Journal of Mathematics Instruction, Social Research and Opinion
Vol. 4, No. 1, March 2025, pp. 195 - 204, https://doi.org/10.58421/misro.v4i1.359
ISSN 2962-7842

Decision Tree Methodology (C4.5) for Predicting Students' Reading Interest in the Library SMK Negeri 1 Kota Cirebon

Muhammad Erwanto[1], Kosim[2], Nur Bambang Riyanto[3], Sukmo Banyu Jogo[4]
[1,2,3,4]Sekolah Tinggi Ilmu Komputer Poltek Cirebon, Indonesia

Article Info

Article history:

Received 2025-01-03
Revised 2025-02-24
Accepted 2025-02-28

Keywords:

C4.5 Algorithm
Data Mining
Decision Tree
Library
Reading Interest

ABSTRACT

Reading is one aspect of language skills that is actively receptive. The media used in reading is written language media. Reading is seeing and understanding the content of what is written, either spelling or pronouncing what is written. Reading activities are often socialised in education because reading is a very important activity to support teaching and learning activities at school. The facility provided by the school as a support in socialising reading activities for students is the library. Many students often utilise the SMK Negeri 1 Kota Cirebon library to carry out the borrowing process and read books there. Reading activities are an obligation that students must carry out, but students who carry out reading activities cannot be categorised as students with an interest in reading. The problem faced by the SMK Negeri 1 Cirebon City library is that it has not been able to predict or know the reading interests of students in the school library. This study uses data mining techniques with the C4.5 algorithm to predict student reading interest. This research produces rules to help SMK N 1 Cirebon City predict student reading interest in the school library. This step is done by designing a system model that uses the C4.5 algorithm to form a decision tree to produce a rule for predicting student reading interest. This research will produce valuable information about predicting student reading interest in the SMK Negeri 1 Cirebon City library using the C4.5 algorithm method.

Corresponding Author:

Muhammad Erwanto
Sekolah Tinggi Ilmu Komputer Poltek Cirebon, Indonesia
Email: erwanto@stikompoltek.ac.id

1. INTRODUCTION

The rapid development of digital technology has significantly impacted the dissemination and accessibility of information and communication in society, causing substantial changes in lifestyles, learning patterns, and cultural habits. One notable example of these changes is the shift in book-reading behaviour. With the increasing availability of digital books and online reading platforms, students now have easier access to a wide range of reading materials. Consequently, this convenience has altered students' traditional behaviour in utilising school libraries as a primary source of reading materials [1].

Reading is a fundamental aspect of language skills that is actively receptive. It involves not only recognising written symbols but also comprehending their meaning, either through silent reading or verbal articulation. In the modern era, where access to knowledge is crucial, reading has become essential for academic and professional success. However, reading activities do not necessarily indicate a genuine interest in reading. Reading interest is an intrinsic motivation to consistently and meaningfully engage with written materials. Schools often promote reading activities to enhance student literacy, and one of the primary facilities provided to support this initiative is the school library [2].

The SMK Negeri 1 Kota Cirebon library is an essential resource centre, supporting students and teachers in accessing scientific information and academic materials. Students frequently utilise the library to borrow books, read, and gather reference materials. The library has various facilities to enhance its role, including textbooks, teaching resource books, and digital materials. However, despite these provisions, challenges persist in maximising the library's potential as a hub for fostering students' reading interest.

Despite the recognised importance of reading in education, many students lack enthusiasm for reading. This issue negatively impacts the effectiveness of the learning process and student engagement in academic activities. At SMK Negeri 1 Kota Cirebon, the library primarily functions as a book-lending facility, but it does not yet incorporate mechanisms to assess or cultivate students' reading interests. As a result, opportunities to tailor reading materials to students' preferences and needs are missed, limiting the library's role in fostering a reading culture.

The library's primary challenge is the absence of data-driven tools to analyse and predict students' reading interests. Without such tools, schools struggle to develop strategies and policies that could effectively enhance students' engagement with reading materials. Consequently, this lack of insight may contribute to students' declining interest in reading.

This study addresses this issue by employing data mining techniques with the C4.5 decision tree algorithm to predict students' reading interest at SMK Negeri 1 Kota Cirebon. Data mining enables the extraction of meaningful patterns from large datasets, allowing for the development of predictive models that can enhance decision-making processes. This research aims to generate classification rules that assist schools in identifying students' reading preferences and improving library services accordingly [3], [4].

The C4.5 algorithm is a widely used decision tree method that facilitates classification and prediction by generating rules based on observed data patterns. Its ability to handle categorical data and produce interpretable decision rules makes it a suitable approach for analysing students' reading behaviour. By implementing this algorithm, the study seeks to provide a systematic framework for predicting reading interest, ultimately contributing to optimising school library services [5].

2. METHOD

The author used the descriptive-analytic method as the research methodology in this study. This method systematically analyses and describes the observed data's characteristics, patterns, and relationships. Based on information acquired naturally or precisely from the subject or item being studied, the descriptive analysis approach aims to characterise, explain, or describe the subject or object under investigation, such as an institution, society, or specific phenomenon [6]. The following research methods are employed: (1) Observation, which serves as the cornerstone of scientific inquiry, is crucial in data collection, as scientists rely on empirical facts gathered from real-world observations [7], and (2) Interviews, which involve structured or semi-structured interactions between two or more individuals to exchange ideas, gather information, and develop an in-depth understanding of a particular topic [7].

Data mining is a multidisciplinary field that addresses the challenge of extracting meaningful patterns and insights from large datasets. It integrates techniques from various domains, including machine learning, pattern recognition, statistics, database management, and visualisation. Through these approaches, data mining facilitates the discovery of significant trends that might remain unnoticed. On the other hand, data mining is also defined as employing pattern recognition software, statistical methods, and mathematics to identify patterns in massive quantities of data stored in digital repositories [8], [9].

The following are characteristics of data mining [9], [10]: (1) It involves identifying both hidden and previously recognised patterns within datasets; (2) It is typically applied to large-scale datasets to enhance the reliability and generalizability of results; (3) It aims to maximise the utility of extensive data sources to improve analytical outcomes; and (4) It is instrumental in supporting critical decision-making processes, particularly in strategic planning and business intelligence.

Data mining, often called Knowledge Discovery in Databases (KDD), is an activity that involves systematically analysing and utilising historical data to find regularities, patterns, or relationships in large datasets. As an advanced analytical technique, KDD enables organisations and researchers to transform raw data into actionable knowledge. Although pattern recognition remains a fundamental component of data mining, the term "data mining" has become more widely adopted, while KDD is used less frequently in contemporary research [9], [11], [12], [13].

The following are the steps in the data mining process [9]: (1) Data cleaning, which involves removing noise, inconsistent data, and irrelevant entries to ensure high-quality datasets; (2) Data integration, where data from multiple sources are merged into a unified repository for analysis; (3) Data selection, which identifies and extracts relevant data for analytical processing; (4) Data transformation, which converts raw data into a structured format suitable for mining techniques; (5) Mining, the core process where computational techniques uncover hidden patterns and valuable insights within the dataset; (6) Pattern evaluation, which assesses the discovered knowledge to ensure its validity and relevance; and (7) Knowledge presentation, where the extracted insights are visualised and communicated effectively to end-users. At this stage, the effectiveness of the applied data mining techniques is evaluated to determine whether the formulated hypothesis is supported by empirical evidence [13], [14]. The procedure for data mining is depicted in Figure 1 below:

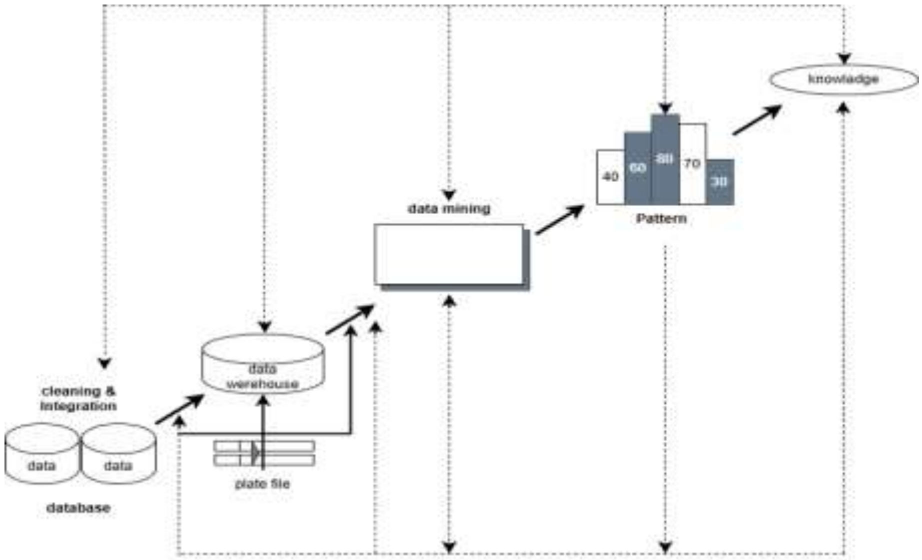

Figure 1. Data Mining Process

The art and science of identifying future events is called forecasting. This can be accomplished by using a mathematical model to project historical data into the future. Forecasting and prediction are typically categorised according to the future time horizon, which is separated into multiple groups, including the following: (1) Short-term: Usually shorter than three months but encompassing up to a year. (2) Medium-term, usually spanning a few months to three years, this forecast is used to plan buying, work schedules, labour quantities, work assignments, and output levels. (3) Long-term, usually for forecasting three years or more; and (3) Cash budgeting, sales planning, manufacturing budgets, and various operating plans can all benefit from prediction. Research and development, capital expenditures, site selection or design, and new product development are all planned using long-term forecasting [15], [2], [16].

The C4.5 algorithm is one of the algorithms used in data mining to build decision tree models. The C4.5 algorithm uses machine learning to generate prediction models based on the data provided. In general, the C4.5 Algorithm accuracy formula i One of the techniques applied in machine learning to create decision tree models is the C4.5 algorithm. Based on the given data, the C4.5 algorithm creates prediction models using the machine learning idea. The accuracy formula for the C4.5 Algorithm is generally:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

with:

TP = True positive

TN = True negative

FP = False positive

FN = False negative

A decision tree is a decision-making method that organises each option into a branching form. This method is called a decision tree because it looks like a tree with many branches in its twigs and roots [14].

Modelling stages using the C4.5 algorithm include [15]: (1) Prepare training data; namely, at this stage, the dataset will be used as training data, and the dataset in the previous process has been grouped into certain classes; (2) Determine the root of the tree. At this stage, the entropy value is calculated to determine the tree's root based on the highest information gain value. Entropy is used to measure the level of diversity of a data set. The formula for Entropy is as follows [14], [17], [18]:

$$Entropy(S) = -\sum_{i=0}^{n} p_i \times (\log p_i)^2 \quad (2)$$

With:
S = Set of cases
n = number of partitions
p_i = Total sample proportion for the i-th class

(3) Determine the value of information gain, namely information gain, which is one way to determine the selection of attributes at each node in the tree. Calculation of information gain uses the following formula:

$$Gain\ (S, A) = S - \sum_{i=0}^{n} \frac{|S_i|}{|S|} \times S_i \quad (3)$$

Remarks:
S_i = Entropy for the case that has the i-th value
S = Set of cases
n = Number of attribute partitions A
$|S_i|$ = number of i-th value cases
$|S|$ = Total Number of cases in S

(4) Repeat the following procedure. That is, continue until all the records have been divided. (5) determination of the decision tree, which is based on the outcomes of modelling using the c4.5 algorithm, is based on the feature with the greatest Entropy at the top. This stage will end if all records in node N receive the same class, if no attributes in the record can be partitioned again, and if there are no records in the unused tree branch [19], [20].

3. RESULTS AND DISCUSSION

Carrying out the reading interest prediction process requires data on student visits to the SMK Negeri 1 Cirebon City library. The C4.5 algorithm is one of the classification techniques. Each data set must have labels and ratings that indicate attributes. The C4.5 algorithm also can classify extensive data very quickly.

This part provides a thorough discussion and explains the research findings. Figures, graphs, tables, and other easily comprehensible formats can be used to convey results [15],[21],[22]. The conversation can be divided into a number of smaller portions.

All the existing data attributes will be managed and classified into several categories to search for entropy, gain and decision trees. After going through analysis, data processing and data filtering, the following table is produced:

a) Prepare Attribute Data

The following are the attributes used in the calculation of the C4.5 algorithm in Table 1:

Table 1. Data Attribute

Number	Name of Attribute	Status Attributes
1	Collection	Known
2	Quality	Known
3	Comfort	Known
4	Service	Known
5	Interest	Wanted

Table 1 describes the attributes that will be used as a reference for assessing students' interest in reading. The first is a collection of the availability and diversity of reading materials that greatly affect students' interest in reading. The quality of reading materials, including content, language and presentation of information, can influence interest in reading. Comfort A comfortable reading environment and good services, such as easy access to reading materials, librarian assistance, or an efficient lending system, can enhance the reading experience and attract more students to read. All attributes that have been described are already known, which are asked, or the results of these attributes will predict/classify students' interest in reading, which is not yet known.

b) Preparing Data Analysis

Table 2 below displays the results of the analysis of the available data:

Table 2. Data Training

No	Activity	Type	Collection	Quality	Comfort	Service	Interest
1	Read	Fiction	Complete	Good	Comfortable	Good	Interest
2	Read	Non- fiction	Complete	Good	Comfortable	Enough	Interest
3	Borrow	Fiction	Complete	Good	Comfortable	not good enough	not interested
4	Return	Non- fiction	Complete	Good	Comfortable Enough	Good	Interest
5	Borrow	Fiction	Complete	Good	Comfortable Enough	Enough	Interest
6	Read	Non- fiction	Complete	Good	Comfortable Enough	not good enough	not interested
7	Return	Fiction	Complete	Good	Less Comfortable	Good	Interest
8	Borrow	Non- fiction	Complete	Good	Less Comfortable	Enough	not interested
				. . .			
81	Return	Fiction	Incomplete	not good enough	Less Comfortable	not good enough	not interested

Data analysis is a step in examining the information required to create the system's design. In this instance, the writers gather material from literature relevant to the research topic and compile discussed hypotheses to create a complicated blend between them [23]. Table 2 shows data training with attributes determined with 81 data points. This data training is for data testing / new data. Data Training has been prepared, and then preprocessing will be done for the next step. The steps of operating the student reading interest prediction application program at the SMK Negeri 1 Cirebon City Library that has been made are as follows:

a) Entropy and Gain Calculation Results

Amount of data = 81
Total interest = 43
Number of disinterests = 38
Entropy all = 0.997

Table 3. Entropy and Gain

Attribute Value	Amount of Data	Number of Interests	Number of No Interest	Entropy	Gain
collection='Complete”	27	16	11	0.975	
collection='Enough'	27	15	12	0.991	
collection='Less'	27	12	15	0.991	0.011
type='fiction'	41	24	17	0.979	
type='non-fiction'	40	19	21	0.998	0.009
quality='good'	27	15	12	0.991	
quality='sufficient'	27	16	11	0.975	
quality='less'	27	12	15	0.991	0.011
comfort='cozy'	27	16	11	0.975	
comfort='enough'	27	14	13	0.999	

Table 3 shows that the attribute with the highest Gain is the best for the first split in the decision tree. For example, quality = 'less' has a Gain of 0.011, meaning it is slightly better at reducing uncertainty than other attributes. Table 3 shows the calculation of Entropy and information gain in the decision tree algorithm (4.5) to determine the best attribute in classifying data based on interest. Each attribute is evaluated based on how it splits the data into groups with lower Entropy. The higher the Gain value, the more effectively the attribute splits the data.

b) Decision Tree Result

Table 4 shows the classification results using a Decision Tree to determine whether someone is interested in books based on class, major, activity, type of book, book collection, book quality, comfort, and service. Classification results are: Of the 8 data, six predictions are true, and two predictions are false. Prediction errors occurred in the third and sixth rows, where the model predicted interest but no interest. This Decision Tree model is accurate as it only has two errors out of 8 data (about 75% accuracy). Factors influencing classification are user activity, book type, convenience, and service.

Table 4. Decision Tree Result

No	Name	Class	Major	Activity	Book type	Book collection	Book Quality	Convenience	Service	Ori class	forecast	result
1	Abdul Muchlisin	X	Mechanical Engineering	reading	fiction	Complete	good	convenient	convenient	interest	interest	True
2	Sholikan	X	Electrical Engineering	reading	non-fiction	Complete	good	convenient	Enough	interest	interest	True
3	Istiqoma	XI	Automotive Engineering	borrow	fiction	Complete	good	convenient	less	no interest	interest	false
4	Mochammad Yhusroni	XI	Computer Network Engineering	Return	non-fiction	Complete	good	Enough	convenient	interest	interest	True
5	Ali Ashari	XII	Software Engineering	borrow	fiction	Complete	good	Enough	Enough	interest	interest	True
6	Luvenaya	XII	Electrical Engineering	reading	non-fiction	Complete	good	Enough	less	no interest	interest	false
7	Salimun	XI	Mechanical Engineering	Return	fiction	Complete	good	less	convenient	interest	interest	True
8	Mohammad Ridwan	XI	Software Engineering	borrow	non-fiction	Complete	good	Enough	Enough	interest	interest	True

4. DISCUSSION

The results of this study use the Decision Tree algorithm to classify a person's interest in books based on several attributes such as activity type, book type, book quality, convenience, and service. From the classification results, the model has an accuracy rate of about 75%, with six correct predictions and two incorrect predictions from 8 test data. This research is in line with previous studies that use machine learning methods in user preference analysis, such as studies that apply Naïve Bayes, K-Nearest Neighbors (KNN), or Support Vector Machine (SVM) algorithms in reading interest classification [19], [24], [25], [26]. In some studies, the Decision Tree method is often used because it is easily understood and highly interpretable. However, compared to other studies that use deep learning or ensemble learning-based methods (Random Forest, Gradient Boosting), the Decision Tree algorithm has limitations in handling more complex data and is prone to overfitting, especially if the amount of data used is small [23], [10].

While the Decision Tree algorithm provides transparency in decision-making, its predictive accuracy remains lower than that of more sophisticated models. Previous research utilising Random Forest or deep learning techniques has reported higher accuracy rates, often reaching 80-90%. The lower accuracy in this study could be attributed to various factors, including the relatively small dataset, the limited number of features used for classification, or potential biases in data collection. Additionally, the model does not account for external variables such as students' digital reading habits or personal motivations, which might significantly influence their reading interests.

These findings highlight the potential for integrating machine learning-based classification into school library management systems. By leveraging predictive models, libraries can tailor their book collections and services based on students' reading preferences. Furthermore, implementing a personalised recommendation system based on classification results could enhance students' engagement with library resources and foster a stronger reading culture. However, ensuring the effectiveness of such a system requires continuous refinement of the model, incorporating additional features such as reading frequency, subject preferences, and borrowing history.

Despite its promising application, this study has several limitations. The small dataset may have affected the model's ability to generalise across a broader population of students. Additionally, while interpretable, the Decision Tree algorithm tends to overfit when dealing with complex datasets. Future research should consider employing ensemble learning approaches like Random Forest or Gradient Boosting and expanding the dataset to include more diverse reading behaviours. Integrating other machine learning techniques, such as hybrid models combining Decision Trees with deep learning, may also improve prediction accuracy and provide deeper insights into students' reading habits.

5. CONCLUSION

This study demonstrates that the Decision Tree (C4.5) algorithm can be effectively applied to predict students' reading interest in school libraries. By utilising key attributes such as book collection, quality, comfort, and library services, the model could classify students' reading interests with an accuracy of 75%. These findings highlight the potential of data mining techniques in enhancing library management by enabling data-driven decisions regarding book selection and service improvements.

Compared to other classification methods such as Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM), the Decision Tree method offers high interpretability, making it a suitable choice for educational settings. However, its predictive accuracy is still lower than that of ensemble learning methods like Random Forest or Gradient Boosting, which can handle complex data more effectively. The main limitation of this study lies in the relatively small dataset used, which may have affected the model's generalizability. Additionally, the classification attributes did not include external factors such as students' digital reading habits and personal motivations, which could further refine the model's predictive capabilities.

Future research should incorporate a larger dataset with more diverse reading behaviour attributes to improve accuracy and applicability. Exploring advanced machine learning techniques, such as ensemble learning or hybrid models, may enhance predictive performance. By integrating these improvements, school libraries can develop more accurate and adaptive systems to support students' reading engagement, ultimately fostering a stronger reading culture.

REFERENCES

[1] Luluk Elvitaria, Salamun, and Esni Malau, "Penerapan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Pengguna E-Learning Pada Mahasiswa Universitas Abdurrab," *JEKIN - J. Tek. Inform.*, vol. 4, no. 2, pp. 391–397, 2024, doi: 10.58794/jekin.v4i2.898.

[2] A. W. Wicaksono and T. Setiadi, "Penerapan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar di Kecamatan Juai," *Format J. Ilm. Tek. Inform.*, vol. 12, no. 2, p. 151, 2023, doi: 10.22441/format.2023.v12.i2.008.

[3] N. Nazifah, C. Prianto, and C. A. Id, "Decision Tree Algoritma C4.5 dengan algoritma lainnya: Sistematic Literature Review," *J. Inform. dan Teknol. Komput.*, vol. 04, no. https://ejurnalunsam.id/index.php/jicom/, pp. 57–64, 2023, [Online]. Available: https://ejurnalunsam.id/index.php/jicom/.

[4] D. Muriyatmoko, A. Musthafa, and M. H. Wijaya, "Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree," *Semin. Nas. Sains dan Teknol. 2024 Fak.*, no. April, 2024.

[5] S. Yunita and V. N. Alaeyda, "Penerapan Algoritma C4.5 Untuk Prediksi Penerimaan Beasiswa di SD 4 Pelangsian," *ICIT J.*, vol. 8, no. 2, pp. 181–193, 2022, doi: 10.33050/icit.v8i2.2408.

[6] Sudaryono and N. Sukmadinata Syaodih, *Metode Penelitian Pendidikan*. 2006.

[7] F. Nurdiansyah and H. S. Rugoyah, "Strategi Branding Bandung Giri Gahana Golf Sebelum dan Saat Pandemi Covid-19," *J. Purnama Berazam*, vol. 2, no. 2, p. 159, 2021.

[8] L. Fajarita *et al.*, "Penerapan Forecasting Stright Line Method Dalam Pengadaan Stok Barang Mendatang," pp. 310–317, 2018.

[9] A. J. P. Sibarani, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 262–276, 2020, doi: 10.35957/jatisi.v7i2.195.

[10] Y. Wanli Sitorus, P. Sukarno, S. Mandala, F. Informatika, and U. Telkom, "Analisis Deteksi Malware Android menggunakan metode Support Vector Machine & Random Forest," *e-Proceeding Eng.*, vol. 8, no. 6, pp. 12500–12518, 2021.

[11] Anggi Trifani, Agus Perdana Windarto, and Hendry Qurniawan, "Penerapan Data Mining Klasifikasi C4.5 dalam Menentukan Tingkat Stres Mahasiswa Akhir," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 2, pp. 91–105, 2022, doi: 10.55606/jurritek.v1i2.414.

[12] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *J. Teknol. Rekayasa*, vol. 3, no. 1, p. 89, 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.

[13] I. Iddrus and D. W. Sari, "Penerapan Data Mining Menggunakan Algoritma Decision Tree C4.5 Untuk Memprediksi Mahasiswa Drop Out Di Universitas Wiraraja," *J. Adv. Res. Inform.*, vol. 1, no. 02, pp. 1–7, 2023, doi: 10.24929/jars.v1i02.2684.

[14] A. Nurzaman and T. N. Suharsono, "Implementasi Metode Algoritma C4.5 Untuk Prediksi Gejala Autisme Pada Anak," *Innov. J. Soc. Sci. Res.*, vol. 3, no. 5, pp. 848–863, 2023.

[15] 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.

[16] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, [Online]. Available: http://jesik.web.id/index.php/jesik/article/view/91.

[17] S. Mulyanda, S. Defit, and Sumijan, "Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas," *J. KomtekInfo*, vol. 10, pp. 116–121, 2023, doi: 10.35134/komtekinfo.v10i3.427.

[18] N. Gusrialni Fitri, S. Adilya, and F. Azizi, "Comparison of the Naive Bayes Classification System and C4.5 for the Diagnosis of Stroke Perbandingan Sistem Klasifikasi Naive Bayes dan C4.5 Untuk Diagnosa Penyakit stoke," *SENTIMAS Semin. Nas. Penelit. dan Pengabdi. Masy.*, pp. 49–55, 2023.

[19] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 193–199, 2023, doi: 10.37859/coscitech.v4i1.4755.

[20] N. Attamami, A. Triayudi, and R. T. Aldisa, "Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 262–269, 2023, doi: 10.35870/jtik.v7i2.756.

[21] A. B. Yamantri and A. Ahmad, "Penerapan Algoritma C4 . 5 Untuk Prediksi Faktor Risiko Obesitas Pada Penduduk Dewasa," vol. 2, no. c, pp. 118–125, 2024.

[22] Y. Indah, "Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C4.5," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 59–64, 2022, doi: 10.37034/infeb.v4i2.99.

[23] D. Yani, "G-Tech : Jurnal Teknologi Terapan Feature Management Approaches and Predictive Methods in Machine Learning," vol. 9, no. 1, pp. 530–539, 2025.

[24] F. Pirmansyah and T. Wahyudi, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Evaluasi Anggota Satuan Pengamanan Studi Kasus Pt. Yimm Pulogadung," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1566–1580, 2023, doi: 10.35870/jimik.v4i3.370.

[25] S. Novita, P. Harsani, and A. Qur'ania, "Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga," *Komputasi*, vol. 15, no. 1, pp. 118–125, 2018.

[26] I. M. D. P. Asana, I. G. I. Sudipa, A. A. T. W. Mayun, N. P. S. Meinarni, and D. V. Waas, "Aplikasi Data Mining Asosiasi Barang Menggunakan Algoritma Apriori-TID," *INFORMAL Informatics J.*, vol. 7, no. 1, p. 38, 2022, doi: 10.19184/isj.v7i1.30901.