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 Decision Tree Algorithm to Determine Student Productivity Based on Sleep Patterns

Tri Fuji Mandala [1]*, Haida Dafitri [2]

[1,2]Universitas Harapan Medan
trifujimandalaa@gmail.com l[1], aida.stth@gmail.com[2]*

Abstract

Sleep patterns refer to an individual's habits in managing sleep and wake times, including duration, quality, and regularity. Students, particularly those in the Informatics Engineering Program at Universitas Harapan Medan, often experience irregular sleep patterns due to heavy academic workloads such as assignments, projects, and practical activities. This condition can reduce academic productivity in terms of concentration, memory, and the ability to complete tasks on time. Therefore, this study aims to develop a classification model to predict student productivity levels based on sleep patterns using the Decision Tree C4.5 algorithm. This algorithm was chosen for its advantages in interpretability, ability to handle both numerical and categorical data, and efficient attribute selection, which contribute to generating an accurate and transparent classification model. The study involved 30 respondents from the 8th semester of the Informatics Engineering Program at Universitas Harapan Medan in the 2024/2025 academic year who filled out questionnaires regarding their sleep patterns and productivity. The results showed that 15 respondents (41.2%) had low productivity, 9 respondents (35.3%) had medium productivity, and 6 respondents (23.5%) had high productivity. These findings indicate a significant relationship between sleep pattern regularity and student productivity levels. The model generated using the C4.5 algorithm is expected to serve as a foundation for developing decision support systems aimed at improving the balance between sleep patterns and academic productivity among students. .

Keywords: *Student Productivity; Sleep Patterns; Classification; Decision Tree; C4.5 Algorithm.*

1. Introduction

Productivity is an important aspect of human life, whether in the workplace, education, or daily activities. It can be defined as a person's ability to produce output efficiently from the available inputs, such as time, discipline, energy, and resources [1]. The rapid development of technology and modern lifestyles has greatly influenced daily activities, one of which is poor sleeping habits. In the field of education, increasing academic pressure often forces students to sacrifice their rest time in order to complete assignments. This becomes one of the causes of inadequate and low-quality sleep, which in turn affects physical and mental health as well as critical concentration. Higher education requires students to manage their time and habits effectively so they can handle lectures, assignments, and academic responsibilities optimally. One of the factors affecting student productivity is sleep patterns. Adequate and good-quality sleep has been proven to improve concentration, memory, and physical well-being. Conversely, staying up late and lack of sleep can reduce motivation to study and increase the risk of stress and fatigue [2]. Sleep is one of the best forms of total rest that restores energy for the body to carry out the next activities [3]. Students with regular sleep patterns tend to have higher levels of focus and productivity compared to those with irregular sleep patterns [4]. Research also shows that sufficient sleep plays a crucial role in improving student productivity. With proper sleep patterns, students can maintain their cognitive abilities and concentrate more effectively on ongoing lectures, making it easier for them to complete assignments efficiently.

Many students often neglect their sleep patterns due to academic pressures and other activities. Irregular sleep patterns are a common phenomenon among university students, especially those in the Informatics Engineering study program. These students frequently deal with academic assignments with tight deadlines, and working on projects often results in the habit of staying up late, leading to widespread sleep disturbances. The prevalence of insomnia in Indonesia is estimated to be around 67%, with most cases categorized as moderate [5]. Students in this field often engage in programming, debugging, and technology-based projects that require long working hours. As a result, they willingly sacrifice their rest time and stay awake late into the night to complete their tasks. Based on a survey previously conducted by the author during the Final Project Proposal course with Informatics Engineering students at Universitas Harapan Medan, it was found that many students experienced a decline in their academic transcript (KHS) results. Some students were unable to focus during class, with several even feeling sleepy or falling asleep. Despite this, many students continue to maintain poor sleeping habits without understanding their impact on health and academic productivity. There is a significant relationship between sleep quality and learning concentration

among students, with findings showing that 80% of students have poor sleep quality and 90% experience difficulty concentrating while studying [6].

To address this issue, an analytical approach is needed to identify and predict the relationship between students' sleep patterns and their level of academic productivity. One method that can be applied is the C4.5 Decision Tree algorithm, a classification algorithm in data mining used to build decision tree models based on datasets. The C4.5 algorithm is one of the data-mining algorithms capable of grouping or segmenting and performing predictive classification [7]. It is widely used in social classification, as it can process both numerical and categorical data while producing models that are easy to understand [8]. The C4.5 data-mining algorithm is often utilized to classify or group predictive data [9]. The advantage of the C4.5 algorithm is its ability to generate decision trees that are easy to interpret, have acceptable accuracy levels, and are efficient [10]. A decision tree consists of roots, nodes, and leaves, each representing specific meanings [11]. Each branching point in the tree represents a data attribute that divides the data into smaller groups [12]. The main benefit of using decision trees is their ability to simplify complex decision-making processes, making it easier for decision makers to understand the solutions [13]. The decision tree method is a tree structure consisting of nodes that represent decision branches [14]. It is a data processing technique used to predict future outcomes by building classification or regression models in the form of a tree structure [15]. Decision trees are well-established and popular tools for predicting and classifying data [16]. They are used to classify predictive data based on given attributes [17]. A classification algorithm known as a decision tree is described as the recursive partitioning of the instance space [18]. Comprehensive facts are represented as rules using the decision-tree approach, making them easy for people to understand [19].

A previous study titled *"Application of Decision Trees in Detecting Healthy Sleep Patterns Based on Lifestyle Habits"* analyzed the C4.5 decision tree algorithm in detecting healthy sleep patterns based on lifestyle factors. The researchers predicted whether a person had a healthy sleep pattern using the decision-tree method due to its ability to provide accurate and easily interpretable predictions. The study demonstrated that decision-tree methods can be used to predict healthy sleep patterns and help individuals increase awareness of maintaining good sleep quality through healthy lifestyle choices [13].

Another previous study titled *"Prediction of Student Dropout Using the C4.5 Decision Tree Algorithm"* concluded that applying data-mining techniques with the C4.5 decision tree method can effectively predict student dropout rates, where results showed that the number of students who dropped out was significantly lower than those who remained active [20].

A related earlier study titled *"Application of the Decision Tree Method in Determining Students' Study Programs"* concluded that classification using decision-tree models provides a comprehensive understanding of the variables influencing students' study-program selections based on report card scores and practical assessments [21].

2. Research Method

This study employs a quantitative research method because the data used are numerical and can be processed statistically. The method applied is a classification approach using the C4.5 decision tree algorithm, which aims to construct a model for assessing student productivity based on their sleep patterns.

2.1. Problem Analysis

The designed system aims to model the productivity level of Informatics Engineering students based on their sleep patterns by applying the C4.5 decision tree algorithm. The system consists of four main functions. The first function is the questionnaire data input, which records information related to students' sleep patterns, such as sleep time, sleep duration, sleep quality, and frequency of nighttime awakenings. The second function is the preprocessing function, which organizes and categorizes the data into formats suitable for classification. The third function is the classification function, which utilizes the C4.5 algorithm to generate a decision tree model from the trained dataset. The final function is the prediction function, which enables users to obtain productivity classification results based on new input data. This system is designed as a desktop application built using Python and Tkinter, intended to run offline on a computer device. The research findings indicate that Python programs can be used to model functions easily and efficiently [22].

2.2. Data Collection

The training data is used to develop the machine learning model in the application being built. This data was obtained from questionnaire responses collected from students of the Informatics Engineering Study Program at Universitas Harapan. A total of 30 respondents participated in this study, consisting of eighth-semester students.

a. Bedtime, with attribute values classified as early (before 22:00) and late (after 22:00).
b. Sleep duration, with attribute values divided into three groups: less than 7 hours, adequate (7–9 hours), and more than 9 hours.
c. Sleep quality, with attribute values categorized as good (waking up refreshed), average, and poor (waking up tired).
d. Sleep disturbances, with attribute values present or absent.
e. Activities before sleeping, categorized into three types: work (doing assignments or completing tasks), entertainment (watching or playing games), and relaxation (daydreaming, reading non-academic books, praying, etc.).

The label used is productivity, which is divided into three categories: high, medium, and low. Productivity is determined based on the students' ability to complete assignments given by lecturers, their focus during learning, and their cumulative grade index.

2.3. Algorithm Analysis

To evaluate the influence of each attribute, entropy and information gain calculations were performed as the basis for constructing the decision tree. Entropy is an information theory measure used to determine the characteristics of impurity and homogeneity within a dataset [23]. The entropy value is used to measure the level of uncertainty in the training data, and the formula used is as follows:

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log_2 pi \quad (1)$$

Description: S is the set of cases, n is the number of partitions of S, and pi represents the proportion of Si to S. After obtaining the entropy value, the next step is to calculate the information gain to determine the best attribute based on the reduction of entropy. Information gain is effective in filtering out irrelevant features, thereby improving accuracy in the classification process [24]. The attribute with the highest information gain will become the root of the decision tree. The formula used to calculate the information gain is as follows:

$$Gain\ (A,S) = Entropy\ (S) = \sum_{i=1}^{n} - \frac{|S_i|}{|S|} * Entropy\ (S) \quad (2)$$

Both formulas are applied to the dataset that was previously prepared. This dataset was constructed based on survey results obtained from 30 respondents who completed a questionnaire regarding their sleep patterns and productivity levels. From the collected data, 17 unique training data combinations were identified and used in the classification process. Each data combination contains attributes such as bedtime, sleep duration, sleep quality, sleep disturbances, and pre-sleep activities, all of which contribute to determining the productivity level of the students. Furthermore, the dataset is evaluated using the C4.5 algorithm to calculate attribute priority levels and to construct a decision tree model capable of classifying data into productivity categories: high, medium, or low. The dataset used in this study is presented in the following table.

Table 1: Student Productivity Dataset

No	Hours Sleep	Duration Sleep	Quality Sleep	Disruption Sleep	Activities Before Bed	Productivity
1	Beginning	Enough	Good	No	Work	Height
2	Beginning	Enough	Good	No	Entertainment	Height
3	Beginning	Enough	Ordinary	Yes	Relaxation	Medium
4	Beginning	Enough	Ordinary	No	Work	Height
5	Beginning	Read More	Ordinary	No	Relaxation	Medium
6	Beginning	Read More	Ordinary	No	Entertainment	Medium
7	Beginning	Enough	Ordinary	No	Work	Medium
8	Soluble	Enough	Ordinary	Yes	Relaxation	Low

3. Discussion Results

3.1. Calculation of Total Entropy Value

The dataset contains a total of 17 data entries, with label distribution consisting of 4 high, 6 medium, and 7 low productivity instances. Therefore, the total entropy value can be calculated using the entropy formula as follows:

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log_2 pi$$

$$Entropy\ (S) = (-\frac{4}{17} Log_2 \frac{4}{17}) + (-\frac{6}{17} Log_2 \frac{6}{17}) + (-\frac{7}{17} Log_2 \frac{7}{17})$$

$$Entropy\ (S) = 0{,}490 + 0{,}531 + 0{,}527 = 1{,}548$$

Hasil dari entropy total adalah 1,548

3.2. Data Trial

In the testing phase, functional testing was conducted on the application. The data used as input for the testing process was obtained from questionnaire responses collected from student respondents. This questionnaire data was selected because it represents real usage scenarios and provides varied input data, allowing the system to be tested for its robustness and the correctness of its data processing logic. Afterward, the user selects the data mining menu and clicks the data mining process button, after which the results will be displayed as follows.

Table 2: Data Trial

Nama	Hours Sleep	Duration Sleep	Quality Sleep	Disruption Sleep	Activities Before Bed	Productivity
Student 1	Beginning	Beginning	Good	No	Work	Height
Student 2	Beginning	Beginning	Good	No	Entertainment	Height
Student 3	Beginning	Beginning	Ordinary	Yes	Relaxation	Medium

Student 4	Beginning	Beginning	Ordinary	No	Work	Height
Student 5	Beginning	Read More	Ordinary	No	Relaxation	Medium
Student 6	Beginning	Read More	Ordinary	No	Entertainment	Medium
Student 7	Beginning	Less	Ordinary	No	Work	Medium

The functional testing phase of this application was carried out systematically using test data derived directly from student questionnaire responses, ensuring that the input used represents real usage scenarios and contains sufficient variation to evaluate the robustness of the application.

The test data presented, as shown in Table 2, consists of samples containing student sleep pattern attributes (such as bedtime, sleep duration, sleep quality, sleep disturbances, and pre-sleep activities), which are then correlated with their productivity levels. Most of the sample data indicates that students fall into the Early bedtime category. The core testing procedure involves entering each data entry from the questionnaire into the application through the training data menu, running the data mining function, and then comparing the output generated by the application with manually verified or theoretically expected results.

The initial interpretation of Table 2 shows a relationship between sleep pattern variables and productivity. For example, students with an Early bedtime and Good sleep quality tend to demonstrate High productivity (Students 1 and 2). However, there are also cases in which an Early bedtime combined with Average sleep quality can still result in High productivity (Student 4) or Medium productivity (Students 3, 5, 6, and 7). These cases highlight the complexity of the relationships among variables, which the application must accurately process and classify. The focus of this testing phase is to verify the accuracy of the algorithm implemented within the application, particularly in the data mining process, by ensuring that the application's output consistently matches the expected results based on correct data processing logic.

Based on the data, several students fall into the Early bedtime category while others have a Late bedtime. These data entries were then entered into the application through the training data menu.

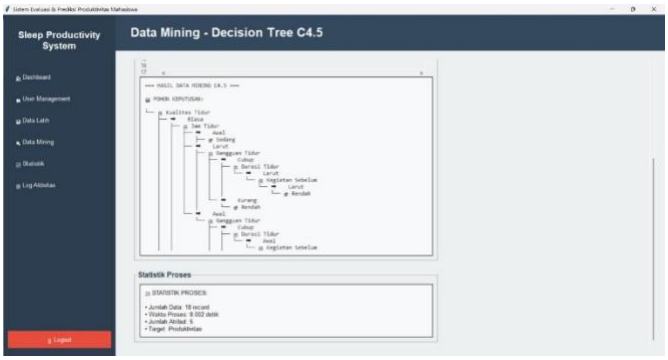

Fig. 1: Data Mining Results

Based on the data mining analysis using the C4.5 algorithm, it can be concluded that bedtime is the most critical factor influencing students' productivity levels. Data mining is a process of managing data with the aim of discovering specific patterns or information [25]. The resulting decision tree reveals a clear pattern indicating that students who go to bed early have a higher likelihood of achieving high productivity compared to those who sleep late.

For students with an early bedtime, the combination of adequate sleep duration and good sleep quality consistently leads to high productivity. Conversely, students who sleep late face a greater risk of low productivity, especially when late bedtime is paired with poor sleep quality or the presence of sleep disturbances. This pattern indicates that maintaining an early bedtime is more important than sleep duration or sleep quality alone. The statistical display of these findings is shown as follow.

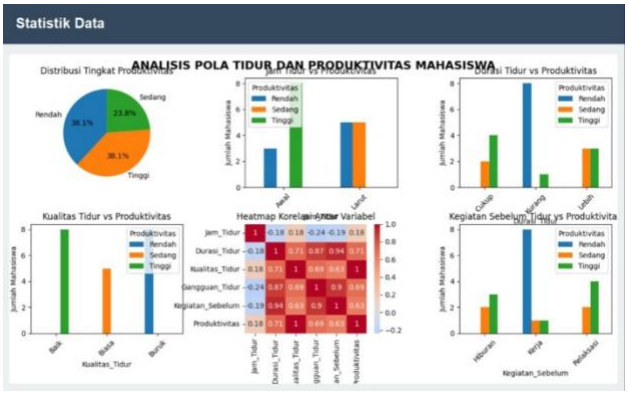

Fig. 2: Data Mining Statistics Results

The statistical analysis of data from 17 students reveals a significant relationship between sleep patterns and productivity. The data distribution shows that high productivity is generally associated with early bedtime, adequate sleep duration, and good sleep quality. Specifically, among the 7 students with an early bedtime, 3 achieved high productivity and 4 were in the medium category, while none recorded low productivity.

Conversely, among the 10 students with a late bedtime, 6 experienced low productivity, 3 were in the medium category, and only 1 achieved high productivity. The correlation heatmap reinforces these findings by indicating a negative relationship between late bedtime and productivity. The data also shows that pre-sleep activities such as work or entertainment do not have as strong an impact on productivity as bedtime and sleep quality.

4. Conclusion

Based on the results of the study involving 30 eighth-semester students from the Informatics Engineering Study Program at Universitas Harapan Medan, it was found that 15 respondents (41.2%) had low productivity, 9 respondents (35.3%) had medium productivity, and 6 respondents (23.5%) demonstrated high productivity. These findings indicate that most students tend to have low productivity due to poor sleep patterns.

The C4.5 Decision Tree algorithm proved to be effective in classifying student data and constructing a model that illustrates the relationship between sleep patterns and student productivity levels. Its implementation in the form of a desktop application provides predictive insights regarding student productivity based on the input data. The application can be operated offline and is easy to use.

References

[1] B. Setiawan and N. Nuridin, "Pengaruh Lingkungan Kerja Dan Disiplin Kerja Terhadap Produktivitas Kerja Karyawan Bagian Operator Spbu Bekasi Pt Pertamina Retail," *J. Manaj. Bisnis Krisnadwipayana*, vol. 9, no. 1, 2021, doi: 10.35137/jmbk.v9i1.520.

[2] S. Wulandari and R. Pranata, "Deskripsi Kualitas Tidur dan Pengaruhnya terhadap Konsentrasi Belajar Mahasiswa," *J. Pendidik. Kesehat. Rekreasi*, vol. 10, no. 1, pp. 101–108, 2024, doi: 10.59672/jpkr.v10i1.3414.

[3] K. Pangestu and A. Dwiana, "Hubungan kualitas tidur dengan memori jangka pendek pada mahasiswa Fakultas Kedokteran Universitas Tarumanagara Angkatan 2017," *Tarumanagara Med. J.*, vol. 2, no. 1, pp. 98–103, 2020, doi: 10.24912/tmj.v2i2.7844.

[4] D. Y. P. Manbait and M. Sitorus, "Dampak pola tidur terhadap konsentrasi belajar mahasiswa," no. March, pp. 0–6, 2025.

[5] F. Fadillah *et al.*, "Insomnia pada Mahasiswa Kedokteran: Sebuah Tinjauan Pustaka," *Med. Prof. J. Lampung*, vol. 14, no. September, pp. 1819–1822, 2024.

[6] N. M. Arifin, N. Al-atsariyah, A. P. Pradani, N. N. Latif, and S. Supriyono, "Pengaruh Kualitas Tidur terhadap Konsentrasi Belajar Mahasiswa Pendidikan Ekonomi 2024 Universitas Pendidikan Indonesia," *J. Pendidik. Tambusai*, vol. 8, no. 3, pp. 49725–49731, 2024, [Online]. Available: http://jptam.org/index.php/jptam/article/view/23683

[7] R. Mantopani, I. L. Ramadhan, R. P. Samsara, and R. Samsinar, "Analisis Perbandingan Decision Tree Dan Algoritma C4.5 Untuk Mengklasifikasikan Penerimaan Mahasiswa Elektro," *Semin. Nas. Call Pap. Hubisintek 2024*, pp. 295–302, 2024.

[8] R. M. Khair, M. E. Setiawan, and M. Rosaensi, "Klasifikasi Tingkat Kekerasan dalam Rumah Tangga Menggunakan Algoritma Decision Tree C4 . 5," *Semin. Nas. CORISINDO*, no. September 2025, pp. 32–39, 2024.

[9] A. Huday and Z. Fatah, "Penerapan Decision Tree C4.5 Dalam Memprediksi Predikat Terbaik Di Madrasah Ta'Hiliyah Ibrahimy," *J. Ilm. Multidisiplin Ilmu*, vol. 2, no. 1, pp. 61–68, 2025, doi: 10.69714/be4q6n31.

[10] H. Syafputra, H. L. Sari, and K. Khairil, "Klasifikasi Penjualan Perhiasan Menggunakan Metode Decision Tree Algoritma C4.5 (Studi Kasus: Toko Emas Berkat Famili)," *J. Media Infotama*, vol. 20, no. 2, pp. 563–569, 2024, doi: 10.37676/jmi.v20i2.6517.

[11] A. A. Ansyah, T. M. Fahrudin, and D. A. Prasetya, "Penerapan Metode Decision Tree C4.5 Untuk Klasifikasi Data Kandidat Tenaga Kerja Pada Perusahaan Outsourcing," *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 6, no. 1, pp. 41–48, 2024, doi: 10.26905/jasiek.v6i1.12670.

[12] R. Mashitapasha, F. Damayanti, and D. Abdul Fatah, "Penerapan Metode Decision Tree Dalam Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma C4.5," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4016–4023, 2025, doi: 10.36040/jati.v9i3.13532.

[13] I. Nawawi and Z. Fatah, "Penerapan Decision Trees dalam Mendeteksi Pola Tidur Sehat Berdasarkan Kebiasaan Gaya Hidup," *J. Ilm. Sains Teknol. dan Inf.*, vol. 2, no. 4, pp. 34–41, 2024.

[14] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.

[15] E. S. Palupi, "Klasifikasi Kualitas Air Bersih Di Jakarta Menggunakan Algoritma Decision Tree Dan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1259–1265, 2025, doi: 10.36040/jati.v9i1.12666.

[16] A. N. Najah, A. N. Amalina, and R. Hidayat, "Penerapan Algoritma Decision Tree C4.5 Untuk Prediksi Cuaca Di Kota Semarang," *Indexia Inform. Comput. Intell. J. Nama Akhir dari Penulis Pertama, Penu*, vol. 7, no. 1, p. 46, 2025, doi: 10.30587/indexia.v7i1.9344.

[17] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024.

[18] A. D. Putri, F. Sholekhah, E. Dadynata, L. Efrizoni, and N. Sapina, "The Application of C4.5 Decision Tree Algorithm for Predicting the Survival Rate of Thyroid Cancer Patients Penerapan Algoritma Decesion Tree C4.5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1485–1495, 2024.

[19] M. Rogib, N. Rahaningsih, and R. danar Dana, "Penerapan algoritma C4.5 untuk seleksi penjurusan siswa baru pada Sekolah Menengah Kejuruan (Studi Kasus: Smk Plus Al-Hilal Arjawinangun)," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 861–866, 2024.

[20] Y. L. Fatma and N. Rochmawati, "Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4.5," *J. Informatics Comput. Sci.*, vol. 5, no. 04, pp. 486–493, 2024, doi: 10.26740/jinacs.v5n04.p486-493.

[21] Y. N. Aini, A. Faqih, and G. Dwilestari, "Penerapan Metode Decision Tree dalam Penentuan Jurusan Siswa INFORMASI ARTIKEL ABSTRACT KATA KUNCI," *JIF J. Ilm. Inform.*, no. 10, 2025.

[22] N. M. Surbakti *et al.*, "Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel," *Algoritm. J. Mat. Ilmu Pengetah. Alam, Kebumian dan Angkasa*, vol. 2, no. 3, pp. 98–107, 2024, doi: 10.62383/algoritma.v2i3.67.

[23] E. Giawa, "Penerapan Metode C 45 Untuk Memprediksi Keuntungan Dari Penjualan Sarang Wallet," *KETIK J. Inform.*, vol. 03, no. 01, pp. 12–21, 2025.

[24] N. T. Khair, I. Afrianty, F. Syafria, E. Budianita, and S. K. Gusti, "Penerapan Information Gain Untuk Seleksi Fitur Pada Klasifikasi Jenis Kelamin Tulang Tengkorak Menggunakan Backpropagation," *Bull. Comput. Sci. Res.*, vol. 5, no. 4, pp. 666–678, 2025, doi: 10.47065/bulletincsr.v5i4.637.

[25] M. M. Nau, V. N. Fathya, and O. P. Martadireja, "Implementasi Data Mining Pada Analisis Karakteristik Pelanggan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 5209–5215, 2025, doi: 10.36040/jati.v9i3.13725.