Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 7, No. 1, Juni 2023
pp. 43-49

Classification of Toddler Nutritional Status Using a Binary Classification Tree With AlgorithmsQuick, Unbiased, Efficent, Statistical Tree

Cahya Dwi Ramadhani [1], Machrani Adi Putri Siregar[2],
[1,2]Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Keywords:

Classification
Toddler nutritional status
QUEST

ABSTRACT

Determination of nutritional status is very important in helping to monitor the state of nutritional health growth in toddlers every time. In this study, there were 70 identity data for toddlers for the 2022 period obtained from the KB Counseling Center in the Pegajahan sub-district, in Sukasari Village. There are four independent variables used, namely gender, health insurance, weight, and height. The purpose of this study is to determine the classification that is formed and the accuracy of the resulting classification on the nutritional status of toddlers. Classification which is part of data mining can make decisions on the nutritional status of toddlers faster and more efficiently. QUEST method *(Quick, Unbiased, Efficient, Statistical Trees)* is one of the statistical methods that can be used to form a decision tree and classify an object using a separator algorithm that produces a binary tree. From the results of the classification there is a variable height (x_3) as the limiting variable. in the early stages of insulation, the parent node which consists of 70 toddler data. Variables are partitioned based on height into two nodes, namely node (1) and node (2). Node (1) is a node containing 25 children under five with a height of more than 85.79 cm, while node (2) is a node for 45 children under five with a height less than or equal to 85.79 cm. in the next process, the blocking is terminated. the overall value of the accuracy of the classification of trees formed is 95.7%. Thus, the probability of misclassification of the tree is 4.3%, which means that this classification tree is optimal.

Corresponding Author:

Cahya Dwi Ramadhani,
Department of Mathematics,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: cahya0703191056@uinsu.ac.id.com

1. INTRODUCTION

Currently, Indonesia is still facing nutritional problems that have a serious impact on the quality of human resources (Rahayu & et al, 2018). One of the problems in nutrition is stunting or malnutrition in toddlers which is still a major obstacle in the population system (Ayu, 2021). According to the results of the 2022 Indonesian Nutrition Status Survey (SSGI), the stunting rate has fallen from 24.4% in 2021 to 21.6% in 2022. However, hard work is still needed to reach the 14% target. (SSGI Ministry of Health RI., 2023

For this reason, the determination of nutritional status is very important in helping to monitor the state of nutritional health growth in toddlers at any time. Nutritional status can be a measure of success in fulfilling nutrition for children as indicated by the child's weight and height. Nutritional status is also defined as health status. (Adzani, 2021).

Classification which is part of data mining can make decisions on the nutritional status of toddlers faster and more efficiently. QUEST method *(Quick, Unbiased, Efficient, Statistical Trees)* is one of the statistical methods that can be used to form a decision tree and classify an object using a separator algorithm that produces a binary tree.

The advantage of the QUEST Algorithm is that it has speed in computing *(quick),* produce unbiased independent variables *(unbiased),* and efficient for complex data, that is, it can use independent variables of categorical and numeric types *(efficient).*(Maharani & et al., 2017). So it is hoped that this research can help the process of classifying data easily computationally, and can provide information that is analyzed statistically. Therefore, a Binary classification tree with the QUEST Algorithm is used to classify the nutritional status of toddlers.

2. RESEARCH METHOD

2.1 Data source

The source of the data used in making binary classification trees with the QUEST algorithm is the identity format data for toddlers in 2022 in Sukasari Village, Pegajahan sub-district, obtained from the KB Counseling Center, sub-district. Pegajahan.

2.2 Research variable

The variables in this study consist of predictor variables and response variables. Response variable(AND) in this study was the nutritional status of toddlers, which in this study were grouped into two categories namely, status 0 = A is normal nutritional status and status 1 = B is nutritional status at risk of stunting. Predictor variable(X) in this study is the toddler identity format, namely gender (X_1) in the form of categorical data with 0 denoting female gender and 1 denoting male gender, current body weight(X_2) in the form of continuous data, current body height(X_3) in the form of continuous data, health insurance(X_4) in the form of categorical data.

2.3 Data analysis method

a. Selection of baffle variables
   The selection of insulating variables is carried out by means of statistical tests to calculate significant values. Statistical tests for numerical variables were carried out using the F test, while the chi-square test was carried out for categorical variables.
b. Selection and comparison of p-values
   The selection of p-values is done based on the variable that has the smallest p-value.
c. Determination of insulation nodes
   The results of numerical p-value variables and categorical variable p-values were compared. If the selected variable is categorical, then the variable is converted into a numeric variable by transforming it into an I-dimensional dummy vector. But if the selected variable is continuous, then the quadratic discriminant analysis steps are carried out as follows:
   1. For example, $\bar{x}_0$ and s_0^2 is the mean and the variance X^* from the observation of the first category response variable, meanwhile $\bar{x}_1$ and s_1^2 is the mean and the variance X^* from the second category of observations. For example, $P(t) = \frac{N_{j,t}}{N_j}$ is the probability of each category in the response variable, with N_j is the number of observations at the initial node of the response group j.
   2. Quadratic discriminant analysis blocks the X into three intervals viz$(-\infty, d_1), (d_1, d_2)$ and(d_2, ∞) Where$d$ are the roots of the equation

$$P(t)s_A^{-1}\phi\left(\frac{x-\bar{x}_A}{s_A}\right) = P(t)s_B^{-1}\phi\left(\frac{x-\bar{x}_B}{s_B}\right) \quad (2.1)$$

   Log both sides to obtain a quadratic equation $ax^2 + bx + c = 0$,

   Were

$$a = s_0^2 - s_1^2 \tag{2.2}$$

$$b = 2(\bar{x}_0 s_1^2 - \bar{x}_1 s_0^2) \tag{2.3}$$

$$c = (\bar{x}_0 S^2{}_1)^2 - \left(\bar{x}_1 S^2{}_0\right)^2 + 2S^2{}_0 S^2{}_1 \,.\, log\left(\frac{P(0|t)\,.\,S_1}{P(1|t)\,.\,S_0}\right) \tag{2.4}$$

If $a = 0$ and, $\underline{x}_A \neq \underline{x}_B$ there is only one root

$$x = \left(\frac{\bar{x}_0 + \bar{x}_1}{2}\right) - \left(\frac{s_0^2}{\bar{x}_0 + \bar{x}_1}\right) log\left(\frac{P(0|t)}{P(1|t)}\right) \tag{2.5}$$

The equation has no roots if $a = 0\ and\ \underline{x}_A = \underline{x}_B$.

3. A knot is split $X^* = d$, Where$d$ defined as follows:

If $a = 0$, for

$$d = \begin{cases} x = \frac{\bar{x}_0 + \bar{x}_1}{2} - \frac{s_0^2}{\bar{x}_0 + \bar{x}_1}\ log\left(\frac{P(0|t)}{P(1|t)}\right), & \bar{x}_0 \neq \bar{x}_1 \\ \bar{x}_0 \qquad\qquad\qquad\qquad\qquad\qquad , \bar{x}_0 = \bar{x}_1 \end{cases} \tag{2.6}$$

For $a \neq 0$,

If $b^2 - 4ac < 0$,for

$$d = \frac{\bar{x}_0 + \bar{x}_1}{2} \tag{2.7}$$

If $b^2 - 4ac \geq 0$, then define

$$d = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{2.8}$$

Which is closer$\underline{x}_A$, provided it produces two non-empty nodes, QUEST only use one of the two roots of the equation, namely the root whose value is closest to the sample average of each class (Handayani & DKK, 2013).

2.4 Termination of Node Blocking

Tree The blocking process is carried out until the nodes cannot be partitioned anymore with the rules for terminating the tree formation process that are determined, including the following:

a. If a node becomes pure, that is, all objects/cases belong to the same class of bound variables at that node, then the node will not be partitioned.
b. If all objects/cases in a node have identical values for each predictor variable, then the node will not be partitioned.
c. If the tree depth at that time has reached the specified maximum point, then the tree insulation process is stopped.
d. If the partition of a node results in a child node whose node size is less than the specified minimum child node size value, the node will not be partitioned. (Handayani & DKK, 2013)

2.5 Classification Tree Accuracy

The correct classification of the observations in binary logistic regression must determine the value*cutpoint* (c) and compared with the estimated probabilities $\pi(x)$. If $\pi(x)$ greater or equal to c then the estimated value is included in the response $y = 1$ and besides $y = 0$. The correct classification consists of specificity and sensitivity. Specificity and sensitivity can be calculated using table 2.1 below:

Table 2.1 Response Classification

Observation	Guess	
	1	**0**
1	a	b
0	c	d

Specificity or classification accuracy in predicting the event that the response does not have the expected criteria, namely on $(y = 0)$. as big$\left(\frac{d}{c+d}\right).100\%$, to evaluate the accuracy of the classification in predicting the event that the response has the expected criteria ie$(y = 1)$ or also called *sensitivity* which is worth $\left(\frac{a}{a+b}\right).100\%$, while the accuracy of classification, namely the accuracy of classification in predicting events accurately can be predicted by a model whose value $\left(\frac{a+d}{a+b+c+d}\right).100\%$. (Rizki & Setyawan, 2018).

3. RESULT AND ANALYSIS

3.1. Selection of Insulating Variables

The determination of the insulating variable begins by finding the smallest p-value variable. For categorical data, the chi square test was carried out, and for numerical data, the ANOVA F test was carried out. Then with the help of SPSS the results of the Anova test and chi square test, can be seen in Table 3.1 below.

Table 3.1 Chi Square Test Results and ANOVA

Variable	Test statistic value	p-value
x_1	$x^2 = 1{,}346$	0,246
x_2	$x^2 = 0{,}210$	0,647
x_3	$F = 53{,}247$	0,000
x_4	$F = 5{,}166$	0,026

The smallest p-value is on the variablex_3, height. Then when used$\alpha = 0.05$ andM_1 is the number of independent variables, obtained $\frac{\alpha}{M_1} = \frac{0{,}05}{4} = 0{,}0125$. Then the smallest p-value$0{,}000 < 0{,}0125$. So the height variable(x_3) is the partitioning variable or initial node. Furthermore, because the selected ones are numerical variables, quadratic discriminant analysis is used to determine the insulating nodes.

3.2. Determination of Insulating Knots

a. Define $\bar{x}_0$ and S_0^2 are the mean and the rangeX_3 from observation to response 0, whereas $\bar{x}_1$ and S_1^2 are the mean and the rangeX_3 from observation to response1. For example, $P(t) = \frac{N_{k,t}}{N_k}$ is the probability of each category of response variables. $N_{k,t}$ is the amount of data on the node$t$ for response$k$ andN_k is the amount of data at the initial node for the response k.

b. Variables x_3 quadratic discriminant analysis was carried out to obtain insulating nodes or insulating points. From the data obtained:

$\bar{x}_0 = 97{,}18$ $\quad\quad$ $\bar{x}_1 = 74{,}40$

$S_0^2 = 29{,}16$ $\quad\quad$ $S_1^2 = 234{,}09$

$P(0|t) = \frac{N_{k,t}}{N_k} = \frac{26}{70} = 0{,}37$ $\quad\quad$ $P(1|t) = \frac{N_{k,t}}{N_k} = \frac{44}{70} = 0{,}62$

The blocking node is the root of the equation$ax^2 + bx + c = 0$,
Where

$a = s_0^2 - s_1^2 = 29{,}16 - 234{,}09 = -204{,}93$

$b = 2(\bar{x}_0 s_1^2 - \bar{x}_1 s_0^2) = 2(97{,}18\,.\,234{,}09 - 74{,}40\,.\,29{,}16) = 41158{,}7$

$c = (\bar{x}_0 S_1^2)^2 - (\bar{x}_1 S_0^2)^2 + 2S_0^2 S_1^2 \left(\frac{P(0|t)\,.\,S_1^2}{P(1|t)\,.\,S_0^2}\right) = (97{,}18\,.\,234{,}09)^2 - (74{,}40\,.\,29{,}16) +$
$2(29{,}16\,.\,234{,}09)\,in\left(\frac{0{,}37\,.\,234{,}09}{0{,}62\,.29{,}16}\right) = 5{,}16281$

Because$a \neq 0$ then a determination is made

$b^2 - 4ac = (41158{,}7)^2 - 4\,(-204{,}93\,.\,5{,}16281) = 1694042815{,}9$

then define,

$$d = \frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{-41158{,}7 \pm \sqrt{41158{,}7^2-4\,(-204{,}93.5{,}16281)}}{2\,(-204{,}93)} = -0{,}0001239$$

Until the roots of the imaginary are obtained. therefore, taken the average from $d = \frac{\bar{x}_0+\bar{x}_1}{2} = 85{,}79$.

And found an insulating knot that is 85.79. knott_1 blocking point ie $\leq 85{,}79$, while node t_2 blocking point ie $> 85{,}79$.

3.3. Classification Tree

By using the QUEST method, the classification tree is obtained as follows:

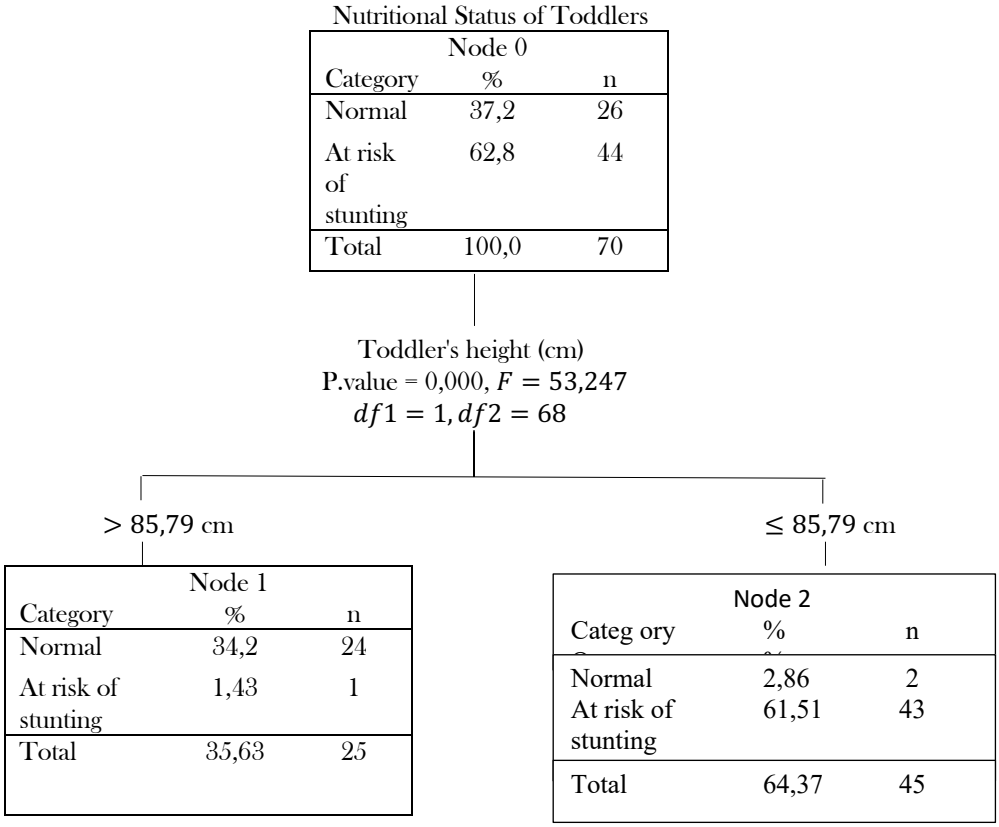

Figure 3.1 QUEST Classification Tree

Information:

1) Observed variables
   Normal category height
   Height category at risk for stunting
2) x_3 = body height
   t_1 (Nodes 1) ie $> 85{,}79$ cm
   t_2 (Nodes 2) ie $\leq 85{,}79$ cm

in the early stages of insulation, the parent node which consists of 70 toddler data. Variables are partitioned based on height into two nodes, namely node (1) and node (2). Node (1) is a node containing 45 children under five with a height of more than 85.79 cm, while node (2) is a node for 25 children under five with a height less than or equal to 85.79 cm. in the next process, the blocking is terminated.

3.3 Classification Tree Accuracy

Classification accuracy consists of *sensitivity, specificity* and accuracy of classification accuracy. Classification accuracy can be calculated based on Table 3.2.

Table 3.2 Accuracy of Classification

Observation	Guess		Total
	Normal	At risk of stunting	
Normal	24	2	26
At risk of stunting	1	43	44
Total	25	45	70

The percentage accuracy of the normal category classification can be calculated as a value*sensitivity,* as follows :

$Sensitivity = \frac{24}{25} \times 100 = 96\ \%$

The percentage of accuracy in the classification of stunting risk categories can be calculated as a value*specificity,* as follows :

$Specificity = \frac{43}{45} \times 100 = 95{,}5\ \%$

The accuracy of the overall classification accuracy can be calculated as follows:

$APER = \frac{24+43}{70} \times 100 = 95{,}7\ \%$

In summary, the overall value of the classification accuracy of the trees formed is 95.7%. Thus, the probability of misclassification of the tree is 4.3%, which means that this classification tree is optimal.

4. CONCLUSION

Classification which is part of data mining can make decisions on the nutritional status of toddlers faster and more efficiently. The QUEST (Quick, Unbiased, Efficient, Statistical Trees) method is a statistical method that can be used to form decision trees and classify an object using a separator algorithm that produces a binary tree.

From the results of the classification there is a variable height (x_3) as the limiting variable. in the early stages of insulation, the parent node which consists of 70 toddler data. Variables are partitioned based on height into two nodes, namely node (1) and node (2). Node (1) is a node containing 25 children under five with a height of more than 85.79 cm, while node (2) is a node for 45 children under five with a height less than or equal to 85.79 cm. in the next process, the blocking is terminated. The overall value of the accuracy of the classification of trees formed is 95.7%. Thus, the probability of misclassification of the tree is 4.3%, which means that this classification tree is optimal.

REFERENCES

[1] Adzani, wildan azka. (2021).*Classification of Toddler Nutritional Status Using Backpropagation Method with Levenberg-Marquardt Algorithm and Nguyen Widrow Initialization*. *12*.

[2] Akbar, F., & et al. (2021).*Strategies to Reduce the Prevalence of Malnutrition in Toddlers* (1st ed.; Syamsidar, ed.). Yogyakarta: Deepublish Publisher (CV BUDI UTAMA Publishing Group.

[3] Come on. (2021). Efforts to Prevent Malnutrition in Toddlers.*Indonesian Ministry of Health*, *11*, 52–57.

[4] Baihaki, E. S. (2017).*Malnutrition in an Islamic Perspective: Theological Responses to the Problem of Malnutrition*. *2*(2).

[5] Candra, A. (2020). *Nutrition Status Check*. Semarang: Faculty of Medicine, Diponegoro University.

[6] Fikri, H., Rahmi HG, I., & Devianto, D. (2017).*APPLICATION OF THE QUEST METHOD IN HOUSEHOLD POVERTY*. *WE*(2), 25–33.

[7] Handayani, S. R., & DKK. (2013).*Application of the QUEST Algorithm (Quick, Unbiased, Efficient, Statistical Tree) in the Establishment of a Classification Tree of Public Interest in Purwakarta Regency for Sharia Banks*.

[8] Ispriyanti, D., Prahutama, A., & Mustafid. (2019). Poverty Classification Analysis in Semarang City Using the Quest Algorithm,*Journal Statistics*, *7*(1).

[9] Indonesian Ministry of Health. (2017).*Public health news; nutrition investment for the nation's future*. Jakarta.

[10] Health, J., Sai, M., Islamiyati, I., Sadiman, S., Midwifery, J., Ministry of Health, P., & Karang, T. (2018).*Family Psychosocial Stimulation by Parents on the Development of Children Aged 48-60 Months*. *11*(2), 65–70.

[11] Maharani, A., & et al. (2017).*IMPLEMENTATION OF BINARY CLASSIFICATION TREE USING QUICK, UNBIASED, EFFICIENT STATISTICAL TREE (QUEST) ALGORITHM ON DATA OF MARCH ELEVEN UNIVERSITY TRANSFER STUDENTS*. UNS (Eleven March University).

[12] Mariam, S. (2022).*Analysis of the factors that influence the death of Covid-19 patients using a classification with a binary tree structure and a quest algorithm*. *2*(1), 32–41.

[13] Muhammad Rosyid Abdurrahman, Dwi Ispriyanti, A.P. (2014).*ESTABLISHMENT OF BINARY CLASSIFICATION TREE WITH QUEST ALGOORHYTHM (QUICK, UNBIASED, AND EFFICIENT STATISTICAL TREE) ON LIVER PATIENT DATA*. *3*, 731–739.

[14] Nofriansyah, D., & Nurcahyo, G. W. (2015).*Data Mining Algorithm And Testing* (1st ed.). Yogyakarta: Deepublish Publisher (CV BUDI UTAMA Publishing Group.

[15] Novela, V., & Apriza, C. (2021).*Digital Book - Basic Science of Public Health*. Bandung: CV. INDONESIAN SCIENCE MEDIA.

[16] Purhanta. (2019). DATA COLLECTION AND RESEARCH INSTRUMENTS Research.*Uin-Malang*, 1–7.

[17] Rahayu, A., & et al. (2018).*ONLINE GROUP GUIDANCE: UTILIZATION OF TECHNOLOGY IN GROUP GUIDANCE*.

[18] Rahmasari, Y., Anggraeni, S., Rahman, E., Studies, P., Society, K., & Society, F. K. (2022).*FACTORS RELATED TO THE NUTRITIONAL STATUS OF CHILDREN IN CEMAKA PUTIH HEALTH CENTER, BANJARMASIN CITY, 2022*.

[19] Rapar, V. L., Rompas, S., Ismanto, A. Y., Studies, P., Nursing, I., Medicine, F., ... Nutrition, S. (2014).*The relationship between maternal parenting and the nutritional status of toddlers in the working area of the Ranotana Weru Health Center, Wanea District, Manado City*. 1–7.

[20] Rizki, V. D., & Setyawan, Y. (2018). Application of the Quick, Unbiased, Efficient Statistical Trees (Quest) Method to Determine the Factors Affecting Diarrhea in Toddlers in Indonesia,*Journal of Industrial and Computational Statistics*, *3*(1), 1–10.

[21] Rokach, L., & Maimon, O. (2008).*Data Mining With Decision Trees*. Singapore: World Scientific Publishing.

[22] Setiawan, R., & Triayudi, A. (2022).*Classification of Nutritional Status of Toddlers Using Naïve Bayes and Web-Based K-Nearest Neighbor*. *6*(2), 777–785. https://doi.org/10.30865/mib.v6i2.3566

[23] Setyawati, V. A. V., & Hartini, E. (2018).*Textbook of Basic Science of Public Health Nutrition*.

[24] SSGI Ministry of Health RI. (2023).*Hasil Survei Status Gizi Indonesia (SSGI) 2022*.

[25] Sugiyono. (2017).*Educational Research Methods (Quantitative, Qualitative, and R&D Approaches) 25th printing*. Bandung: Alphabeta, Cv.

[26] Sumarjaya, I. W., Ayu, I. G., & Srinadi, M. (2015).*QUEST AND CHAID METHODS ON CLASSIFICATION*. *4*(November), 163–168.