Application of Machine Learning in Predicting Children's Nutritional Status with Multiple Linear Regression Models

Mutamimul Ula, Ananda Faridhatul Ulva, Mauliza, Muhammad Abdullah Ali, Yumna Rilasmi Said

Author Affiliation
1 Universitas Malikussaleh Aceh Utara

Author Email
a) Corresponding: mutamimul@unimal.ac.id

Keywords: Machine Learning, linear Regression, Child Nutrition, Forecasting

ABSTRACT
Forecasting is an important part of making plans and making decisions that can predict future events. Forecasting techniques in this study used multiple linear regression. This study aims to predict the number of cases of child nutritional status in children in each region. The purpose of this study was to see the results of predicting the number of children's nutritional status in each region and to make it easier to predict children's nutrition. The research method includes the analysis of the system built and the design of machine learning applications using the Multiple Linear Regression method. Then the system built can help predict the nutritional status of children in Aceh quickly, precisely, and accurately. The data used is data on the nutritional status of children in 2018, 2019, and 2020. Based on the results of forecasting for 2021 based on data obtained in previous years, the predicted results of total nutritional status in 2021 are 449,091. The results of this study indicate that the linear regression method obtains the best model results by being able to predict the implementation of machine learning.

1. Introduction

Nutritional problems are one of the main problems in human life. Indonesia is a country that has a dense population and very diverse citizens. Therefore, Indonesia is faced with the problem of malnutrition. When viewed from statistical data, there are still many problems that need to be resolved, especially those concerning the problem of under-five malnutrition. Gradually, Indonesia has actually been able to reduce the prevalence of malnourished children under five. Nutritional status in children will affect the physical growth of children. The causes include nutritional factors, exclusive breastfeeding, education level, the mother's knowledge and history of LBW. Support for good nutritional intake for children requires the mother's ability to provide care for the child. The relationship between maternal parenting patterns and the incidence of nutritional status of children aged 12–59 months was examined.

The achievement of weight for age is a sign that the nutritional status of children has succeeded in meeting these needs. The goal of reducing the nutritional status of children under five has not been too substantial in recent years. This is a challenge to complete nutrition in children and is very difficult to detect in adults. The prevalence of malnutrition and malnutrition in Indonesia in 2018 was 17.7%. One of the factors that affect a person's health is the lack of nutritional intake in children. As a person ages, the body's metabolic rate tends to decrease.

Many factors affect the nutritional status of children. One of them is that socio-economic variables play an important role in determining nutritional status. It is hoped that the nutritional status will be better if the socio-economic conditions support it. The socio-economic conditions of the family (parents), including education level, occupation, number of
children, parenting patterns, and the general financial condition of the family, will have a significant impact on the nutritional health of children under the age of five [6]. Forecasting is critical for understanding how interconnected decision-making is in predicting future events. Forecasting results fluctuate and trends decline, and forecasts can see events in certain months for status up, down, or significantly increased. [7], nutritional status will affect the poor because of the inability to meet the minimum standard of living. This situation can be seen from the low income in meeting basic needs, both in the form of needs that must be fulfilled [8], [9].

School-age children need more nutritional intake, which is necessary for growth and development into adulthood. If exercise is not balanced with food consumption, it can cause nutritional problems. The consumption of good food but the condition of the child is sick or has an infectious disease can cause the child to experience poor nutritional status. [10]. Malnutrition usually occurs in children under the age of 5 years. The occurrence of malnutrition in children under five years old ranges from 12–59 months of age. The nutritional status of these children is an age group that is vulnerable to health and nutrition problems. At this age, the needs increase, while the need for nutritional intake cannot ask for and find food on their own, and often at this age they are no longer taken care of and the management is left to others, so the risk of malnutrition will be even greater. Children who are malnourished will experience a decrease in their immune systems, making them vulnerable to infectious diseases[11]. Forecasting in the development of plants to meet nutritional needs is a need that must always be met for development, where each plant requires different nutrients. Forecasting in supplementing plant nutrient intake is used at the nutritional valve openings every day, so that nutrients can be monitored using a forecasting system. [12].

The quality of nutrition in toddlers greatly affects the future. this is a problem and concern for the parties concerned [13]. The government must immediately evaluate the quality of nutrition in every region in Indonesia. One way to find out the extent to which the nutritional quality of toddlers is affected by grouping and classification with machine learning. [14]. The extent to which nutritional quality in Indonesia and for the purposes of government evaluation is known from the processed data. From the data obtained and the clustering, it was found that the government does not need to study in depth the improvement of nutrition in malnutrition areas [15]. Next is the need for forecasting to be able to predict the nutritional needs of each region are met or not so that there are no provinces / regions that lack nutritional intake.

2. Literature Review

Some of these factors have a negative effect and some have a positive effect on the number of malnourished children in Central Java Province. GRDP and the poor have a positive effect on increasing the value of malnutrition. While the population and per capita expenditure have a negative effect on the value of malnutrition [6]. Further research on parenting patterns in 2019 found that the status of children under five had an effect on the pattern of the main caregiver for children, the number of family members and the pattern of health care for children under five, related to the nutritional status of children under five. The number of family members is related to the distribution of nutritional intake in the family. Parenting is related to daily child care in a healthy environment that supports the fulfillment of child nutrition. The pattern of health care for toddlers, especially when sick and afterward, affects the recovery of the child’s body. These results further confirm that the nutritional status of children is largely determined by the quality of family care. [16]

Further research, seen from the results of multivariate analysis, found 2 (two) risk factor variables that were significantly related to the incidence of nutritional status, namely the mother’s attitude towards bad food with an OR of 6.98 and poor environmental sanitation with an OR of 5.03. Based on the results of multivariate analysis). Variable risk factors that are not related to the incidence of malnutrition and malnutrition are not given exclusive breastfeeding, low energy intake, low protein intake, ARI three times in the last two months, ARI diarrhea three times in the last two months, number of children greater than two, education level, low motherhood, low family income, infrequent access to health services, and TV viewing frequency greater than two hours per day [17]. Malnutrition or malnutrition is still an important problem in children under the age of five. In addition, maternal, socioeconomic, and environmental factors all play a role in childhood malnutrition [18].

The main cause of malnutrition in children is poverty. Specific and targeted interventions are needed to address this and should include efforts to prevent low birth weight and reduce health inequalities and require clustering for each region. [19]. Clinical control and growth monitoring should also be encouraged. However, further studies are needed on the timing and use of information on growth disorders to prevent severe forms of malnutrition in the future [20]. Prediction serves to make plans for needs that must be met which are expressed in quantity or amount as a function of time. Prediction is needed by making comparisons between forecasted needs [21].

Malnutrition or malnutrition One method to analyze forecasts or predictions is to use the Multiple Linear Regression model. Regression analysis is a statistical analysis that utilizes the relationship between two or more variables. Regression analysis is a statistical analysis that utilizes the relationship between two or more variables. Multiple regression analysis is a method for predicting the value of the influence of two or more independent variables on one dependent variable. Nutrition is still an important problem in children under the age of five. In addition, maternal, socioeconomic, and environmental factors all play a role in childhood malnutrition [22] [23]. In this study, the multiple linear regression method was used to predict what factors influence malnutrition in children.
3. Method

A. Research Stages
The research stages of Machine Learning Application in Predicting Children's Nutritional Status with Multiple Linear Regression Model are:
1. Problem Identification: The initial stage is to identify problems regarding the nutritional status of children to collect data in each region.
2. A literature study is a literature review related to the problem to be studied, such as theoretical references that are relevant to the research conducted.
3. Theoretical Foundation: taking references and related materials about children's nutritional status and using them as variables/symptoms in hybrid machine learning
4. Data Collection: we collect all data for forecasting and as initial data, which is then used as training data.
5. Data Processing: Processing data to be used as variables to be processed into hybrid machine learning in the form of data for prediction.
6. Design Phase: At this stage, the prototype system is used to prepare for the program implementation phase.
7. Implementation Stage: This stage is in the form of coding and connecting the database server with the prototype that has been built.
8. Testing Stage: This stage performs testing on each menu from the prototype and implementation. This stage also tests the hybrid machine learning model in classification and prediction.

B. System Development Process
The process of developing the Machine Learning Application system in Predicting Children's Nutritional Status with Multiple Linear Regression Models, namely:

C. Forecasting Research Flow
The research flow of the Application of Machine Learning in Predicting Children's Nutritional Status with Multiple Linear Regression Models is as follows:
4. Results and Discussion

A. System Analyst

The use of weight and height is an important parameter in determining the health status of humans, especially those related to nutritional status. The use of the BW/U and BW/U indexes is an indicator of nutritional status to see any disturbances in growth function and body composition, such as being short, tall, thin, or fat. Then the system will process the data with a machine learning process, and finally the system will continue with the forecasting process. This research is an applied technological innovation in the application of machine learning for monitoring the nutritional status of children. The purpose of the machine learning model in classifying malnutrition in children and being able to monitor areas affected by malnutrition is to be able to predict areas that are prone to malnutrition in children. This is to make it easier for governments, health services, and hospitals to take quick action with the implementation of machine learning. In the application of the Multiple Linear Regression method for cases of child nutritional status, it can be completed with the following calculation steps:

\[
\begin{align*}
X_1 &= \text{Total Malnutrition} \\
X_2 &= \text{Amount of Malnutrition} \\
Y &= \text{Result}
\end{align*}
\]

B. Initial Data Analysis

The data for the initial analysis of the application of machine learning in predicting the nutritional status of children with multiple linear regression models are as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>176</td>
<td>199</td>
<td>188</td>
</tr>
<tr>
<td>2019</td>
<td>196</td>
<td>278</td>
<td>221</td>
</tr>
<tr>
<td>2020</td>
<td>187</td>
<td>358</td>
<td>337</td>
</tr>
<tr>
<td>2021</td>
<td>298</td>
<td>489</td>
<td>453</td>
</tr>
<tr>
<td>Count</td>
<td>857</td>
<td>1.324</td>
<td>1.199</td>
</tr>
</tbody>
</table>

Calculation of each variable using machine learning in predicting children's nutritional status using multiple linear regression models

<table>
<thead>
<tr>
<th>Year</th>
<th>(X_1^2)</th>
<th>(X_2^2)</th>
<th>(X_1 \times X_2)</th>
<th>(X_1 \times Y)</th>
<th>(X_2 \times Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>30.976</td>
<td>39.601</td>
<td>35.024</td>
<td>33.088</td>
<td>37.412</td>
</tr>
<tr>
<td>2019</td>
<td>38.416</td>
<td>77.284</td>
<td>54.488</td>
<td>43.316</td>
<td>61.438</td>
</tr>
<tr>
<td>2020</td>
<td>34.969</td>
<td>128.164</td>
<td>66.946</td>
<td>63.019</td>
<td>120.646</td>
</tr>
<tr>
<td>2021</td>
<td>88.804</td>
<td>239.121</td>
<td>145.722</td>
<td>134.994</td>
<td>221.517</td>
</tr>
<tr>
<td>Count</td>
<td>193.165</td>
<td>484.170</td>
<td>302.180</td>
<td>274.417</td>
<td>441.013</td>
</tr>
</tbody>
</table>

To obtain the regression coefficient \(a, b_1,\) and \(b_2\) can be obtained simultaneously from the following three equations:

\[
\begin{align*}
4a + b_1 (857) + b_2 (1.324) &= 1.199 \\
857a + b_1 (193.165) + b_2 (302.180) &= 274.417 \\
1.324a + b_1 (302.180) + b_2 (484.170) &= 441.013
\end{align*}
\]

Table 3. Koefesien regresi

<table>
<thead>
<tr>
<th>n</th>
<th>(\Sigma x_1^2)</th>
<th>(\Sigma x_2^2)</th>
<th>(\Sigma x_1 y)</th>
<th>(\Sigma x_2 y)</th>
<th>(\Sigma x_1 x_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>9552.75</td>
<td>45926</td>
<td>17531.25</td>
<td>44144</td>
<td>18513</td>
</tr>
</tbody>
</table>

Then the three equations above are solved to obtain the value of \(a, b_1,\) dan \(b_2.\) Then the three equations above are solved to obtain the value of:

1. Equations 1 and 2 are eliminated

\[
\begin{align*}
4a + b_1 (857) + b_2 (1.324) &= 1.199 \\
857a + b_1 (193.165) + b_2 (302.180) &= 274.417
\end{align*}
\]

\[
\begin{align*}
3.428a + 734.449b_1 + 1.134.668b_2 &= 1.027.543 \\
3.428a + 772.660b_1 + 1.208.720b_2 &= 1.097.668
\end{align*}
\]
2. Enter the value of into equation 4 by substituting.

\[-38.211 \, b_1 + 74.052 \, b_2 = -70.125\]  
\[b_1 = \frac{-70.125}{38.211}\]  
\[b_1 = -0.12603274\]  
\[b_2 = \frac{70.125 - 74.940}{4.815}\]  
\[b_2 = -0.14965887\]

Table 4. Equation Result

<table>
<thead>
<tr>
<th>b1</th>
<th>b2</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.12603274</td>
<td>-0.14965887</td>
<td>-8.220437667</td>
</tr>
</tbody>
</table>

2. Then enter the values of b_1 and b_2 into equation 1 by substituting.

\[4a + b_1 (857) + b_2 (1.324) = 1.199\]  
\[4a + 0.12603274 \times 857 + 1.012002877 \times 1.324 = 1.199\]  
\[4a = 1.199 - (0.12603274 \times 857 + 1.012002877 \times 1.324)\]  
\[4a = 1.199 - 1.23181750968\]  
\[a = \frac{1.199 - 1.23181750968}{4}\]  
\[a = -8.220437667\]

From the search above, it has been found that the values a, b_1, b_2, where the values are as follows:

a = -8.220437667
b_1 = -0.12603274
b_2 = 1.012002877

Forecasting

\[Y = a + b_1 X_1 + b_2 X_2\]  
\[Y = -8.220437667 + (-0.12603274 \times 298) + (1.012002877 \times 489)\]  
\[Y = -8.220437667 - 37.5578 + 494.8694\]  
\[Y = 449.0912126\]

Based on the data obtained in previous years, the results of the prediction of the number of nutritional status in 2021 are as much as 449,091,212.

4.4 Implementation of regresi multiple linear

The application of machine learning in predicting the nutritional status of children with multiple linear regression models is as follows:

Figure 3. The results of the implementation of the Multiple Regression method
5 Conclusion

The conclusions of the application of machine learning in predicting the nutritional status of children with multiple linear regression models are as follows:
1. Can find out the process of predicting the nutritional status of children with multiple linear regression models
2. Can assist the Health Office and Doctors in seeing the nutritional status of children with multiple linear regression models
3. The implementation of multiple linear regression predictions 2021 X1 is 298, X2 is 489 and Y 453. The results of the forecasting are based on the data obtained in previous years, the results of the prediction of the number of nutritional status in 2021 are 449,091216.

Acknowledgement

In future research, the forecasting data in each region is more complete for better results in prediction and . application of machine learning in predicting nutritional status with other models to see which method is the best stage

References

