Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

Toddlers' Nutritional Status Prediction Using the Multinomial Logistics Regression Method

Rendra Gustriansyah[1)*], Nazori Suhandi[2)], Shinta Puspasari[3)], Ahmad Sanmorino[4)], Dewi Sartika[5)]
[1,2,3,4,5)]Universitas Indo Global Mandiri, Indonesia
[1)]rendra@uigm.ac.id, [2)]nazori@uigm.ac.id, [3)]shinta@uigm.ac.id, [4)]sanmorino@uigm.ac.id, [5)]dewi.sartika@uigm.ac.id

ABSTRACT

Malnutrition is one of the foremost health problems experienced by children under five in many countries, especially in low and middle-income countries. Meanwhile, the target of Sustainable Development Goals (SDGs) 2.2 is that all forms of malnutrition must end by 2025. Therefore, this study aims to predict the toddlers' nutritional status (malnutrition, undernutrition, overnutrition, and normal nutrition) based on age, body mass index (BMI), weight, and length using the Multinomial Logistic Regression (MLR) classification method. The dataset consists of two hundred toddlers obtained from the Kaggle site. Following pre-processing, the dataset is divided, with 80 percent of the data for training and the remaining 20 percent for testing. The model was trained using 10-fold cross-validation (CV). In Addition, the MLR model performance was evaluated using the confusion matrix (CM), the area under the curve (AUC), and the Kappa coefficient (KC). The evaluation results using CM show that the accuracy, sensitivity, and specificity values are 0.9412, 0.9375, and 0.9790, respectively. AUC and KC also show excellent results. It indicates that the MLR method is an esteemed and recommended method for predicting the nutritional status of toddlers. Therefore, this research can contribute to providing early information so that the Government can immediately determine the necessary treatment.

Keywords: classification; logistic regression; nutritional status; prediction; toddlers

INTRODUCTION

Toddlers' general growth and development may be impacted by malnutrition. According to the Indonesian Nutrition Status Survey (SSGI) statistics from 2022, 7.7 percent of toddlers in Indonesia suffer from wasting, while 17.1 percent are underweight. The rate of malnutrition rose by around 0.1 percent in comparison to 2021. Malnutrition rose by 0.6 percent throughout this time (Kementerian Kesehatan RI, 2023). The percentage of malnourished children under-five in Indonesia is higher than the norm for both Africa (5.8%) and the world (6.8%) (UNICEF, 2023). This condition is exacerbated by the community's lack of knowledge about the importance of nutrition, lack of access to nutritious food, cases of poverty, and limited food availability. In fact, the target of Sustainable Development Goals (SDGs) 2.2 is that all forms of malnutrition must end by 2025. This target is an international consensus regarding wasting and stunting in children under-five (WHO, 2023).

Toddler malnourishment or undernutrition carries a risk of serious health complications, including an increased chance of developing diabetes, hypertension, stroke, behavioral difficulties, and even mortality. Therefore, an early and efficient method of determining the toddlers' nutritional condition is required so that the government (via the District Health Office) can start treating them right once. By categorizing children according to their age, body length, weight, and BMI, the machine learning (ML) approach can

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

rapidly predict their nutritional condition. It will affect how simple it is to find out how nourished the toddlers are.

The k-nearest neighbor (Aryuni et al., 2023; Lonang et al., 2022; Setiawan et al., 2022), naïve Bayes (Arumi et al., 2023; Lasarudin et al., 2022; Setiawan et al., 2022), ordinal logistic regression (Kassie et al., 2020), support vector machines (Ramon et al., 2022; Sanmorino et al., 2022), and decision trees (Lasarudin et al., 2022; Nazir et al., 2022; Pinaryanto et al., 2021; Ula et al., 2022) methods are some of the machine learning approaches used to identify or predict malnutrition or nutritional status in toddlers. Additionally, several research (Bitew et al., 2021; Gustriansyah et al., 2024; Hemo et al., 2021; Momand et al., 2020; Rahman et al., 2021; Shahriar et al., 2019; Talukder et al., 2020) compare several machine learning approaches for categorizing malnutrition in toddlers. Nevertheless, no studies utilizing the MLR technique to predict toddlers' multi-class nutritional status - that is, malnutrition, undernutrition, normal, and overnutrition - have been conducted in Indonesia. Therefore, the purpose of this study is to use the MLR method to identify the multi-class nutritional status of toddlers. Research (Kassie et al., 2020) that employed the ordinal logistic regression method served as inspiration for the MLR method selection. Furthermore, the predictive performance of the MLR method will be evaluated based on its accuracy, sensitivity, specificity, area under the curve, and Kappa coefficient.

LITERATURE REVIEW

Multinomial Logistic Regression Classification Method

A machine learning approach called Multinomial Logistic Regression (MLR) is used for classifying multi-class targets using a logistic function. The formulation is shown in (1) (Umaña-Hermosilla et al., 2020).

$$RL(x) = \frac{1}{(1+e^{-x})} \qquad (1)$$

Where: x refers to a linear combination of associated variables (x) and coefficient values (β), $x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$. Under-five status, a multi-class variable that includes the classes of malnutrition, undernutrition, overnutrition, and normal nutrition, is the target variable. The MLR model will use the normal nutrition class as a comparison.

Multiclass Confusion Matrix (CM)

Accuracy, sensitivity, and specificity metrics are measured using the CM, a square matrix made up of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (PN). Each metric's formula is shown in (2)-(4) (Yao, 2022).

$$accuracy = TP + TN / (TP + TN + FP + FN) \qquad (2)$$
$$sensitivity = TP / (TP + FN) \qquad (3)$$
$$specificity = TN / (TN + FP) \qquad (4)$$

The accuracy metric is known as the ratio of the number of correct predictions to the total number of predictions (Puspasari et al., 2022). The accuracy is crucial because it shows how closely the predicted value matches the actual value . The most widely used statistic to assess overall prediction performance is this one. The accuracy value in this study is the percentage of all toddlers who were properly predicted to have normal nutrition, undernutrition, overnutrition, or malnutrition.

The ratio of the number of positive predictions to the total of valid positive predictions is known as recall or sensitivity (Gustriansyah et al., 2024). This metric illustrates how well the model predicts positive

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

cases. Out of all the toddlers who genuinely suffer from malnutrition, undernutrition, overnutrition, or normal nutrition, the sensitivity value shows the proportion of toddlers who are exactly predicted to have these conditions.

The specificity metric calculates the ratio between the number of negative predictions and the total of valid negative predictions (Gustriansyah et al., 2024). The percentage of toddlers who are exactly predicted to not suffer malnutrition, undernutrition, overnutrition, or normal nutrition out of all toddlers who do not suffer any of these conditions is represented by the specificity value.

Area Under Curve (AUC)

AUC is an assessment statistic that evaluates a multi-class classification method's overall performance by establishing a threshold value. AUC quantifies a model's capacity to distinguish between two classes or groups. One class is assigned as the "positive" class in this multi-class research, and the others are allocated as the "negative" classes. The more accurate the model is at predicting class, the higher its AUC value (closer to 1) (Kim et al., 2023).

Kappa Coefficient (KC)

The degree of inter-rater reliability, or agreement between annotators, on a classification task is measured by the Kappa coefficient. The KC can be used for multi-class or unequal-class classification. 0 represents a low agreement, and 1 represents a perfect agreement for the coefficient values (Rafieyan et al., 2023). Only qualitative (categorical) data measurement outcomes are subject to KC.

METHOD

Research Stages

The stages of the research are depicted in Fig. 1, from data collection to evaluation. The data processing technology employed in this study is R programming.

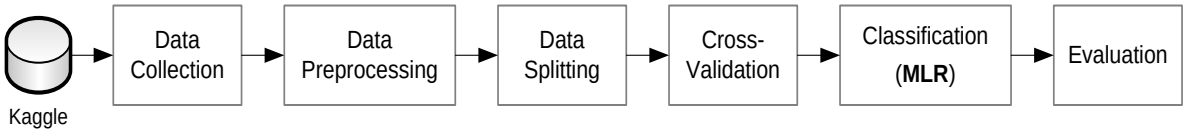


Fig. 1 Research Stages

Data Collection

The Kaggle website (Kaggle, 2022) provided the dataset used in this study. Name, gender, age, weight, height, BMI, and toddler status are the seven variables that make up the 200 toddlers in the dataset. However, since they are not included in the computation, two variables - name and gender - are not incorporated into the classification procedure. Consequently, in this work, training and testing were performed using just five variables from the dataset.

Data Preprocessing and Splitting

Outliers, duplicate data, and null data are removed from the dataset. Boxplot diagrams are used to identify outliers (Gustriansyah et al., 2022b), while Histograms are used to determine each variable's normal distribution. If the amount of data for each class on the target variable (toddler status) is not balanced, the modest amount of data in the class will serve as the foundation for regulation. Besides, the correlation between variables is calculated using the rank correlation coefficient. All variables are normalized by centering (subtracting data from the mean) and scaling (z-score), with the target variable encoded as a

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

category (Gustriansyah et al., 2022a). The dataset is then randomly split, with 80% of the data going toward training and the remaining 20% for testing.

Cross-Validation

Cross-validation seeks to methodically arrange model training for the best possible prediction accuracy by splitting datasets into distinct and varying sizes. The study employed the k-fold cross-validation (CV) method with a value of k=10 for the distribution of sample data. To reduce the amount of prediction bias, this approach will resample the dataset for every feasible situation. As a result, this approach is frequently employed in different classification research (Chen et al., 2024).

Evaluation

The final performance of the MLR model will be measured using the metrics of accuracy, sensitivity, specificity (Rafieyan et al., 2023), area under the curve (Kim et al., 2023), and Kappa coefficient (Rafieyan et al., 2023).

RESULT

Data Exploration

Table 1 presents a descriptive statistical analysis of 200 toddler data in the dataset. This dataset consists of 14 percent of children under five who suffer from malnutrition, 29 percent of children who are undernourished, and 13 percent who suffer from overnutrition. The remaining 44 percent are children under five in normal nutritional condition. The standard deviation indicates that the data is less diversified because the distribution is less than the mean.

Table 1
Descriptive Statistics of the Dataset

Variable	Observation	Minimum	Maximum	Mean	Std. Deviation
Age (month)	200	11.00	54.00	32.77	8.47
Weight (kg)	200	3.40	14.00	8.79	2.13
Height (cm)	200	49.00	98.00	73.24	10.75
BMI	200	6.04	40.80	17.34	6.59

Data Pre-Processing and Splitting Results

Following the verification of duplicates and null data, there is no data reduction at the pre-processing stage. However, as Fig. 2 illustrates, there are outlier values in the weight, height, and BMI variables. After the outlier data was removed, 188 data were left, representing the malnutrition, undernutrition, overnutrition, and normal nutrition classes, respectively, with 24, 56, 20, and 88 data. Each class's data volume demonstrates how unbalanced the dataset is. Therefore, the dataset was balanced by random data reduction to minimize calculation bias.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

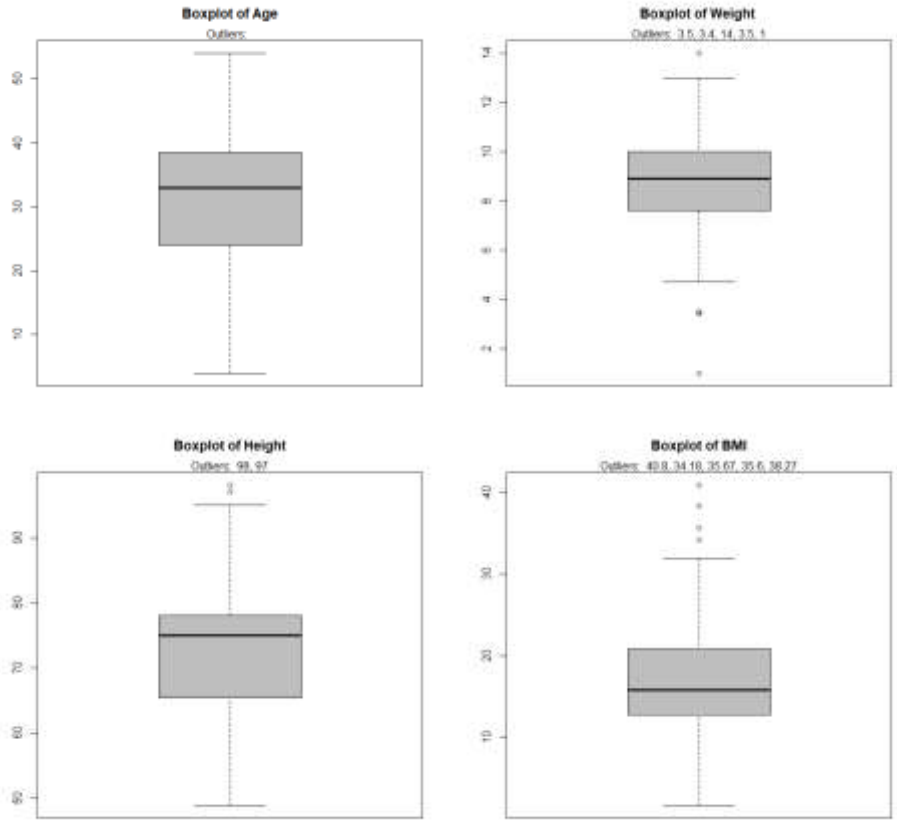
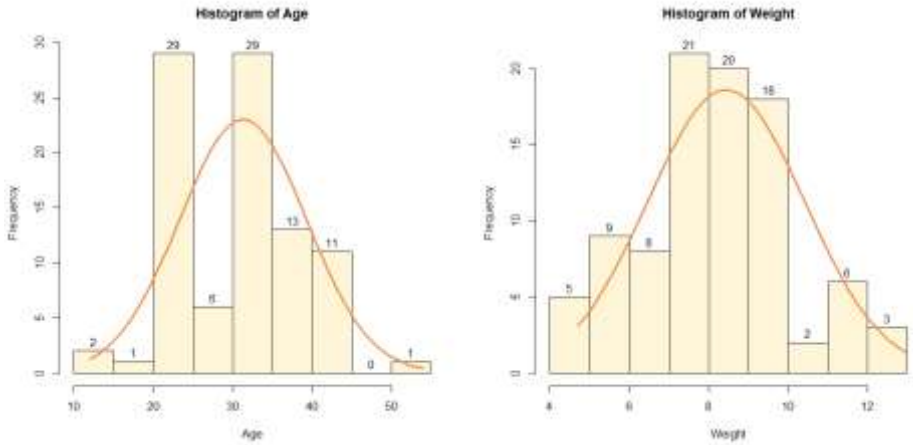


Fig. 2 Boxplot for each variable

There were 92 data in the final dataset for the classes of malnutrition, undernutrition, overnutrition, and normal nutrition, which are 24, 23, 20, and 25, respectively. Meanwhile, the data distribution for each variable is relatively normal except for the BMI variable, which has a little rightward skew (Fig. 3). However, this does not significantly affect the data distribution.



* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

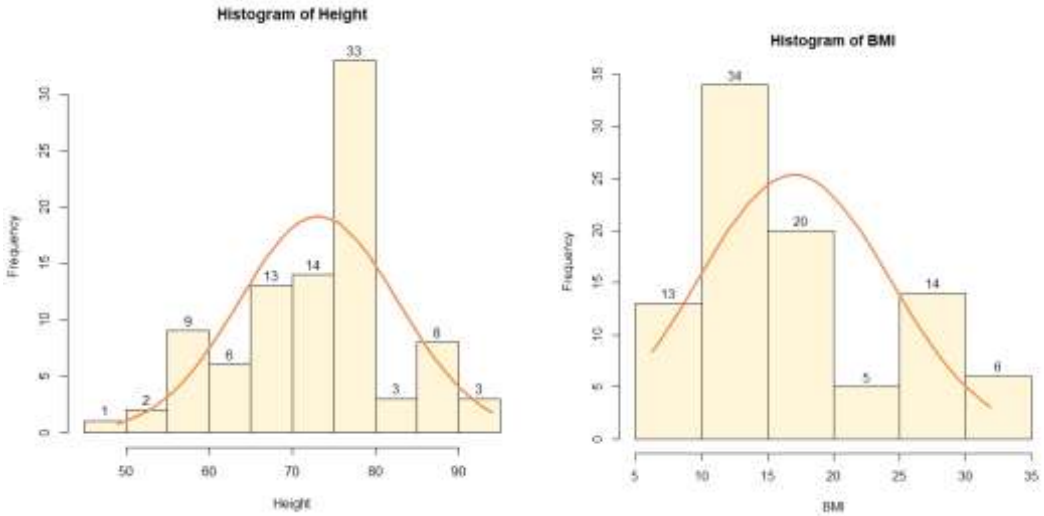


Fig. 3 Histogram of data distribution for each variable

Figure 4 shows there is commonly a low-rank correlation coefficient value between the variables. It demonstrates how well the variables are independent of one another. Nonetheless, there is a strong correlation between the target variable (status of children under five) and the weight and BMI variables. Furthermore, the dataset was split into 75 training data (80%) and 17 testing data (20%).

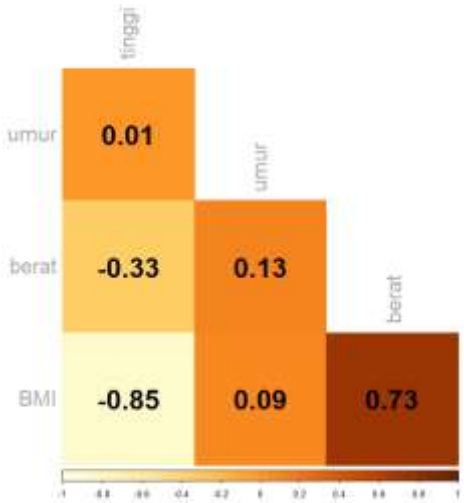


Fig. 4 Coefficient values between variables

MLR Classification Results with 10-Fold CV

Figure 5 displays the confusion matrix of classification results using the MLR model with a 10-fold CV for test data. CM visualizes the performance of the MLR prediction model. Based on the CM data, accuracy, sensitivity, and specificity can be calculated using (2)-(4).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

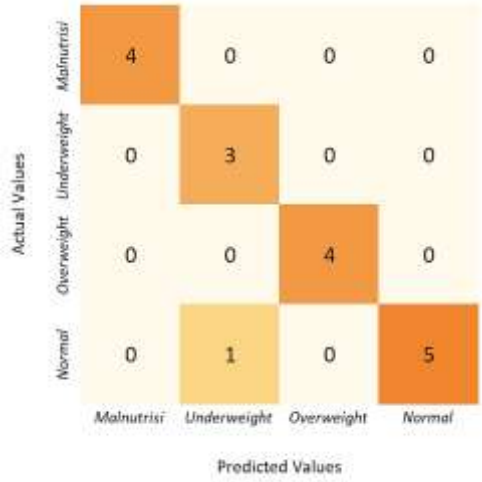


Fig. 5 Confusion matrix of the MLR model

DISCUSSIONS

Based on the CM, high accuracy, sensitivity, and specificity metrics were produced, namely 0.9412, 0.9375, and 0.9792, respectively. Table 2 further shows that this MLR model has an outstanding level of discrimination, with an AUC value of more than 0.91, and a near-perfect reliability level, with a KC value of more than 0.81. The test's outcomes demonstrate the MLR model's excellent performance.

However, the dataset used for training in this study is small, with only four variables (age, weight, height, and BMI) based on Indonesian anthropometric standards. It poses a restriction to the research. In order to achieve better performance, future studies can use a larger dataset and incorporate additional variables like demographics and others.

Table 2
MLR Model Performance

Metrics	Value
Accuracy	0.9412
Sensitivity	0.9375
Specificity	0.9792
AUC	1.0000
KC	0.9209

CONCLUSION

In many nations, malnutrition is one of the primary health issues experienced by toddlers. Thus, a method for promptly determining a toddler's nutritional state is required so that the government (via the District Health Office) can immediately treat them as soon as necessary. This study employs the Multinomial Logistic Regression approach to determine the toddlers' nutritional status based on age, BMI, weight, and height. Test results using a confusion matrix featuring 10-fold CV, AUC, and KC reveal that the MLR model performs exceptionally well.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

REFERENCES

Arumi, E. R., Subrata, S. A., & Rahmawati, A. (2023). Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(2), 201–207. https://doi.org/10.29207/resti.v7i2.4438

Aryuni, M., Miranda, E., Kumbangsila, M., Richard, Zakiyyah, A. Y., Sano, A. V. D., & Bhatti, F. M. (2023). Comparison of Nutritional Status Prediction Models of Children Under 5 Years of Age Using Supervised Machine Learning. *3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics*, 265–277. https://doi.org/10.1007/978-981-99-0248-4_19

Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2021). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, *25*(3), 269–280. https://doi.org/10.1017/S1368980021004262

Chen, Y., Li, L., Li, W., Guo, Q., Du, Z., & Xu, Z. (2024). Fundamentals of neural networks. In *AI Computing Systems* (pp. 17–51). Elsevier. https://doi.org/10.1016/B978-0-32-395399-3.00008-1

Gustriansyah, R., Alie, J., Sanmorino, A., Heriansyah, R., & Noor, M. N. M. M. (2022a). Machine Learning for Regencies-Cities Clustering Based on Inflation and Poverty Rates in Indonesia. *Indonesian Journal of Information Systems (IJIS)*, *5*(1), 64–73. https://doi.org/10.24002/ijis.v5i1.5682

Gustriansyah, R., Alie, J., & Suhandi, N. (2022b). Hierarchical clustering for crime rate mapping in Indonesia. *ILKOM Jurnal Ilmiah*, *14*(3), 275–283. https://doi.org/10.33096/ilkom.v14i3.1135.275-283

Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine Learning Method to Predict the Toddlers' Nutritional Status. *INFOTEL*, *16*(1), 1–6.

Hemo, S. A., & Rayhan, M. I. (2021). Classification tree and random forest model to predict under-five malnutrition in Bangladesh. *Biom Biostat Int J*, *10*(3), 116–123. https://doi.org/10.15406/bbij.2021.10.00337

Kaggle. (2022). *The Baby Nutrition Dataset*. Kaggle. https://www.kaggle.com/datasets/mjalaluddinassuyuti/baby-nutrition-classification

Kassie, G. W., & Workie, D. L. (2020). Determinants of under-nutrition among children under five years of age in Ethiopia. *BMC Public Health*, *20*(1), 1–11. https://doi.org/10.1186/s12889-020-08539-2

Kementerian Kesehatan RI. (2023). *Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Kementerian Kesehatan RI. https://kesmas.kemkes.go.id/assets/uploads/contents/attachments/09fb5b8ccfdf088080f2521ff0b4374f.pdf

Kim, T., & Lee, J.-S. (2023). Maximizing AUC to learn weighted naive Bayes for imbalanced data classification. *Expert Systems with Applications*, *217*, 1–17. https://doi.org/10.1016/j.eswa.2023.119564

Lasarudin, A., Gani, H., & Tomayahu, M. (2022). Perbandingan Metode Naïve Bayes dan C4.5 Klasifikasi Status Gizi Bayi Balita. *SPECTA Journal of Technology*, *6*(3), 273–283. https://doi.org/10.35718/specta.v6i3.789

Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, *6*(1), 49–56. https://doi.org/10.30865/mib.v6i1.3312

Momand, Z., Mongkolnam, P., Kositpanthavong, P., & Chan, J. H. (2020). Data Mining Based Prediction of Malnutrition in Afghan Children. *2020 12th International Conference on Knowledge and Smart Technology (KST)*, 12–17. https://doi.org/10.1109/KST48564.2020.9059388

Nazir, A., Akhyar, A., Yusra, Y., & Budianita, E. (2022). Toddler Nutritional Status Classification Using C4.5 and Particle Swarm Optimization. *Scientific Journal of Informatics*, *9*(1), 32–41. https://doi.org/10.15294/sji.v9i1.33158

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 1, January 2024
https://doi.org/10.47709/cnahpc.v6i1.3372

Submitted : Dec 27, 2023
Accepted : Dec 31,2023
Published : Jan 1, 2024

Pinaryanto, K., Nugroho, R. A., & Basilius, Y. (2021). Classification of Toddler Nutrition Using C4.5 Decision Tree Method. *International Journal of Applied Sciences and Smart Technologies*, *3*(1), 131–142. https://doi.org/10.24071/ijasst.v3i1.3366

Puspasari, S., Ermatita, E., & Zulkardi, Z. (2022). Machine Learning for Exhibition Recommendation in a Museum's Virtual Tour Application. *International Journal of Advanced Computer Science and Applications*, *13*(4), 404–412. https://doi.org/10.14569/IJACSA.2022.0130448

Rafieyan, S., Vasheghani-Farahani, E., Baheiraei, N., & Keshavarz, H. (2023). MLATE: Machine learning for predicting cell behavior on cardiac tissue engineering scaffolds. *Computers in Biology and Medicine*, *158*, 1–11. https://doi.org/10.1016/j.compbiomed.2023.106804

Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLOS ONE*, *16*(6), 1–11. https://doi.org/10.1371/journal.pone.0253172

Ramon, E., Nazir, A., Novriyanto, N., Yusra, Y., & Oktavia, L. (2022). Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Sistem Informasi Dan Informatika (Simika)*, *5*(2), 143–150. https://doi.org/10.47080/simika.v5i2.2185

Sanmorino, A., Gustriansyah, R., & Alie, J. (2022). DDoS Attacks Detection Method Using Feature Importance and Support Vector Machine. *JUITA : Jurnal Informatika*, *10*(2), 167–171. https://doi.org/10.30595/juita.v10i2.14939

Setiawan, R., & Triayudi, A. (2022). Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web. *Jurnal Media Informatika Budidarma*, *6*(2), 777–785. https://doi.org/10.30865/mib.v6i2.3566

Shahriar, M. M., Iqubal, M. S., Mitra, S., & Das, A. K. (2019). A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh. *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 145–149. https://doi.org/10.1109/ICIAICT.2019.8784823

Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, *78*, 1–22. https://doi.org/10.1016/j.nut.2020.110861

Ula, M., Ulva, A. F., Mauliza, M., Ali, M. A., & Said, Y. R. (2022). Application of Machine Learning in Determining The Classification of Children's Nutrition with Decision Tree. *Jurnal Teknik Informatika (Jutif)*, *3*(5), 1457–1465. https://doi.org/10.20884/1.jutif.2022.3.5.599

Umaña-Hermosilla, B., de la Fuente-Mella, H., Elórtegui-Gómez, C., & Fonseca-Fuentes, M. (2020). Multinomial Logistic Regression to Estimate and Predict the Perceptions of Individuals and Companies in the Face of the COVID-19 Pandemic in the Ñuble Region, Chile. *Sustainability*, *12*(22), 1–20. https://doi.org/10.3390/su12229553

UNICEF. (2023). *Child Malnutrition*. https://data.unicef.org/topic/nutrition/malnutrition/

WHO. (2023). *SDG Target 2.2 Malnutrition*. The Global Health Observatory. https://www.who.int/data/gho/data/themes/topics/sdg-target-2_2-malnutrition

Yao, C. (2022). Hearing loss classification via stationary wavelet entropy and cat swarm optimization. In *Cognitive Systems and Signal Processing in Image Processing* (pp. 203–221). Elsevier. https://doi.org/10.1016/B978-0-12-824410-4.00014-3

* Corresponding author