

CART Classification on Ordinal Scale Data with Unbalanced Proportions using Ensemble Bagging Approach

Luthfia Hanun Yuli Arini¹, Solimun¹, Achmad Efendi¹, Mohammad Ohid Ullah²

¹Department of Statistics, Brawijaya University, Indonesia

²Department of Statistics, Shaljalal University of Science and Technology, Bangladesh

luthfiaeerini@student.ub.ac.id

ABSTRACT

Article History:

Received : 16-11-2023

Revised : 26-03-2024

Accepted : 27-03-2024

Online : 01-04-2024

Keywords:

Unbalanced;

CART;

Bagging.



CART is one of the algorithms in data exploration techniques with decision tree techniques. Unbalanced class proportions in the classification process can cause classification results of minor data to be incorrect. One way to overcome the problem of data imbalance is to use an ensemble bagging algorithm. The bagging algorithm utilizes the resampling method to carry out classification so that it can reduce bias in imbalanced data. The data used is secondary data from Fernandes and Solimun's 2023 research report. The number of sample are 100 respondents that has been valid and reliable. The sample for this research was mothers with toddlers in Wajak village, Malang Regency. The results showed that the ensemble bagging CART method is better at overcoming the problem of imbalance in the proportion of classes with a performance value of accuracy, sensitivity, specificity, and F1-Score values of 85%, 94.1%, 66.7%, and 78%. This research is limited to the Sumberputih Village area. So, the results of this research are only representative for the Wajak District area.



<https://doi.org/10.31764/jtam.v8i2.20201>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. INTRODUCTION

Classification and Regression Trees (CART) is one of the algorithms in data exploration techniques with decision tree techniques. CART was developed to classify nominal, ordinal, and continuous response variables. CART can also select the variables that are most important in determining the results (Siahaan et al., 2017). The main problem that often becomes a challenge in classification analysis is unbalanced class proportions. Unbalanced class proportions is a condition where there is an unbalanced proportion between classes in the data. Unbalanced class proportions can be defined as a condition in a data set where there is very small proportion of one or more categories of sample from the original sample (Li & Zhang, 2021). Unbalanced class proportions in the classification process can cause the classification results on minor data to be covered by the prediction of major data or in other words the classification results of minor data to be incorrect. Unbalanced class proportions can lead to biased classification models that may ignore the minority class altogether, resulting in poor predictive performance for the minority class (Fitriani et al., 2021; Kumar et al., 2021; Luque et al., 2019).

One way to overcome the problem of data imbalance is to use an ensemble algorithm. An ensemble approach is an algorithm that combines various predictions into one final prediction. A study published compared the performance of bagging and boosting ensemble classifiers for the classification of multispectral, hyper-spectral, and polarimetric synthetic aperture radar data. The study found that both bagging and boosting can improve the accuracy of classification, but the performance of the two methods depends on the specific dataset and the choice of classifier (Jafarzadeh et al., 2021). Another study that proposed two novel ensemble-learning-based fine-tuning approaches, boosting fine-tuning (BF) and Bagging and boosting fine-tuning (BBF), which can improve the performance of ensemble learning models (Zhao et al., 2023). One of the most commonly used ensemble methods is ensemble bagging. Bagging is an ensemble method for training data on a subset of random samples from the original dataset (Ngo et al., 2022). This subset is generated through a bootstrap resampling process that involves random sampling with returns from the original data. A subset of the resampling data will be analysed using CART to obtain boundaries for each decision which are then applied to the testing data to obtain classification results. So, ensemble bagging is a widely used technique in ensemble learning, and it can be used with various types of models. Its performance depends on the specific dataset and the choice of classifier.

One of the classification problems where class imbalance occurs is the classification of stunting data. The expected proportion of stunted toddlers is much smaller than the proportion of toddlers with normal or high height. Therefore, there will be a class imbalance in the data with stunting cases. Stunting is a condition of child growth failure due to malnutrition in the first thousand days of a child's life since being in the mother's womb. Stunting can have immediate and long-term impacts that result in decreased productivity, decreased intellectual ability, increased risk of infection and infectious diseases in adulthood, and even death. The government is trying to carry out a stunting prevention program by appointing areas to become pioneers of the stunting prevention acceleration program. One of the areas that became a pioneer of the accelerated stunting prevention program is Malang District. Bappeda Malang District designated 32 villages as priority villages for accelerating stunting prevention in 2021, one of which is Sumberputih Village, Wajak District. For example, a study that evaluated the performance of machine learning classifiers in predicting stunting among children under five in Zambia. The study found that the random forest classifier outperformed other classifiers in terms of accuracy, sensitivity, and specificity (Chilyabanyama et al., 2022).

This research develops previous research where in previous research, the categories used were binary categories, whereas in this research there were three ordinal scale categories. This research is applied to under-five nutritional status data on Height/Age assessment with three categories namely stunting, normal, and high. The CART method is suitable for classifying ordinal scale data because this method is a machine learning algorithm that can classify and find important variables that influence the response variable. So this method is more flexible because there are no assumptions that must be met. The class imbalance in the nutritional status data of toddlers occurs because the proportion of toddlers with stunting or tall status is much lower. This is good, but when classification is carried out it requires appropriate methods to avoid misclassification. Misclassification can cause errors in nutritional management of toddlers with special status such as stunting because stunted toddlers must receive appropriate

treatment to support their development. Therefore, this research developed the CART classification method for ordinal scale data with unbalanced proportions using an ensemble bagging approach in the case of nutritional status of toddlers in Sumberputih Village so that the best classification method for classifying nutritional status data for toddlers as unbalanced and imbalanced can be determined class.

B. METHODS

1. Classification and Regression Trees (CART)

CART is a machine learning that is used to perform classification analysis for categorical and continuous response variables. The results of CART itself depend on the scale of the response variable (Krzywinski & Altman, 2017). If the response variable is continuous, the resulting tree model is regression trees, while if the response variable is categorical, the resulting tree model is classification trees (Breiman et al., 1984). The purpose of CART itself is to get an accurate group of data to characterize a classifier (Bramer, 2016). There are three stages of the classification tree formation process with CART, namely:

a. Node Breaking

The selected variables and threshold values are selected based on criteria that maximize data cleaning or reduce impurity in each resulting group. The selected variables and threshold values are chosen based on criteria that maximize data cleaning or reduce impurity in each resulting group. In this study, the Gini Impurity Index criterion was used to decide which variables to use as separators. Gini Impurity is calculated by equation (1) (Daniya et al., 2020).

$$i(t) = 1 - \sum_{j=1}^n p^2(j|t) \quad (1)$$

where $i(t)$ is the gini impurity at node y and $p(j|t)$ is the probability of class j at node t .

b. Class Labelling

Class labelling is the process of identifying vertices to determine the dominant class of a vertex. Class labelling is done to find out the characteristics of each vertex formed. The largest class probability indicates that the class dominates the node. The calculation of the dominant class probability is presented in equation (2).

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2)$$

with $p(j_0|t)$ is probability of class j_0 at node t (dominant class probability); $p(j|t)$ is probability of class j at node t ; $N_j(t)$ is the number of observations of class j at node t ; and $N(t)$ is the number of observations at vertex t .

c. Pruning

Tree pruning is done to prevent large trees from forming. Large classification trees cause high complexity. After pruning is done, an optimal classification tree will be formed. The pruning method performed is Minimal Cost-Complexity Pruning in equation (3).

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \tag{3}$$

where the value of $R(T)$ is shown in equation (4).

$$R(T) = \sum_{t \in \tilde{T}} r(t)p(t) = \sum_{t \in \tilde{T}} R(t) \tag{4}$$

with $R_\alpha(T)$ is re-substitution of tree T at complexity α or cost-complexity pruning value; $R(T)$ is re-substitution estimate; α is complexity parameter; $|\tilde{T}|$ is the number of terminal or leaf nodes in tree T ; $r(t)$ is probability of making a wrong classification at node t ; and $p(t)$ is probability of node t .

2. Ensemble Bagging

Ensemble is a machine learning algorithm where several weak models are trained to solve a problem and combined to get better results (Cendani & Wibowo, 2022). Ensemble approach combines various predictions from each iteration into one final prediction (Siringoringo & Jaya, 2018). Ensemble techniques are able to provide predictions with very good accuracy (Efendi et al., 2020). The main idea of ensemble is to combine several sets of models that solve the same problem to get a more accurate model (Friedman et al., 2000).

Bagging is an ensemble used to improve classification stability. Bagging uses based-models by performing parallel and independent learning on each based-model which is then combined to obtain the best results. The bagging process is depicted in Figure 1 which is a redrawing of (Cendani & Wibowo, 2022). This method is used as a tool to improve stability and predictive power by reducing the variance of a predictor of classification and regression methods whose use is not limited to improving estimators. One of the ensemble bagging methods is bootstrap aggregating (bagging). Bagging works by combining models trained using randomly generated data using bootstrap resampling. Bagging can be used with various types of models, including decision trees, neural networks, and support vector machines (Altman & Krzywinski, 2017).

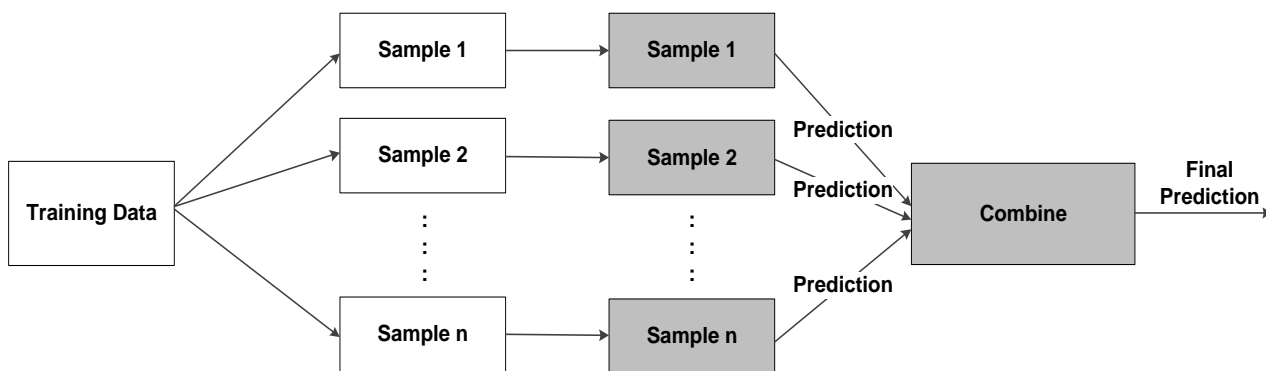


Figure 1. Bagging Ensemble Learning

Resampling is used as a tool to improve predictive consistency by reducing the variance of a predictor in classification. The basic idea of bagging with bootstrap resampling is to generate multiple versions of predictors which, when combined, produce better results for solving the same problem (Breiman, 1996). The idea behind ensemble bagging is to reduce the variance of

the predictive model by training several similar predictive models on different subsets of data obtained from resampling (De Prado, 2018; du Plooy & Venter, 2021). The main steps in the ensemble bagging algorithm are as follows.

a. Resampling with returns (bootstrap)

Bootstrap sampling of \mathcal{L}_B as many as n from dataset \mathcal{L} .

b. Independent model training

Each subset generated from the resampling process is used to train an independent predictive model. In this study, the bagging algorithm will be applied in logistic regression and CART analysis.

c. Aggregate prediction

After all models are trained, the final prediction is done by combining the prediction results of each model with majority voting for classification or average for regression where each model has the same weight. The majority voting used is argmax in equation (5).

$$\hat{y}_B = \text{argmax}_j f(\mathbf{x}, \mathcal{L}_B) \quad (5)$$

3. Performance

Classification performance is measured based on three criteria, including: accuracy, sensitivity, model specificity, and F1-Score (Erickson & Kitamura, 2021). Accuracy measures how correctly a diagnostic test identifies and excludes certain conditions (Chen et al., 2020). In other words, accuracy is used to measure the goodness of the model. In diagnostic tests, the terms sensitivity and specificity are also known. Sensitivity and specificity in diagnostic tests are measures of the ability to identify objects precisely according to reality (Wong & Lim, 2011). Sensitivity is the percentage of answers given by the system that can be classified from information on all requested data, while specificity is the success rate of the system in recovering information data that can be classified correctly. In addition, the calculation of the F1-Score value used gives an idea of how well the model identifies the positive class correctly without giving many positive errors or negative errors. For example, in the classification of stunting toddlers, there are three categories, namely (1) Stunting; (2) Normal; and (3) high. The most common way to show classification results is by presenting them in the form of a confusion matrix to get accuracy, sensitivity, specificity, and F1-Score values as in Table 1.

Table 1. Confusion Matrix

Actual	Predict		
	Stunting	Normal	High
Stunting	a	b	c
Normal	d	e	f
High	g	h	i

Classification accuracy is calculated through accuracy using formula (6), sensitivity is calculated using formula (7), specificity is calculated using formula (8), and F1-Score is calculated using formula (9).

$$Accuracy = \frac{a+e+i}{a+b+c+d+e+f+g+h+i} \quad (6)$$

$$Sensitivity = \frac{\frac{a}{a+(d+g)} + \frac{e}{e+(b+h)} + \frac{i}{i+(c+f)}}{3} \quad (7)$$

$$Specificity = \frac{\frac{a}{a+(b+c)} + \frac{e}{e+(d+f)} + \frac{i}{i+(g+h)}}{3} \quad (8)$$

$$F1 - Score = 2 \left(\frac{Sensitivity \times Specificity}{Sensitivity + Specificity} \right) \quad (9)$$

With a is number of observations from group 1 that are correctly classified to group 1, b is number of observations from group 1 that are classified to group 2, c is number of observations from group 1 that are classified to group 3, d is number of observations from group 2 that are classified to group 1, e is number of observations from group 2 that are correctly classified to group 2, f is number of observations from group 2 that are classified to group 3, g is number of observations from group 3 that are classified to group 1, h is number of observations from group 3 that are classified to group 2, and i is number of observations from group 3 and correctly classified to group 3.

4. Research Data

The data used is secondary data from the research of Fernandes & Solimun (2023) which examines the factors that cause stunting in Wajak District. The sample in Fernandes & Solimun's research was mothers who had toddlers in Sumberputih Village. Sampling was carried out using stratified random sampling technique with a sample obtained of 100 respondents, all of which were used as samples in this study. Data on economic conditions (X_1), health services (X_2), children's diet (X_3), and environment (X_4) are predictor variables in the form of community perceptions that are assessed with Likert-scale indicators. The Toddler Nutritional Status variable (Y) is an ordinal scale response variable with categories 1 (stunting), 2 (normal), and 3 (high).

5. Steps

Secondary data processing uses CART with ordinal response and bagging algorithms CART with ordinal response.

a. Stages of CART with ordinal response

- 1) Split the data into training data and testing data with a ratio of 80:20.
- 2) Node Breaking for CART analysis as in equation (1).
- 3) Giving labels to classes according to equation (2).
- 4) Tree pruning to get an optimal tree with equation (4).
- 5) Testing data with the formulas Accuracy, Sensitivity, Specificity, and F1 - Score using Equation (6) to Equation (9).

b. Stages of bagging algorithms CART with ordinal response

- 1) Split the data into training data and testing data with a ratio of 80:20.
- 2) Perform resampling bootstrapping \mathcal{L}_B as many as n from the training data with bootstrap.
- 3) Node Breaking for CART analysis as in equation (1).
- 4) Giving labels to classes according to equation (2).
- 5) Tree pruning to get an optimal tree with equation (4).
- 6) Testing data with the formulas Accuracy, Sensitivity, Specificity, and F1 - Score using Equation (6) to Equation (9).

C. RESULT AND DISCUSSION

1. CART Classification Results

The important variables in the case of classifying toddler nutritional status based on height/age using CART showed in Figure 2.

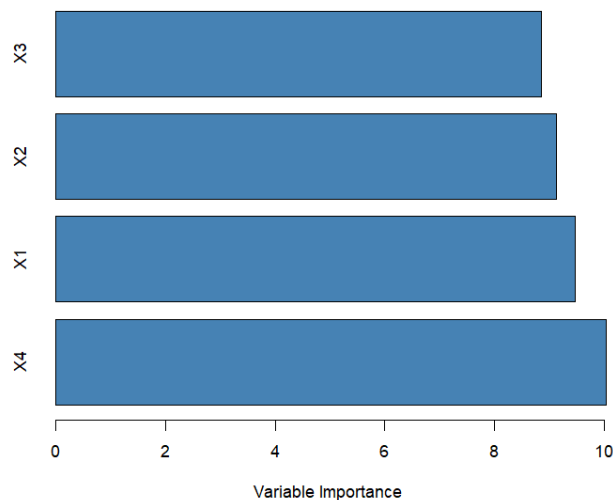


Figure 2. Variable Important of CART with Ordinal Response

Based on Figure 2, the important variables in the case of classifying toddler nutritional status based on height/age are variables environment (X_4), economic conditions (X_1), health services (X_2), and children's diet (X_3), respectively. The importance values of the four variables are not too far apart, which means that the variables economic conditions (X_1), health services (X_2), children's diet (X_3), and environment (X_4) are equally important in classifying the nutritional status of toddlers. The boundary values of the node breaking results to obtain classification results on unbalanced data with CART are presented in Figure 3.

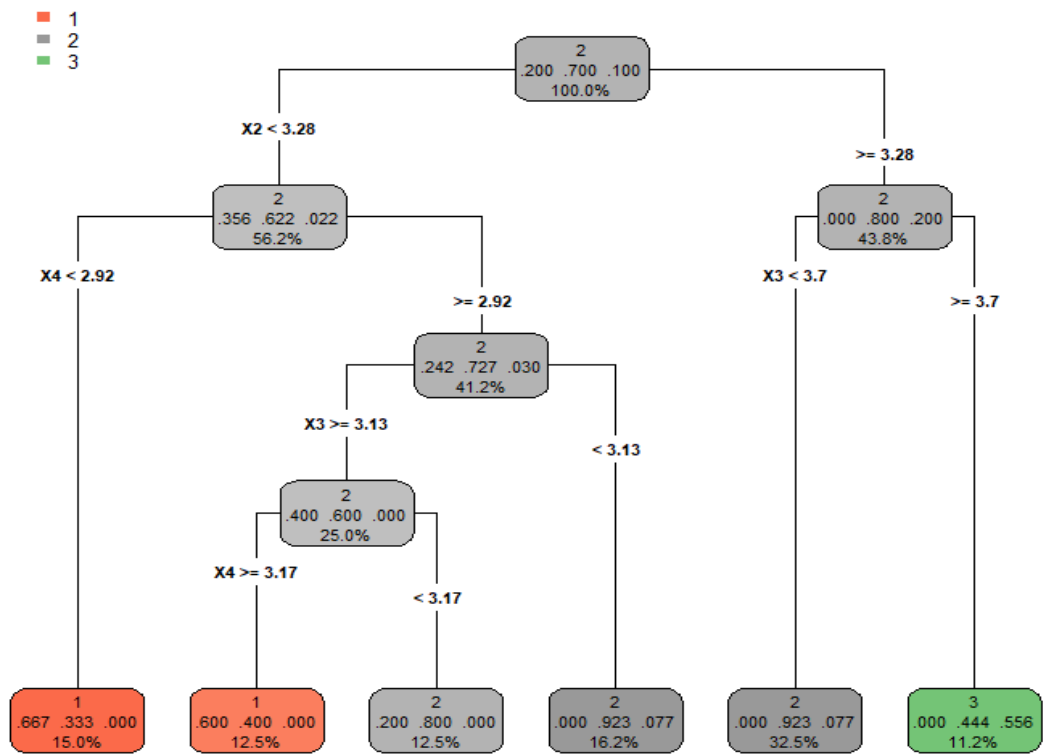


Figure 3. Classification Tree CART with Ordinal Response

Based on the classification tree formed in Figure 3, toddlers are categorized in category 3 or high toddlers when the value of the variable $X_2 \geq 3.28$ and the value of the variable $X_3 \geq 3.7$. Meanwhile, toddlers are categorized as category 1 or stunted toddlers when the value of variable $X_2 < 3.28$ and the value of variable $X_4 < 2.92$ or when the value of variable $X_4 < 2.92$ and the value of variable. On the other hand, other combinations will determine toddlers in category 2 or the normal toddler category. The results of classification of ordinal-scale data with CART are presented in Table 2.

Table 2. Confusion Matrix CART

Actual	Predict		
	Stunting	Normal	High
Stunting	1	3	0
Normal	6	7	1
High	1	0	1

Based on Table 2, the results show that CART is able to classify the nutritional status of 9 out of a total of 20 toddlers correctly. Based on the results in Table 2, the accuracy value is 45%, sensitivity is 44.3%, specificity is 41.7%, and F1-Score is 42.9%.

2. Bagging CART Classification Results

The important variables in the case of classifying toddler nutritional status based on height/age using bagging CART showed in Figure 4.

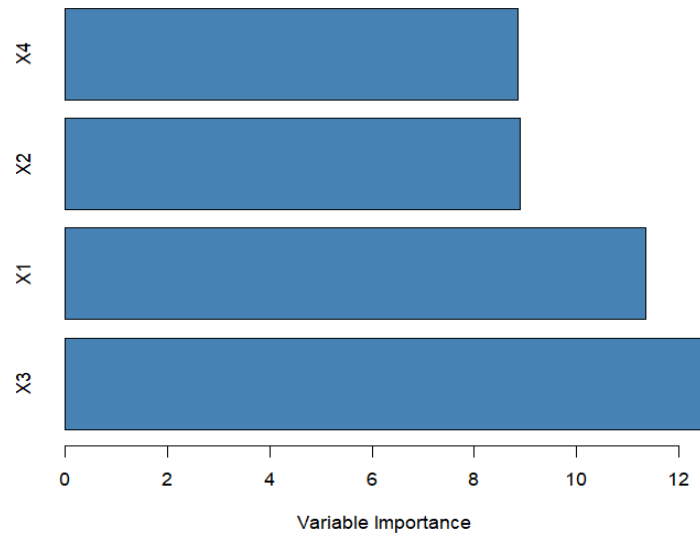


Figure 4. Variable Important of Bagging CART with Ordinal Response

Based on Figure 4, the important variables in the case of classifying toddler nutritional status based on height/age are variables children’s diet (X_3), economic conditions (X_1), health services (X_2), and environtment (X_4), respectively. The greatest importance value is found in variable children’s diet (X_3) so it can be said that variable children’s diet (X_3) is the most important variable in classifying the nutritional status of toddlers. The boundary values of the node breaking results to get classification results on unbalanced data using bagging CART showed in Figure 5.

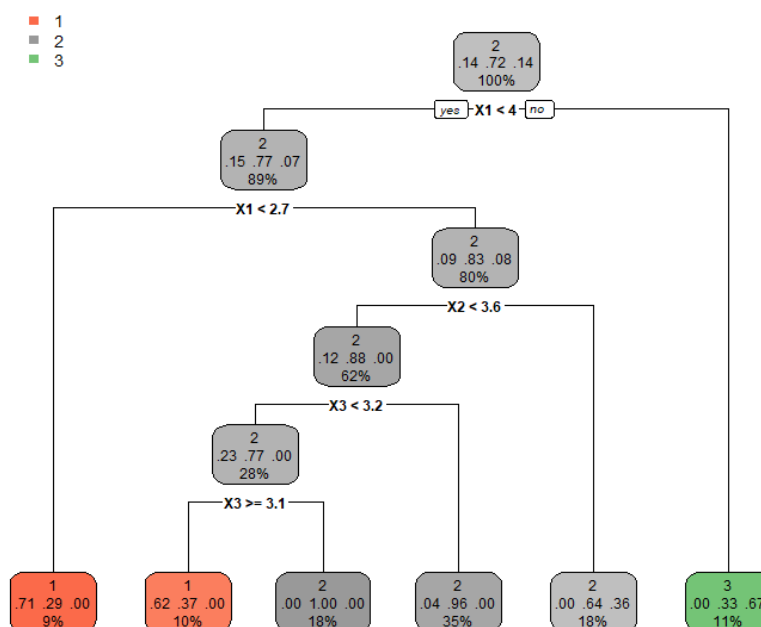


Figure 5. Classification Tree Bagging CART with Ordinal Response

Based on the classification tree formed in Figure 5, toddlers are categorized in category 3 or high toddlers if the variable value $X_1 \geq 4$. Meanwhile, toddlers are categorized as category 1 or stunted toddlers if the value of variable $X_1 < 2.7$ or if the value of variable $X_2 < 3.6$ and the value of variable $X_3 < 3.1$. On the other hand, other combinations will determine the toddler to be in category 2 or the normal toddler category. The results of classification of ordinal-scale data with Bagging CART are presented in Table 3.

Table 3. Confusion Bagging CART

Actual	Predict		
	Stunting	Normal	High
Stunting	2	2	0
Normal	0	14	0
High	0	1	1

Based on Table 3, the results show that CART is able to classify the nutritional status of 17 out of a total of 20 toddlers correctly. Based on the results in Table 3, the accuracy value is 85%, sensitivity is 94.1%, specificity is 66.7%, and F1-Score is 78%.

3. Performance of Classification

Based on Table 2 and Table 3 for testing data category stunting. There are 14 toddlers in the normal category. Using the CART method, the results of the CART classification showed that there were only 7 toddlers who could be classified correctly in the normal category. Meanwhile, using the CART bagging method, of the 14 toddlers in the normal category, the prediction results show that all toddlers in the normal category can be classified correctly in the normal category. This has an impact on the performance of each method. Performance value of bagging CART is much greater than the conventional CART method. The accuracy value is 85% which means that the Bagging CART method is able to classify cases of nutritional status of toddlers correctly by 85%. Sensitivity is 94.1% that shows that the Bagging CART method can correctly classify the nutritional status category of positive toddlers by 94.1%. Specificity is 66.7% this shows that the Bagging CART method can correctly classify the nutritional status category of negative toddlers by 66.7%. F1-Score is 78% which mean bagging CART has balance measurement between a model's ability to correctly identify positive cases (sensitivity) and its ability to avoid classifying negative cases as positive cases (specificity) by 78%. So, it can be concluded that the bagging CART method is better for classifying data with proportion imbalance problems.

This is in line with the research of Kumari et al. (2021) which states that the ensemble method is better at classifying data with unbalanced proportions compared to conventional classification methods. In the research Arrahimi et al. (2019) was found that the CART bagging method has a higher accuracy value than the conventional CART method in the classification of student study period with binary and unbalanced categories. This study uses ordinal category data with three categories. The results of the study are linear with the results in the binary category so that the CART bagging ensemble method is better used on unbalanced data in various categories, both binary and multi-category.

The increase in the performance value of ensemble bagging CART results from the use of bootstrap resampling method. Performing bootstrapping on imbalanced data can reduce variance caused by over-fitting. This is because each tree in the ensemble allows only looking at a random subset of the existing data, resulting in a more general model. In addition, in the bootstrap process, each sample from the minority class may appear several times in the training data set for each tree thereby strengthening the representation of the minority class in the ensemble model. The combined results of all trees can expand the classification possibilities represented by the model thereby increasing the model's ability to generalize on unseen data so that it can improve overall performance, especially on imbalanced data where minority classes are difficult to predict accurately by a single model such as CART.

The results of classification using the best method, namely Bagging CART, show that the most important variable in classifying the nutritional status of toddlers is children's diet (X_3). This is in line with research by Mahmudiono, et al. 2017 who confirmed the widely observed protective relationship between dietary diversity and stunting in children. Therefore, population interventions should focus on increasing food groups that are currently less available in maternal and child diets, including groups rich in growth-promoting nutrients such as milk, meat/poultry. It is hoped that improving the quality of children's diet will accelerate the prevention of stunting and improve nutrition in toddlers. This can start with the mother, as the party who contributes greatly to the child's growth and development, should start paying attention to the nutritional intake that the child eats.

In fact, there are 5 categories of nutritional status for toddlers based on the Regulation of the Minister of Health of the Republic of Indonesia Number 2 of 2020 concerning Anthropometric Standards for children, namely very stunted, stunted, normal, tall, and very tall. However, in this study only 3 categories were taken with a minimum threshold of stunting and height, because no toddlers in this study area were included in the very stunting and/or very tall categories. Therefore, in future research, especially in research in the smallest areas where it is possible to have toddlers in the very stunted category, 5 categories can be used because the possibility of data imbalance will be much greater.

D. CONCLUSION AND SUGGESTIONS

Bagging CART is better at classifying data with unbalanced proportions compared to ordinary CART. This is because the performance value produced by Bagging CART is the highest with accuracy, sensitivity, specificity, and F1-Score values of 85%, 94.1%, 66.7%, and 78%, respectively. The most important variable in classifying the nutritional status of toddlers is children's diet (X_3). So, it is important to measure toddler nutritional consumption and children's eating patterns to prevent stunting. Because this research is limited to the Sumberputih Village area with, Wajak District, the results of this research are only representative of the Wajak District area, Malang Regency. To generalize to a wider area, reassessment is needed with a more representative sample. In future research, it is expected that simulation studies can be carried out with various unbalanced proportions and different sample sizes. Besides that, it can also be compared with other ensemble methods such as boosting or stacking so that it can be seen which method is more suitable for classifying unbalanced data for three-category ordinal data.

ACKNOWLEDGEMENT

Thank you to all parties for their support and input in the preparation of this research. All the support and input given is very useful for the perfection of this research.

REFERENCES

- Altman, N., & Krzywinski, M. (2017). Ensemble methods: bagging and random forests. *Nature Methods*, 14(10), 933–934. <https://doi.org/10.1038/nmeth.4438>
- Arrahimi, A. R., Ihsan, M. K., Kartini, D., Faisal, M. R., & Indriani, F. (2019). Teknik Bagging Dan Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa. *Jurnal Sains Dan Informatika*, 5(1), 2598–5841. <https://doi.org/10.34128/jsi.v5i1.171>
- Bramer, M. (2016). *Principles of Data Mining* (3rd ed.). Springer London. <https://doi.org/10.1007/978-1-4471-7307-6>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Cendani, L. M., & Wibowo, A. (2022). Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes. *Jurnal Masyarakat Informatika*, 13(1), 33–44. <https://doi.org/10.14710/jmasif.13.1.42912>
- Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1), 52. <https://doi.org/10.1186/s40537-020-00327-4>
- Chilyabanyama, O. N., Chilengi, R., Simuyandi, M., Chisenga, C. C., Chirwa, M., Hamusonde, K., Saroj, R. K., Iqbal, N. T., Ngaruye, I., & Bosomprah, S. (2022). Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia. *Children*, 9(7), 1082. <https://doi.org/10.3390/children9071082>
- Daniya, T., Geetha, M., & Kumar, K. S. (2020). Classification and regression trees with gini index. *Advances in Mathematics: Scientific Journal*, 9(10), 8237–8247. <https://doi.org/10.37418/amsj.9.10.53>
- De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.
- du Plooy, R., & Venter, P. J. (2021). A Comparison of Artificial Neural Networks and Bootstrap Aggregating Ensembles in a Modern Financial Derivative Pricing Framework. *Journal of Risk and Financial Management*, 14(6), 254. <https://doi.org/10.3390/jrfm14060254>
- Efendi, A., Fitriani, R., Naufal, H. I., & Rahayudi, B. (2020). Ensemble Adaboost In Classification And Regression Trees To Overcome Class Imbalance In Credit Status Of Bank Customers. *Journal of Theoretical and Applied Information Technology*, 98(17), 3428–3437.
- Erickson, B. J., & Kitamura, F. (2021). Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology: Artificial Intelligence*, 3(3), e200126. <https://doi.org/10.1148/ryai.2021200126>
- Fernandes, A. A. R., & Solimun. (2023). *Menelisik Faktor-Faktor Penyebab Stunting pada Anak di Kecamatan Wajak: Integrasi Cluster dengan Path Analysis dengan Pendekatan Statistika dan Sains Data*.
- Fitriani, R. D., Yasin, H., & Tarno. (2021). Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2). <https://doi.org/10.1214/aos/1016218223>
- Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. *Remote Sensing*, 13(21), 4405. <https://doi.org/10.3390/rs13214405>
- Krzywinski, M., & Altman, N. (2017). Classification and regression trees. *Nature Methods*, 14(8), 757–758. <https://doi.org/10.1038/nmeth.4370>

- Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of Imbalanced Data: Review of Methods and Applications. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012077. <https://doi.org/10.1088/1757-899X/1099/1/012077>
- Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 2, 40–46. <https://doi.org/10.1016/j.ijcce.2021.01.001>
- Li, X., & Zhang, L. (2021). Unbalanced data processing using deep sparse learning technique. *Future Generation Computer Systems*, 125, 480–484. <https://doi.org/10.1016/j.future.2021.05.034>
- Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- Mahmudiono, T., Sumarmi, S., & Rosenkranz, R. R. (2017). Household dietary diversity and child stunting in East Java, Indonesia. *Asia Pacific Journal of Clinical Nutrition*, 26(2), 317–325. <https://search.informit.org/doi/10.3316/ielapa.688058173877148>
- Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, 510, 1–14. <https://doi.org/10.1016/j.neucom.2022.08.055>
- Siahaan, D., Wahyuningsih, S., & Amijaya, F. D. T. (2017). Aplikasi Classification and Regression Tree (CART) dan Regresi Logistik Ordinal dalam Bidang Pendidikan. *EKSPONENSIAL*, 7(1), 95–104. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/46>
- Siringoringo, R., & Jaya, I. K. (2018). Ensemble Learning Dengan Metode Smotebagging Pada Klasifikasi Data Tidak Seimbang. *Information System Development*, 3(2), 75–81. <https://ejournal-medan.uph.edu/index.php/isd/article/view/204>
- Wong, H. B., & Lim, G. H. (2011). Measures of Diagnostic Accuracy: Sensitivity, Specificity, PPV and NPV. *Proceedings of Singapore Healthcare*, 20(4), 316–318. <https://doi.org/10.1177/201010581102000411>
- Zhao, C., Peng, R., & Wu, D. (2023). Bagging and Boosting Fine-tuning for Ensemble Learning. *IEEE Transactions on Artificial Intelligence*, 1–15. <https://doi.org/10.1109/TAI.2023.3296685>