Published online on the page : https://journal.makwafoundation.org/index.php/intellect

I n t e l l e c t :

Indonesian Journal of Learning and Technological Innovation

| ISSN (*Online*) 2962-9233 |

Analysis of Lifestyle Classification Using a Decision Tree Approach

Afriosa Syawitri[1], Siti Rahmi Hidayatullah[2], Miranda Nuraini[3]*

[1,2,3]Universitas Negeri Padang, Padang, Indonesia

Informasi Artikel

Sejarah Artikel:
Submit : 23 Desember 2025
Revisi : 29 Desember 2025
Diterima : 30 Desember 2025
Diterbitkan: 31 Desember 2025

Kata Kunci

Decision Tree, Lifestyle Classification, Diet, Sleep Patterns, Physical Activity, Mental Health

Correspondence

E-mail: afriosasyawitri@unp.ac.id*

A B S T R A C T

Lifestyle related health issues continue to increase, highlighting the need for data-driven approaches that not only classify lifestyle patterns but also provide interpretable insights to support health-related decision making. This study aims to develop an interpretable lifestyle classification model using the Decision Tree algorithm, with a specific analytical focus on identifying dominant behavioral factors and their hierarchical relationships in distinguishing healthy and unhealthy lifestyles. The dataset was collected through a questionnaire survey involving 130 respondents representing diverse lifestyle behaviors. Initially, 23 attributes measured using Likert scales were used to capture multiple aspects of lifestyle. To improve analytical clarity and reduce data complexity, the attributes were transformed by grouping conceptually related items into four main behavioral domains: diet, physical activity, sleep patterns, and mental health. Personal demographic attributes were excluded from the modeling process due to their limited relevance to lifestyle behavior and their potential to introduce classification bias. Within each domain, sub-attributes were aggregated using mean values to generate stable composite scores, a methodologically appropriate approach given the non-parametric and threshold-based characteristics of the Decision Tree algorithm. The applied to reduce overfitting. The results indicate that the proposed model achieved an accuracy of 84.62% and a weighted average F1-score of 0.84, demonstrating balanced classification performance. The model showed strong recall in identifying healthy lifestyles, while limitations related to generalizability remain. Transformed dataset was divided into training and testing sets using a 70:30 hold-out validation strategy. Model construction employed the entropy criterion and information gain for attribute selection, with complexity control.

Abstrak

Masalah kesehatan terkait gaya hidup terus meningkat, menyoroti perlunya pendekatan berbasis data yang tidak hanya mengklasifikasikan pola gaya hidup tetapi juga memberikan wawasan yang dapat ditafsirkan untuk mendukung pengambilan keputusan terkait kesehatan. Penelitian ini bertujuan untuk mengembangkan model klasifikasi gaya hidup yang dapat diinterpretasikan menggunakan algoritma Decision Tree, dengan fokus analitis khusus untuk mengidentifikasi faktor perilaku dominan dan hubungan hierarkisnya dalam membedakan gaya hidup sehat dan tidak sehat. Kumpulan data dikumpulkan melalui survei kuesioner yang melibatkan 130 responden yang mewakili beragam perilaku gaya hidup. Awalnya, 23 atribut yang diukur menggunakan skala Likert digunakan untuk menangkap berbagai aspek gaya hidup. Untuk meningkatkan kejelasan analitis dan mengurangi kompleksitas data, atribut diubah dengan mengelompokkan item yang terkait secara konseptual menjadi empat domain perilaku utama: diet, aktivitas fisik, pola tidur, dan kesehatan mental. Atribut demografis pribadi dikecualikan dari proses pemodelan karena relevansinya yang terbatas dengan perilaku gaya hidup dan potensinya untuk memperkenalkan bias klasifikasi. Dalam setiap domain, sub-atribut dikumpulkan menggunakan nilai rata-rata untuk menghasilkan skor komposit yang stabil, pendekatan yang sesuai secara metodologis mengingat karakteristik non-parametrik dan berbasis ambang batas dari algoritma Pohon Keputusan. Himpunan data yang diubah dibagi menjadi set pelatihan dan pengujian menggunakan strategi validasi penahanan 70:30. Konstruksi model menggunakan kriteria entropi dan perolehan informasi untuk pemilihan atribut, dengan kontrol kompleksitas diterapkan untuk mengurangi overfitting. Hasil menunjukkan bahwa model yang diusulkan mencapai akurasi 84,62% dan skor F1 rata-rata tertimbang 0,84, menunjukkan kinerja klasifikasi yang seimbang. Model ini

menunjukkan ingatan yang kuat dalam mengidentifikasi gaya hidup sehat, sementara keterbatasan yang terkait dengan generalisasi tetap ada.

1. Introduction

Change in lifestyle of modern society have increasingly shifted away from healthy living behaviors. High calorie dietary patterns, insuffient physical activity, irregular sleep habits, and rising phsychological stress have been identified by the World Health Organization (WHO) as major risk factors contributing to declining helath outcomes. Globally, WHO reports that physical inactivity contributes to appromaximately 27% of diabetes cases and nearly 30% of ischemic heart diase cases, while obesity rates have nearly tripled over the past decades [1]. In indonesia, data from the Ministry of Health indicate a continuous increas in lifestyle related diseases, with the prevalence of obesity and diabetes showing a significant upward trend in recent national health surveys. This condition is particularly concerning because lifestyle factors play a crucial role in the prevention of non-communicable diseases such as obesity, diabetes, hypertension, and other metabolic disorders. Consequently, an objective and data-driven analytical approach is needed to better understand lifestyle patterns and the key factors influencing individual health behaviors [2]

In social and behaviour sciences, lifestyle is not only understood as a set of habits or physical activities, but also as a reflection of an individual's values, preferences, and choices in managing daily life[3]. This perspective is consistent with established health behavior theories, such as the Health Belief Model and Social Cognitive Theory, which emphasize that health-related behavior is shaped by cognitive, psychological, and environmental factors [4]. Lifestyle includes food consumption patterns, physical activity intensity, sleep duration and quality, and mental health. All of these factors require an analytical approach capable of processing data efficiently. Current technological developments have created a technology called machine learning that allows lifestyle analysis to be conducted objectively through predictive modeling [5]. Machine learning uses algorithms that are able to learn complex patterns and relationships in data, rather than relying on rule-based approaches. This allows users to make decisions with a higher degree of accuracy [6] One of the algorithms that is often used for classification is the Decision Tree.

Decision Tree is a classification method that represents decision rules in a tree-like structure, where selected attributes iteratively partition the data into increasingly homogeneous subsets. This approach enables the construction of relatively simple models with clear decision paths and minimal redundancy, facilitating the division of heterogeneous populations into smaller and more homogeneous groups based on objective variables [7]. The algorithm operates using a top-down strategy, starting from a root node and recursively splitting the dataset according to the most informative features until all instances within a node belong to the same class or no further meaningful splits can be performed. Decision Tree uses a top-down approach to start from a single node and recursively divide the data into smaller subsets using the most informative features. This division continues until all instances in the subset belong to the same class or when no more divisions are possible. Overall, the decision tree technique performs well on tabular-style datasets with individually meaningful features and without a strong multi-scale temporal or spatial structure. Decision trees are considered explainable because of their ability to represent the decision-making process of a model in a clear and transparent way [8].

Various studies have shown the effectiveness of decision trees in analyzing health behaviors such as Monasari and Pratiwi (2025) obtaining an accuracy of 95.73% accuracy in classifying the nutritional status of toddlers into four categories, namely normal, stunted, severely stunted, and high. The findings show that the decision tree algorithm is effective in identifying stunting conditions and supporting a digital nutrition monitoring system for children under five [6]. Sari et al. (2025) in their study compared the decision tree, random forest and naïve bayes algorithms in classifying sleep disorders and concluded

that the decision tree algorithm has optimal performance compared to the random forest and naïve bayes algorithms. This is shown by the highest accuracy results obtained in the decision tree algorithm which is 94.49%, as well as good precision and recall ability in distinguishing between people with sleep disorders and those without sleep disorders [10] Nawawi and Fatah (2024) also used this algorithm to identify healthy sleep patterns based on lifestyle variables and obtain consistent results. These findings confirm that the decision tree is a superior method for structurally and accurately mapping the dynamics of health behaviors [11]. show that there is a gap, namely the lack of studies that analyze lifestyle in a multidimensional manner using a comprehensive classification approach based on several attribute groups at once. In fact, lifestyle is a holistic construct and involves complex interactions between diet, physical activity, sleep patterns and emotional states. Khoirunnisa and Soleh (2024) only looked at the influence of a healthy lifestyle on physical and mental health, and showed results that a healthy lifestyle had a strong correlation with improved physical and mental health. A healthy lifestyle is also related to increased productivity and overall quality of life [12]. Sementra septiani et al. (2025) highlighted the relationship between the amount of food, type of food, meal schedule, and physical activity on the status of virgin sugar levels of people with Type 2 DM [13]. Although previous studies have shown the effectiveness of decision trees in analyzing various aspects of health behaviors, the focus of analysis is generally still limited to one specific dimension, such as nutritional status, sleep disorders, or healthy sleep patterns. In contrast to this approach, this study adopts a multidimensional lifestyle perspective by integrating four main domains, namely diet, sleep patterns, physical activity, and mental health in one classification framework. The results showed that diet was the most dominant factor in determining lifestyle classification, followed by sleep patterns, physical activity, and mental health, which were further visualized hierarchically through the structure of the decision tree. These findings not only reinforce the results of previous studies that emphasized the role of single factors, but also broaden understanding by showing the linkages and sequences of influence between lifestyle factors. Thus, this study places the decision tree not only as an accurate classification tool, but also as an analytical tool that is able to reveal lifestyle dynamics comprehensively and interpretively.

Based on this gap, this study offers the development of a lifestyle classification model based on a decision tree algorithm by utilizing 23 attributes grouped into four main domains, namely physical activity, diet, sleep patterns, and mental health. The main contribution of this research lies in the application of the decision tree as a classification model that is not only oriented to accuracy, but also to the ability to interpret through the structure of the tree and the resulting decision rules. This approach allows the identification of dominant factors as well as hierarchical relationships between lifestyle domains, thus providing a deeper understanding of the mechanisms of classification of healthy and unhealthy lifestyles. Thus, this research contributes to the development of a more transparent data-driven lifestyle assessment.

2. Method

This study employs an applied machine learning research design with an exploratory classification approach to analyze individual lifestyle patterns. This design is considered appropriate for lifestyle analysis as it enables the identification of complex relationships among behavioral attributes while maintaining model interpretability. The use of a decision tree algorithm allows the classification process to be transparent and explainable, making it suitable for understanding the contribution of each lifestyle factor to the classification outcome. The research method is systematically structured to ensure that the classification process is measurable, transparent and replicable. The overall workflow consists of data collection, data preprocessing, feature transformation and selection, dataset partitioning, model construction using the decision tree algorithm, and performance evaluation. The entire process was implemented using predefined parameter setting and standard machine learning libraries to support reproducibility.

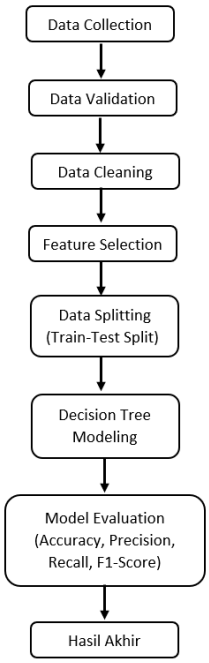

Figure 1. Research Stages

2.1. Data Collection

The research data was obtained from a questionnaire distributed to 130 respondents of various ages using the independent participation method. Respondents represent individuals with diverse lifestyle characteristics. The questionnaire consisted of 23 attributes measured using a likert scale, which were further grouped into five main categories, namely personal information, physical activity, dietary patterns, sleep patterns, and mental health conditions. The entire process of data processing and analysis is carried out using the python programming language. Table 1 lists all the features in the dataset.

Table 1. Attribute Description

Attribute Groups	Description
Personal Data	Demographic information including age, gender and occupation, collected for descriptive purposes and excluded from model training.
Physical Activity	Measures the frequency, duration and intensity of physical activities performed by respondents, assessed using an likert scale based on activity regularity.
Diet	Assesses dietary behavior including food categories consumed, meal frequency, portion size, breakfast habits, and beverage intake, measured using a likert scale reflection consumption patterns.
Sleep patterns	Captures sleep duration and sleep habits, including regularity and perceived sleep quality, measured using a likert scale
Mental health	Evaluates emotional conditions such as stress levels, mood stability, and psychological well-being, measured using self reported likert-scale items.

2.2. Pre-Processing Data

Data preprocessing was performed to ensure dataset quqlity before the modeling stage. This stage included data validation and data cleaning, as well as numerical transformation of questionnaire responses. All questionnaire items were originally measured using likert scale values and were retained in numerical form without further normalization, as decision tree algorithms are not sensitive to feature scalling. To represent overall behavioral tendencies, sub-attributes eithin each lifestyle category were aggregated using mean values, resulting in continuos numerical features suitable for entropy-based

splitting. This preprocessing strategy ensured consistent data representation while preserving the interpretability of the ordinal measurements.

2.2.1. Data Validation

Data validation was conducted to ensure the completeness and logical consistency of the dataset prior to modeling. Missing values were identified programmatically by checking for null or empty entries in each attribute. Instances with incomplete responses were excluded from further analysis to prevent bias in the classification process. In addition, logical boundary checks were applied to ensure that all attribute values fell within the predefined likert-scale ranges used in the questionnaire. Values outside the valid range were flagged and reviewed to maintain data consistency. These validation procedures were implemented using Phyton-based data inspection techniques, ensuring transparency and reproducibility of the preprocessing stage.

2.2.2. Data Cleaning

Data cleansing was performed by checking for duplicate records, inconsistencies, and irrelevant values. The inspection revealed that no duplicate entries were present in the dataset. Minor inconsistencies related to input formatting were identified and corrected, while no records required removal due to invalid or irrelevant values. As a result, that total number of instances remained unchanged after the data cleansing process, ensuring that the classification results were not influenced by data qulity issues.

2.3. Feature Selection

Feature selection was conducted to retain only attributes that meaningfully contribute to the lifestyle classification task. Personal data attributes were excluded based on both methodological and conceptual considerations. Preliminary exploratory modeling indicated that personal data variables did not contribute to information gain in the decision tree and did not appear in the resulting tree structure, suggesting a negligible influence on classification outcomes. In addition, these attributes do not directly represent lifestyle behaviors and were therefore considered less relevant for the analytical objective of this study. Removing personal data attributes helped reduce model complexity and improved interpretability without compromissing classification performance.

2.4. Data Splitting

The dataset was divided into training and testing sets using a 70 : 30 ratio. This ratio was selescted based on preliminary comparative experiments using different splitting schemes (60:40, 70:30, and 80:20), where the 70:30 split produced the most stable classification performance in terms of accuraccy and F1-score. This proportion also provided a sufficient number of instances for model training while preserving an adequate testing set for unbiased performance sevaluation. Given the limited dataset size, a fixed hold-out validation strategy was adopted, and the results are therefore interpreted within this sampling context.

2.5. Application of Methods

The main algorithm used in this study is the Decision tree, a supervised classification method that constructs a tree structure by recursively testing attributes at each node. The model was built using the entropy criterion and information gain to determine the optimal attribute for splitting at each node. During the training process, the dataset was partitioned into increasingly homogeneous subsets based on attibute threshold values. The tree growth process was terminated when all instances within a node belonged to the same class or when no further attributes were available for meaningful splitting. No explicit maximum tree depth constraint was applied, allowing the model to fully capture the underlying patterns in the data. The trained model was then evaluated using the testing dataset to measure classification performance. One of the main advantages of the decision tree algorithm is its ability to produce explicit and interpretable classification rules derived directly from the resulting tree structure.

3. Result and Discussion

In this study, the dataset used contains data collected from the distribution of questionnaires. The initial dataset consisted of 23 attributes representing various aspects of respondents' lifestyles. To improve the clarity of the analysis and reduce the complexity of the data, the attributes were then transformed by grouping them into five main attribute groups: personal data, physical activity, diet, sleep patterns, and mental health. This grouping is carried out based on the similarity of concepts and functional relationships between attributes, so that each group reflects a complete and theoretically relevant lifestyle dimension. This transformation aims to reduce redundancy between attributes, minimize noise in the data, and improve the stability and interpretability of the classification model. In the later stages of data processing, personal data attributes are omitted from the dataset because they do not directly represent lifestyle behaviors and have the potential to cause bias in the classification process. Thus, this process of attribute transformation allows the Decision Tree algorithm to focus on behavioral factors that substantively contribute to the classification of an individual's lifestyle, while generating a simpler, informative, and easier to interpret model. Not only the data on the attributes, the values on each attribute also undergo transformation, namely by adding the scores on each sub-attribute and then calculating the average value to be further processed using the decision tree algorithm. This approach is done because the items in each sub-attribute are designed to measure the same construct, so the aggregation of scores in the form of mean values can represent the characteristics of that domain more stably and comprehensively. In machine learning-based data analysis, particularly using decision tree algorithms, the use of the average value of the Likert scale sub-attributes is methodologically acceptable because this algorithm does not require certain data distribution assumptions and works based on the separation of numerical values to form decision rules. In addition, this transformation helps reduce model complexity, improves the interpretability of decision tree structures, and makes it easier to identify dominant factors in lifestyle classifications. The following is an example of the results of data transformation that has been carried out on the dataset that represents each individual.

Table 2. Dataset Transformation

Physical Activity	Diet	Sleep Patterns	Mental Health
2.4	2.2	1.75	4
2.8	3.6	3	2.2
3	4	2.75	3.2
3.6	3.4	2.75	2.8
1.6	3.8	2.75	2.4
3.6	4.2	3.5	3.4
1.6	4.4	2.5	2.4
1.8	3.4	2.5	3.8
3.6	3.8	2.75	2.2
3.2	4.2	3.25	3
3	3.4	2.25	2.4

After the data transformation process, the modeling process is carried out using the decision tree algorithm. The dataset of the transformation results was then divided into training data and testing data with a ratio of 70 : 30. Training data is used to build classification models, while test data is used to evaluate the performance of the resulting model. At the model formation stage, the decision tree algorithm applies the Entropy measure as the best attribute selection criterion for each node. The attributes with the highest information gain values are selected as separators to divide the dataset into more homogeneous subsets. The process of forming a decision tree is carried out repeatedly until the stop condition is met, which is when all data on a node belongs to the same class or when there are no more attributes that can be used to perform separation. In addition to these termination criteria, model complexity control is also considered to reduce the risk of overfitting, especially given the relatively limited size of the dataset. Complexity limitation is carried out by controlling the maximum depth of

the tree as well as the minimum number of samples at each leaf node, so that the resulting tree structure remains simple and able to generalize well on the test data. This approach allows for the formation of a decision tree that is not only structurally optimal, but also balanced between the classification capabilities and the interpretability of the model. The following is a decision tree generated using the decision tree algorithm.

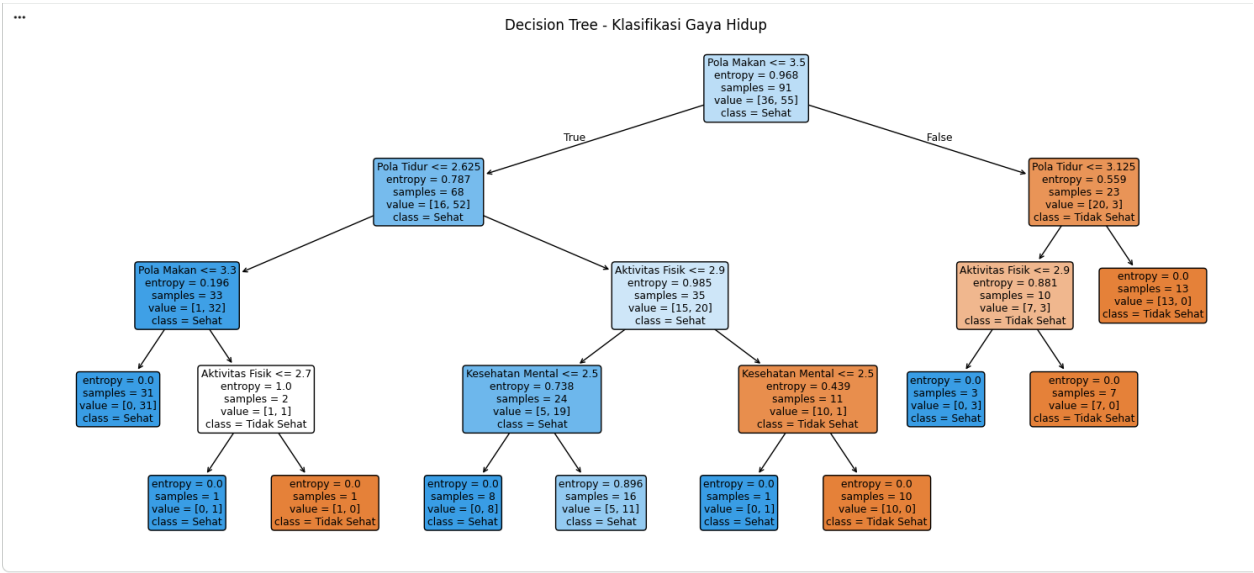

Figure 2. Decision Tree

The decision tree algorithm classifies an individual's lifestyle by building a decision tree structure through the selection of attributes based on the value of information gain. This process begins by selecting the most informative attribute, i.e., diet, as the root node, followed by other attributes such as sleep patterns, physical activity, and mental health as the follow-up nodes. Each attribute is used to divide the data into more homogeneous subsets until a leaf node is formed that represents healthy or unhealthy lifestyle classes. With this mechanism, the decision tree is able to model the hierarchical relationship between attributes and produce lifestyle classifications that are interpretive and easy to understand. From the decision tree in figure 2, the rules for lifestyle classification are obtained as follows:

```
Rules Decision Tree:

|--- Pola Makan <= 3.50
|   |--- Pola Tidur <= 2.62
|   |   |--- Pola Makan <= 3.30
|   |   |   |--- class: 1
|   |   |--- Pola Makan >  3.30
|   |   |   |--- Aktivitas Fisik <= 2.70
|   |   |   |   |--- class: 1
|   |   |   |--- Aktivitas Fisik >  2.70
|   |   |   |   |--- class: 0
|   |--- Pola Tidur >  2.62
|   |   |--- Aktivitas Fisik <= 2.90
|   |   |   |--- Kesehatan Mental <= 2.50
|   |   |   |   |--- class: 1
|   |   |   |--- Kesehatan Mental >  2.50
|   |   |   |   |--- class: 1
|   |   |--- Aktivitas Fisik >  2.90
|   |   |   |--- Kesehatan Mental <= 2.50
|   |   |   |   |--- class: 1
|   |   |   |--- Kesehatan Mental >  2.50
|   |   |   |   |--- class: 0
|--- Pola Makan >  3.50
|   |--- Pola Tidur <= 3.12
|   |   |--- Aktivitas Fisik <= 2.90
|   |   |   |--- class: 1
|   |   |--- Aktivitas Fisik >  2.90
|   |   |   |--- class: 0
|   |--- Pola Tidur >  3.12
|   |   |--- class: 0
```

Figure 3. Decision Tree algorithm rules

The rules on the decision tree show that the process of classifying an individual's lifestyle starts with the diet attribute as the main separator, indicating that this attribute is the most influential factor.

Individuals with a good diet tend to be categorized as having a healthy lifestyle, especially if they are supported by good sleep quality, adequate physical activity levels, and positive mental health conditions. In contrast, individuals with poor diets, especially when accompanied by poor sleep patterns or low mental health, have a greater tendency to be classified as not having a healthy lifestyle. In general, decision tree rules provide a simple and easy-to-understand picture of the relationship between lifestyle attributes, and show the model's ability to classify fairly well.

After the results of the classification and rules are obtained, the next step is to evaluate the decision tree model. This model evaluation is carried out to see how well the model performs classification tasks. Figure 4 below is the result of the evaluation of the decision tree model in conducting lifestyle classification.

```
=== Evaluasi Model Decision Tree ===
Akurasi : 84.62 %

Laporan Klasifikasi:
              precision    recall  f1-score   support

 Tidak Sehat       0.92      0.69      0.79        16
       Sehat       0.81      0.96      0.88        23

    accuracy                           0.85        39
   macro avg       0.87      0.82      0.83        39
weighted avg       0.86      0.85      0.84        39
```

Figure 4. Results of Decision Tree Algorithm Model Evaluation

The results of the evaluation showed that the decision tree model was able to achieve a fairly good classification performance with an accuracy value of 84.62%, which indicates that most of the test data was successfully classified appropriately into the healthy and unhealthy lifestyle classes. The lifestyle classes in this study were determined based on the established thresholds of the empirical composite score distribution, which were obtained by combining the average scores of four main domains, namely diet, physical activity, sleep patterns, and mental health. This approach follows the principle of empirical statistical categorization, where the cut-off is based on the characteristics of the sample data distribution, so that the classes formed represent actual lifestyle patterns in the research dataset.

In more detail, the model showed excellent ability to identify healthy lifestyles, which was reflected in a high recall value of 0.96 and an F1-score value of 0.88. However, the precision value in this class still indicates a prediction error, indicating that some individuals with certain characteristics tend to be classified as healthy lifestyles even though they are near the class limit. In contrast, in the unhealthy lifestyle class, the model produced a high precision value of 0.92, but with a lower recall value of 0.69. This pattern suggests an overlap of behavioral characteristics between the two classes, particularly in individuals with composite scores that are around the classification threshold.

These differences in recall performance between classes also need to be understood in relation to the distribution of classes in datasets that are not fully balanced. The dominance of one class in the data can increase the model's sensitivity to the majority class pattern, while the ability to detect minority classes becomes more limited. Therefore, the observed performance inequality does not solely reflect the intrinsic limitations of the algorithm, but is also influenced by the structure and proportions of the data used in the model training process.

Overall, the weighted average F1-score value of 0.84 indicates that the decision tree model has a good balance of performance in classifying both lifestyle classes. This performance was considered adequate in the context of the study, given that the main goal of modeling was not only oriented towards achieving the highest accuracy, but also on the ability to interpret and be transparent in

explaining the relationship between lifestyle factors. Compared to the black box classification method, the decision tree allows direct tracing of the established decision rules, making it more relevant for health behavior analysis that demands structural understanding and data-driven decision-making support. Nonetheless, further development is still needed to improve the sensitivity of the model to unhealthy lifestyle classes, especially through data enrichment or control of model complexity.

4. Conclusion

This study examines the application of the Decision Tree algorithm in classifying individual lifestyles into healthy and unhealthy categories based on questionnaire data. Utilizing the process of attribute transformation and sub-attribute value aggregation based on the Likert scale, the study assumes that each lifestyle domain—diet, physical activity, sleep patterns, and mental health—can be stably represented through empirical composite scores. This methodological assumption allows for the simplification of the data structure without eliminating the substantive meaning of the measured behavioral construct, and is in line with the characteristics of the Decision Tree algorithm that does not depend on certain data distribution assumptions. The results of the evaluation showed that the model was able to achieve adequate classification performance with a good level of accuracy and performance balance, as well as produce decision rules that are transparent and easy to trace. Nevertheless, the study also has a number of limitations, especially related to the relatively limited size of the dataset and the distribution of the class that is not completely balanced. This condition contributes to differences in the sensitivity of models in detecting healthy and unhealthy lifestyle classes, as well as limiting the ability to generalize results to a wider population with different demographic and behavioral characteristics. Nonetheless, the findings of this study provide important practical implications, particularly in the context of data-driven lifestyle assessments. The resulting Decision Tree model serves not only as a classification tool, but also as an exploratory means to identify the behavioral factors that most affect an individual's lifestyle status. These interpretive characteristics make the proposed approach relevant to support decision-making in the health sector, such as designing preventive interventions, healthy lifestyle education, and developing survey-based decision support systems. This study directly addresses the gap in the literature that is still dominated by high-accuracy-oriented classification models but minimal interpretability, by offering an approach that balances predictive performance and model transparency. For further research, development can be directed not only to the technical improvement of the model, but also to the expansion of the coverage of the dataset, cross-population validation, the integration of contextual variables, as well as the exploration of the impact of the use of interpretive models in the practice of lifestyle assessment and health policy. Thus, this research is expected to be the starting foundation for the development of a more comprehensive, contextual, and real-utilization-oriented approach to lifestyle analytics.

References

[1] WHO, “Physical activity,” https://www.who.int/en/news-room/fact-sheets/detail/physical-activity.
[2] N. Sundari, F. S. T. Dewi, and M. R. Ikhsan, “Perilaku gaya hidup dan obesitas sebagai faktor risiko kejadian diabetes melitus tipe 2 di RSUD. Aji Batara Agung Dewa Sakti Samboja Kabupaten Kutai Kartanegara,” Ber. Kedokt. Masy., vol. 32, no. 12, p. 461, Dec. 2016, doi: 10.22146/bkm.8337.
[3] P. Kotler and K. L. Keller, Marketing Management MARKETING MANAGEMENT Marketing Management, Global Edi., vol. 2, no. 1. 2021. [Online]. Available: http://download.garuda.kemdikbud.go.id/article.php?article=2354118%5C&val=22677%5C&title=The The influence of social media marketing on brand awareness and brand image moderating effect of religiosity
[4] M. A. Plow and J. Chang, “Promoting Healthy Behaviors in Stroke Survivors,” in Stroke Rehabilitation, Elsevier, 2019, pp. 279–290. doi: 10.1016/B978-0-323-55381-0.00020-2.
[5] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. 2021.
[6] S. Jahandideh, G. Ozavci, B. W. Sahle, A. Z. Kouzani, F. Magrabi, and T. Bucknall, “Evaluation of machine learning-based models for prediction of clinical deterioration: A systematic literature review,” Int. J. Med. Inform., vol. 175, p. 105084, Jul. 2023, doi: 10.1016/j.ijmedinf.2023.105084.
[7] A. F. A. Naibaho and A. Zahra, “PREDIKSI KELULUSAN SISWA SEKOLAH MENENGAH PERTAMA

MENGGUNAKAN MACHINE LEARNING," J. Inform. dan Tek. Elektro Terap., vol. 11, no. 3, Jul. 2023, doi: 10.23960/jitet.v11i3.3056.

[8] A. Tursunalieva, D. L. J. Alexander, R. Dunne, J. Li, L. Riera, and Y. Zhao, "Making Sense of Machine Learning: A Review of Interpretation Techniques and Their Applications," Appl. Sci., vol. 14, no. 2, p. 496, Jan. 2024, doi: 10.3390/app14020496.

[9] L. D. H. Monasari and Nunik Pratiwi, "Klasifikasi Stunting Pada Balita Menggunakan Algoritma Decision Tree Pada Machine Learning," Decod. J. Pendidik. Teknol. Inf., vol. 5, no. 2, pp. 416–428, Jul. 2025, doi: 10.51454/decode.v5i2.1193.

[10] I. N. Sari, M. K. L. F. Lidimilah, and M. K. A. Lutfi, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naive Bayes dalam Klasifikasi Gangguan Tidur," Proc. Natl. Conf. Sinesia, vol. 1, no. 1, pp. 192–209, Jun. 2025, doi: 10.69836/ncrcs-sinesia.v1i1.41.

[11] Imam Nawawi and Zaehol 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, Oct. 2024, doi: 10.59024/jiti.v2i4.969.

[12] Lisa Khoirunnisa and Nur Muhammad Soleh, "PENGARUH POLA HIDUP SEHAT TERHADAP KESEHATAN FISIK DAN MENTAL," Cent. Publ., vol. 2, no. 2, pp. 1686–1691, 2024.

[13] J. S. Santi and W. Septiani, "HUBUNGAN PENERAPAN POLA DIET DAN AKTIFITAS FISIK DENGAN STATUS KADAR GULA DARAH PADA PENDERITA DM TIPE 2 DI RSUD PETALA BUMI PEKANBARU TAHUN 2020," J. Kesehat. Masy., vol. 9, no. 5, pp. 711–718, Sep. 2021, doi: 10.14710/jkm.v9i5.30816.