithms, and it is used to construct decision trees. A decision tree is an influential and well-known method for classification and prediction. This method transforms large amounts of factual data into a decision tree representing a set of rules. These rules can be easily understood in natural language. This method is considered suitable for solving classification problems, such as determining the status of outpatient visitors using data directly obtained from the hospital[17].
The advantage of the C4.5 algorithm is that it is a decision tree method capable of avoiding issues by using fewer criteria at each internal node without significantly reducing the quality of the resulting decisions[18]. An algorithm is an effective method expressed as a finite sequence. It is also a set of instructions designed to solve a particular problem, where these instructions can be executed step by step from beginning to end. The problem can be anything, provided each problem has initial criteria to be met before running an algorithm. An algorithm also involves decision-making to reach a final outcome [19].
The C4.5 algorithm uses an information gain theory approach to build a model as a decision tree. The attribute with the highest information gain value will be selected as the parent for the next node. Before calculating the gain, the entropy value must first be determined. Entropy is a parameter used to measure a data sample's heterogeneity (diversity). The more heterogeneous the sample data, the higher the entropy value[20]. The formula used to calculate entropy is:

$$Entrophy(S,A) = \sum_{i=1}^{n} - pi \times log2pi \quad \text{.....(1)}$$

Explanation:
S = Set of Classes
n = Number of Partitions in S
Pi = Proportion of Si to S
After the entropy value is obtained, the next step is to calculate the gain to measure the effectiveness of an attribute in classifying the data. Gain is calculated using the following formula:

$$Gain\ (S,A) = Entrophy(S) \sum_{i=1}^{n} \frac{|si|}{|s|} \times Entrophy(S_i) \quad \text{.....(2)}$$

Explanation:
S = Set of Cases
A = Attribute
n = Number of Partitions of Attribute A
(Si) = Number of Cases in the *i-th* Partition
(S) = Total Number of Cases in S

3. Research Method

This research was conducted at the Dr. Zubir Mahmud Regional General Hospital at JL. Medan – Banda Aceh, Seuneubok Baro, Idi Timur District, East Aceh Regency. This location was chosen because it has aspects that support the needs and completion of this research. The research was conducted throughout 3 (three) months.

4. Results and Discussion

This research uses a web-based system developed with PHP and MySQL to apply the C4.5 algorithm in classifying the visit status of outpatient patients at Dr. Zubir Mahmud Regional General Hospital. The system is designed to provide recommendations for patient visit statuses categorized as "Urgent," "Highly Urgent," and "Not Urgent," based on outpatient data from 2022 and 2023.
The primary focus of this study is to explore the capability of the C4.5 algorithm in producing accurate classifications to improve patient management more effectively and efficiently, as well as to assist the hospital in prioritizing patient services based on the level of urgency. In addition, this study also evaluates the performance of the C4.5 algorithm through accuracy, precision, and recall analysis to measure the consistency and relevance of the results compared to conventional approaches.

4.1. System Analysis

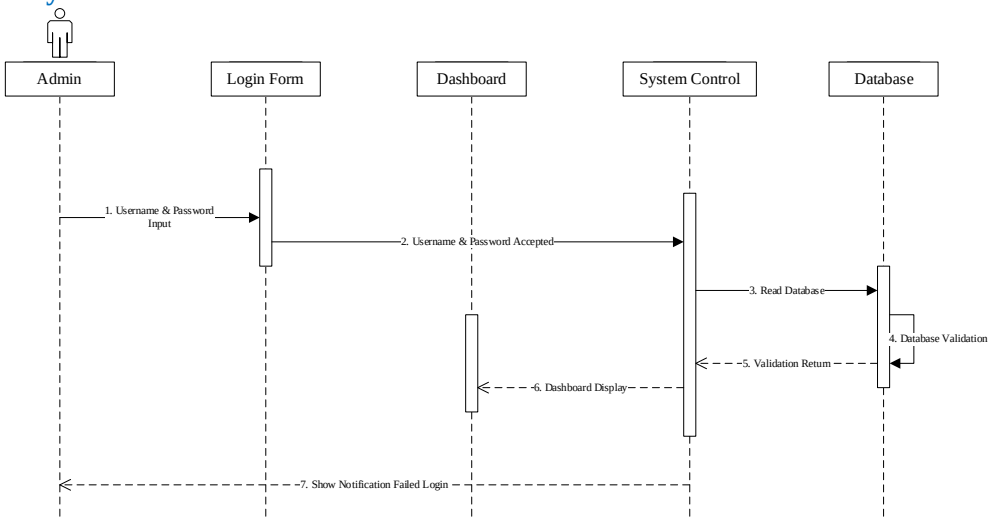


Fig. 1 Sequence Diagram Login

This Login Sequence Diagram describes the flow of the admin authentication process in the system. The process starts when the admin enters the username and password on the Login form. The data is then forwarded to the dashboard to validate the information entered. Next, the dashboard sends an authentication request to the control system, which then forwards it to the database to match the Login data with the stored data. After the validation process, the database returns the validation results to the control system, which returns the information to the dashboard. The admin will be directed to the main page (dashboard) if the validation is successful. This flow ensures that only verified users can access the system.

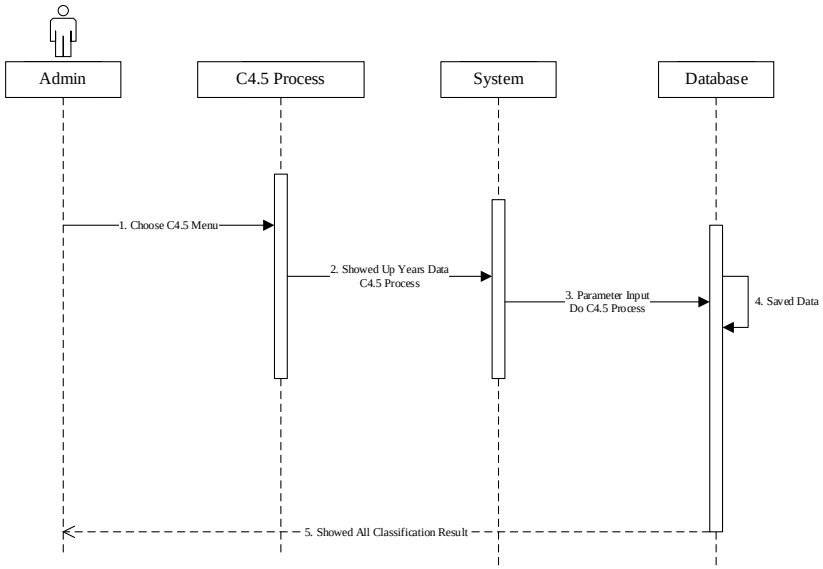


Fig. 2 Sequence Diagram of C4.5 Process

The C4.5 Process Sequence Diagram illustrates the workflow of data processing using the C4.5 algorithm in the system. The process begins when the admin selects the C4.5 Process menu from the system interface. The system then displays the available year data for the admin to choose as part of the classification process. Once the year data is selected, the admin enters the parameters required to run the C4.5 algorithm. The system uses the parameters to ensure the classification process runs according to user requirements and relevant data scenarios. Next, the system forwards the data along with the parameters entered by the admin to the database for processing. The database stores the classification results generated by the C4.5 algorithm. After the classification process is complete, the system displays all classification results to the admin, including the resulting status categories, such as "Urgent," "Very Urgent," or "Not Urgent." This process is designed to ensure that data processing runs efficiently, classification results are appropriately stored, and can be accessed by the admin for analysis or report needs.

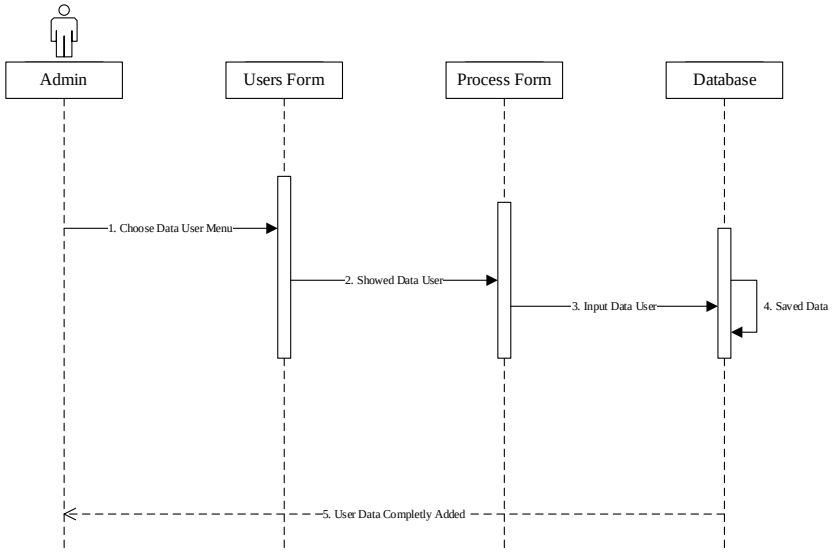


Fig 3. User Data Sequence Diagram

This User Data Sequence Diagram explains the process flow of managing user data in the system. The process starts when the admin selects the User Data menu on the system interface. After that, the system displays the user data form, which allows the admin to view, add to, or update user data. The admin then enters the required user data, such as name, username, or access rights. The data is then forwarded to the process form for further processing.

The process form sends the entered data to the database to be saved or updated. After the database successfully stores the data, the system confirms to the admin that the user data has been added or updated. This flow ensures that user data can be managed efficiently and organized, making managing access and users in the system easier.

4.2. Descriptive Data

This study uses outpatient visit data from Dr. Zubir Mahmud Regional General Hospital for 2022 and 2023. The data includes the number of visits by male (M) and female (F) patients at each polyclinic every month. In total, 16 polyclinics were recorded in 2022 and 17 in 2023, with the addition of the Orthopedic polyclinic in 2023.

Table 1. Research Dataset 2022

Polyclinic	Total Male	Total Female	Total
Internal Medicine	3567	4858	8425
Surgery	2191	2437	4628
Pediatrics	1686	1180	2866
Obstetrics and Gynecology	0	2070	2070
Neurology	2312	3749	6061
Psychiatry	1309	671	1980
Psychology	0	0	0
Ophthalmology	2068	2886	4954
Dental and Oral Health	312	757	1069
……..	……	…….	……
Urology	0	0	0
THT	1051	1149	2200

In 2022, the total number of patient visits reached 44,060, consisting of 19,493 male and 24,567 female patients. The polyclinic with the highest number of visits was the Neurology Polyclinic, serving 6,061 patients, followed by the Internal Medicine Polyclinic with 8,425 visits. On the other hand, the Psychology Polyclinic recorded the lowest number of visits, with zero trips throughout the year. Overall, the number of visits by female patients was higher than that of male patients in most polyclinics.

Table 2. Research Dataset 2023

Polyclinic	Total Male	Total Female	Total
Internal Medicine	3063	4458	7521
Surgery	1692	1956	3648
Pediatrics	1034	946	1980
Obstetrics and Gynecology	0	1515	1515
Neurology	2886	5091	7977
Psychiatry	987	589	1576
Psychology	0	0	0
Ophthalmology	1477	2225	3702
Dental and Oral Clinic	419	1005	1424
…….	……	……..	……
THT	1503	1495	2998
Orthopedics	779	784	1563

In 2023, the total number of patient visits was recorded at 45,227, consisting of 19,252 male and 25,975 female patients, indicating a continuing trend where female patients dominate visits in most polyclinics. The Neurology Polyclinic again had the highest number of visits, totaling 7,977, followed by the Pulmonology Polyclinic. The Orthopedic Polyclinic, which began operations this year, recorded 173 trips. Monthly patient visits showed significant fluctuations, with January and December typically recording high numbers, while April and May tended to be lower. Seasonal illness patterns, service availability, or changes in hospital operations may influence these variations. This descriptive data provides an initial overview of the distribution of patient visits per polyclinic and serves as an essential foundation for further analysis using the C4.5 algorithm to classify visits based on urgency levels, aiming to improve hospital service efficiency.

4.2. Manual C4.5 Calculation

This process begins with collecting patient visit data as the basis for manual calculation of the C4.5 algorithm. Next, the initialization process is carried out by calculating each attribute's entropy and gain ratio values to determine the most relevant attribute to be used as the root of the decision tree. The inference process continues by constructing the decision tree based on the selected attributes, where existing facts are classified into three main categories: "Highly Urgent," "Urgent," and "Not Urgent." The sample data used in the manual C4.5 calculation is the "Internal Medicine" polyclinic data from 2022.

Table 3. Sample Data Generation

Month	Male	Female
January	335	496
February	302	408
March	424	479
April	220	296
May	253	380
June	286	479
July	289	349
August	318	408
September	285	366
October	260	402
November	296	379
December	299	416

Total Male & Female Visits	3567	4858
Total Overall Visits	8425	

Based on the calculation results, the Gender Ratio attribute has a Gain Ratio of 0.603. If this is the highest value compared to other attributes, then this attribute is selected as the root of the decision tree. This process is continued for each branch until all data is classified.

4.3. Manual Calculation of Evaluation Metrics

In this subchapter, manual calculations for evaluation metrics using the confusion matrix generated from the classification of the 2022 data are performed. The calculated evaluation metrics include accuracy, precision, recall, specificity, and F1 score for each category: "Very Urgent," "Urgent," and "Not Urgent." The calculation results show that for the category "Very Urgent," the model has an accuracy of 75%, precision of 100%, recall of 69.23%, specificity of 100%, and F1 score of 81.82%. Meanwhile, for the categories "Urgent" and "Not Urgent," the model has the same accuracy of 93.75%, but with precision, recall, and F1 score values of 0%, due to the absence of accurate optimistic predictions for both categories. This calculation provides an overview of the model's excellent performance in the top priority category but less than optimal in other categories, so that it can be evaluated for further development.

4.4. System Implementation

The researcher discusses the implementation of the C4.5 algorithm in a decision support system for the classification of outpatient visit status at RSUD Dr. Zubir Mahmud, according to the design analysis that has been designed previously. This implementation is the stage after design, where the system allows users, such as admins, to input variable data such as year, polyclinic, and patient visit data (number of men and women). This data is then processed using the C4.5 algorithm to produce a classification of patient visit status into three categories: "Urgent," "Very Urgent," and "Not Urgent." The system also allows efficient validation of the interface and classification results, which are presented as tables or reports for easy analysis.

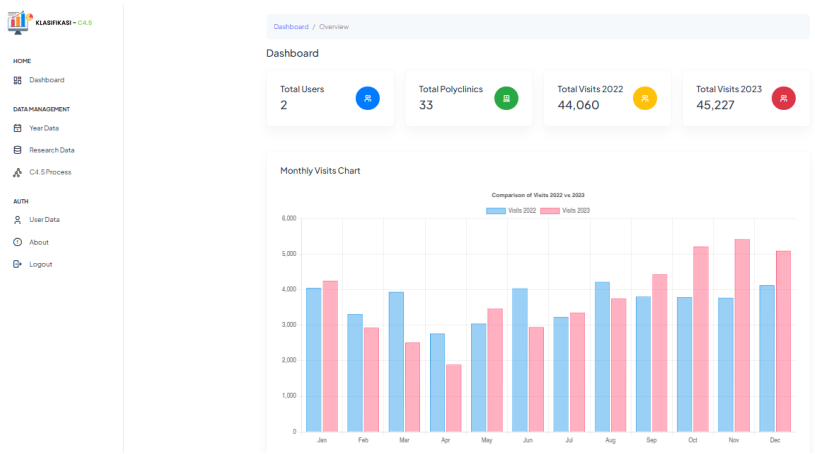


Fig. 4 Dashboard Page Display

The dashboard page is the leading information center of the C4.5 algorithm-based classification application, designed to provide an overall overview. Users can view essential statistics such as total users, number of polyclinics, and total patient visits in 2022 and 2023. The main feature on this page is a bar graph that compares the number of patient visits per month between the two years, making it easy to analyze trends and visit patterns visually. In addition, on the left side of the page, there is a navigation sidebar efficiently designed to facilitate access to various essential sections of the application. Users can quickly move to the Dashboard menu, Data Processing section (including Year Data, Research Data, and C4.5 Process), and system authorization features such as Data Users, About, and Logout. With an organized appearance and intuitive navigation, the dashboard page supports a smooth workflow and helps users manage and analyze data effectively.

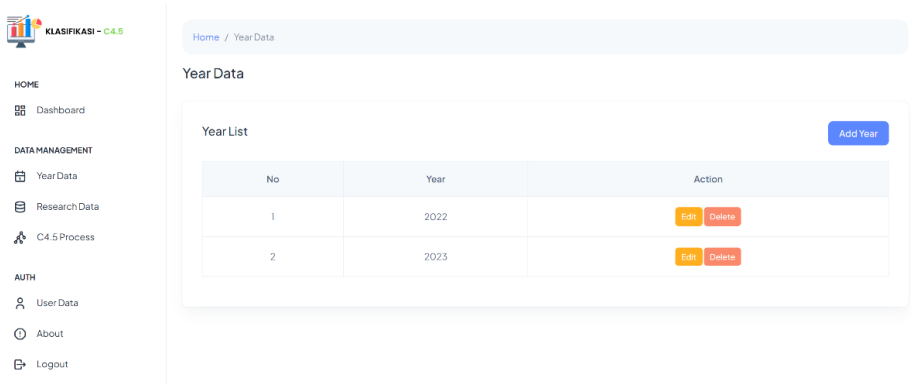


Fig 5. Year Data Page View

The Year Data page is an essential component in the C4.5 algorithm-based classification system, especially in managing the time parameter, one of the main variables in the data analysis process. The interface of this page displays an informative table containing a list of years that have been registered in the system, with a column arrangement in the form of sequence numbers (No), years (Year), and an action column

(Action) that provides two main options: a yellow “Edit” button to update data, and a red "Delete” button to delete unnecessary data. In the top right corner of the page is an "Add Year” button designed to make it easier for users to add new year entries to the database, supporting the completeness of historical data required in the classification process. The layout of this page is organized in a functional yet straightforward manner, allowing users to access, add, change, and delete year data quickly and efficiently. Not only does this page make it easier to manage time in the context of analysis, but it also strengthens the data integration required to produce accurate classification results that are consistent with the algorithmic approach used.

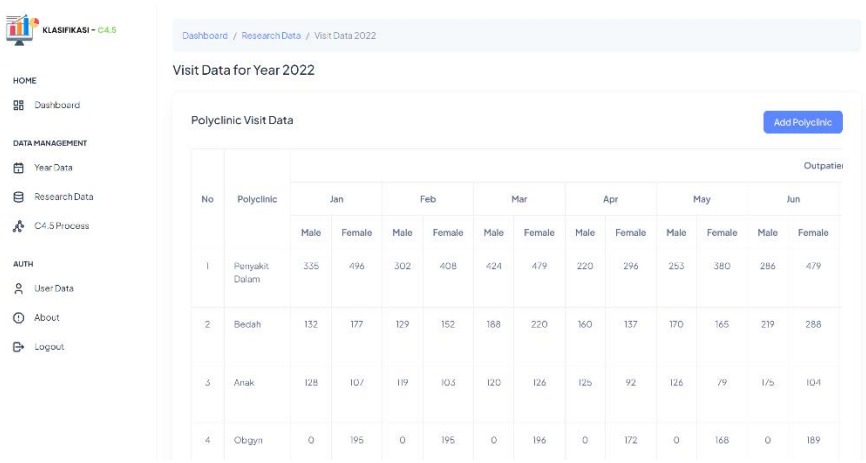


Fig 6. Data Presentation View 2022

The Year 2022 Visit Data page displays outpatient visit data per polyclinic based on the selected year. Users can view, edit, add, or delete visit data through the available CRUD (Create, Read, Update, Delete) features. This page provides an “Add Polyclinic” button to enter new polyclinic data, and an action column on the far right of the table to manage existing data. This page is designed to facilitate the management of visit data according to analysis needs.

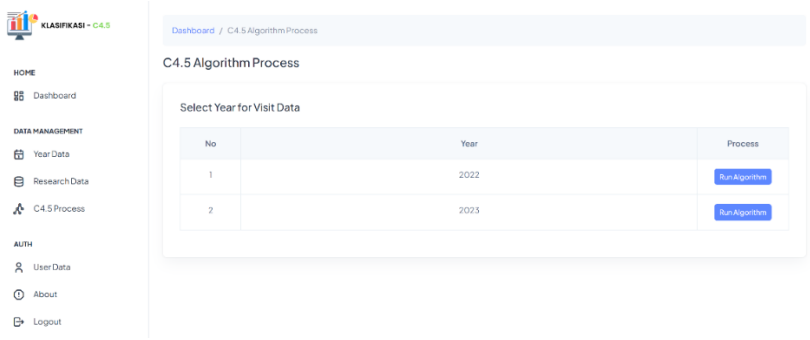


Fig. 7 Display of C4.5 Process Page

The C4.5 Algorithm Process start page is the main entry point for starting the classification analysis of patient visit data by year using the C4.5 method. On this page, the user is presented with an informative table that lists the year periods available in the system. Each year, entry is equipped with an “Algorithm Process” button in the action column, which triggers the classification process based on visit data for that year. This interface is designed to help users determine the temporal context efficiently, allowing the system to perform calculations only on relevant data. With an intuitive design and directed workflow, this page becomes integral in facilitating the initial stages of the data analysis process using the C4.5 algorithm in a more structured and systematic manner.

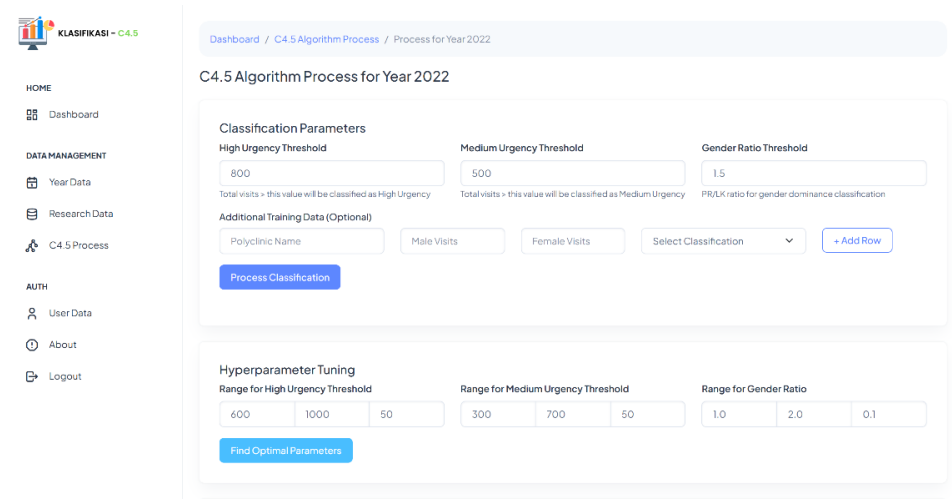


Fig 8. Hyperparameter Tuning or Manual Parameter Optimization

The initial configuration stage of the system provides a comprehensive interface for users to set classification parameters crucial in analyzing patient visit data. Users can freely set threshold values for priority categories such as “Very Urgent” and “Urgent”, as well as set the “Gender Threshold Ratio” according to the hospital's specific needs or policies. Determining these parameters allows health institutions to customize the model to the unique characteristics of the patient population they serve.

The system also has an advanced hyperparameter tuning feature that optimizes parameter combinations within a predefined spectrum of values. This functionality systematically evaluates various parameter configurations to identify the optimal combination that yields the best classification performance. This methodological approach improves the reliability and accuracy of the model in classifying patient visit data, ensuring that data-driven decisions have a strong and validated analytical foundation.

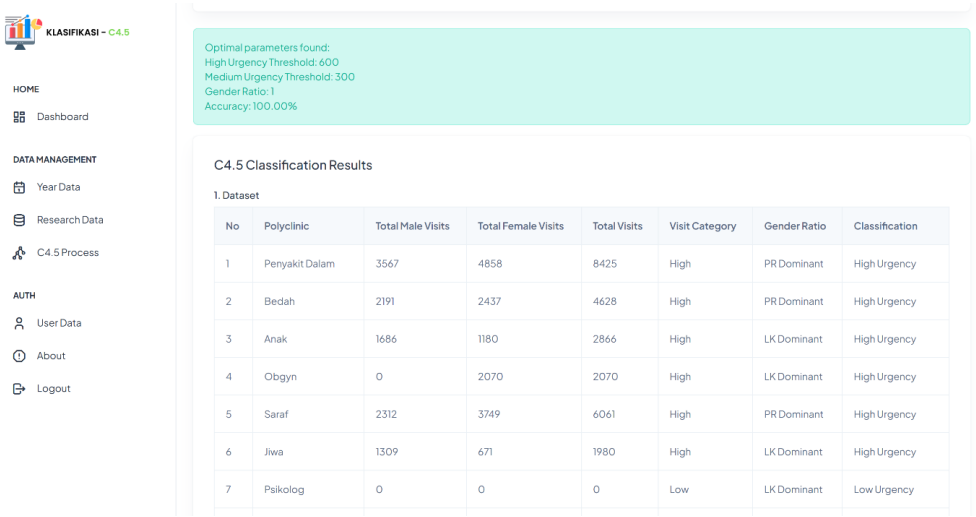
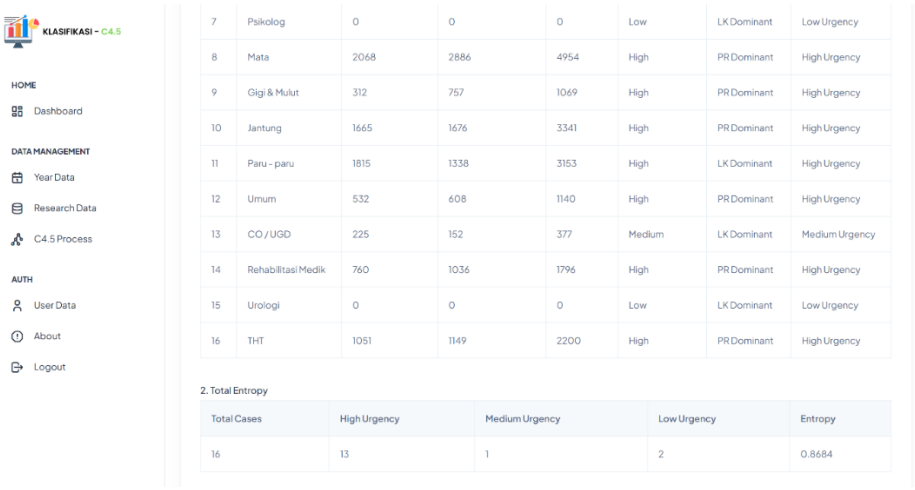


Fig 9. First Presentation of the C4.5 Process 2022

The optimal results of hyperparameter tuning show that the best parameter combination is the “Very Urgent” threshold of 600, the “Urgent" threshold of 300, and the gender ratio of 1. Furthermore, the classification process using the C4.5 algorithm produces several datasets displayed in several main tables. First, the classification results table presents data that includes the polyclinic name, total patient visits, gender ratio, visit category, and classification status. For example, polyclinics with a high number of visits, such as “Internal Medicine" and “Surgery”, fall into the priority category of “Highly Urgent”. In contrast, polyclinics with a low number of visits, such as “Psychologist" and “Urology”, are categorized as “Not Urgent”. This data provides a clear picture of the priority level of services in each polyclinic, which helps the hospital to manage resources more effectively.



7	Psikolog	0	0	0	Low	LK Dominant	Low Urgency
8	Mata	2068	2886	4954	High	PR Dominant	High Urgency
9	Gigi & Mulut	312	757	1069	High	PR Dominant	High Urgency
10	Jantung	1665	1676	3341	High	PR Dominant	High Urgency
11	Paru - paru	1815	1338	3153	High	LK Dominant	High Urgency
12	Umum	532	608	1140	High	PR Dominant	High Urgency
13	CO / UGD	225	152	377	Medium	LK Dominant	Medium Urgency
14	Rehabilitasi Medik	760	1036	1796	High	PR Dominant	High Urgency
15	Urologi	0	0	0	Low	LK Dominant	Low Urgency
16	THT	1051	1149	2200	High	PR Dominant	High Urgency

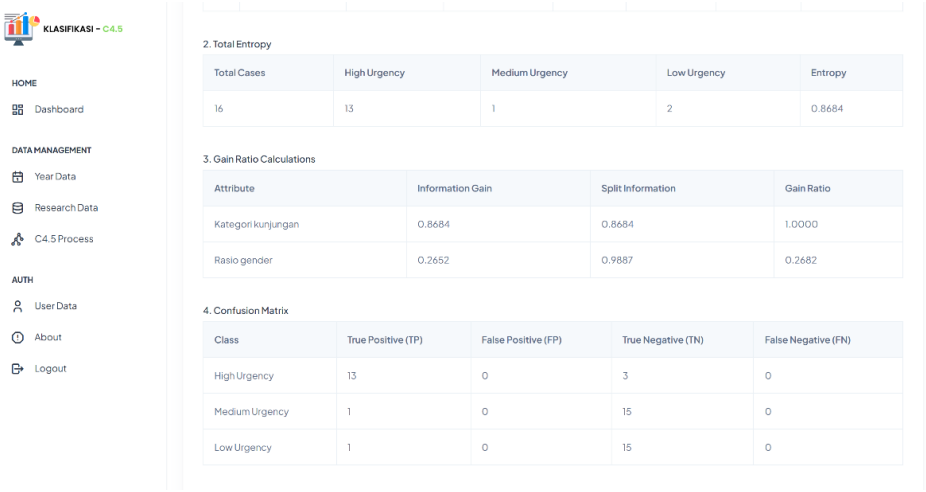
2. Total Entropy

Total Cases	High Urgency	Medium Urgency	Low Urgency	Entropy
16	13	1	2	0.8684

Fig. 10 Second Presentation of the C4.5 Process 2022

In the total entropy calculation table, the system calculates an entropy value of 0.8684 based on the distribution of cases in each category. For example, there are 13 cases in the “Very Urgent” category, 1 case in the “Urgent" category, and 2 cases in the “Not Urgent” category. This entropy value indicates the level of uncertainty in the data, which is the basis for further calculations.

Furthermore, the gain ratio calculation table displays two main attributes: "Visit Category" and "Gender Ratio”. The attribute "Visit Category" has the highest gain ratio value of 1.0000, followed by "Gender Ratio” with a value of 0.2682. Based on these results, the system selects "Visit Category" as the primary attribute in the decision tree formation because it greatly influences classification.



2. Total Entropy

Total Cases	High Urgency	Medium Urgency	Low Urgency	Entropy
16	13	1	2	0.8684

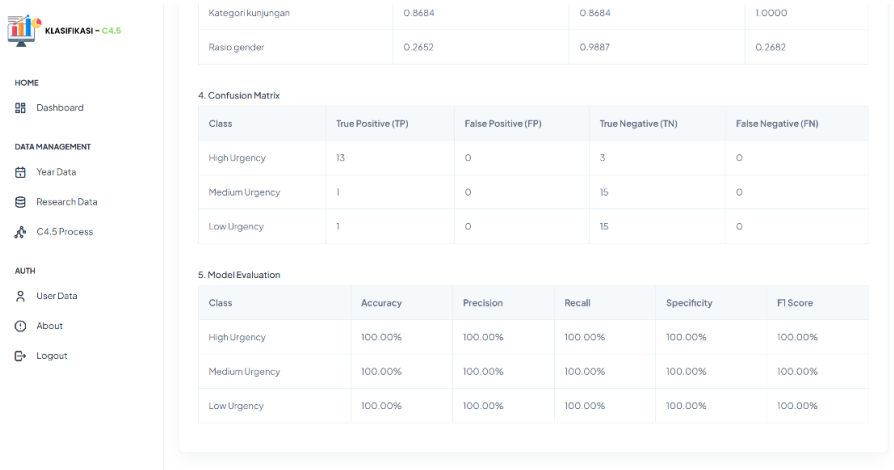
3. Gain Ratio Calculations

Attribute	Information Gain	Split Information	Gain Ratio
Kategori kunjungan	0.8684	0.8684	1.0000
Rasio gender	0.2652	0.9887	0.2682

4. Confusion Matrix

Class	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
High Urgency	13	0	3	0
Medium Urgency	1	0	15	0
Low Urgency	1	0	15	0

Fig. 11 Third Presentation of C4.5 Process 2022



Kategori kunjungan	0.8684	0.8684	1.0000
Rasio gender	0.2652	0.9887	0.2682

4. Confusion Matrix

Class	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
High Urgency	13	0	3	0
Medium Urgency	1	0	15	0
Low Urgency	1	0	15	0

5. Model Evaluation

Class	Accuracy	Precision	Recall	Specificity	F1 Score
High Urgency	100.00%	100.00%	100.00%	100.00%	100.00%
Medium Urgency	100.00%	100.00%	100.00%	100.00%	100.00%
Low Urgency	100.00%	100.00%	100.00%	100.00%	100.00%

Fig. 12 Fourth Presentation of C4.5 Process 2022

In the model evaluation stage, evaluation results are presented through a confusion matrix and other metrics. The confusion matrix shows the classification performance for each category, such as “Very Urgent”, “Urgent”, and “Not Urgent”. For example, in the “Very Urgent” category, the system recorded 9 True Positives (TP), 3 True Negatives (TN), 0 False Positives (FP), and 4 False Negatives (FN). In addition, the evaluation used metrics such as accuracy, precision, and recall. The overall accuracy reached 93.75%, with high precision in the “Very

Urgent" category (100%) but zero for the other categories. These results show that the system is able to classify the top priority categories well, although it requires improvement in the other categories.

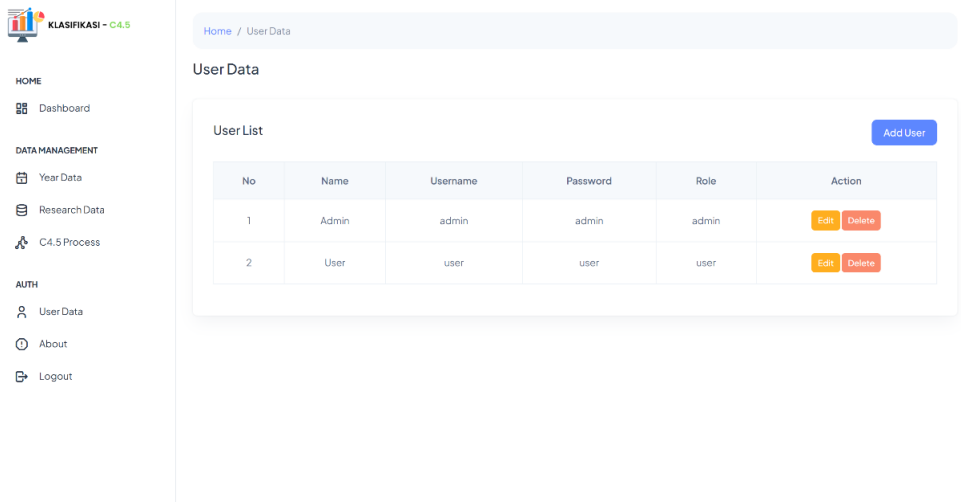


Fig 13. User Data Page Display

The User Data page is a feature in the system used to manage user data that has access to the C4.5 algorithm-based classification application. This page has a user list that includes information such as number, name, username, password, and the role of each user, such as admin or regular user. Admins can add new users with the "Add User" button located in the top right corner, as well as manage existing user data through the "Edit" and "Delete" options provided in the Action column.

5. Conclusion

This study aims to classify the status of outpatient visits at RSUD Dr. Zubir Mahmud into three main categories, namely "Very Urgent", "Urgent", and "Not Urgent", using the C4.5 algorithm. Based on the research results from the 2022 and 2023 data, the C4.5 algorithm successfully classified the status of patient visits with high accuracy, dividing the data into three main categories according to the level of urgency. Polyclinics with high visit rates, such as "Internal Medicine" and "Pediatric", tended to be categorized as "Very Urgent", while polyclinics with low visit rates, such as "Psychologist", were more often classified as "Not Urgent". Implementing a web-based system that implements this algorithm shows high accuracy of 93.75% for 2022 and 100% for 2023, allowing users to adjust parameters and improve classification accuracy according to the hospital's needs. Although the "Very Urgent" category has excellent accuracy, other categories, such as "Urgent" and "Not Urgent," still need improvement, especially in improving the precision and recall for these categories.

References

[1] Febriani, Indra Gunawan, Rafiqa Dewi, Dedy Hartama, and Muhammad Ridwan Lubis, "Klasifikasi Rasa Susu Almond Berdasarkan Minat Pelanggan Menggunakan Algoritma C4.5 di Home Made MamiNia Ice Pematangsiantar," *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 1, no. 2, pp. 71–74, 2021, doi: 10.54259/satesi.v1i2.33.

[2] R. Bayu Prasetyo, Y. Agus Pranoto, and R. Primaswara Prasetya, "Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2144–2151, 2023, doi: 10.36040/jati.v7i4.7419.

[3] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.

[4] T. Widiastuti, K. Karsa, and C. Juliane, "Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5," *Technomedia J.*, vol. 7, no. 3, pp. 364–380, 2022, doi: 10.33050/tmj.v7i3.1932.

[5] T. R. Matondang, Y. Ramadhan Nasution, Armansyah, and M. Furqan, "Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita," *J. Fasilkom*, vol. 14, no. 1, pp. 216–225, 2024, doi: 10.37859/jf.v14i1.6941.

[6] F. A. Aryatama and S. Samsugi, "Sistem Keamanan Kendaraan Bermotor Dengan ESP32 Menggunakan Kontrol Android," *Smatika J.*, vol. 14, no. 01, pp. 167–181, 2024, doi: 10.32664/smatika.v14i01.1267.

[7] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[8] N. Safitri and C. Bella, "Penggunaan Algoritma Apriori Dalam Penerapan Data Mining Untuk Analisis Pola Pembelian Pelanggan (Studi Kasus: Toko Diengva Bandar Jaya)," *J. Portaldata*, vol. 2, no. 1, pp. 1–8, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/72/72

[9] F. O. Dayera, Musa Bundaris Palungan, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229

[10] S. Ahdan and P. Indah Sari, "Pengembangan Aplikasi Web Untuk Simulasi Simpan Pinjam (Studi Kasus : Bmt L-Risma)," *J. Tekno Kompak*, vol. 14, no. 1, pp. 33–40, 2020.

[11] A. Nurkholis, M. Muhaqiqin, and T. Susanto, "Analisis Kesesuaian Lahan Padi Gogo Berbasis Sifat Tanah dan Cuaca Menggunakan ID3 Spasial (Land Suitability Analysis for Upland Rice based on Soil and Weather Characteristics using Spatial ID3)," *JUITA J. Inform.*, vol. 8, no. 2, pp. 235–244, 2020.

[12] V. Y. Hendrawan, S. Winardi, and H. Surbakti, "Sistem Informasi Rekam Medis Rawat Jalan Dan Pemeriksaan Penunjang Diagnosa Berbasis Website (Studi Kasus : Rumah Sakit Khusus Bedah Klinik Sinduadi, Mlati, Sleman, Yogyakarta)," *Respati*, vol. 9, no. 27, pp. 53–58, 2017, doi: 10.35842/jtir.v9i27.82.

[13] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
[14] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
[15] S. Suriana, E. D. Bintari, and D. Kurniawan, "Desain Aplikasi Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes dan Algoritma C4.5," *J. Big Data Anal. Artif. Intell.*, vol. 3, no. 1, pp. 31–41, 2017, [Online]. Available: https://www.neliti.com/publications/308842/
[16] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2018, [Online]. Available: https://politeknikmeta.ac.id/meta/ojs/index.php/inkofar/article/view/29/45
[17] R. Hidayat, A. D. P. Dewi, F. A. Putri, F. D. Anugraheni, and R. Robitoh, "Penerapan Algoritma C4.5 Terhadap Tingkat Kepuasan Pasien pada Pelayanan Puskesmas Kebumen," *Technol. Informatics Insight J.*, vol. 2, no. 1, pp. 24–32, 2023, doi: 10.32639/tiij.v1i1.391.
[18] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Januari*, vol. 2, no. 1, pp. 39–46, 2021.
[19] P. G. S. C. Nugraha and G. S. Mahendra, "Explorasi Algoritma C4.5 Dan Forward Feature Selection Untuk Menentukan Debitur Baik Dan Debitur Bermasalah Pada Produk Kredit Tanpa Agunan (Kta)," *JST (Jurnal Sains dan Teknol.*, vol. 9, no. 1, pp. 39–46, 2020, doi: 10.23887/jstundiksha.v9i1.24627.
[20] R. A. Edwin, "Penerapan Data Mining Untuk Klasifikasi Kinerja Peralatan Pembangkit Listrik Menggunakan Algoritma C4. 5," *J. Teknol. Pint.*, vol. 2, no. 2, pp. 1–16, 2022, [Online]. Available: http://teknologipintar.org/index.php/teknologipintar/article/view/313%0Ahttp://teknologipintar.org/index.php/teknologipintar/article/download/313/298