Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706
P-ISSN: 2810-0557

Research Article

Classification of Favorite Book Borrowing Data at the STIKOM CKI Library Using the Decision Tree Algorithm

Yuma Akbar [1], Untung Surapati [2], and Sutisna [3], Yansen [4]

1 Computer Science College of Informatics (Stikomcki) Jakarta; Email: yumekhan@stikomcki.ac.id
2 Computer Science College of Informatics (Stikomcki) Jakarta; Email: untungsurapati@stikomcki.ac.id
3 Computer Science College of Informatics (Stikomcki) Jakarta; Email: sutisna@stikomcki.ac.id
4 Jakarta Computer Science and Informatics Institute (Stikomcki)
* Corresponding Author: yumekhan@stikomcki.ac.id

Abstract: The library on the STIKOM CKI campus as a means of providing information and has a complete collection of learning media books, but the data processing system for borrowing and returning favorite books in the library is currently still manual, that is, all data collection processes are written on book cards, although it is quite good but the process is rather slow and requires quite a long time because in the process of searching the data must be checked per page one by one so that the data processing is less effective and efficient. To overcome this, it is necessary to develop an application using the decision tree algorithm method which can make it easier to collect borrowing data and return favorite books that are more effective and efficient and display integrated output of student reports that have not been returned so that data processing is more accurate and can speed up officer performance. libraries. Submitting a favorite book classification lending application can make it easier to access loans and returns anywhere and anytime. So that data processing is more accurate and can speed up librarian performance.

Keywords: Classification; Decision Tree Algorithm; Favorite Books; Library Information System; Data Processing.

Received: December 15, 2021
Revised: January 10, 2022
Accepted: January 25, 2022
Published: February 16, 2022
Current Ver.: February 28, 2022

1. Introduction

A university library is an academic facility that plays a vital role in supporting learning, research, and scientific development. Libraries serve not only as reference sources but also as a means for students to access various supporting literature effectively and efficiently. Therefore, good library management is a crucial factor in supporting the quality of information services for the academic community.

In practice, library management often faces various obstacles, particularly in managing book collection data, loan transactions, returns, and library member data. These problems can lead to data discrepancies, recording errors, and delays in reporting. This condition is also found at the STIKOM CKI Library, which still implements manual management processes. An inadequately documented management system has the potential to lead to errors in the book borrowing and return process, thus suboptimal library service quality.

Advances in information technology have enabled the application of data mining techniques to assist in the management and analysis of library data. One technique that can be used is the classification method using the Decision Tree algorithm. This algorithm is capable of grouping data based on specific characteristics, thus helping to identify book borrowing patterns and determine the book categories most in demand by library users. This information can be used as a basis for decision-making in collection management and improving the quality of library services.

This study aims to apply the Decision Tree algorithm to the classification process of book loan data at the STIKOM CKI Library. The application of this method is expected to assist librarians in managing book collection data, member data, loan data, and return data in a more structured and accurate manner. Furthermore, the classification results are expected to provide information on the most popular books, thus supporting more effective library management.

Based on the existing problems, this research focuses on the application of the Decision Tree algorithm to classify favorite book borrowing data at the STIKOM CKI Library. The contribution of this research is to facilitate the processing of book borrowing and returning data more accurately and to improve the work efficiency of librarians in managing library information.

2. Literature Review

Classification

Classification is the process of grouping data into specific classes or categories based on their characteristics. In data mining, classification is used to predict the class of an object based on patterns derived from historical data. This technique is widely used to aid decision-making in various fields, including education, healthcare, business, and libraries.

Decision Tree Algorithm

Decision Tree is a machine learning algorithm used to perform classification by forming a decision tree structure. This algorithm operates based on a set of rules derived from data attributes, resulting in decisions that are easy to understand and interpret. Each branch on the tree represents the test results for an attribute, while each leaf node shows the classification results. Decision Tree's high interpretability makes it widely used in various classification studies.

Data Mining

Data mining is the process of extracting useful information or knowledge from large data sets. This process is part of Knowledge Discovery in Databases (KDD), which includes data cleansing, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation. The application of data mining enables organizations to obtain information that can be used as a basis for decision-making.

RapidMiner

open-source software widely used in data analysis and data mining. It provides a variety of operators for data preprocessing, modeling, evaluation, and visualization. Its ease of use through its graphical user interface (GUI) makes RapidMiner widely used in classification and prediction research.

Dataset

A dataset is a collection of data used as a source of information in the analysis process. This study uses a private dataset derived from book loan data at the STIKOM CKI Library. Private datasets have a higher level of confidentiality because they can only be

accessed and used by parties with specific authority or permission. This data is used as the basis for the classification process to identify the books most in demand by library users.

Data Cleaning

Data cleaning is the process of preparing data before analysis. This stage aims to remove incomplete, inconsistent, duplicate, or inappropriately formatted data. Good data quality will improve the accuracy of the classification model, resulting in more reliable analysis results. Therefore, the data cleaning process is a crucial step in data mining applications.

Related Studies

Various previous studies have shown that the Decision Tree algorithm has good performance in solving classification problems in various domains. Decision Tree implementation has been used for student classification, best-selling products, toddler nutritional status, book recommendations, blood donor eligibility prediction, fruit quality classification, disease classification, air quality analysis, student study period prediction, and library system management. The results of previous studies show a high level of accuracy and good interpretability so that this algorithm is considered suitable for application in classifying book loan data at the STIKOM CKI Library.

Based on the literature review, it was found that the Decision Tree algorithm is effective for classifying various types of data, including data related to libraries and reading interests. Therefore, this study applies the Decision Tree algorithm to classify favorite book borrowing data to help librarians manage borrowing, returning, and book collections more effectively and accurately.

3. Materials and Method

Research Data

In this discussion, data collection techniques are required to complement this discussion, requiring data or information from both inside and outside the home. The author obtained relevant data using the following methods:

Field Research

This is research conducted at the company concerned to obtain data related to writing by means of:

Observation

At this stage, direct observation will be carried out at the STIKOM CKI campus to obtain information regarding the criteria required for borrowing and returning books in the library.

Interview

This is a form of research conducted by interviewing relevant parties in the organization to obtain data related to the discussion material.

Literature study

Data collection through various literature that supports information systems that use the Decision Tree Algorithm Method, with the rapidminer application obtained through sources such as journals related to research.

Implementation of Methodology

The steps of the Decision Tree Algorithm are as follows:

a. Starting from the root node, let us assume it is S, which contains the complete dataset.

b. Select the best attribute in the dataset using the Attribute Selection Measure (ASM). Possible ASMs include Information Gain and Gini Index.

c. Split the set S into subsets containing possible values for the best attribute.

d. Create a decision tree node that contains the best attributes.

e. Create new decision tree nodes recursively using subsets of the data set created in step 3. Continue this process until the final stage where we cannot classify any further nodes. This node becomes the final node, also known as a leaf *node* .

For example, there are several data attributes, such as weather , temperature , humidity , and wind . These attributes will determine whether we can play or not. In this case, the output is yes or no.

Temperature	Outlook	Humidity	Windy	Played?
Mild	Sunny	80	No	Yes
Hot	Sunny	75	Yes	No
Hot	Overcast	77	No	Yes
Cool	Rain	70	No	Yes
Cool	Overcast	72	Yes	Yes
Mild	Sunny	77	No	No
Cool	Sunny	70	No	Yes
Mild	Rain	69	No	Yes
Mild	Sunny	65	Yes	Yes
Mild	Overcast	77	Yes	Yes
Hot	Overcast	74	No	Yes
Mild	Rain	77	Yes	No
Cool	Rain	73	Yes	No
Mild	Rain	78	No	Yes

Figure 1. Decision Tree Algorithm Dataset

Types of Decision Trees

The types of decision trees are based on the type of target variable we have. The types of decision trees are:

a. Categorical variable decision tree: A decision tree that has a categorical target variable. For example, the category can be yes or no. Categorical means that each stage of the decision process falls into exactly one group, and there are no in-betweens.

b. Continuous variable decision tree: A decision tree whose target variable is continuous. For example, the income of an unknown individual can be predicted based on available information such as occupation, age, and other continuous variables.

Advantages of Decision Trees

The following are the advantages of the Decision Tree Algorithm:

a. It is easy to understand because it follows the same process as how humans make decisions in real life.

b. Very useful for solving decision related problems.

c. Helps to think of all possible outcomes for a problem.

d. Data cleaning tends to be less than with other algorithms.

Disadvantages of Decision Trees

The following are the disadvantages of the Decision Tree Algorithm:

a. It contains many layers which makes this algorithm quite complicated.

b. Overfitting problems can occur, but they can be solved by using the Random Forest algorithm.

c. For class labels that tend to be numerous, the computational complexity of

Decision trees can be improved.

Test Design

This application program was tested by the user developers involved, who provided input data. Additionally, users tried out various features within the application.

Tools, Materials, Objects and Testing Locations

Tools and materials

For testing here, you need the necessary equipment and materials such as a PC or laptop, and the application being tested.

Object

In accordance with the purpose of creating this web-based application, namely for the STIKOM CKI library, the object of testing is the STIKOM CKI librarian.

Place

The testing location was at the STIKOM CKI Library.

Software Development Design

The design used in software development is to use the Rapid Miner software application and use the following decision tree algorithm method, namely the research stages for the classification of favorite book loans, namely:

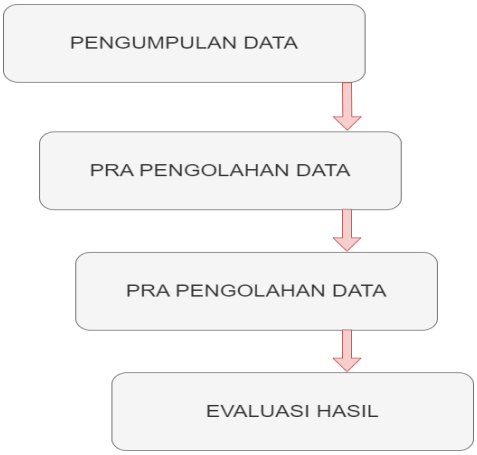

Figure 2. Software Development Stages

Data collection

The data used in this study are data on borrowing and returning books, and book stock obtained from the STIKOM CKI campus library. This data is primary and secondary data obtained directly from the STIKOM CKI campus library. The data is data on borrowing and returning books, and book stock at the STIKOM CKI campus library. The data contains attributes such as transaction number, transaction date, book type, book name, student name, and book stock.

Data Pre-Processing

Data pre-processing involves removing duplicate data, checking for inconsistencies, and correcting errors. This process also includes enrichment, which involves enriching existing data with other relevant data. To obtain high-quality data, several pre-processing techniques are employed:

Data Validation

This is done to identify and remove odd data (outliers/noise), inconsistent data, and incomplete data (missing values). The stock and sales data used in the dataset will be identified for odd data, inconsistent data, and incomplete data.

Data Cleaning

This cleaning process includes, among other things, removing duplicate data, checking inconsistent data, and correcting errors in the collected data. The collected data will be analyzed to determine whether there is inconsistent or irrelevant data that will disrupt the rule patterns of the algorithm to be formed.

Data Cleaning

A decision tree is a tree structure, where each node represents a tested attribute. Each branch represents a division of test results, and leaf nodes represent specific class groups. The top node of a decision tree is the root, which is typically the attribute that has the greatest influence on a particular class. Decision trees generally employ a top-down search strategy for solutions. In the classification process, attribute values are tested by studying the path from the root node to the final node (leaf), and then a new class is determined.

Evaluation of Results

After the classification results are obtained, the results are then evaluated using Cross Validation (confusion matrix) to see the accuracy, precision and recall produced by the proposed model.

4. Results and Discussion

Research Tools

Software Specifications

The following are the software technology requirements used in this research, as follows:

Table 1. Software Specifications

No	Type *Software*	Specification
1	System Operation	*Windows 10*
2	Ms. Excel	2021
3	RapidMiner Studio	10.1

Hardware Specifications

The following is the hardware and its specifications used in this research, as follows:

Table 2 . Specification *Hardware*

No	Type *Hardware*	Specification
1	RAM	4 *GB*
2	CPU	*AMD Ryzen 5*
3	Storage	500 GB
4	Laptop	*Mobile phone*

Test Implementation

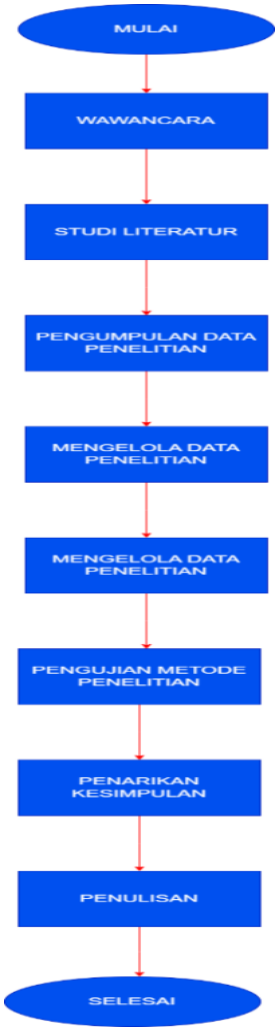

Figure 3. Research Stages

In the research stage image above, the author uses the Decision Tree Algorithm in this THESIS research, from the problems that exist in the STIKOM CKI library where the management of book loan and return data is currently still done manually with these problems, the author proposed a research idea to apply the Decision Tree Algorithm, after reading references and journals with the research theme that will be undertaken, then determining the variables to be studied. The author collects data from interviews with STIKOM CKI Librarians to be processed and applied with the Decision Tree Algorithm Method, the results of which will be implemented in the Rapid Miner Application on the STIKOM CKI campus.

Sample Stage

In the data set there is data on various books from the Stikom CKI campus library according to those used for the data sample, consisting of Description, Prediction, Confidence, Title, Borrowing Date, Borrowing Time, Return Date, Return Time .

Row No.	KETERANG...	prediction(K...	confidence(...	confidence(...	JUDUL	TANGGAL P...	WAKTU PE...	TANGGAL P...	WAK
1	TIDAK	TIDAK	1	0	Akuntansi Pen...	Jan 1, 2022	8:30:00 AM ICT	Jan 2, 2022	9:40:
2	TERLAMBAT	TERLAMBAT	0	1	Biografi Gus D...	Jan 6, 2022	1:30:00 PM ICT	Jan 10, 2022	2:40:
3	TERLAMBAT	TERLAMBAT	0	1	Cedera Kepala	Jan 8, 2022	3:30:00 PM ICT	Jan 12, 2022	4:40:
4	TERLAMBAT	TERLAMBAT	0	1	Dasar Dasar ...	Jan 9, 2022	8:30:00 AM ICT	Jan 13, 2022	9:40:
5	TERLAMBAT	TERLAMBAT	0	1	Etnografi Peng...	Jan 10, 2022	9:30:00 AM ICT	Jan 14, 2022	10:40
6	TERLAMBAT	TERLAMBAT	0	1	Hantu Digoel; ...	Jan 11, 2022	10:30:00 AM I...	Jan 15, 2022	11:40
7	TERLAMBAT	TERLAMBAT	0	1	Islam Jawa; K...	Jan 13, 2022	12:30:00 PM I...	Jan 17, 2022	1:40:
8	TERLAMBAT	TERLAMBAT	0	1	Keanekaraga...	Jan 16, 2022	3:30:00 PM ICT	Jan 20, 2022	4:40:
9	TIDAK	TERLAMBAT	0	1	Kesehjateraan ...	Jan 19, 2022	10:30:00 AM I...	Jan 19, 2022	11:40
10	TIDAK	TIDAK	1	0	Kontroversi Ha...	Jan 21, 2022	12:30:00 PM I...	Jan 21, 2022	1:40:
11	TIDAK	TERLAMBAT	0	1	Menuju Punca...	Jan 28, 2022	11:30:00 AM I...	Jan 28, 2022	12:40
12	TERLAMBAT	TERLAMBAT	0	1	Neo Patriotism...	Jan 31, 2022	2:30:00 PM ICT	Feb 4, 2022	3:40:
13	TERLAMBAT	TIDAK	1	0	NU vis a vis N...	Feb 1, 2022	3:30:00 PM ICT	Feb 5, 2022	4:40:
14	TIDAK	TIDAK	1	0	Patologi I (umu...	Feb 5, 2022	11:30:00 AM I...	Feb 5, 2022	12:40
15	TERLAMBAT	TIDAK	1	0	Runtuhnya Ke...	Feb 12, 2022	10:30:00 AM I...	Feb 16, 2022	11:40
16	TERLAMBAT	TIDAK	1	0	Sejarah Pendi...	Feb 13, 2022	11:30:00 AM I...	Feb 17, 2022	12:40

Figure 4. Sample Data

Explore Stage

Data exploration will be carried out at the explore stage by selecting unexpected trends or anomalies. After the dataset is successfully obtained at this stage, it is used for data understanding and selection . After that, data Understanding is carried out which contains 200 Example data, 3 Special Attributes, 3 Regular Attributes namely Description, Prediction, Confidence, Title, Borrowing Date, Borrowing Time, Return Date, and Return Time to perform Decision Tree. then using Rapidminer software with Title and author labels. finished doing data Understanding After that, the data selection process involves Split data. And the ratio of 0.7 and 0.3 to perform the performance of the model formed by Decision Tree.

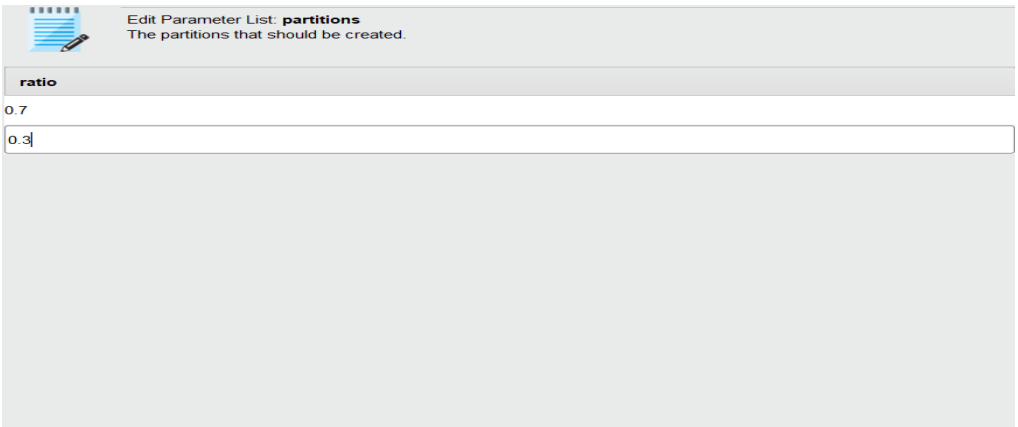

Figure 5. Split Data

In Figure 4.3, Split Data is performed with the 2x setting and parameter as the sampling type. This divides or separates the training and test data to see how well the Decision Tree model performs. The ratio or capacity of the training and test data must then be set. Select Edit Numbering from the menu. The following dialog box will open; enter a value and then select OK. 0.7 indicates that 70% of the comparisons between datasets are taken from the training data, while 0.3 indicates that 30% are taken from the test data.

Model Stage

In the model stage process, Rapidminer Studio 10.1 software was used with a dataset of 200 rows of data with the parameters Description, Prediction, Confidence, Title, Borrowing Date, Borrowing Time, Return Date, and Return Time.

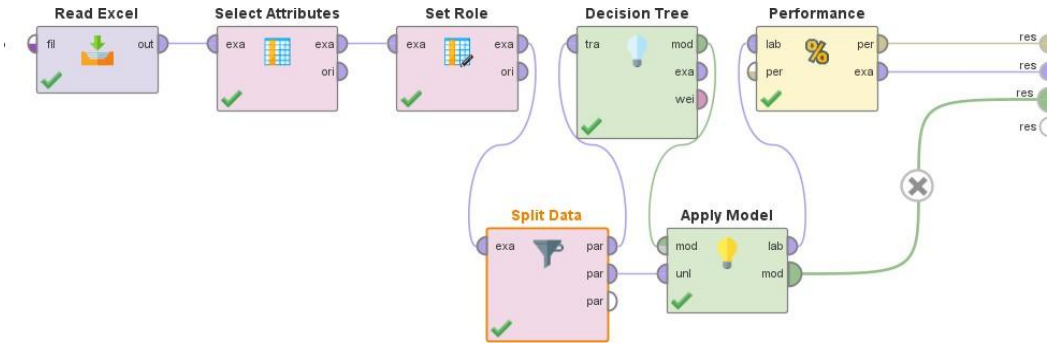

Figure 6. Decision Tree Process in Rapidminer

In figure 4.4 explains the Decision Tree process in rapidminer, the first step is to read excel data with 200 examples, 2 special attributes, 3 regular attributes after that do an example filter to filter or filter empty data or missing attributes. then split the data to separate the two parts, the first part is used to evaluate or test the data and the other data is used to train the model, after that do the decision tree algorithm process the next step is to apply the model to apply the performance that has been considered previously using excel read data. after that do a simulator model to simulate the prediction of total book borrowing on the independent variables that will be input, the last step is to perform performance un evaluate the performance of the model that provides a list of performance criteria values automatically according to the given task.

Assess Stage

At this stage, the accuracy and classification error results were obtained using Rapidminer Studio 10.1. The results are as follows:

classification_error: 20.34%

	true TIDAK	true TERLAMBAT	class precision
pred. TIDAK	20	4	83.33%
pred. TERLAMBAT	8	27	77.14%
class recall	71.43%	87.10%	

Figure 7. Classification error results

accuracy: 79.66%

	true TIDAK	true TERLAMBAT	class precision
pred. TIDAK	20	4	83.33%
pred. TERLAMBAT	8	27	77.14%
class recall	71.43%	87.10%	

Figure 8. Accuracy results

Based on the image above, the performance results obtained a classification error value and accuracy value of 20.34%/79%66, which means that the predicted value is close to the observed value because a lower value indicates a better value.

Prediction Results

Prediction results obtained from a large number of favorite book borrowing data using the decision tree algorithm in Rapidminer Studio Version 10.1 software.

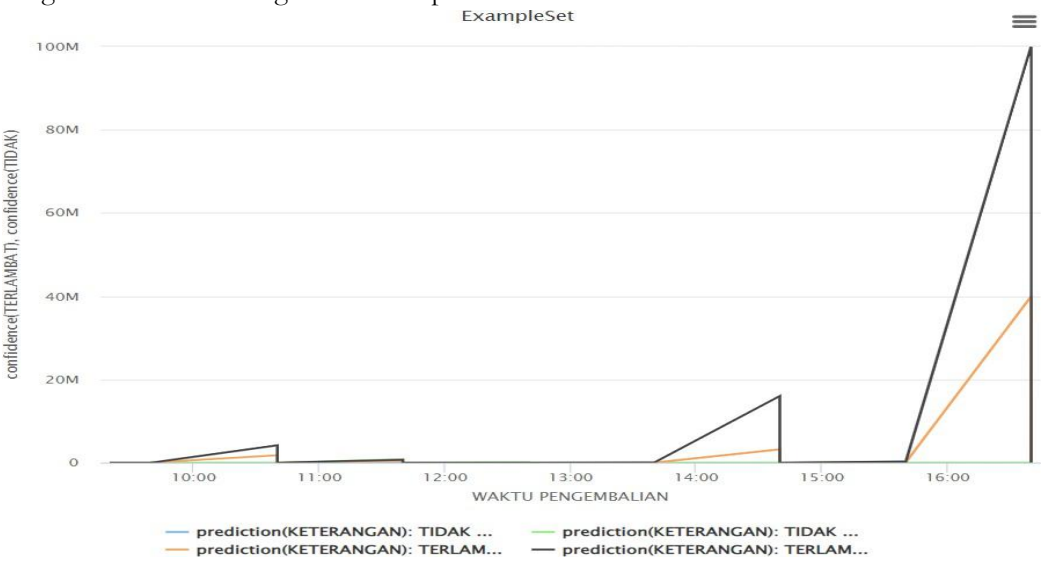

Figure 9. Visualization of return information with prediction

The line chart visualization in the image above shows that the black line represents late predictions and the orange line represents not late. The line chart visualization shows that there is no significant difference between the not late and late prediction values.

Final Test Results

To determine the accuracy of the prediction results in recording the borrowing and return of favorite books, the lower the error value, the better the prediction.

6. Conclusion

Conclusion

Based on the results of the research that has been conducted, the application of the Decision Tree algorithm to the classification of favorite book loan data at the STIKOM CKI Library can help the library data management process to be more effective and efficient. The system built can simplify and accelerate the process of storing and searching data needed by librarians. In addition, the classification process using the Decision Tree algorithm allows book loan data to be grouped in a structured manner so that the required information can be obtained more easily. The implementation of the system also supports the digitization of library data management, so that the data storage process becomes more organized and is able to improve the quality of services at the STIKOM CKI Library.

Suggestions

Based on the research results, future system development can include adding layered security mechanisms to protect book borrowing and return data from unauthorized access. Furthermore, improvements to the data search feature are needed to speed up and accurately retrieve information. Developing the application on a mobile platform could also be an alternative to improve ease of access, flexibility of use, and efficiency of library data management, thereby providing more optimal service to users.

References

1. Sutoyo, I. (2018). Implementation of Decision Tree Algorithm for Student Data Classification. *PILAR Nusa Mandiri* , *14*(2), 217–224.
2. Nasrullah, AH (2021). Implementation of the Decision Tree Algorithm for Classifying Best-Selling Products. *Scientific Journal of Computer Science* , *7*(2), 45–51.
3. Ahmad Zam Zami, ONGD (2022). Classification of Nutritional Conditions of Infants Under Five Years Old at the Melati Integrated Health Post Using the Decision Tree Algorithm. *Journal of Computer Systems and Informatics (JSON)* , *3*(3), 305–310.
4. Maulidah, M., Gunawan, W., & Chandra, RA (2020). Decision Tree Classification Algorithm for Book Recommendations Based on Book Categories. *JOURNAL OF ECONOMICS AND BUSINESS SCIENTIFIC* , *13*(2), 89–96.
5. Qadrini, L., Sirajuddin, A., & Aulia, A. (2021). Decision Tree and Adaboost in Classifying Recipients of Social Assistance Programs. *Journal of Research Innovation* , *2*(17), 1959–1965.
6. Irawan, Y. (2021). Application of the C4.5 Decision Tree Algorithm to Predict the Eligibility of Prospective Donors to Donate Blood Using Data Mining Classification. *JTIM: Journal of Information Technology and Multimedia* , *2*(4), 181–189.
7. Pamungkas, YW, & Qomariyah, AD (2017). Snake Bite Image Classification Using Regionprops and Decision Tree Algorithm. *Journal Speed Center for Engineering and Education Research* , *9*(2), 1–7.
8. Robianto, SH, & Ramadhan, UR (2021). Application of the Decision Tree Method to Classify Orange Fruit Quality Based on Color and Size Features. *Journal of Computers and Applications* , *9*(1), 76–86.
9. Nursyahfitri, R., Maulana, AN, & Firdaus, YU (2021). Classification of Drug Type Determination Using Decision Tree Algorithm. *JIP (Jurnal Informatika Polinema)* , *7*(3), 53–60.
10. Sang, AI, & Syafrudin, E. (2021). Data Mining Analysis for Classification of DKI Jakarta Air Quality Data Using Decision Tree Algorithms and Support Vector Machines. *e-Proceeding of Engineering* , *8*(5), 8954–8963.
11. Pambudi, RE, & Fadlilah, S. (2020). Classification of Stroke Disease Using Decision Tree C Algorithm. 45. *JOURNAL OF TECHNIQUES* , *16*(2), 221–226.
12. Pane, FN (2020). Design and Construction of a Web-Based Kraepelin Test Application Using the Bubble Sort Method. *JOINTECS (Journal of Information Technology and Computer Science)* , *5*(1), 41–48.
13. Yulita, L., Suryadi, A., & Ningsih, S. (2022). Genetic Algorithm Optimization in Predicting Student Reading Interest at the SMK Negeri 1 Gantar Library Using the Decision Tree Method. *Journal of Practical Computer Science (JPCS)* , *2*(1), 15–23.
14. Rahmianti, I. (2022). Analysis of the Feasibility of Providing Cooperative Credit Using the Decision Tree Data Mining Method. *JIRE (Journal of Informatics & Electronic Engineering)* , *5*(2), 153–161.
15. Partogi, Y. (2021). Designing the Decision Tree Method for the STMIK Kuwera Library System. *JOURNAL OF INFORMATION SYSTEMS AND TECHNOLOGY (SINTEK)* , *1*(2), 20–26.
16. Santoso, P., & Sari, R. (2020). Application of Decision Tree Classification Method and C4.5 Algorithm in Predicting Mega Auto Finance Credit Customer Criteria. *JURIKOM (Journal of Computer Research)* , *7*(2), 200–206.
17. Gunawan, I., Santoso, S., & Himawan, T. (2019). The Use of Bubble Sort Algorithm to Determine Student Achievement Grades. *SISTEMASI (Journal of Information Systems)* , *8*(2), 296–304.
18. Hana, FM (2020). Classification of Diabetes Patients Using the C4 Decision Tree Algorithm. 5. *Journal of Computer Systems and Artificial Intelligence* , *4*(1), 32–39.
19. Cynthia, EP, & Iskandar, E. (2018). Implementation of the Bubble Sort Algorithm in Library Systems Using the CodeIgniter Framework. *Journal of Information Systems and Informatics Engineering Research (JURASIK)* , *3*, 1–13.

20. Orpa, EPK, Fendrik, E., & Tinus, R. (2019). Prediction Model for the Initial Study Period of Students Using the Decision Tree C4 Algorithm. 5. *JUSTIN (Journal of Information Systems and Technology)* , *7*(4), 272–278.