Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Implementation of the C4.5 Algorithm in the Website-Based Classification System for Students of Al-Ishlah Islamic Middle School

Ulfa Rifky Awaliyah[1]*, Lilis Nur Hayati[2], Amaliah Faradibah[3]

[1,3]*Informatics Engineering, Universitas Muslim Indonesia, Makassar*
[2]*Information Systems, Universitas Muslim Indonesia, Makassar*
13020220029@student.umi.ac.id[1*], lilis.nurhayati@umi.ac.id[2], amaliah.faradibah@umi.ac.id[3]

Abstract

Violations of school regulations constitute an issue that requires systematic management to support students' character development; however, recording violations manually using books often creates difficulties in monitoring violation histories and determining appropriate sanctions. Therefore, this study aims to design and develop a web-based student violation classification system using the C4.5 algorithm. The classification process is carried out by constructing a decision tree based on entropy and information gain calculations to determine the best attribute. System performance evaluation is conducted using a confusion matrix along with accuracy, precision, recall, and F1-score metrics, with a focus on the test data. The results, based on 500 datasets with a 90:10 data split, indicate that the system is able to classify student violation levels very effectively, achieving an accuracy of 98.86% on training data and 97.96% on test data, with a precision of 0.91, recall of 1.00, and an F1-score of 0.95. With this strong performance, the system can serve as a supporting tool for recording violations and assisting decision-making in managing student violations more effectively and accurately.

Keywords: *Classification, Student Violations, C4.5 Algorithm, Decision Tree, Web-Based System*

1. Introduction

Digital transformation in Indonesia has changed the way individuals communicate, work, and access information, thus influencing social interactions and daily life patterns [1], [2]. This change has also impacted the world of education, both in the learning process and school administration [3], [4]. Schools are required to be able to adapt to technological developments in order to shape students who excel academically and non-academically, and have good character and discipline [5], [6].

In an effort to strengthen student character and discipline, schools establish various mandatory regulations, with each violation subject to appropriate sanctions [7]. However, in practice, recording and managing student violations is still done manually by Guidance and Counseling (BK) teachers, creating obstacles in monitoring each student's violation history and complicating the process of determining sanctions quickly and accurately [8]. This situation is also experienced at SMP IT Al-Ishlah, where inconsistent violation recording results in suboptimal student evaluation and guidance processes.

Therefore, this study developed a website-based student violation classification system using the C4.5 algorithm. The C4.5 algorithm was chosen because it can form decision trees through entropy and information gain calculations to determine the best attributes, and can effectively process categorical and numerical data [9], [10]. This system is expected to assist schools in managing, monitoring, and evaluating student violations more efficiently, while also representing an innovative step in digitizing school governance at SMP IT Al-Ishlah.

2. Theoretical Foundation

2.1. Classification

Classification is a process used to find a set of patterns or functions that can describe and separate one class of data from another [11]. This classification process involves training (learning) data with known results (training data), so that the system can learn and understand certain patterns. Once the model is trained, the system can be used to predict the class of new, uncategorized data [12].

2.2. C4.5 Algoritm

The C4.5 algorithm is a method for building a decision tree based on provided training data. This algorithm is an extension of the ID3 algorithm. Some of the improvements implemented in the C4.5 algorithm include the ability to handle data containing missing values,

manage continuous data, and apply pruning to simplify the decision tree and reduce overfitting [13]. The steps in this algorithm include determining the attribute that will become the root node using the basic entropy formula:

$$\text{Entropy(S)} = \sum_{i=1}^{n} - P_i \ \log 2 \ P_i$$

Keterangan:
S = Himpunan kasus
N = jumlah kelas atau subset
Pi = Proporsi data kelas ke-i

The formula above will be used to calculate the entropy for each attribute. After calculating, the objects will be separated using the information gain formula. The information gain formula is as follows.

$$\text{Gain(S,A)} = \text{Entropy(S)} - \sum_{i=1}^{n} - \frac{|Si|}{|S|} * \text{Entropy(Si)}$$

Keterangan:
S = himpunan Kasus
Si = Subset dari S berdasarkan atribut A
A = Atribut pemisah
N = jumlah kelas atau subset

2.3. Website

A website is a collection of pages used to display various types of information in the form of text, data, still or moving images, animated data, sound, video, or a combination of these, whether static or dynamic. Each page on a website is connected through a network of pages or hyperlinks, forming a structure that allows users to easily access information via the internet [14] .

2.4. Analysis

The classification process using the C4.5 algorithm requires training data as a basis for determining the rules for the decision tree to be formed. This manual calculation uses 30 datasets as training data with four attributes: violation type, violation points, frequency, and total points. The manual calculation process is performed using Microsoft Excel to calculate the entropy, gain, and gain ratio values for each attribute.

Jenis Pelanggaran	Poin	Frekuensi	Total	Tingkat Pelanggar
Terlambat	2	5	10	Ringan
Berkelahi	10	7	70	Sedang
Menghasut teman	30	3	90	Berat
Memakai perhiasan berlebihan	2	2	4	Ringan
Tidak menjaga kebersihan toilet	2	3	6	Ringan
Menggunakan fasilitas sekolah tanpa izin	5	10	50	Ringan
Berduaan dengan yang bukan mahramnya	30	3	90	Berat
Merendahkan martabat sesama	25	3	75	Sedang
Tidak sholat berjamaah	10	8	80	Berat
Terlambat	2	2	4	Ringan
Meletakkan perlengkapan sekolah tidak pada tempatnya	2	2	4	Ringan
Membawa Rokok	50	2	100	Berat
Main kartu (permainan terlarang)	15	5	75	Sedang
Membawa rokok	50	1	50	Ringan
Menggangu dan Mengancam teman	30	3	90	Berat
Tidak memakai seragam sesuai hari	2	2	4	Ringan
Tidak sholat berjamaah	10	8	80	Berat
Tidak Sholat berjamaah	10	9	90	Berat
Tidak memakai seragam sesuai hari	2	2	4	Ringan
Tidak sholat berjamaah	10	8	80	Berat
pacaran	30	3	90	Berat
Mengganggu dan mengancam teman	30	3	90	Berat
Membawa HP	15	5	75	Sedang
Tidak sholat berjamaah	10	7	70	Sedang
Tidak sholat berjamaah	10	8	80	Berat
Makan dan minum berdiri	2	2	4	Ringan
Tidak sholat berjamaah	10	7	70	Sedang
Tidak sholat berjamaah	10	2	10	Ringan
Terlambat	2	5	10	Ringan
menghina atau merendahkan martabat sesama teman	25	1	25	Ringan

Fig. 1: Test Data

The training data is then subjected to entropy and gain calculations to form decision tree rules. Here's an example of a manual calculation.

Ringan	Sedang	Berat	Jumlah Data
13	6	11	30

Fig. 2: Distribution of Training Data

$$\text{Entropy (S)} = \sum_{i=1}^{n} P_R \ Log_2 \ P_R + \ P_S \ Log_2 \ P_S + \ P_B \ Log_2 \ P_B$$

$$\text{Entropy (S)} = 1{,}51791596$$

The entropy value (S) obtained was 1.51791596, then this value was used as the basis for calculating the gain ratio value for each attribute in the training data value.

Node	Atribut	Nilai	Jumlah kasus	Ringan	sedang	Berat	Entropy	Gain
1	Jenis Pelanggaran	Berduaan dengan yang bukan mahromnya	1	0	0	1	0	1,03824
		Berkelahi	1	0	1	0	0	
		Main kartu (permainan terlarang)	1	0	1	0	0	
		Makan dan minum berdiri	1	1	0	0	0	
		Menggunakan fasilitas sekolah tanpa izin	1	1	0	0	0	
		Meletakkan peralatan sekolah tidak pada tempatnya	1	1	0	0	0	
		Memakau perhiasan berlebih	1	1	0	0	0	
		Membawa Hp	1	0	1	0	0	
		Membawa rokok	2	1	0	1	1	
		Menggangu dan mengancam teman	2	0	0	2	0	
		Menghasut teman	1	0	0	1	0	
		Menghina dan merendahkan martabat sesama teman	2	1	0	1	1	
		Pacaran	1	0	0	1	0	
		Terlambat	3	3	0	0	0	
		Tidak menjaga kebersihan toilet	1	1	0	0	0	
		Tidak sholat berjamaah	8	1	2	5	1,29879	
		Tidak memakai seragam sesuai hari	2	2	0	0	0	
2	Poin	2	10	10	0	0	0	0,97909
		10	9	1	3	5	1,35164	
		15	2	0	2	0	0	
		25	2	1	1	0	1	
		30	5	0	0	5	0	
		50	2	1	0	1	1	
3	Frekuensi	1	2	2	0	0	0	1,11652
		2	8	8	0	0	0	
		3	7	1	1	5	1,14883	
		5	4	2	2	0	1	
		7	3	0	3	0	0	
		8	4	0	0	4	0	
		9	1	0	0	2	0	
		10	1	1	0	0	0	
4	Total	4	6	6	0	0	0	
		6	1	1	0	0	0	1,51792
		10	3	3	0	0	0	
		25	1	1	0	0	0	
		50	2	2	0	0	0	
		70	3	0	3	0	0	
		75	3	0	3	0	0	
		80	4	0	0	4	0	
		90	6	0	0	6	0	
		100	1	0	0	1	0	

Fig. 3: Manual calculation

Based on the data in figure 3, the Total Points attribute has the highest gain value, namely 1.51792, and is therefore selected as the root node in the decision tree. The calculation process then continues based on this attribute. However, because all branches formed from the Total Points attribute have an entropy value of 0, no further calculations are necessary. Thus, the decision tree formation process stops at the Total Points attribute. The following is the resulting decision tree:

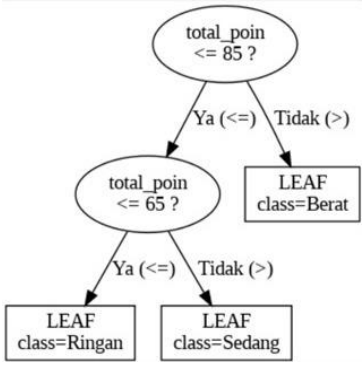

Fig. 4: Decision Tree Rules

Based on the decision tree rules, the system on the website will classify the level of student violations based on the total_points value, namely if the total_points is more than 85 then it is categorized as a Serious violation, if the total_points is less than or equal to 85 but more than 65 then it is categorized as Moderate, and if the total_points is less than or equal to 65 then it is categorized as Light, so that the category determination process is carried out automatically based on the predetermined value limits.

2.5. Desaign

The system design was developed using several Unified Modeling Language (UML) diagrams, namely use case diagrams, class diagrams, activity diagrams, and sequence diagrams.

2.5.1. Use Case Diagram

The use case diagram in Figure 4 depicts several actors, namely the admin, the guidance counselor, the homeroom teacher, and the students. Each actor has different access rights, such as the admin can manage all report data, the principal can only view monitoring data and reports, the guidance counselor manages violations and classifications, the homeroom teacher monitors the students under his/her care and sends warning messages to students who violate, while students can only view their profiles and view their violation history.

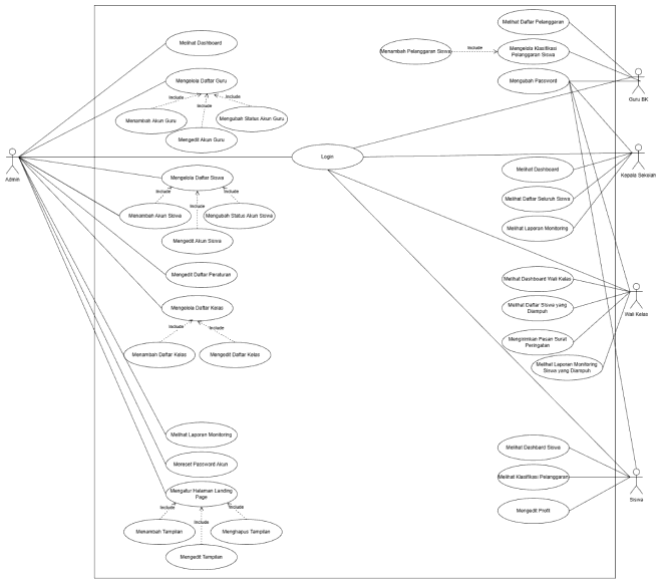

Fig. 5: Use Case Diagram

2.5.2. Database Schema

The following database schema design describes the class structure in the student violation classification system along with the relationships between classes that form the data processing flow in the student violation classification system.

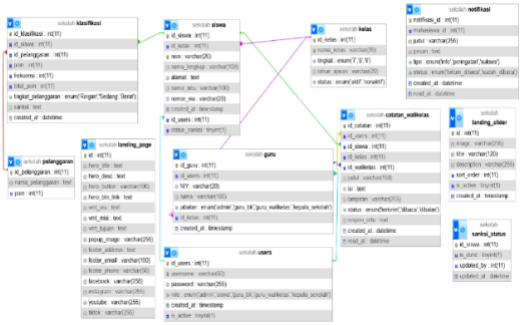

Fig. 6: Database Schema

3. Research Result

This stage presents the design and implementation results of a website-based student violation classification system developed to meet user needs. The system also implements data processing using the C4.5 algorithm through the formation of a decision tree to determine the level of student violations. The system can be run online or locally and accessed through various devices, with the main interface adapting to each user's access rights.

1. Landing Page

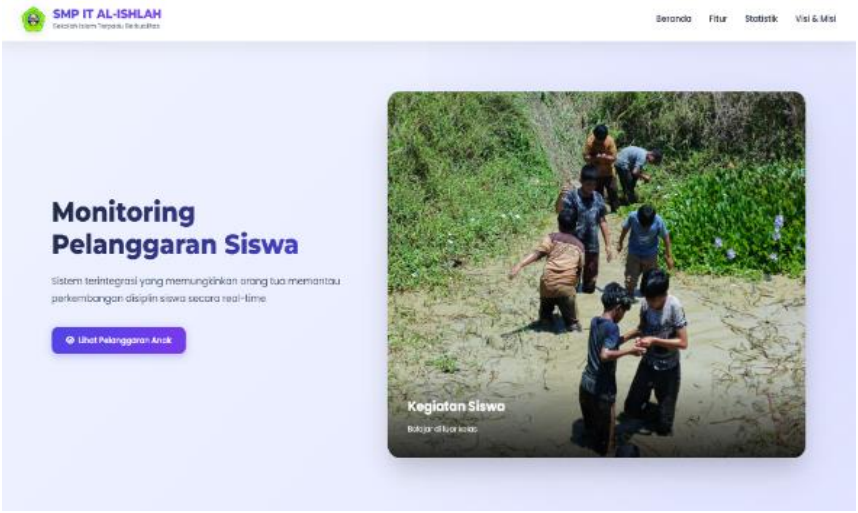

Fig. 7: Website Landing Page

Figure 7. above displays the landing page of the Al-Ishlah IT Middle School student violation classification system which displays general information regarding the system's function in monitoring student violations and discipline.

2. Admin Login page, principal, guidance counselor, and homeroom teacher

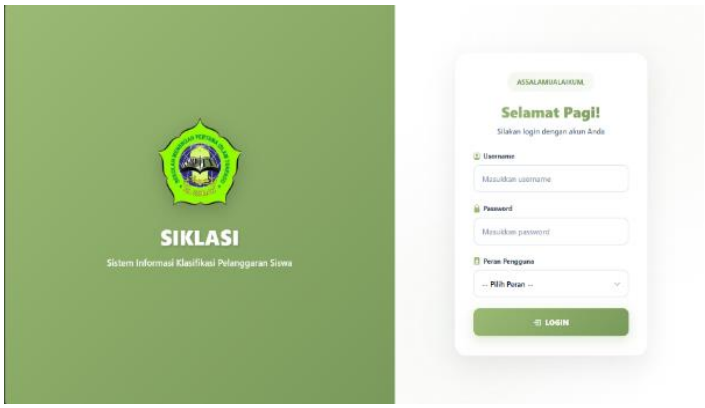

Fig. 8: Website Login View

Figure 8. Displays the login page used by the admin, Principal, Guidance and Counselor, and Homeroom Teacher to enter the system by entering the username, password, and selecting the user role according to their respective access rights.

3. Student Login Page

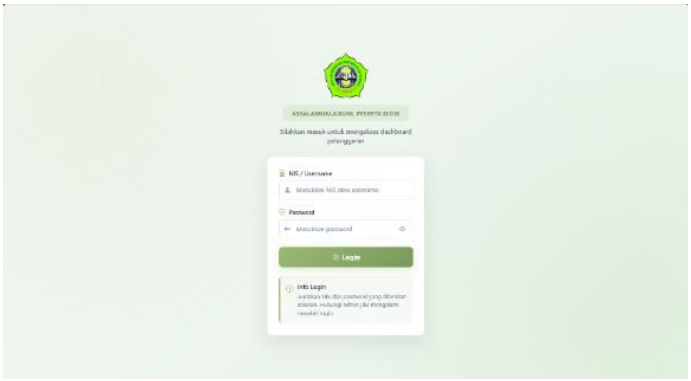

Fig. 9: Website Student Login view

Figure 9. The image above displays the login page used by students to enter the system by entering the username and password that have been provided.

4. Student List page

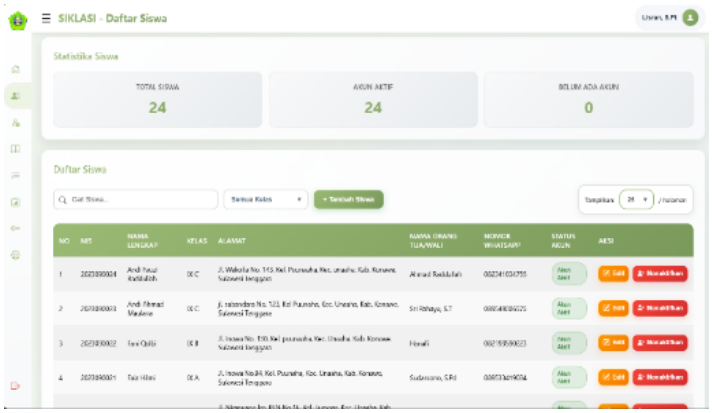

Fig. 10: Student List Page

Figure 10. The above image shows the student list page in the system used by the admin to manage student data. This page displays student identity information and account status, as well as features for searching, filtering by class, adding student data, and editing student accounts.

5. Classification page

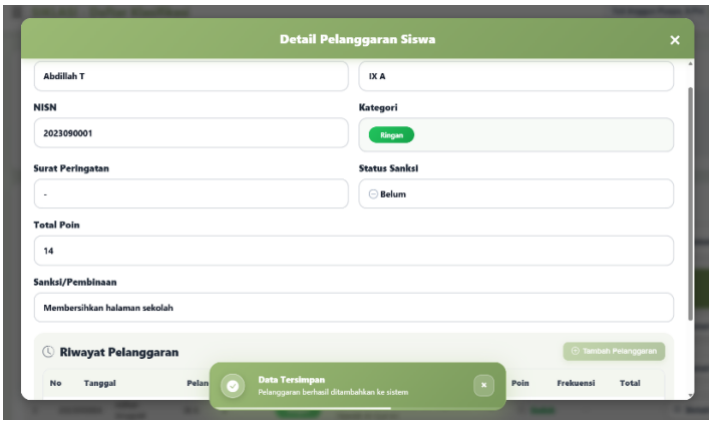

Fig. 11: Clasification Page

Figure 11 displays the student violation addition form used to classify violation levels. On this page, the guidance counselor can input violation data, including the date, type, points, and frequency. The system then processes the data to determine the violation classification (Minor, Moderate, or Severe) based on the total points earned. A "Success" notification indicates that the violation addition and classification process has been successful.

6. Monitoring Page

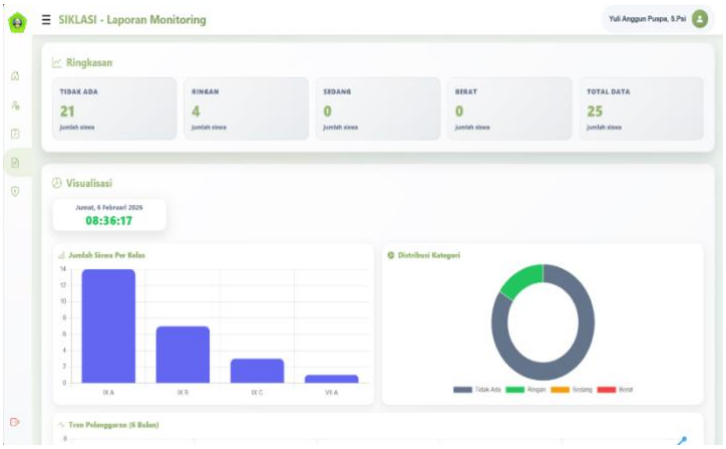

Fig. 12: Monitoring Page

Figure 12. Displays the monitoring page, which displays a summary of student violation data in graphical and statistical form, including the number of students, the distribution of violation categories, and violation trends over time to facilitate monitoring. This page can be accessed by the Admin, principal, guidance counselor, and homeroom teacher.

3.1. Blackbox (Alpha) Test Results

Based on the alpha test results for the add student, add teacher, add class, and specifically the add student violation feature, all system functions performed as designed. In normal testing, the system successfully displayed the input form, saved data correctly, and displayed a success message. In error testing, the system successfully validated incomplete or non-compliant data. Specifically, the add student violation feature demonstrated optimal performance, as it accurately stored violation data and supported the system's violation classification process.

3.2. Blackbox (Beta) Test Result

Based on the beta test results, the website achieved an average score of 4.36, representing 87.3%. Referring to the Likert scale assessment criteria, this result falls into the "Very Good" category. This indicates that the system's features function well, are easy to use, and meet user needs, resulting in a positive reception from all parties involved.

4. Conclusion

Based on the research results and the implementation of a website-based student violation classification system using the c4.5 algorithm, the following conclusions can be drawn.

1. This research successfully designed and built a website-based student violation classification system that can assist schools in managing and monitoring student violations more effectively and in a structured manner. This system provides key features: student data management, violation recording, a violation level classification process, and a presentation of violation history per student. Based on beta testing results, the system achieved an accuracy of 87.3%, categorized as "Very Good,"

making it suitable for use in supporting student evaluation and development processes. This system minimizes the problem of bookkeeping, thereby making the student evaluation and development process more efficient.

2. The application of the C4.5 algorithm to the student violation classification system is able to classify the level of violation based on the available data through the formation of a systematic decision tree. The results of the study using 500 datasets with a 90:10 data split show that the system built is able to classify the level of student violations very well, with an accuracy value of 98.86% on the training data and 97.96% on the test data. The evaluation uses a precision of 0.91, a recall of 1.00, and an f1-Score with a value of 0.95 on the test data. Thus, the system built is acceptable and suitable for use as a tool in decision-making related to student violation management in schools.

Acknowledgement

The author expresses all praise to Allah and expresses his deepest appreciation and thanks to Mrs. Lilis Nur Hayati, S.Kom., M.Eng., MTA., and Mrs. Amaliah Faradibah, S.Kom., M.Kom., MTA., MCF, as Supervisors for all the guidance, direction, suggestions, constructive criticism, and patience that have been given during the process of preparing this thesis. Thanks to the support, attention and time spent, the author was able to understand each stage of the research better and was able to complete this thesis as expected. The guidance provided not only helps in preparing this scientific work, but also provides valuable experience and insight for the author in developing academic and professional skills. Hopefully all the goodness and knowledge that has been given will be rewarded in kind and become a valuable good deed in His sight.

References

[1] D. A. Rabbani and F. U. Najicha, "Pengaruh Perkembangan Teknologi terhadap Kehidupan dan Interaksi Sosial Masyarakat Indonesia," *Researchgate.Net*, no. November, pp. 0–13, 2023, [Online]. Available: https://www.researchgate.net/profile/Dana-Rabbani/publication/375525102_Pengaruh_Perkembangan_Teknologi_terhadap_Kehidupan_dan_Interaksi_Sosial_Masyarakat_Indonesia/links/654dcc8dce88b87031d8db65/Pengaruh-Perkembangan-Teknologi-terhadap-Kehidupan-dan-Inte

[2] H. Azis, Purnawansyah, and N. Alfiyyah, "Multiclass Classification on Nominal Value of Rupiah Banknotes Based on Image Processing," vol. 16, no. 01, pp. 87–99, 2024.

[3] R. Junita, S. Anraeni, and Irawati, "Perancangan Aplikasi Pengingat Jadwal Perkuliahan Menggunakan Metode Push Notification Berbasis Mobile," vol. 2, no. 1, pp. 52–55, 2021.

[4] D. Indra and E. I. Alwi, "Pemanfaatan Informasi dan Teknologi Komunikasi di Pesantren Mizanul Ulum," vol. 15, no. 1, pp. 70–75, 2024.

[5] F. Zeki Yusuf, "Intergrasi Teknologi Digital Dalam Lembaga Pendidikan Islam: (Tantangan Dan Peluang)," *Reli. J. Kaji. Agama dan Multikulturalisme Indones.*, vol. 7693, pp. 118–125, 2024.

[6] A. Andhika and Samsoni, "Perancangan Sistem Informasi Monitoring Kegiatan Santri Di Pondok Pesantren Technopreneur As-Shofa Berbasis Web," *Biner J. Ilmu Komput. , Tek. dan Multimed.*, vol. 1(4), no. 4, pp. 947–956, 2023, [Online]. Available: https://journal.mediapublikasi.id/index.php/Biner

[7] A. A. Mallaena, H. Hasbi, and M. Yusuf, "Kinerja guru bimbingan konseling dan implikasi penerapan sistem poin pelanggaran terhadap kedisiplinan siswa," *J. Konseling dan Pendidik.*, vol. 11, no. 3, p. 183, 2023, doi: 10.29210/199100.

[8] E. Novianti, F. Annas, S. Derta, and Y. E. Yuspita, "Perancangan Sistem Informasi Pencatatan Poin Pelanggaran Siswa Menggunakan Framework Codeigniter (Ci) Di SMPN 2 Kamang Magek," vol. 10, no. 1, pp. 1–14, 2024, doi: 10.31980/jpetik.v10i1.547.

[9] S. Romlah and Z. F. Fatah, "Penerapan Data Mining Untuk Memprediksi Hafalan Santri Tahfidz Menggunakan Metode Decision Tree (Studi Kasus MIK Sarina)," *Gudang J. Multidisiplin Ilmu*, vol. 2, no. November, pp. 160–166, 2024.

[10] D. Widyawati, A. Faradibah, and P. L. L. Belluano, "Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine," *Indones. J. Data Sci.*, vol. 4, no. 2, pp. 80–89, 2023, doi: 10.56705/ijodas.v4i2.76.

[11] 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.

[12] A. Srirahayu and L. S. Pribadie, "Review Paper Data Mining Klasifikasi Data Mining," *J. Ilm. Inform. Glob.*, vol. 14, no. 1, 2023, doi: 10.36982/jiig.v14i1.2981.

[13] A. Prasetio, M. H. Hasibuan, and P. Sitompul, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 3, pp. 209–214, 2021, doi: 10.32672/jnkti.v4i3.2983.

[14] L. Nurhayati, W. Astuti, and A. R. Pratama, "PkM Penerapan Website Desa Binaan UMI di Desa Lanna Kecamatan Parangloe Kab. Gowa," *Ilmu Komput. Untuk Masy.*, vol. 2, no. 1, pp. 30–36, 2021.