

Classification of High Risk of Obesity in Women using Decision Tree Methods

Riftika Rizawanti*, Rajiansyah Rajiansyah

Fakultas Teknik, Prodi Informatika, Universitas Mulawarman, Samarinda, Indonesia

Email: ¹*riftikariza@ft.unmul.ac.id, ²rajiansyah@ft.unmul.ac.id

Correspondence Author Email: riftikariza@ft.unmul.ac.id

Submitted: 05/06/2025; Accepted: 30/06/2025; Published: 30/06/2025

Abstract—Excessive body fat accumulation characterizes obesity, a medical condition primarily caused by an energy imbalance. This excess fat is stored throughout the body, including the abdomen, thighs, and arms. Obesity is a global health concern, prevalent in Indonesia, impacting physical, psychological and social well-being. Women are more susceptible to obesity due to a combination of biological and lifestyle factors. A community health center study of 156 patients revealed that 71.20% exhibited central obesity, with women comprising 76.60% of this group and men 31.6%. This study focuses on the disproportionate impact of obesity on women. To better understand and address obesity, classification is crucial. This study uses a Decision Tree method to classify 898 women based on 14 assessments for high obesity risk, comparing its performance using three attribute selection criteria. The Decision Tree (Gini Index) model achieved 77.22% overall accuracy (Figure 12). The Normal category has 83% precision and 88.30% recall. The Overweight category had 62.50% precision and 63.83% recall. The Obese category had 75% precision and 66.67% recall. The Underweight category achieved 100% precision and recall. While the model demonstrates good classification performance, particularly for Normal and Underweight categories, it requires further refinement to better differentiate between Overweight and Obese individuals.

Keywords: Classification; Decision Tree; Obesity; Data Mining; Overweight

1. INTRODUCTION

Obesity is a medical condition characterised by excessive fat accumulation in the body. An imbalance between calorie intake and the amount of calories burnt is the main cause of this condition. The excess fat is then stored in various parts of the body, such as the abdomen, thighs and arms. Obesity has become a worldwide public health problem, including in Indonesia. This condition not only adversely affects physical health, but can also cause psychological and social problems. Women have a higher risk of developing obesity due to complex biological and lifestyle factors. A study at a community health centre showed that out of 156 patients examined, 71.20% were centrally obese. Interestingly, the majority of those with central obesity were women (76.60%), while men were only 31.6%[1]. This dominance of women in obesity cases is the main focus of this study, which will specifically analyse data from women affected by obesity.

Given the complexity of the obesity problem, classification is a crucial step to understanding and addressing it more effectively. Grouping data on obese women based on factors such as obesity severity, associated risk factors (e.g., family history, lifestyle, or other medical conditions), and demographic characteristics (such as age, ethnicity, or socioeconomic status) allows for the identification of specific patterns and significant differences between groups. This information is invaluable in developing more targeted prevention and intervention strategies. One classification method that can be used is Decision Tree.

Decision Tree is one of the methods in machine learning that has undergone development since the mid-20th century. Algorithms such as CART, ID3, random forests, and C4.5 have been the foundation of modern decision trees. Key concepts such as gini index and binary splitting play an important role in the formation of Decision Tree models. Today, Decision Tree has become a very important tool in machine learning[2][3][4].

M Riski Qisthiano, et al (2023) conducted research using datasets from several universities in Palembang City. The model used is Decision Tree, which functions as one of the classification methods. The dataset used is alumni data that has been collected from universities in Palembang City. The classification criteria include major, college, type of elective class, student grades from the first semester to semester 4, graduation year, and student entry year. After determining the data attributes to be used in the classification process, researchers used Rapidminer tools to manage classification data with the Decision Tree model. Furthermore, testing was carried out using 5 times the K-Fold Validation test process by dividing the dataset into training data and testing data. The results of this study show that the accuracy of the classification results against the predictions obtained from the Rapidminer tool and the Decision Tree method is 87.93%[5].

Aldi Tangkelayuk and Evangs Mailoa conducted research in 2022, this study used water quality datasets and compared three algorithms including Naive Bayes, Decision Tree, and K-Nearest Neighbors to find out which method is most accurate in classifying data. The highest level of accuracy became the benchmark for comparison. The results show that the K-Nearest Neighbors method has the highest accuracy rate, which is 86.88%, compared to Decision Tree (80.84%) and Naive Bayes (63.60%). Therefore, it can be concluded that the K-Nearest Neighbors method is the best for data classification in this context[6].

Eplia Triwira Lestari, et al (2022) conducted research on the nutritional status of a person who generally uses the Body Mass Index (BMI) formula. However, this formula has limitations because it does not consider the size of

the body frame, even though this factor greatly affects a person's weight, as does the abdominal circumference. Two people with the same weight and height can have different nutritional statuses due to differences in their body frame size. This study aims to classify the nutritional status of obesity in children with disabilities in SLB Sri Mujinab Pekanbaru. The algorithms used in this research are Naive Bayes Classifier (NBC) and K-Nearest Neighbor (KNN). The attributes used include name, gender, type of need, education level, height, weight, and abdominal circumference. The results show that the KNN algorithm provides the highest accuracy, which is 68%, compared to NBC which only reaches 48%. Therefore, in this study, classification using the KNN algorithm proved to be better than NBC. The KNN algorithm was then implemented on a web-based information system. The system was tested using Blackbox Testing with a 100% success rate, and User Acceptance Test (UAT) testing resulted in a success rate of 89.6%[7].

A study conducted by Cici Emilia Sukmawati, et al. in 2024 evaluated the performance of the AdaBoost and XGBoost algorithms in classifying obesity diseases. They used a dataset obtained from Kaggle. After the data was preprocessed, models were built and evaluated using accuracy, precision, and recall metrics. The results showed that XGBoost showed superior performance compared to AdaBoost, with accuracy, precision, and recall rates reaching 92%[8].

Based on several previous studies that have been described, this study will classify obesity in women using the Decision Tree method with several parameters used to select attributes that will be used in each node when dividing data, including gini index, information gain and gain ratio. The performance of the three parameters will be compared to find out which parameters are best used for the decision tree method in the classification of obesity.

2. RESEARCH METHODOLOGY

Obesity classification research using Gain Ratio, Information Gain and Gini Index parameters in the Decision Tree method involves several stages, from data collection to model application. The selection of the right parameters will affect the performance of the model in classifying obesity, the following stages of research that will be carried out in this study can be seen in Figure 1 below:

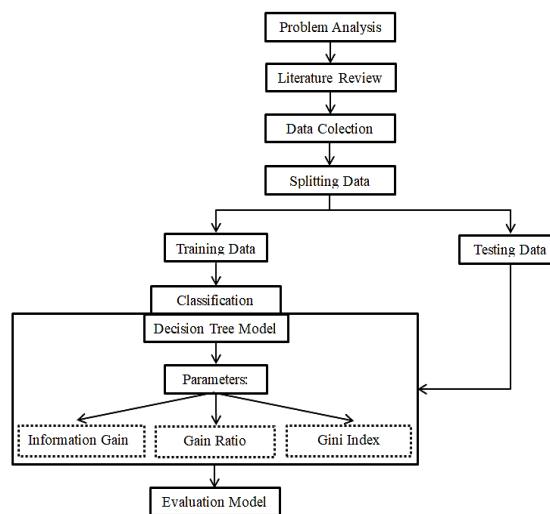


Figure 1. Research Stages

Based on Figure 1, the following is a detailed explanation of each stage of the research carried out from problem analysis to evaluation.

a. Problem Analysis

The problem analysis stage is a crucial step in understanding and identifying the root cause of a problem. The main objective of this stage is to find the true source of the problem, so that an appropriate and effective solution can be found. The problem analysis process involves collecting, analysing and interpreting data to identify the factors causing the problem as well as its impact on the affected system. Problem analysis aims to gain a comprehensive understanding of the problem at hand and the reasons why the problem has arisen. By understanding the root of the problem, relevant steps can be taken to solve or address the problem at hand. Problem analysis has a very important role in directing problem-solving efforts appropriately and effectively by focusing on addressing the source of the problem, not just the symptoms.

b. Literature Review

The next stage in the research is to conduct a literature study or literature review, which is an important step to produce quality research. This process involves collecting, analysing, evaluating and compiling various sources of information or relevant literature, such as books, journals, articles, reports and other documents related to the topic or problem under study. The main purpose of a literature review is to gain a deeper understanding of the topic or problem being researched, as well as identify areas of knowledge that are lacking or require further research.



c. Data Collection

The third step is to collect data related to obesity. This may include medical data (such as body mass index (BMI), age, gender), lifestyle data (such as diet, physical activity), or other relevant data. The data sample used is an obesity dataset taken from kaggle, the dataset used has 14 variables required to determine obesity, where the dataset contains male and female obesity data. This research focuses on female obesity data so that only 898 data are used as data samples from the total obesity dataset of 1610[9]. the following table 1 contains the data samples used in this study.

Table 1. Sample Data

No.	Age	Height	Overweight_O bese Family	Consumption_ of Fast Food	Type_of_Transp ortation Used	Class
1	18	155	2	2	4	Normal
2	18	158	2	2	3	Normal
3	18	159	2	2	4	Normal
4	18	162	2	2	4	Normal
5	18	165	2	1	2	Normal
6	18	176	1	1	4	Normal
7	19	152	2	2	2	Normal
8	19	158	2	2	3	Normal
9	19	159	2	2	4	Normal
10	19	162	2	2	4	Overweight
...
...
448	53	166	2	1	1	Obesity
449	54	169	1	1	1	Obesity
450	18	156	2	2	4	Normal
451	18	161	2	2	3	Underweight
452	18	160	2	2	4	Normal
...
...
897	53	168	2	1	1	Obesity
898	54	170	1	1	1	Obesity

d. Splitting Data

The data samples that have been determined will then be divided into two parts, namely training data and test data. In this study, 80% of the data samples will be used as training data and 20% will be used as test data. The outcomes of splitting data are presented in Table 1.

Table 1. Data Splitting Outcomes

Data Training	Data Testing
718	180

e. Decision Tree Model Building

At this stage, the Decision Tree model will be built using the processed data. Decision Tree is one of the popular supervised learning methods used for classification and regression tasks. It works by building a tree structure that resembles a decision-making pipeline, where each node in the tree represents an input feature, each branch represents the possible values of that feature, and each leaf represents the resulting class or target value. The decision tree building process is done recursively by selecting the most informative features to split the data into more homogeneous groups. This separation is repeated until all the data in a group has the same class or target value. Once the decision tree is formed, we can use it to predict the class or target value of new data by following the path of the branches on the tree that correspond to the values of the data input features. Decision trees have the advantage of being easy to interpret and visualise, making it easier to understand how the model makes decisions. In addition, it can also handle data with various types of features, both numerical and categorical[10][11][12][13][14][15]. In this context, we will use three parameters to select the best attributes to build the decision tree:

1. Information Gain

Measures the entropy reduction after separating data based on a particular attribute. The attribute with the highest Information Gain will be selected as the root or the node of the tree[16][17][18].

2. Gain Ratio

An improvement of Information Gain that considers the number of branches generated by an attribute. Gain Ratio is calculated by dividing the Information Gain by the intrinsic value of the attribute[19][20].

3. Gini Index

Measures the impurity of a data set. The attribute with the lowest Gini Index will be selected as the separator[21][22].

The Decision Tree model's calculation is presented in equations (4), (5), and (6) [23].

$$Entropy = - \sum_{v \in values(A)} \left(\frac{|S_v|}{|S|} \times \log_2 \left(\frac{|S_v|}{|S|} \right) \right) \tag{1}$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \left(\frac{|S_v|}{|S|} \times Entropy(S_v) \right) \tag{2}$$

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \tag{3}$$

f. Model Evaluation

Once the Decision Tree model is formed, it needs to be evaluated to measure its performance. Evaluation can be done using metrics such as accuracy, precision, recall, or F1-score. Accuracy is a calculation to measure the proportion of total correct identifications, Precision is a calculation to measure the proportion of truly positive identifications, Recall is a calculation to measure the proportion of all positive cases that were predicted correctly, and the F1-score is a metric that is used to determine the optimal balance between Precision and Recall. It is calculated by averaging precision and recall. The data used for evaluation is data that was not used during model building (test data). The classification report assesses model performance by calculating True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [24][25][13][27][27].

1. Accuracy

$$Accuracy = \frac{TP+TN}{TPP+FP+TN+FN} \tag{4}$$

2. Precision

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

3. Recall

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

4. F1-score

$$F1 - Score = 2 * \frac{(Precision*Recall)}{Precision+Recall} \tag{7}$$

3. RESULT AND DISCUSSION

This study aims to classify 898 female data consisting of 14 assessments that can be used to determine High Risk of Obesity. This research will compare the performance of the Decision Tree method against 3 criteria as a parameter to select the best attribute as a separation node. The following is a trial using rapidminer tools to see the accuracy of the Decision Tree algorithm based on each of these parameters in classifying High Risk of Obesity, as for the operators used by researchers in this study can be seen in Figure 2 below.

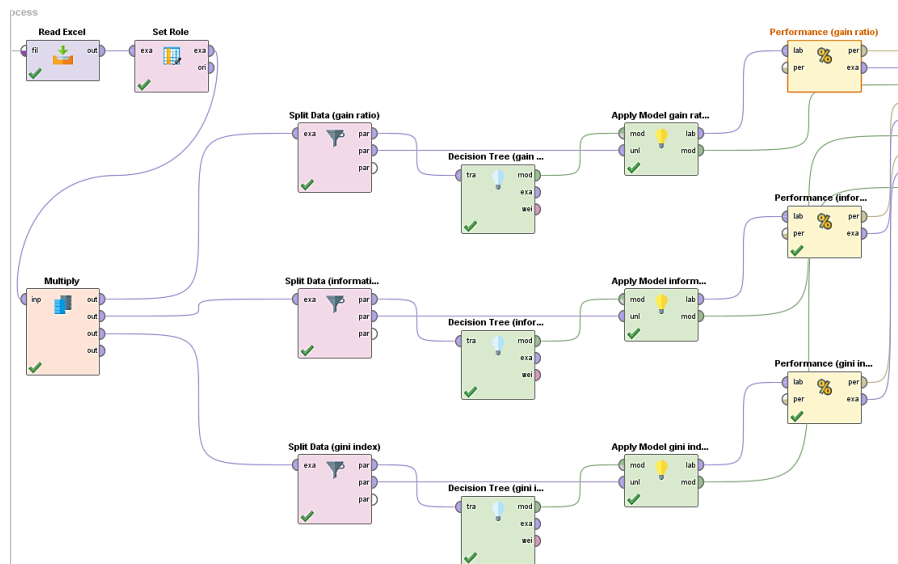


Figure 2. Building models with Rapidminer

Based on Figure 2 above, it can be explained that before carrying out the classification process, the data samples to be processed and the necessary operators need to be adjusted. In Figure 1, it can be seen that the initial tool in the classification process is Read Excel, where in this study the data sample file to be classified is in excel format so that researchers have these tools. Besides read Excel. There are other tools that can be used as dataset input to the classification process such as read CSV, read URL and others. After the data is input, then use the ‘Set Role’ tool to determine the attributes that are used as labels because in classification there must be one attribute that becomes the label, in this study the attribute used as a label is the ‘Class’ attribute, where there are 4 labels including Underweight, Normal, Overweight and Obesity. Furthermore, the ‘Multiply’ tool is used as a link for the 3 classification processes that will be carried out starting from splitting data to assessing the performance of each parameter. The following figure 3 is each label of the data sample used.

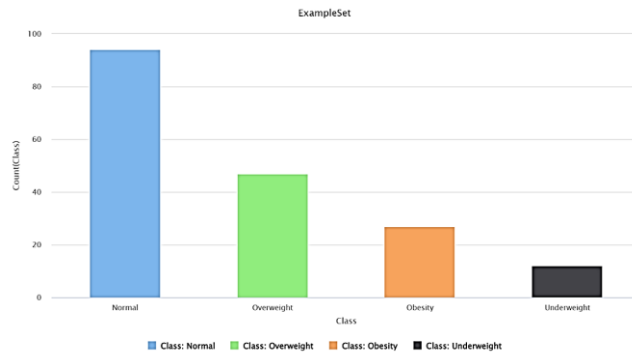


Figure 3. Distribution of sample data

Based on Figure 3, it can be seen that the normal class is more than other classes, namely 94 data, while the Overweight class is 47 or half of the normal class. Meanwhile, the obese class is only 27 and the underweight class is quite small at 12 data.

3.1 Decision Tree (Information Gain)

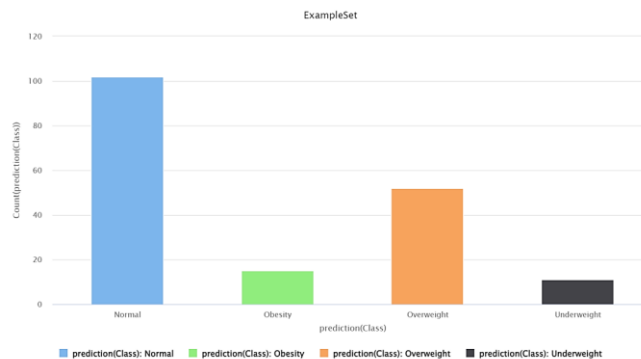


Figure 4. Class Prediction Distribution with Information Gain Parameter

The bar chart in Figure 4 presents the ‘ExampleSet’ data with four class prediction categories. The ‘Normal’ category is shown in blue and has the highest count of 102, indicating that the majority of the data is normal. The ‘Overweight’ category is represented by the orange colour with a count of 52, indicating a significant amount of overweight data. The green-coloured ‘Obesity’ category has a count of 15, indicating a small portion of the data is obese. Finally, the ‘Underweight’ category coloured black has the lowest number of 11, indicating the smallest proportion of underweight data.

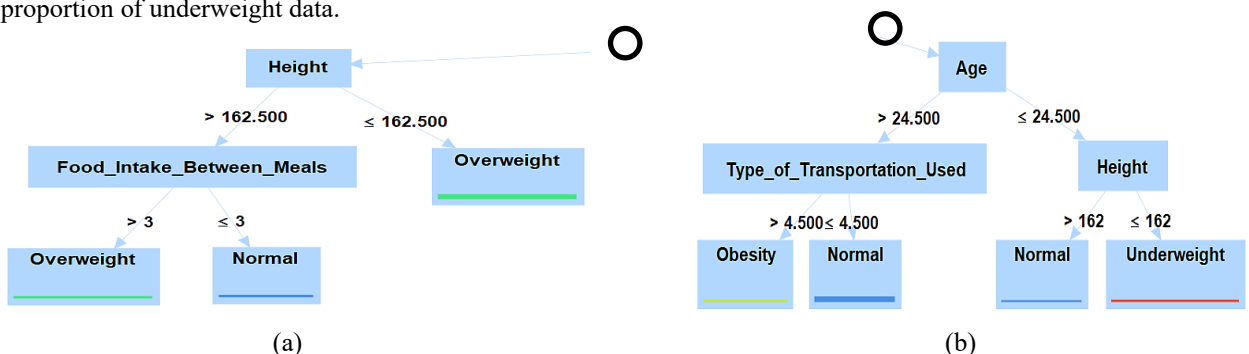


Figure 5. Visualisation of Decision Tree for Obesity Classification with Information Gain Parameters

Figure 5 above is a slice of the decision tree in the classification process using the Information Gain parameter. Tree diagram (a) shows that to predict whether someone is overweight or not, two criteria can be used, namely height and food intake between meals. If a person has a height of more than 162.5 and food intake between meals is more than 3 times, then the person is predicted to be overweight. If the height is less than or equal to 162.5, then the person is also predicted to be overweight regardless of food intake between meals. On the other hand, decision tree (b) shows that to classify a person into obese, normal, or underweight categories, three criteria can be used: age, type of transport used, and height. If a person is more than 24.5 years old and the type of transport used is more than 4.5, then the person is classified as obese. If a person is more than 24.5 years old and the type of transport used is less than or equal to 4.5, then the person is classified as normal. If a person is less than or equal to 24.5 years old and their height is more than 162, then the person is classified as normal. If a person is less than or equal to 24.5 years old and their height is less than or equal to 162, then the person is classified as underweight.

accuracy: 80.56%

	true Normal	true Overweight	true Obesity	true Underweight	class precision
pred. Normal	89	10	2	1	87.25%
pred. Overweight	4	34	13	1	65.38%
pred. Obesity	0	3	12	0	80.00%
pred. Underweight	1	0	0	10	90.91%
class recall	94.68%	72.34%	44.44%	83.33%	

Figure 6. Evaluation of Classification Model with Information Gain Parameters Using Confusion Matrix

Figure 6 presents a confusion matrix used to evaluate the performance of the high risk of obesity classification model with Decision Tree (Information Gain). The numbers in the matrix show that the rows represent the actual label or condition of the patient while the columns represent the prediction results of the model. The diagonal numbers (89, 34, 12, 10) show the number of correct predictions for each category while the off-diagonal numbers show the misclassification, i.e. how many samples from a category were incorrectly predicted as another category. For example, 89 patients who were actually Normal were predicted as Normal by the model, 10 patients who were actually Normal were mispredicted as Overweight, 13 patients who were actually Overweight were mispredicted as Obese, and so on. The overall accuracy of the model is 80.56%, although it is high, but from this matrix, it can be seen that the model has the best performance in classifying Normal and Underweight patients, but not good in classifying Overweight and Obese patients. This can be seen from the lower precision and recall values for both categories.

3.2 Decision Tree (Gain Ratio)

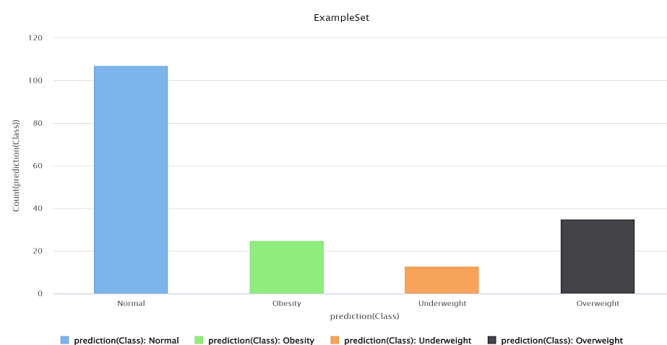


Figure 7. Class Prediction Distribution with Gain Ratio Parameter

Figure 7 presents the data in the form of a bar chart showing the class predictions with quantitative details. The blue colour representing 'Normal' has the highest number of 107, indicating that the majority of the data is normal. It is followed by the black colour representing 'Overweight' with a count of 35, indicating a significant amount of overweight data. The green colour representing 'Obesity' has a count of 25, indicating a small portion of the data is obese. Finally, the orange colour representing 'Underweight' has the lowest number of 13, indicating the smallest proportion of underweight data.

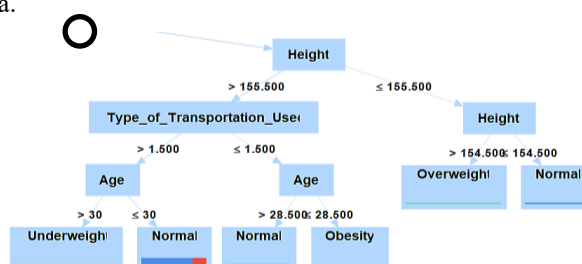


Figure 8. Visualisation of Decision Tree for Obesity Classification with Gain Ratio Parameters

Figure 8 is a decision tree used to classify a person based on several attributes. Starting from the top, with the question ‘Height less than or equal to 1500?’, if yes, then the person is categorised as ‘Normal’. If no, then the next question is ‘Height greater than 155,500?’. If yes, then the next question is ‘Type of Transport Used is more than 1,500?’. If yes, then the next question is ‘Age over 30?’. If yes, then the person is categorised as ‘Underweight’, and if no, then categorised as ‘Normal’. If in the question ‘Type of Transport Used’ the answer is no, then the next question is ‘Age over 28,500?’. If yes, then the person is categorised as ‘Normal’, and if no, then categorised as ‘Obese’. Going back to the question ‘Height greater than 155,500?’, if the answer is no, then the next question is ‘Height greater than 154,500?’. If yes, then the person is categorised as ‘Overweight’, and if no, then categorised as ‘Normal’.

accuracy: 72.78%

	true Normal	true Overweight	true Obesity	true Underweight	class precision
pred. Normal	85	13	6	3	79.44%
pred. Overweight	5	23	7	0	65.71%
pred. Obesity	0	11	14	0	56.00%
pred. Underweight	4	0	0	9	69.23%
class recall	90.43%	48.94%	51.85%	75.00%	

Figure 9. Evaluation of Classification Model with Gain Ratio Parameters Using Confusion Matrix

The confusion matrix in Figure 9 in the classification of high risk of obesity with Decision Tree (Gain Ratio) at the top shows the number of correct and incorrect predictions for each category. For example, out of 107 actual data labelled ‘Normal’, the model correctly predicted 85 as ‘Normal’, but incorrectly predicted 13 as ‘Overweight’, 6 as ‘Obesity’, and 3 as ‘Underweight’. The overall accuracy of the model was 72.78%. In addition, the table below presents more detailed evaluation metrics. ‘Class Precision’ measures the accuracy of positive predictions for each class. For example, if the model predicts that a person is ‘Normal’, there is a 79.44% chance that the person is indeed ‘Normal’. ‘Class Recall’ measures the ability of the model to find all positive data from a class. For example, the model managed to find 90.43% of all people who were actually ‘Normal’. These metrics provide a more comprehensive picture of how the model performs in classifying each weight category.

3.2 Decision Tree (Gini Index)

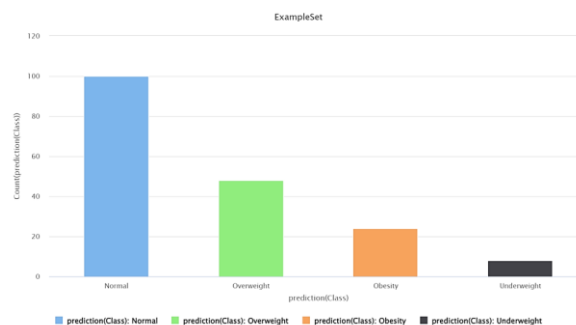


Figure 10. Class Prediction Distribution with Gini Index Parameter

Figure 10 presents the data in the form of a bar chart showing the class predictions with quantitative details. The blue colour representing ‘Normal’ has the highest number of 100, indicating that the majority of the data is normal. It is followed by the green colour representing ‘Overweight’ with a count of 48, indicating a significant amount of overweight data. The orange colour representing ‘Obesity’ has a count of 24, indicating a small portion of the data is obese. Finally, the black colour representing ‘Underweight’ has the lowest number of 8, indicating the smallest proportion of underweight data.

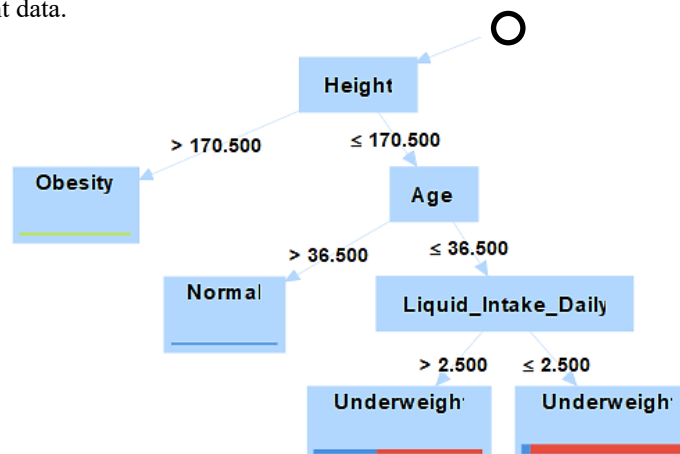


Figure 11. Visualisation of Decision Tree for Obesity Classification with Gini Index Parameters

Based on Figure 11 above, this decision tree works by dividing the data based on the values of these attributes in each branch. Starting from the root of the tree, which is Height, if someone has a height of more than 170.5 cm, then they are categorised as Obese. If not, the Age attribute will be checked. If the age is more than 36.5 years, then it is categorised as Normal. Otherwise, the last attribute checked is Daily Fluid Intake. If the daily fluid intake is more than 2.5, then it is categorised as Underweight, and if it is less than or equal to 2.5, it is also categorised as Underweight.

accuracy: 77.22%

	true Normal	true Overweight	true Obesity	true Underweight	class precision
pred. Normal	83	11	2	4	83.00%
pred. Overweight	11	30	7	0	62.50%
pred. Obesity	0	6	18	0	75.00%
pred. Underweight	0	0	0	8	100.00%
class recall	88.30%	63.83%	66.67%	66.67%	

Figure 12. Evaluation of Classification Model with Gini index Parameters Using Confusion Matrix

In the classification of high risk of obesity with Decision Tree (Gini Index) in Figure 12 above, the model achieved an overall accuracy of 77.22%. The evaluation is presented in a table that includes precision and recall metrics for each category. For the Normal category, the model showed a precision of 83% and a recall of 88.30%, meaning that the model was good at identifying individuals who were truly Normal and did not misclassify individuals from other categories as Normal. The Overweight category has a precision of 62.50% and recall of 63.83%, indicating that the model has challenges in distinguishing between Overweight and other categories. The Obese category has a precision of 75% and recall of 66.67%, showing better performance than Overweight, but still needs improvement. Finally, the Underweight category showed perfect performance with 100% precision and recall, meaning the model was very accurate in identifying individuals who were Underweight. Overall, the model performed well in classifying medical conditions, but needs further improvement, especially in distinguishing between Overweight and Obesity.

4. CONCLUSIONS

This study aims to classify the high risk of obesity in 898 female data using 14 judgements. The performance of the Decision Tree method was compared using three criteria: Information Gain, Gain Ratio, and Gini Index. The results showed that Decision Tree with Information Gain achieved an accuracy of 80.56%, but had the best performance in classifying Normal and Underweight patients, but less good in classifying Overweight and Obese patients. Decision Tree with Gain Ratio achieved 72.78% accuracy and Decision Tree with Gini Index achieved 77.22% accuracy. Overall, the Decision Tree model shows good performance in classifying medical conditions, but needs further improvement, especially in distinguishing between Overweight and Obese.

REFERENCES

- [1] P. Kecamatan, K. Muka, K. Depok, dan J. Barat, "Gizi indonesia," vol. 47, no. 2, hal. 195–208, 2024, doi: 10.36457/gizindo.v47i2.1066.
- [2] I. D. Mienye dan N. Jere, "A Survey of Decision Trees : Concepts , Algorithms , and Applications," *IEEE Access*, vol. 12, no. May, hal. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
- [3] M. Arenas, P. Barceló, M. Romero Orth, dan B. Subercaseaux, "On computing probabilistic explanations for decision trees," *Adv. Neural Inf. Process. Syst.*, vol. 35, hal. 28695–28707, 2022.
- [4] E. Demirović *et al.*, "Murtree: Optimal decision trees via dynamic programming and search," *J. Mach. Learn. Res.*, vol. 23, no. 26, hal. 1–47, 2022.
- [5] P. A. Prayesy dan I. Ruswita, "G-Tech : Jurnal Teknologi Terapan," vol. 7, no. 1, hal. 21–28, 2023.
- [6] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, hal. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
- [7] E. T. Lestari dan J. Adhiva, "Implementation Naive Bayes Classifier Algorithm and K-Nearest Neighbor For Obesity Nutritional Status of Children with Disabilities Implementasi Algoritma Naive Bayes Classifier dan K-Nearest Neighbor Untuk Klasifikasi Status Gizi Obesitas Anak Disabilitas," hal. 1–11, 2022.
- [8] C. E. Sukmawati, A. Fitri, N. Masruriyah, dan A. R. Juwita, "Efektivitas algoritma AdaBoost dan XGBoost pada dataset obesitas populasi dewasa," vol. 6, no. 2, hal. 101–111, 2024, doi: 10.37905/jji.
- [9] S. Uni, J. N. Sci, dan A. Sulak, "Sinop Üniversitesi Fen Bilimleri Dergisi Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities," vol. 9, no. 1, hal. 217–239, 2024.
- [10] V. G. Costa dan C. E. Pedreira, "Recent advances in decision trees: An updated survey," *Artif. Intell. Rev.*, vol. 56, no. 5, hal. 4765–4800, 2023.
- [11] M. Bansal, A. Goyal, dan A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, hal. 100071, 2022.
- [12] D. Muriyatmoko, M. Aziz, dan M. H. Wijaya, "Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree C4.5," *Pros. Semnastek*, 2024.
- [13] G. Kanugrahan, V. H. C. Putra, dan Y. Ramdhani, "Analisis Sentimen Aplikasi Gojek Menggunakan SVM, Random Forest



- dan Decision Tree,” *J. Infortech*, vol. 6, no. 2, hal. 171–178, 2024.
- [14] W. Bagaskara, N. N. Pusparini, dan I. Irwansyah, “Klasifikasi Penjadwalan Kerja Perawatan Air Conditioner (Ac) Menggunakan Algoritma Decision Tree (C4. 5) Pada PT Xyz,” *Infotech J. Technol. Inf.*, vol. 10, no. 1, hal. 11–20, 2024.
- [15] M. R. Hidayatullah dan Warih Maharani, “Depression Detection on Twitter Social Media Using Decision Tree,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, hal. 677–683, Agu 2022, doi: 10.29207/resti.v6i4.4275.
- [16] O. M. Mustafa, O. M. Ahmed, dan V. A. Saeed, “Comparative analysis of decision tree algorithms using gini and entropy criteria on the forest covertypes dataset,” in *The International Conference on Innovations in Computing Research*, 2024, hal. 185–193.
- [17] S. Abolhosseini, M. Khorashadizadeh, M. Chahkandi, dan M. Gotalizadeh, “A modified ID3 decision tree algorithm based on cumulative residual entropy,” *Expert Syst. Appl.*, vol. 255, hal. 124821, 2024.
- [18] K. Gao, Y. Wang, dan L. Ma, “Belief entropy tree and random forest: Learning from data with continuous attributes and evidential labels,” *Entropy*, vol. 24, no. 5, hal. 605, 2022.
- [19] Y. Lu, T. Ye, dan J. Zheng, “Decision tree algorithm in machine learning,” in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, 2022, hal. 1014–1017.
- [20] U. Ramasamy dan S. Sundar, “An Illustration of Rheumatoid Arthritis Disease Using Decision Tree Algorithm,” *Informatica*, vol. 46, no. 1, 2022.
- [21] H. Åkesson dan H. Fridlund, “A Comparative Study on the Effects of Removing the Most Important Feature on Random Forest and Support Vector Machine.” 2023.
- [22] O. Rahmati, M. Avand, P. Yariyan, J. P. Tiefenbacher, A. Azareh, dan D. T. Bui, “Assessment of Gini-, entropy- and ratio-based classification trees for groundwater potential modelling and prediction,” *Geocarto Int.*, vol. 37, no. 12, hal. 3397–3415, 2022.
- [23] Abdussalam Sulaiman Olainiyi, Saheed Yakub Kayode, Hambali Moshood Abiola, Salau-Ibrahim Taofeekat Tosin, dan Akinbowale Nathaniel Babatunde, “Student’s Performance Analysis Using Decision Tree Algorithms,” *Ann. Comput. Sci. Inf. Syst.*, 2017.
- [24] C. Intelligence and Neuroscience, “Retracted: Psychological Analysis for Depression Detection from Social Networking Sites,” *Comput. Intell. Neurosci.*, vol. 2023, no. 1, Jan 2023, doi: 10.1155/2023/9796187.
- [25] J. G. Choi, K. O. Inhwon, dan S. Han, “Depression Level Classification Using Machine Learning Classifiers Based on Actigraphy Data,” *IEEE Access*, vol. 9, hal. 116622–116646, 2021, doi: 10.1109/ACCESS.2021.3105393.
- [26] D. Kurniasari, R. Nurul Hidayah, dan R. Khoirun Nisa, “CLASSIFICATION MODELS FOR ACADEMIC PERFORMANCE: A COMPARATIVE STUDY OF NAÏVE BAYES AND RANDOM FOREST ALGORITHMS IN ANALYZING UNIVERSITY OF LAMPUNG STUDENT GRADES,” *J. Tek. Inform.*, vol. 5, no. 5, hal. 1267–1276, 2024, doi: 10.52436/1.jutif.2024.5.5.2066.
- [27] D. Valero-Carreras, J. Alcaraz, dan M. Landete, “Comparing two SVM models through different metrics based on the confusion matrix,” *Comput. Oper. Res.*, vol. 152, Apr 2023, doi: 10.1016/j.cor.2022.106131.