International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 8, No. 2, (2024): pp. 110-116

Implementation of Machine Learning Classification of Obesity Weight using Dicision Tree

Fajar Rahardika Bahari Putra[1], Muhammad Surahmanto[2], Haris[3]
[1,2,3]Program Studi Teknik Informatika, Universitas Muhammadiyah Sorong, Indonesia
Email: fajar_rbp@um-sorong.ac.id[1], surahmanto@um-sorong.ac.id[2], harisaja2303@gmail.com[3]

Abstract

This work presents the application of the Decision Tree algorithm in the classification of obesity status using a machine learning approach. Indonesia is faced with three nutrition problems at once: stunting, wasting, and obesity or overnutrition. Obesity is a condition with excessive accumulation of body fat, which can lead to diseases and reduce quality of life. This study uses a dataset of 500 respondents and aims to classify obesity status early using the Decision Tree algorithm. The findings show that the developed Decision Tree model has an accuracy of 82%, with high precision and recall values, demonstrating the effectiveness of the algorithm in classifying obesity status. In conclusion, this study demonstrates the significant potential of the Decision Tree algorithm in supporting the early detection of obesity and facilitating more focused health interventions.

Keywords: *Obesitas, Klasifikasi, Machine Learning, Decision Tree, Indonesia*

1. Introduction

Indonesia is a country with three nutrition problems at once: stunting, wasting and obesity or overnutrition [1]. Obesity has become an increasing global health problem in recent decades [2], [3]. Obesity or overweight is a condition of excessive body fat accumulation so that a person's weight is far above normal, this occurs because of an imbalance between the energy that enters the body is greater than the energy used by the body [4]. The worldwide increase in obesity has an important impact on health problems and reduced quality of life [5]. Specifically in Indonesia, the prevalence of nutritional status in adults aged 18 and over in 2018 who were overweight was 12.8% compared to the proportion of obesity in 2013 of 14.8% [6]. North Sulawesi (30.2%), DKI Jakarta (29.8%), East Kalimantan (28.7), West Papua (26.4%), Riau Islands (26.2), followed by other provinces [7]. The prevalence continues to increase to 10.2% in 2030 and 10.9% in 2045 [8].

Factors contributing to obesity include changes in diet that increasingly rely on fast food and low nutrition, lack of physical activity among children, and lack of education on healthy eating. In addition, limited access to healthy food is also a problem, where many families prefer cheap and easily accessible foods that tend to be high in calories but low in nutrients. The cause of many health problems and even death is due to late detection. Classification makes it possible to find patterns in obesity body weight which allows early detection of obesity diseases.

One of the widely used classification techniques is Decision Tree In overcoming the problem of obesity weight classification, Decision Tree can be used as an effective algorithm in classifying obesity status [9]. By using Decision Tree, a deeper analysis of the factors that influence obesity status can be done and produce more accurate predictions. Therefore, this research aims to apply the Decision Tree algorithm to the classification of obesity status in Indonesia, especially the city of Sorong.

ISSN : 2580-7250

Thus the author seeks to make research, namely "IMPLEMENTATION OF MACHINE LEARNING CLASSIFICATION OF OBESITY WEIGHT USING DICISION TREE". In order to help the community in knowing a person's weight.

2. Research Methodology

This research uses a qualitative approach with data processing from classification machine learning with the decision tree method. The research design is exploratory and predictive, aiming to classify obesity weight using decision tree.

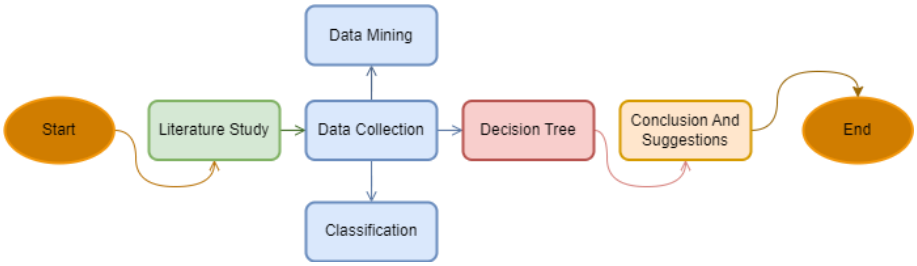

Figure 1. Research Flow [10]

2.1. Literature Study

Literature study is an important step in this research. The researcher traced and reviewed various scientific articles related to obesity and the use of machine learning for its classification. The main focus was on research published in the last 2 years.

2.2. Data Collection

Data collection is a crucial early stage in any research. This process aims to collect relevant and high-quality data that will be used in further analysis.

a) Data Mining

Data mining is a technique in which a large group of data is collected to gain knowledge, understanding the patterns in the data collection so that a general understanding of a particular data model can be achieved [11], [12]. Data mining allows managers to extract valuable information from big data that includes factors such as customer preferences, peak periods, and the effects of promos [13]. Through the application of complex algorithms and statistical methods, data mining also allows researchers and practitioners to unearth hidden information that can be used for better decision making [14]. The data mining process uses various techniques such as techniques in statistical, mathematical, and machine learning processes that are used in identifying and processing various data into useful information.

b) Classification

Classification is one of the most commonly used techniques for predicting disease occurrence [15]. To produce the best classification model, an effective algorithm is required [16]. One of the main advantages of using machine learning algorithms to classify obesity weight is the ability to process large amounts of data and identify patterns that may not be easily visible through traditional methods [17]. Therefore, a decision tree algorithm model is used for the implementation of classifying obesity in the community which is able to provide accurate classification results [18].

2.3. Decision Tree

Decision tree is the most favorite classification algorithm because it is easy for people to interpret [19][20]. When certain categorical groupings or classifications do not exist, regression analysis techniques can still predict possible outcomes based on independent

variables thanks to a hierarchical or divided tree structure [21]. This model can also find a collection of patterns or functions that describe and separate one data class from another which is used to predict data that has not selected a particular data class [22]. Apart from being relatively fast to build, the results of the model are easy to understand. In a decision tree, there are 3 types of nodes, namely:

1. Root Node, is the topmost node, in this node there is no input and can have no output or have more than one output.
2. Internal Node, is a branching node, in this node there is only one input and has a minimum of two outputs.
3. Leaf node or terminal node, is the final node, in this node there is only one input and has no output.

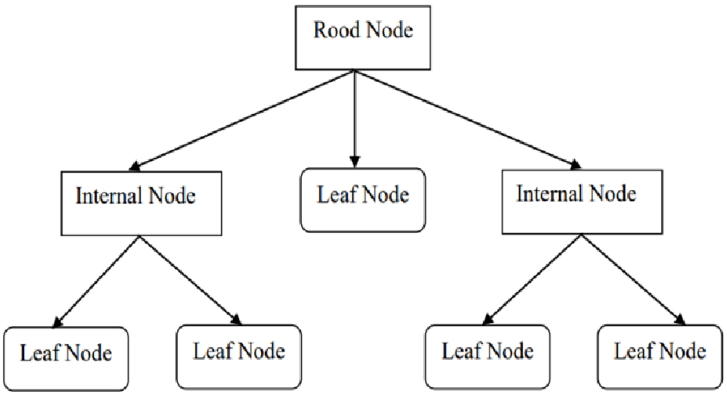

Figure 2. Decision Tree Model

2.4. Dataset

In this study, we used an obesity dataset taken from the Kaggle platform. This dataset consists of 500 respondents with various characteristics related to weight obesity.

Table 1. Dataset Characteristic

Characteristics	Description
Total Number of Samples	500
Input Variables	3 (Gender, Height, Weight)
Output Variable	1 (Label)
Gender	Male: 253 (50.6%), Female: 247 (49.4%)
Height Range	140 cm – 199 cm
Weight Range	50 kg – 160 kg

The dataset consists of three input variables and one output variable, namely;

a. Gender, a categorical variable (Male/Female).
b. Height, a continuous numeric categorical variable (in cm).
c. Body Weight, a continuous numeric variable (in kg)
d. Label, ordinal categorical variable distribution label weight classification(0-5)

2.4.1. Label Distribution

The distribution of labels in the dataset is presented in table 2.

Table 2 Weight Classification Label Distribution

Label	Frequency	Percentage
0	10	2%
1	15	3%
2	53	10.6%

Label	Frequency	Percentage
3	62	12.4%
4	135	27%
5	225	45%

Based on this distribution, it can be interpreted that label 0 may represent 'Very Thin', label 1 'Thin', label 2 'Lower Normal', label 3 'Upper Normal', label 4 'Overweight', and label 5 'Obese'. However, this interpretation needs to be further verified.

2.4.2. Dataset Initial Analysis

The initial analysis revealed some important characteristics of the dataset, namely:

a. Average height of about 170.5 cm ± 15.3 cm
b. Average body weight of approximately 104.3 kg ± 31.2 kg
c. The sex ratio is almost balanced, providing good representation for both genders.
d. There is a significant imbalance in the label distribution, with label 5 dominating 45% of the total sample.

3. Result and Discussion

This study aims to classify obesity status based on anthropometric data using the Decision Tree model. The dataset used consists of 500 samples with predictor variables including gender, body mass, and height. The analysis process includes the stages of Data Understanding, Data Cleaning, Exploratory Data Analysis (EDA), Data Preparation, Modeling, Evaluation, and Testing.

The EDA results show the distribution of height and gender as shown in Figure 3.

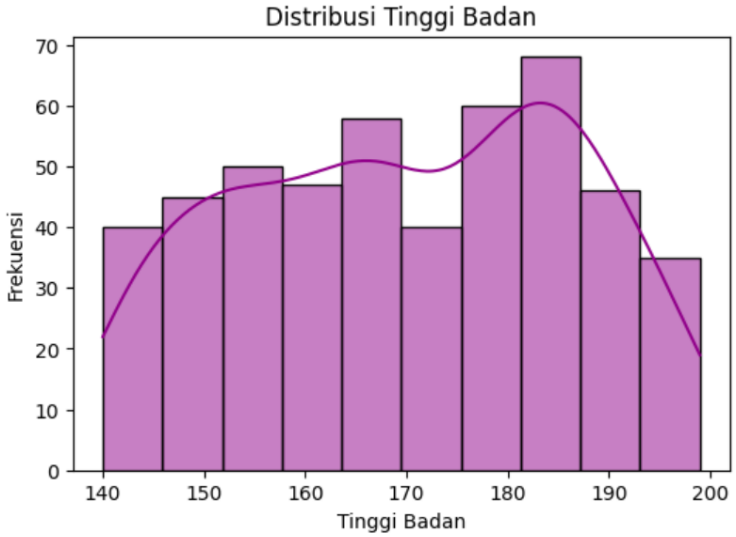

Figure 3. Height and Gender Distribution

Analysis of height distribution using the seaborn library showed the highest frequency was in the 180-190 cm range, while the lowest frequency was in the 190-200 cm range. Gender visualization using plotly.express displays the proportion of males (blue) and females (pink) against weight labels (0-5), with label 0 indicating the lowest frequency and label 5 the highest frequency.

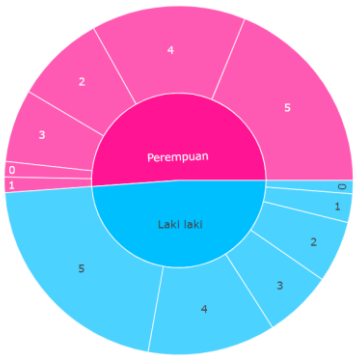

Figure 4. Visualization of Gender and Weight

The Decision Tree algorithm was selected for classification based on its high interpretability and ability to handle non-linear data. The model evaluation used accuracy, precision, recall, F1-score, and confusion matrix metrics for comprehensive analysis. The model evaluation results show:

1. Accuracy: 82%
2. Precision: 0.87
3. Recall: 0.94
4. F1-score: 0.91

The constructed Decision Tree model achieved an accuracy of 81.63%, indicating good classification ability. The high precision and recall values indicate the effectiveness of the model in identifying obesity categories with a relatively low error rate. The decision tree structure of the classification results can be seen in Figure 5.

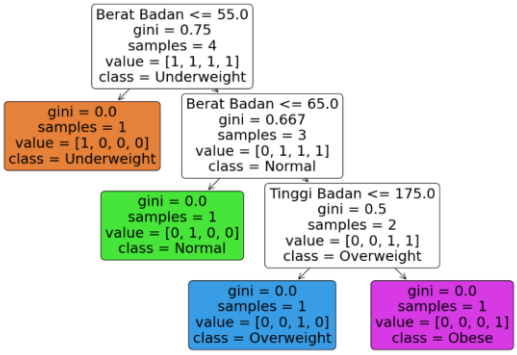

Figure 5. Classification result decision tree structure

This model demonstrates significant potential in supporting early detection of obesity and facilitating more targeted health interventions. However, it should be noted that interpretation of the results should consider the clinical context and other factors that may influence obesity status. The results of this study highlight the effectiveness of the Decision Tree algorithm in the classification of obesity status based on anthropometric data. These findings may contribute to the development of clinical decision support systems for more effective obesity management.

4. Conclusion

This research successfully implemented the Decision Tree algorithm for obesity weight classification using machine learning techniques. Analysis of a dataset consisting of 500 samples with anthropometric variables showed that the developed Decision Tree model performed well in classifying obesity status, with accuracy reaching 82%, precision 0.87, recall 0.94, and F1-score 0.91. Exploratory Data Analysis (EDA) revealed

informative height and gender distribution patterns, providing initial insights into the dataset characteristics. The decision tree visualization of the classification results presents a transparent decision-making structure, improving the interpretability of the model. The results of this study demonstrate the significant potential of the Decision Tree algorithm in supporting early detection of obesity and facilitating more targeted health interventions. Although the model showed promising performance, it should be noted that interpretation of the results should consider the clinical context and other factors that may influence obesity status. This study makes an important contribution to the development of clinical decision support systems for more effective obesity management, however further research is needed to improve the accuracy of the model and explore its applicability in a broader clinical context. These results also open up opportunities for the development of practical applications, such as integration with mobile platforms for real-time monitoring of weight status. For future research, it is recommended to expand the dataset by including additional variables such as age, family medical history, and diet to improve the accuracy and reliability of the model. In addition, the application of ensemble learning techniques or other machine learning algorithms such as Random Forest or Gradient Boosting can be explored to compare performance with the Decision Tree model. Cross-validation is also recommended to test the consistency of the model on various subsets of data. Further research can focus on developing practical applications based on this model, such as clinical decision support systems or mobile applications for real-time monitoring of weight status.

Acknowledgments

The authors would like to thank Universitas Muhammadiyah Sorong for their support and guidance in the preparation of this journal.

References

[1] W. Purnomo, T. Susanto, and A. T. Afandi, “Studi Literature Pola Makan dan Pola Aktivitas Fisik pada Remaja dengan Obesitas di Indonesia,” *J. Sahabat Keperawatan*, vol. 6, no. 01, pp. 115–134, 2024, doi: 10.32938/jsk.v6i01.6437.

[2] A. Subandi, D. Noerjoedianto, and N. L. Parawansa, “Jurnal PEDAMAS (Pengabdian Kepada Masyarakat) Volume 1 , Nomor 4 , November 2023 ISSN : 2986-7819,” vol. 1, no. November, pp. 677–686, 2023.

[3] Adilah Pradipta Syahri and Ratih Kurniasari, “Pengaruh Pemberian Konseling Gizi terhadap Pemilihan Makan Remaja Obesitas : Literature Review,” *Media Publ. Promosi Kesehat. Indones.*, vol. 7, no. 2, pp. 307–312, 2024, doi: 10.56338/mppki.v7i2.4490.

[4] N. Nurhidayati, I. R. Irawan, and Sudikno, “Hubungan Obesitas Dengan Profil Lipid Pada Remaja Di Indonesia,” *Penelit. Gizi dan Makanan (The J. Nutr. Food Res.*, vol. 45, no. 1, pp. 35–46, 2022, doi: 10.22435/pgm.v45i1.6081.

[5] R. Dianah, E. A. Andari, Elvira Anjani Putri, Cahya Chita Dwinanti, and D. N. Nafisah, “Penyuluhan Cara Mencegah Obesitas Pada Remaja dengan Pola Makan Yang Sehat,” *J. Abdimas ADPI Sains dan Teknol.*, vol. 3, no. 3, pp. 41–50, 2022, doi: 10.47841/saintek.v3i3.220.

[6] T. G. A. D. Pemayun, A. A. G. Budhitresna, and P. A. N. K. Permatananda, “Gambaran Tingkat Aktivitas Fisik dan Kejadian Obesitas pada Civitas Akademika Universitas Warmadewa , Bali,” *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 8526–8532, 2022.

[7] M. M. Putra, N. N. I. Saraswati, and N. M. Raningsih, “Pola Hidup Dengan Kejadian Obesitas: Literature Review,” *J. Ilmu Keperawatan Med. Bedah*, vol. 5, no. 1, pp. 15–35, 2022, doi: 10.32584/jikmb.v5i1.1166.

[8] M. R. Kuna, M. Ananda, O. Manika, and T. Pobela, "Analisis Penggunaan Obat Pada Pasien Diabetes Melitus Di Puskesmas Gogagoman Tahun 2022," *J. Cakrawala Ilm.*, vol. 2, no. 4, pp. 1631–1638, 2022, doi: 10.53625/jcijurnalcakrawalailmiah.v2i4.4421.

[9] A. Z. Zami, O. Nurdiawan, and G. Dwilestari, "Klasifikasi Kondisi Gizi Bayi Bawah Lima Tahun Pada Posyandu Melati Dengan Menggunakan Algoritma Decision Tree," *J. Sist. Komput. dan Inform.*, vol. 3, no. 3, p. 305, 2022, doi: 10.30865/json.v3i3.3892.

[10] F. R. B. Putra, A. Fadlil, and R. Umar, "Application of Forward Chaining Method , Certainty Factor , and Bayes Theorem for Cattle Disease," vol. 14, no. 1, pp. 365–374, 2024, doi: https://doi.org/10.18517/ijaseit.14.1.18912.

[11] M. S. Tahsin, S. Abdullah, M. Al Karim, M. U. Ahmed, F. Tafannum, and M. Y. Ara, "A comparative study on data mining models for weather forecasting: A case study on Chittagong, Bangladesh," *Nat. Hazards Res.*, vol. 4, no. 2, pp. 295–303, 2024, doi: 10.1016/j.nhres.2023.12.014.

[12] M. Gheisari *et al.*, "Data Mining Techniques for Web Mining: A Survey," *Artif. Intell. Appl.*, vol. 1, no. 1, pp. 3–10, 2022, doi: 10.47852/bonviewaia2202290.

[13] N. Istiqomah, M. A. Ridla, and N. Azise, "Gudang Jurnal Multidisiplin Ilmu Data Mining : Tingkat Penghuni Kamar Hotel Di Aceh Dari Tahun 2018-2022 Menggunakan Aplikasi Zaitun," vol. 2, pp. 9–12, 2024.

[14] X. Meng, "Research on the Development of Modern Design Through Data Mining Technology," *Inform.*, vol. 48, no. 6, pp. 59–70, 2024, doi: 10.31449/INF.V48I6.5241.

[15] V. S. Souza and D. A. Lima, "Identifying Risk Factors for Heart Failure: A Case Study Employing Data Mining Algorithms," *J. Data Sci. Intell. Syst.*, vol. 2, no. July 2023, pp. 161–173, 2023, doi: 10.47852/bonviewjdsis32021386.

[16] N. Afiatuddin, Mt. Wicaksono, V. Rezky Akbar, and D. Wulandari, "Komparasi Algoritma Machine Learning dalam Klasifikasi Kanker Payudara," *J. Media Inform. Budidarma*, vol. 8, no. 2, pp. 889–899, 2024, doi: 10.30865/mib.v8i2.7457.

[17] A. H. RS and R. A. Permata, "Analisis Metode Klasifikasi Penyakit Bell's Palsy Menggunakan Machine Learning," *Empiricism J.*, vol. 5, no. 1, pp. 127–139, 2024, doi: 10.36312/ej.v5i1.1610.

[18] N. Arminarahmah, G. Mahalisa, and T. Informatika, "Implementasi Model Machine Learning pada Klasifikasi Status Penyakit Diabetes Berbasis Streamlit Implementation of Machine Learning Models in Diabetes Disease Status Classification Based on Streamlit," vol. 13, no. 105, pp. 470–475, 2024.

[19] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.

[20] Z. Azam, M. M. Islam, and M. N. Huda, "Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree," *IEEE Access*, vol. 11, no. July, pp. 80348–80391, 2023, doi: 10.1109/ACCESS.2023.3296444.

[21] R. Zhou, Y. Tang, H. Li, and Z. Liu, "Predicting the compressive strength of ultra-high-performance concrete using a decision tree machine learning model enhanced by the integration of two optimization meta-heuristic algorithms," *J. Eng. Appl. Sci.*, vol. 71, no. 1, pp. 1–17, 2024, doi: 10.1186/s44147-023-00350-1.

[22] A. R. Raharja, Jayadi, A. Pramudianto, and Y. Muchsam, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang," *Technol. J.*, vol. 1, no. 1, 2024, doi: 10.62872/cwgzp962.