Journal of Digital Health Innovation and Medical Technology
Vol 1 No 2 Februari 2025
ISSN: XXXX-XXXX (Electronic)
Open Access: https://journals.ai-mrc.com/jdhimt

Evaluating the Accuracy of Obesity Classification with Tree-Based Models: Decision Trees, Random Forests, and Gradient Boosting

Viviana Herlita Vidiasari [1], B. Nadila Nuzululnisa [2], Neny Sulistianingsih[3]
[1]Universitas Bumigora, Indonesia
[2]Universitas Bumigora, Indonesia
[3]Universitas Bumigora, Indonesia
[1]24010820004@universitasbumigora.ac.id *, [2]24010820006@universitasbumigora.ac.id, [3]neny.sulistianingsih@universitasbumigora.ac.id

Article Info

Article history:
Received 18-01-2025
Revised 07-02-2025
Accepted 12-02-2025

Keyword:
Obesity; Classification; Tree Based Model; Accurancy; Data Mining.

ABSTRACT

Obesity is a condition of excess weight that can have a negative impact on health. This condition increases the risk of various diseases, such as heart disease, type 2 diabetes, hypertension, and other metabolic disorders. According to the World Health Organization (WHO) report, in 2022 there will be 2.5 billion adults aged 18 years and over who are overweight, including more than 890 million people who are obese. Seeing this problem, this study aims to develop an obesity classification model based on age, gender, body mass index (BMI), physical activity, and obesity category attributes. The models used in this study are decision tree, random forest, and gradient boosting, which are included in tree-based methods. The research stages include data collection, dataset processing, model building, and performance evaluation. In the final stage, the best method was selected from the three models used. The results showed that decision tree has an accuracy of 99.5%, random forest 99.7%, and gradient boosting 99.8%, making it the method with the best accuracy in obesity classification. With these results, gradient boosting can be used as a tool in health decision making, especially in detecting and categorizing individuals at risk of obesity more accurately. In addition, this model can help in developing more effective prevention and intervention strategies in dealing with obesity.

INTRODUCTION

Health problems are one of the main issues that often arise in society. One of them is obesity. Obesity is caused by the accumulation of fat due to the consumption of foods high in sugar and fat without burning enough energy. This condition has a negative impact on health and increases the risk of various diseases (Azahra, 2024). Based on epidemiological studies, obesity is caused by several factors, such as an unbalanced diet, lack of physical activity, and genetic influences that also increase the risk (Sukmawati et al., 2024). In addition, obesity is a condition of being overweight that can have a negative impact on health, as measured by the Body Mass Index (BMI). Obesity is influenced by various factors, such as gender, level of knowledge, education, type of work, marital status, family history, physical activity, smoking habits, and calorie intake. Obesity not only occurs in adults and adolescents, but can also occur in children (Setiyani et al., 2023).

According to the World Health Organization (WHO) report, in 2022 there will be 2.5 billion adults aged 18 years and over who are overweight, including more than 890 million who are obese. This figure covers 43% of the adult population (43% men and 44% women), a significant increase compared to 1990, where only 25% of adults were overweight (Nadira & Farida Utami, 2024).

Obesity poses various serious health risks, such as heart disease, type 2 diabetes, hypertension, and other metabolic disorders. In addition, obesity also has an impact on mental health

`
and reduces the overall quality of life (Yamantri & Ahmad, 2024).. By looking at these problems, this study aims to develop an obesity classification model such as age, gender, BMI, physical, and obesity category using data mining, including using a tree-based model with a decision tree, random forest, and gradient boosting approach. In addition, this study aims to analyze the performance of each algorithm in predicting obesity based on accuracy, precision, recall, and F1 score and provide recommendations for the best method for obesity classification that can be applied in decision making in the health sector and further research.

Decision tree is a tree-based structure used to predict the numerical outcome of a dependent variable through a learning process, where new data is grouped based on existing training samples. Random forest, as an extension of the bagging method, aims to reduce the variance of a statistical model by randomly sampling boothstraps from the training data and combining predictions for new data. This algorithm is one of the most widely used tree-based ensemble classification methods. Gradient Boosting is a decision tree-based supervised learning technique that starts by building an initial classification tree, then adjusting subsequent trees by minimizing the loss function (Rahardika et al., 2024).

Some related studies, namely (Alifah et al., 2024) the difference from previous studies is in terms of novelty, this study classifies obesity using attributes such as age, gender, BMI, physical activity, and obesity category. While the previous study only used the attributes of age, height, weight, and BMI. (Rahardika et al., 2024) conducted a study by applying the decision tree method which showed an accuracy of 82% previous research indicated that the decision tree algorithm has great potential in supporting early detection of obesity. (Santoso et al., 2024) the use of the random forest model can also provide more precise recommendations in efforts to prevent and treat obesity with an accuracy of 95%. (Setiyani et al., 2023) in another study also stated that gradient boosting is an algorithm that achieves the highest AUC, precision, and recall with a value of 86.6%. (Arta et al., 2024) explained that the random forest algorithm showed the highest accuracy of 71.09% compared to other algorithms with a random forest sensitivity of 37.50%.

With the description of the problem, this study provides benefits in the health sector and can increase public awareness to maintain health, especially obesity which is a factor in the occurrence of other diseases. In addition, it can be a reference for further research to develop models and can increase the accuracy of the available models (Hardwis et al., 2024)

RESEARCH METHODS

The following are the research stages used before conducting obesity classification modeling. There are 6 steps starting from identifying the type of research, data collection, data loading, data evaluation stage.

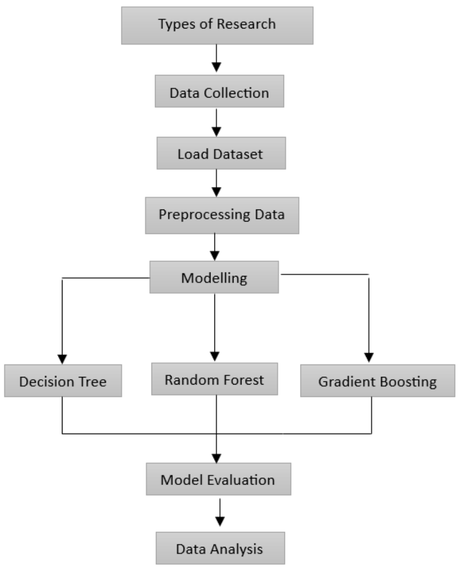

Figure 1. Research Flowchart

Type of Research. This study uses a quantitative approach with an experimental method, which aims to apply and compare tree-based models in classifying obesity. The models tested in this study are decision trees, random forests, and gradient boosting.

Data Collection. The data collection stage is the first step in the data analysis and modeling process, where the data to be used in the study is collected (Nida Winarti et al., 2022). The data used in this study is a dataset from the Kaggle.com site which contains information related to factors related to obesity consisting of 7 attributes. These attributes include age, gender, height, weight, BMI (body mass index), physical activity level, obesity category (obesity category: normal, overweight, obese). Taking data from trusted sources such as Kaggle is very important to ensure that the dataset used has adequate diversity and representation, so that the trained model can accurately identify obesity classifications.

Load Dataset. Once the data is obtained, the next step is to load the dataset into the programming process so that it can be accessed and processed further. Obesity datasets taken from Kaggle are usually loaded into DataFrame format using the pandas library in Python. DataFrame allows data to be arranged in a table format, making it easier to explore, filter, and manipulate. Loading this dataset is an important step to ensure that the data is available in a format that can be understood and used by machine learning algorithms. With DataFrame, we can perform initial analysis such as displaying the first few rows, arranging data in a table format, and checking the completeness of the data. This step is an important part of the preparation before proceeding to the pre-processing and data modeling stages.

Data Preprocessing. The data preprocessing stage is one of the important steps in the machine learning pipeline that functions to clean and prepare the data so that it is ready to be used by the model. This process includes a series of steps to ensure that the data entered into the model is of high quality, free from inconsistencies, and arranged in a format that is easy for the algorithm to understand. The first step is to remove the height and weight attributes because both of these data have been included in the BMI data. BMI is the Body Mass Index obtained from the results of dividing the weight and height values (Alifah et al., 2024). Furthermore, after the attribute selection process, the missing value process is carried out, namely checking for empty data, for example using the imputation method, namely replacing it with the average value for numeric attributes and mode for categorical attributes. (Sudrajat & Cholid, 2023). Furthermore, the categorical variable encoding process, namely categorical features such as gender and obesity categories, will be converted into numeric variables using the label encoder method to facilitate processing by the machine learning model. (Herdian et al., 2024). Then the last stage is the division of training data and testing data. In this study, 80% training percentage was used to train the model and 20% to test the model performance. The 80:20 ratio is quite commonly used in machine learning because it provides a balance between enough data to train the model and enough data to test its performance.

Modeling. After the data has gone through the preprocessing stage, the next step is to build a model for classification using three different algorithms, namely decision tree, random forest, and gradient boosting. These three models were chosen because this study wanted to test the three tree-based models that have their respective advantages. At this stage, the pre-processed dataset is used to train third models, which are then evaluated based on their respective performances. To compare performance using accuracy metrics, where the model with the highest accuracy will be the best model. This step aims to ensure that the classification results are reliable and useful in real application scenarios. In the accuracy metric, several final results will be displayed, namely the level of accuracy, precision, recall, F1 score and support.

Decision tree is a model that studies the classification and prediction of patterns from data and describes the relationship between attribute variables x and target variables y in the form of a tree (Nasrullah, 2021). Random forest is an ensemble algorithm that uses many decision trees to improve accuracy and reduce variance. (Rifai et al., 2024). Gradient boosting is a boosting algorithm that combines decision trees iteratively to minimize prediction errors (Sabili & Umbara, 2024)

Model evaluation is the final stage in the dataset classification process. In this process, an evaluation of the performance of the three algorithms used is carried out. To see the accuracy results, several performance matrices can be used, namely accuracy, which is the proportion of correct predictions from the total predictions (Dayera, Musa Bundaris Palungan, 2024), precision, which is the ratio between the number of correct positive predictions and the total number of positive predictions. (Putri et al., 2023), recall, which is the ratio between the number of correct positive predictions and the total number of actual positive cases. (Al-afghoni et al., 2025). F1 score is the harmonic mean between precision and recall which provides a balance between the two. Support is the number of actual data (instances) for each class in the dataset being evaluated. (Kartika Sari et al., 2024)

Data analysis is to analyze the evaluation results that have been obtained. Data analysis needs to be done to provide a detailed description of the results of the dataset classification. Some important points that can be done in data analysis include comparing the performance between decision trees, random forests, and gradient boosting, concluding the model that gives the best results based on the accuracy, precision, recall, F1 score and support matrices. And analyzing the factors that influence obesity classification based on the attributes used in the model

RESULTS AND DISCUSSION

After going through the research methods stages above, the results and discussions are obtained which will be explained below.

Data Collection. The data collection stage is carried out through the kaggle.com site, where data is obtained with a total of 7 features and a total of 1000 data. Table 1 shows several attributes contained in the dataset with different types, namely int (integer), cat (categorical) and float (decimal).

Table 1. Obesity Data Attributes

No.	Attribute	Type	Description
1	Age	Int	Respondent's age in years
2	Gender	Cat	Gender (male/female)
3	Height	Float	Respondent's height in meters
4	Weight	Float	Respondent's weight in kilograms
5	BMI	Float	Body mass index, calculated from weight and height
6	Physical activity level	Cat	Physical activity level (range 1 - 4)
7	Obesity category	Cat	Obesity category based on BMI (Underweight, Normal, Overweight, Obese)

Load dataset. The dataset that has been obtained is then imported to Google Drive to be further processed using Google Colab. The data consists of various attributes and types of data used to support the obesity classification process. Table 2 will display some of the top 5 data to ensure that the imported data matches the data to be processed. By loading this dataset, you can start exploring the data to understand the obesity classification pattern and then as preparation for the next analysis stage.

Table 2. Dataset Load Process

No.	Age	Gender	Height	Weight	BMI	Physical Activity Level	Obesity Category
0	56	Male	173.575262	71.982051	23.8971783	4	Normal weight
1	69	Male	164.127306	89.959256	33.395209	2	Obese
2	46	Female	168.072202	72.930629	25.817737	4	Overweight
3	32	Male	168.459633	84.886912	29.912247	3	Overweight
4	60	Male	183.56568	69.038942	20.487903	3	Normal weight

Data Preprocessing. The first step is to determine what attributes will be used in processing this data. For height and weight data, it will be deleted because the two data have been included in the BMI data. There are 5 attributes used. The following shows the distribution results of the data used

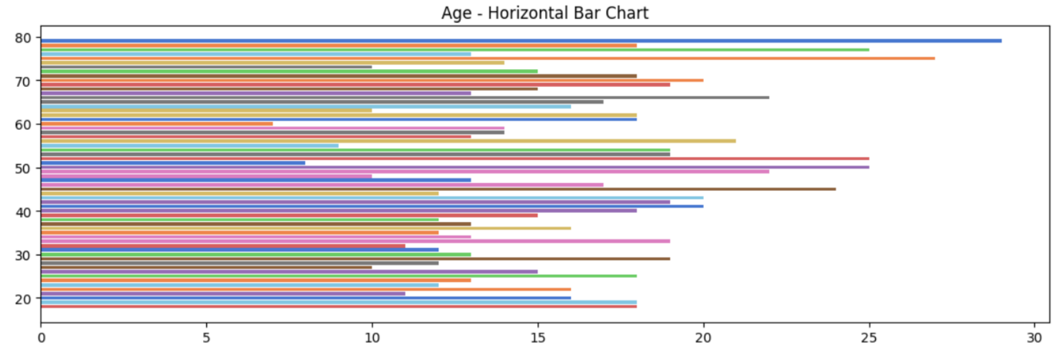

Figure 2. Data Age

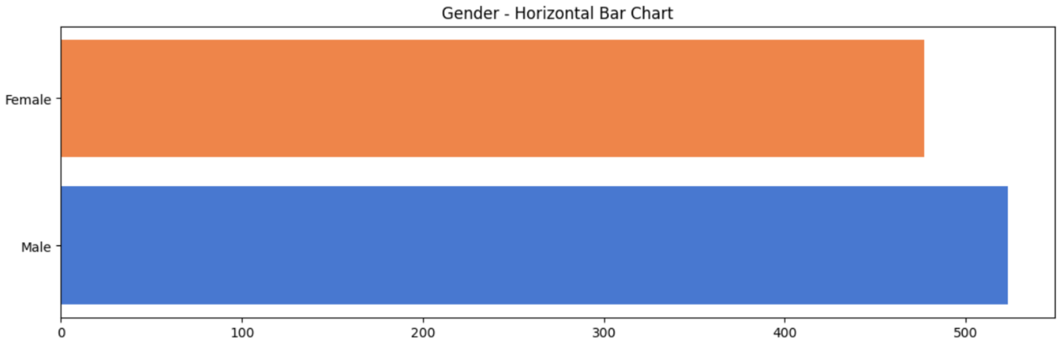

Figure 3. Data Gender

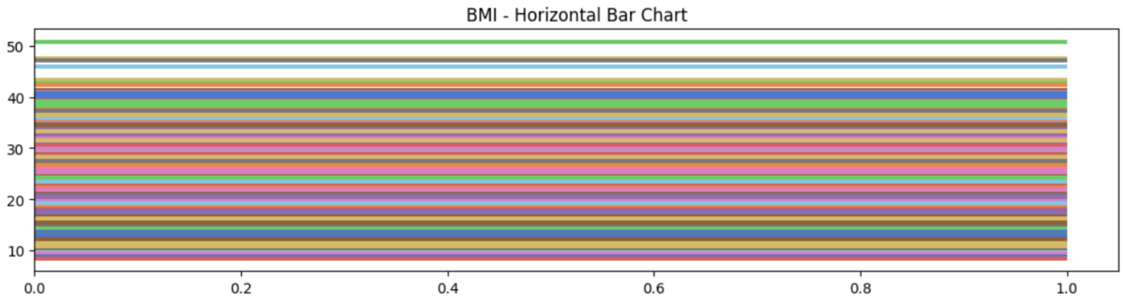

Figure 4. Data BMI

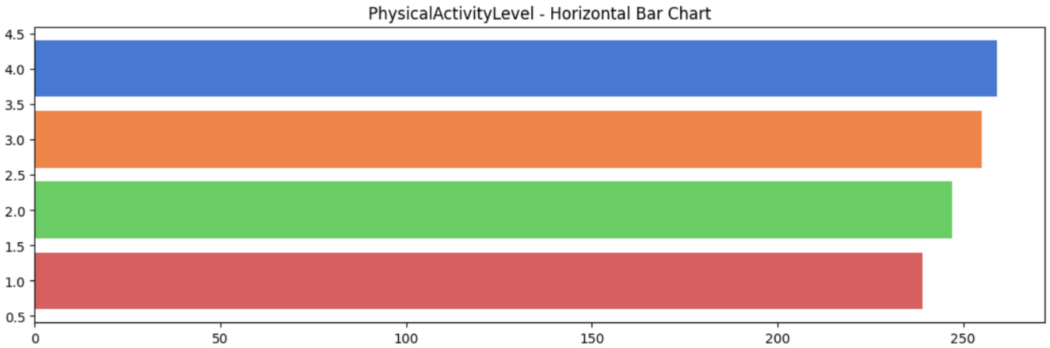

Figure 5. Data Physical Activity Level

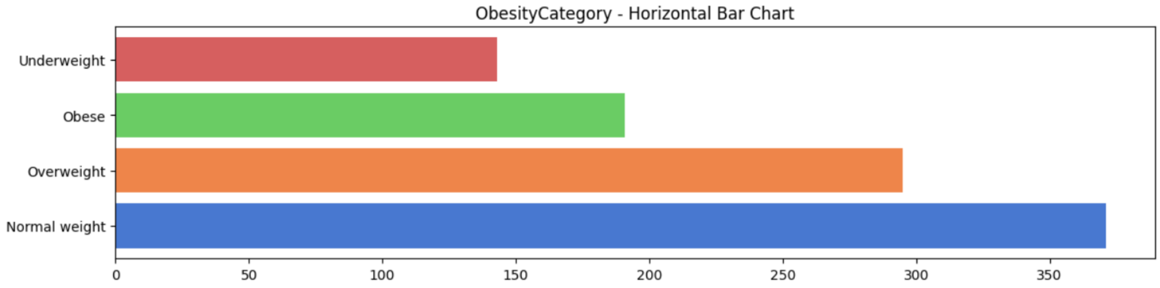

Figure 6. Data Obesity Category

After deleting the height and weight data, distribute the data in the form of a horizontal bar chat. Then, check the missing value to check for empty data. Based on the results of the check, all attributes produce a value of 0 which indicates that there is no empty data.

Table 3. Missing Value

Attribute	Missing Value
Age	0
Gender	0
BMI	0
Physical Activity Level	0
Obesity Category	0

The data has been confirmed to be empty, then the variable encoding process by changing categorical data into numeric, namely in gender and obesity category data using the label encoder method. Both data have been changed into numbers according to the previous categorical description.

Table 4. Encoding Variabel

No.	Age	Gender	BMI	Physical Activity Level	Obesity Category
0	56	1	23.8971783	4	0
1	69	1	33.395209	2	1
2	46	0	25.817737	4	2
3	32	1	29.912247	3	2
4	60	1	20.487903	3	0

The last stage is the division of datasets for training data and testing data. The amount of data is 1000 data, so the percentage of 80% training data is 800 data and the percentage of 20% testing data is 200 data.

Modeling. The next process is to test the model using a tree-based model, namely decision tree, random forest and gradient boosting to find out the best model for classifying obesity. In the next process, namely model evaluation, an evaluation matrix will be displayed as a result of the performance of each model in processing data for obesity classification.

Evaluation. In this process, the results of 3 classification models will be displayed which can be seen in tables 5, 6 and 7.

`

Table 5. Decision Tree Model Results

	Precision	Recall	F1-Score	Support
0	0.998	0.997	0.998	100
1	0.997	0.996	0.997	120
2	0.996	0.997	0.997	110
3	0.998	0.998	0.998	90
Accurancy			0.997	420
Macro Avg	0.997	0.997	0.997	420
Weighted Avg	0.997	0.997	0.997	420

In the decision tree model, namely table 5 in class 0, with precision is 0.998, which means 99.8% of positive predictions for this class are correct. Meanwhile, the recall for class 0 reaches 0.997, which indicates that 99.7% of all class 0 data is correctly identified by the model. The F1-Score for class 0 is 0.998, which shows the right balance between precision and recall. However, for precision, recall and F1-Score class 1 are slightly lower, namely 0.997 for precision, namely 99.7%, 0.996 for recall, namely 99.6% and F1-Score 0.997, namely 99.7%. However, although it decreased, the decrease was only 0.001. For class 2, the precision value is lower than class 2, namely 0.996 indicating 99.6%, but recall has increased, namely 0.997 where 99.7% of the values obtained. For class 3 all values increased, namely the precision value of 0.998 which indicates 99.8%. The recall value of 0.998 which indicates 99.8% and the F1-Score value of 0.998 which indicates 99.8%

Overall, the accuracy of the model is 99.7%, which means the model successfully classifies 99.7% of the data correctly. Macro Average shows the average of Precision, Recall, and F1-Score without considering the class distribution, which are 0.997, 0.997, and 0.997 respectively. Meanwhile, Weighted Average takes into account the class distribution and gives a value of 0.997 for Precision, 0.997 for Recall, and 0.997 for F1-Score, indicating that the model performs well despite the difference in the amount of data between class 0 to class 3. Overall, the decision tree model shows very good performance in recognizing and classifying data from all four classes as a whole.

Table 6. Random Forest Model Results

	Precisio n	Recall	F1-Score	Support
0	0.996	0.994	0.995	95
1	0.995	0.996	0.995	110
2	0.994	0.995	0.994	105
3	0.997	0.996	0.996	90
Accurancy			0.995	400
Macro Avg	0.996	0.995	0.995	400
Weighted Avg	0.995	0.995	0.995	400

Table 6 in the performance evaluation results of the random forest model displays the results of the matrix, namely precision, recall, F1-Score, and support. The random forest model shows good performance in four classes (class 0 and class 3). For class 0, the precision is 0.996, which means that 99.6% of positive predictions for this class are correct. Meanwhile, the recall for class 0 reaches 0.994, which shows that 99.4% of all class 0 data are correctly identified by the model. The F1-Score for class 0 is 0.995, which shows a good balance between precision and recall. However, for class 1, the precision is slightly lower, which is 0.995, which is 99.5% of positive predictions for class 1 are correct. The recall for class 1 shows a value of 0.996, which means that 99.6% of all class 1 data are correctly identified. The F1-Score for class 1 is 0.995, reflecting a good balance even though the precision is slightly lower. For class 2, the precision, recall and F1-Score values are lower than class 1, with values of 0.994%, which is 99.4, 0.995, which is 99.5% and 0.994, which is 994% respectively. However, although all the values are lower, the difference is not too significant. For class 3, all the values increase, namely the precision value of 0.997 which indicates 99.7%. The recall value is 0.996 which indicates 99.6% and the F1-Score value is 0.996 which indicates 99.6%

`

Overall, the accuracy of the model is 99.5%, which means the model successfully classifies 99.5% of the data correctly. Macro Average shows the average of Precision, Recall, and F1-Score without considering the class distribution, which are 0.996, 0.995, and 0.995, respectively. Meanwhile, Weighted Average takes into account the class distribution and gives a value of 0.995 for Precision, 0.995 for Recall, and 0.995 for F1-Score, indicating that the model performs well despite the difference in the amount of data between class 0 to class 3. Thus, the model shows very good performance in recognizing and classifying data from all four classes overall.

Table 7. Gradient Boosting Model Results

	Precision	Recall	F1-Score	Support
0	0.999	0.998	0.998	100
1	0.998	0.999	0.999	120
2	0.997	0.998	0.998	110
3	0.998	0.999	0.998	90
Accurancy			0.998	420
Macro Avg	0.998	0.998	0.998	420
Weighted Avg	0.998	0.998	0.998	420

In the last model, namely gradient boosting, table 7 shows the results in class 0, with a precision of 0.999, which means that 99.9% of positive predictions for this class are correct. Meanwhile, the recall for class 0 reached 0.998, which indicates that 99.8% of all class 0 data were correctly identified by the model. The F1-Score for class 0 is 0.998, which shows a suitable balance between precision and recall. However, for class 1, the precision is slightly lower, namely 0.998, namely 99.8% of positive predictions for the class are correct, although the precision value has decreased, the recall value has increased, namely 0.999 with a value of 99.9%. So that the data is still very good. For class 2, the precision, recall and F1-Score values have decreased but do not have a big impact because the results are still relatively high. For class 3, all values increase, namely a precision value of 0.998 which indicates 99.8%. The recall value is 0.999 which indicates 99.9% and the F1-Score value is 0.998 which indicates 99.8%.

Overall, the accuracy of the model is 99.8%, which means the model successfully classifies 99.8% of the data correctly. Macro Average shows the average of Precision, Recall, and F1-Score without considering the class distribution, which are 0.998, 0.998, and 0.998 respectively. Meanwhile, Weighted Average takes into account the class distribution and gives a value of 0.998 for precision, 0.998 for recall, and 0.998 for F1-Score, indicating that the model performs well despite the difference in the amount of data between class 0 to class 3. Overall, the gradient boosting model shows the best performance among the two models.

Gradient boosting excels in obesity classification because it is able to gradually correct errors, capture non-linear patterns, and has better generalization ability than decision trees and random forests. However, major challenges such as high computational cost, risk of overfitting, and interpretability need to be addressed. For future research, the use of optimization techniques such as LightGBM, XGBoost, early stopping, and model interpretability can improve efficiency while maintaining the accuracy advantage of gradient boosting.

CONCLUSION

Based on the research conducted, it was found that decision trees, random forests, and gradient boosting each have advantages in classifying obesity. Among the three models, gradient boosting showed the best performance with an accuracy of 99.8%, compared to decision trees and random forests. This study concludes that models based on gradient boosting, decision trees, and random forests can be effective tools in detecting and classifying obesity. With high accuracy, this model can be applied in decision support systems in the health sector, for example in early obesity screening, intervention recommendations, or patient monitoring in health facilities.Can add additional features, namely dietary data, calorie intake and health history and genetic data to improve accuracy in future research

`

CONFLICT OF INTEREST

That there is no conflict of interest.

REFERENCES

Al-afghoni, J. M. H. Y., Setiawan, W., Dwi, Y., Negara, P., Informasi, S., Madura, U. T., Matrix, C., Tree, D., & Rusdiana, L. (2025). KLASIFIKASI JENIS BENIH KACANG MENGGUNAKAN SMOTE DAN DECISION TREE C4. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(1), 462–469. https://doi.org/https://doi.org/10.36040/jati.v9i1.12366

Alifah, R. N., Najib, M. K., Nurdiati, S., Sari, A. P., Herlambang, K., Putri, T., Ginting, B., & Sya'adah, S. N. (2024). Perbandingan Metode Tree Based Classification untuk Masalah Klasifikasi Data Body Mass Index. *Indones. J. Math. Nat. Sci*, *47*(1), 2024. https://doi.org/https://doi.org/10.15294/m2k97436

Arta, M. C., Anwar, N., Putri, Y. A., Suharjito, S., & Asroll, M. (2024). Implementasi Prediksi Penyakit Jantung Menggunakan Data Mining Untuk Dunia Kesehatan. *Jurnal Optimalisasi*, *10*(1), 42. https://doi.org/10.35308/jopt.v10i1.9075

Azahra, R. A. (2024). Komparasi Metode Naïve Bayes dan Kolmogorov Arnold Networks dalam Klasifikasi Tingkat Kemungkinan Obesitas. *TEKNOLOGI NUSANTARA*, *6*(2), 77–84. http://ojs.uninus.ac.id/index.php/TEKNOLOGINUSANTARA/article/view/3334

Dayera, Musa Bundaris Palungan, F. O. (2024). Optimalisasi Akurasi Algoritma C4.5 dengan Metode Adaptive Boosting Memprediksi Siswa dalam Menerima Dana Pendidikan. *G-Tech : Jurnal Teknologi Terapan*, *8*(1), 186–195. https://doi.org/https://doi.org/10.70609/gtech.v8i4.5612

Hardwis, S., Studi, P., Keolahragaan, I., & Indonesia, U. P. (2024). Analisis Resiko Obesitas Berdasarkan Aktivitas Fisik : Implementasi Metode Artificial Intelligence Machine Learning. *Jurnal Keolahragaan*, *10*(November), 29–36. https://doi.org/10.25157/jkor.v10i2.16884

Herdian, C., Kamila, A., & Agung Musa Budidarma, I. G. (2024). Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi. *Technologia : Jurnal Ilmiah*, *15*(1), 93. https://doi.org/10.31602/tji.v15i1.13457

Kartika Sari, A., Akhmad Irsyad, Dinda Nur Aini, Islamiyah, & Stephanie Elfriede Ginting. (2024). Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, *3*(1), 64–73. https://doi.org/10.30872/atasi.v3i1.1373

Nadira, A., & Farida Utami, S. (2024). Implementasi Data Mining Dalam Mengidentifikasi Faktor Pasien Yang Berpotensi Mengalami Obesitas Menggunakan Algoritma C4.5. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(4), 7872–7876. https://doi.org/10.36040/jati.v8i4.10498

Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, *7*(2), 217. https://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/70

Nida Winarti, Maula, L. H., Amalia, A. R., Pratiwi, N. L. A., & Nandang. (2022). Penerapan Model Pembelajaran Project Based Learning Untuk Meningkatkan Kemampuan Berpikir Kritis Siswa Kelas Iii Sekolah Dasar. *Jurnal Cakrawala Pendas*, *8*(3), 552–563. https://doi.org/10.31949/jcp.v8i3.2419

Putri, A., Hardiana, C. S., Novfuja, E., Siregar, F. T. P., Rahmaddeni, R., Fatma, Y., & Wahyuni, R. (2023). Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(1), 20–26. https://doi.org/10.57152/malcom.v3i1.610

Rahardika, F., Putra, B., & Surahmanto, M. (2024). Implementation of Machine Learning Classification of Obesity Weight using Dicision Tree. *IJISTECH (International Journal of Information System and Technology)*, *8* (158), 110–116. https://ijistech.org/ijistech/index.php/ijistech/article/view/354/352

Rifai, A., Permana, S. E., & Hamonangan, R. (2024). Optimalisasi Klasifikasi Indeks Desa Membangun Menggunakan Metode Ensemble Dan Algoritma Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(4), 8226–8234. https://doi.org/10.36040/jati.v8i4.10971

Sabili, N. L., & Umbara, F. R. (2024). Klasifikasi Penyakit Diabetes Menggunakan Algoritma

Categorical Boosting Dengan Faktor Risiko Diabetes. *JATI(Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 11391–11398. https://doi.org/https://doi.org/10.36040/jati.v8i6.11447

Santoso, H. T., Felmidi, F. A., Nur, A., Ristyawan, A., & Daniati, E. (2024). Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, *8*, 113–122. https://doi.org/https://doi.org/10.29407/inotek.v8i1.4917

Setiyani, L., Indahsari, A. N., & Roestam, R. (2023). Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma Machine Learning dan Deep Learning. *JTERA (Jurnal Teknologi Rekayasa)*, *8*(1), 139. https://doi.org/10.31544/jtera.v8.i1.2022.139-146

Sukmawati, C. E., Fitri, A., Masruriyah, N., & Juwita, A. R. (2024). *Efektivitas algoritma AdaBoost dan XGBoost pada dataset obesitas populasi dewasa*. *6*(2), 101–111. https://doi.org/10.37905/jji.

Yamantri, A. B., & Ahmad, A. (2024). Penerapan Algoritma C4.5 Untuk Prediksi Faktor Risiko Obesitas Pada Penduduk Dewasa. *Jurnal Komputer Antartika*, *2*(c), 118–125. https://doi.org/https://doi.org/10.70052/jka.v2i3.341