Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 2 (2023) 201 - 207 ISSN Media Electronic: 2580-0760

Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City

Endah Ratna Arumi[1], Sumarno Adi Subrata[2], Anisa Rahmawati[3]
[1,3]Informatics Engineering, Engineering, Universitas Muhammadiyah Magelang
[2]Nursing Science, Health Sciences, Universitas Muhammadiyah Magelang
[1]arumi@ummgl.ac.id, [2]adisubrata@ummgl.ac.id, [3]anisarahmawati.kokid@gmail.com

Abstract

Nutritional status is an important factor in assessing the growth and development rate of babies and toddlers. Cases of malnutrition are increasing, especially in magelang city. Because nutritional problems (Malnutrition) can affect the health of toddlers. Therefore, this study aims to predict the level of prevalence of malnutrition with the Naïve Bayes method. This research uses an observational design, a single center study at the Magelang City Office, using the Naïve bayes method which is used as an application of time series data, and is most widely used for prediction, especially in data sets that have many categorical or nominal type attributes. The Naïve bayes method is used to predict such cases of malnutrition. The results of this study show that the Naïve Bayes method succeeded in predicting the magnitude of cases of malnourished toddlers in Magelang City with an accuracy percentage of 75% due to the very minimal amount of training data, and the areas that have the most malnutrition are in three areas, namely Magersari, North Tidar and Panjang.

Keywords: *Malnutrition in a child; information technology; predictor; naïve bayes; health problems*

1. Introduction

The phenomenon of malnutrition that occurs in Indonesia is triggered by an increase or low quality of human resource health (HR) in general caused by unbalanced food consumption characterized by physical and psychological disruption of children. Lack of nutritional needs will lead to malnutrition leading to the highest mortality in developing countries, so currently malnutrition is a national problem that needs regular review [1]–[4]. Nutritional status is included in the important factor in assessing the growth and development rate of babies and toddlers [4]. Nutritional problems (Malnutrition), have adverse effects on the health conditions of infants, toddlers, and women of productive age, because they are caused by several factors including food availability, infectious diseases, lack of knowledge about parenting, difficulty in accessing environmental health, healthy living behaviors, the environment, and poverty [2].

Based on the results of the 2021 Indonesian Nutritional Status Survey (SSGI) conducted by the Ministry of Health, the prevalence rate of stunting in Indonesia in 2021 was 24.4%, or decreased by 6.4% from 30.8% in 2018 [5]. Therefore, the Government is trying to reduce the prevalence of malnutrition by monitoring nutritional status periodically once a year, Nutritional Status Monitoring has been carried out in 34 provinces and 514 regencies/cities [3]. The National Medium-Term Development Plan (RPJMN) in 2021 of 24.4% of toddlers experiencing nutritional problems and is targeted to decrease to 14% by 2024 [5]. There are several ways that can be used as a way to assess nutritional status in children, namely by taking body measurements, including weight, height, upper arm circumference, head circumference, chest circumference, and subcutaneous fat layer [4]. This body measurement is often referred to as Anthropometry.

Toddlers with malnutrition status based on the correlation of body weight with age and using references based on the WHO NCHS book [6]. Toddlers experience malnutrition usually caused by lack of protein energy intake consumed every day for a long time [7].

The current Health Data owned by the Ministry of Health of the Republic of Indonesia amounts to 270 million people who have medical record records stored on individual-based Health data and managed by many applications [8]. Health data prediction and integration

Accepted: 21-08-2022 | Received in revised: 05-02-2023 | Published: 03-03-2023

can support strategic decision-making in the health sector.

There are many predictive processing methods that can be applied, with C4.5, KNN, and Naïve Bayes algorithms. This study used Naïve Bayes, because it can take into account the chances of results and the accuracy of information with a small amount of data [9]. The Naive Bayes classifier in machine learning is used as an application of time series data, and is most widely used for prediction especially in data sets that have many categorical or nominal type attributes [9].

Several studies on predictions in the field of health and prediction of malnutrition status using Machine Learning-based algorithms have been widely published in the literature. Research conducted by M.M. Islam, et al in 2022, resulted in the detection and prediction of Malnourished women with potential risk factors for Malnutrition extracted using multinomial logistic regression (MLR), with results showing that 9 risk factors for thin women and 11 factors for overweight/obese women. These selected factors are used as input features in the classification of predictions of malnourished women with Random Forest (RF). Experimental results show that the RF-based MLR-RF-based classifier system provides an accuracy of 81.4% and 0.837 AUC for underweight and an accuracy of 82.4% and 0.853 AUC for overweight/obese [10].

Research on the prediction of nutritional status of toddlers by Wulandari in 2018, it can be concluded that with the Tsukamoto and Mamdani methods, the data processing criteria that have been used are Age Weight (T / T), Height (T / T), Weight Height (BB / T). The Tsukamoto method uses the implications of MIN on the inference evaluation engine and the Centralized Average for Defuzzification. The measurement results show that the Fuzzy Tsukamoto method obtained 82.35% from Fuzzy Mamdani [3]. The machine learning method researched by Afifaturahman in 2021 for prediction uses a comparison of the K-Nearest Neighbour (KNN) and Naive Bayes algorithms, the Naïve bayes method to measure the optimal level of accuracy with fewer training datasets, while the KNN algorithm can group datasets with higher-feature space [11].

Research by Ramanathan in 2022 by predicting type 2 diabetes with a hybrid system can integrate the Naïve bayes classification and the Fuzzy Hybrid System, this Fuzzy method is used as a way to minimize the rule size in the dataset before data classification is carried out to predict the risk level of type 2 diabetes patients with a certain range (0-10), with an accuracy value of 90.26% when tested using the Pima diabetes dataset [12]. One method for improving the analysis of cancer prognosis prediction with the Artificial Algae Algorithm (AAA) method, namely by combining multidimensional features through an early and late fusion approach and extracting resilient features to select effective features from several data modalities, and the results are obtained more accurately with faster processing time [13]. In a study by Thakkar in 2021 produced diabetes predictions with the Data Mining method to classify, with the Fuzzy Logic method for data mining and the Random Forest method as a classification process [14].

Currently, stunting cases occur a lot, ranging from East Nusa Tenggara, West Sulawesi, Aceh, West Kalimantan, South Kalimantan, Central Sulawesi, and West Nusa Tenggara which have a high prevalence, it is also necessary to pay attention to areas that have a large number of stunted children, such as in West Java, East Java, Central Java, Banten and North Sumatra. These areas that need attention [5]. Magelang City is one of the regions in Central Java so it requires special attention in handling 10 percent of stunting cases, but these results still need to be looked at again and are waiting for the results of the Indonesian nutritional status survey (SSGI) which has not been released [15].

Based on prediction references in the health sector, this article discusses the prediction of the rate of spread of toddlers experiencing malnutrition status in Magelang City using the Naïve bayes method with a time span of 2017 – 2021. The Naïve bayes method was chosen to perform probability classification as a continuous value evaluation and prediction step associated with each feature containing numerical values. The Naïve Bayes method will perform predictions with Anthropometric factors, so that it can analyze the accuracy and validation values obtained for each stunting prediction region.

This study aims to make parents pay attention to the growth and development of children, especially at the age of toddlers by paying attention to the nutrition consumed by children. In accordance with the policy of the Magelang City Health Office (Dinkes), namely policies in the Health Sector regarding disease prevention and control, as well as health resources. So this research is expected to help the Magelang City Health Office to find out how much the distribution of malnourished toddlers is so that they can control and provide socialization to parents of toddlers to reduce the percentage of distribution.

2. Research Methods

The stages of the research method to solve the problem of predicting malnutrition status in Magelang City are based on the Anthropometric Measurement Indicators in Table 1, as the main requirement in measuring nutritional adequacy and child growth and development using the Naïve bayes method.

DOI: https://doi.org/10.29207/resti.v7i2.4438

Table 1. Categories and Thresholds for Child Nutritional Status

Index	Category Nutritional Status	Threshold (Z-Score)
Weight Loss by Age (BB/U) children aged 0-60 months	Very underweight	<- 3 SD
	Underweight	-3 Sd s.d <-2 SD
	Normal weight	-2 SD s.d +1 SD
	Overweight	>+1 SD
Length or Height by Age (PB/you or TB/U) of children aged 0-60 months	Very short	<-3 SD
	Short	-3 Sd s.d <-2 SD
	Usual	-2 SD s.d +3 SD
	Tall	>+3 SD
Weight by Length or Height (BB/PB or BB/TB) children aged 0-60 months	Malnutrition	<-3 SD
	Malnutrition	-3 SD s.d <-2 SD
	Good nutrition	-2 SD s.d +1 SD
	More nutritional risks	>+2 SD s.d +2 SD
	More nutrition	>+2 SD s.d +3 SD
	Obesity	>+3 SD
Body Mass Index by Age (BMI/U) of children aged 0-60 months	Malnutrition	<-3 SD
	Malnutrition	-3 SD s.d <-2 SD
	Good nutrition	-2 SD s.d +1 SD
	More nutritional risks	>+1 SD s.d +2 SD
	More nutrition	>+2 SD s.d +3 SD
	Obesity	>+3 SD
Body Mass Index by Age (BMI/U) of children aged 5-18 years	Malnutrition	<- 3 SD
	Malnutrition	-3 SD s.d <-2 SD
	Good nutrition	-2 Sd s.d +1 SD
	More nutrition	>+1 SD s.d +2 SD
	Obesity	>+2 SD

The stages of research carried out are in figure 1.

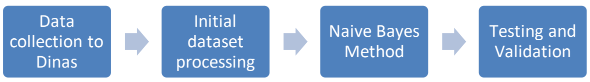

Figure 1. Stages of research

2.1 Data Collection

The method of collecting datasets that were processed and calculated with this study was obtained from open data in Magelang City[16], with the title Toddlers Weighing and Nutritional Status of Toddlers in Magelang City, in the time span from 2017 to 2021 there were 3 sub-districts and each sub-district was affected by several villages with a total of 17 villages in Magelang City. The total attributes in the data obtained are 11 attributes consisting of Subdistrict / Kelurahan, Toddlers Weighed by Men, Toddlers Weighed by Women, Good Nutritional Status of Men, Good Nutritional Status of Women, Malnutrition Status of Men, Malnutrition Status of Women, Nutritional Status Very Less Men, Nutritional Status Very Less Women, Nutritional Status More Men, Nutritional Status More Women, but in this study only 6 attributes were selected that are related to the calculation of malnutrition predictions is in accordance with Anthropometric measurements, and the final target of this study is related to the areas affected by the malnutrition. The dataset of 6 attributes used includes village, year, toddlers weighed by men, toddlers weighed by women, nutritional status less men and malnutrition status of women. In this study, it used the attributes of kelurahan because it was to be more specific to go to the village and for the health office to focus more on handling cases of malnutrition in toddlers in each village that has high cases.

2.2 Initial Dataset Processing

At this stage, researchers carry out data cleaning (pre-processing). In this process, it is like choosing the data to be used. This stage is done manually, namely by using excel by selecting the data used. After sorting out the attributes to be used, then combine 5 datasets into 1 dataset containing data that is clean and ready to be used for the calculation process. The dataset processed in this study is consistent data. The database is divided into two, namely training data and test data, henceforth if the data has been pre-processed, then the next stage is the implementation of the Naïve bayes method for prediction.

2.3 Naïve Bayes

Naïve bayes belongs to supervised learning algorithms that can be used in making predictions using probability theory with Bayes Theorem and statistical methods [17], [18]. This method uses probability classification with the Bayes Theorem which is based on strong conditionally independent assumptions between all dataset variables, with class y labels with attributes X={X1, X2, X3, Xd}[12], [19]. The Naïve bayes algorithm is implemented with low complexity, since it does not require a lot of training data, and does not require model optimization.

The stages of the Naïve bayes algorithm are[19]: The first stage is preparing the dataset for training data. The next stage calculates the average value (mean) of each attribute with the entire number of classes in the dataset. The mean formula is in equation 1.

$$\mu = \frac{\sum_{i=1}^{n} xi}{n} \qquad (1)$$

μ is an average (mean) of the population, xi is midpoint of value in an attribute, and n is an amount of data.

The next stage calculates the standard deviation of each attribute with the entire number of classes present in the dataset. The standard deviation formula is in equation 2.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(xi-\mu)^2}{n}} \qquad (2)$$

σ is a variety variants for the population, μ is an average (mean) of the population, xi is a midpoint of value in an attribute, and n is an amount of data.

The processed dataset must be represented as an n-dimensional vector, namely X=X1, X2, X3, … Xn. The stages of the classification process use a probabilistic model for attribute X with equation 3.

$$P(Ci|X) = \frac{P(X|Ci).\, P(Ci)}{P(X)} \qquad (3)$$

DOI: https://doi.org/10.29207/resti.v7i2.4438

P(C|X) is an odds attribute X is classified as class C, P(C) is the opportunity for the appearance of class C on all training data, P(X) is the chance of the appearance of attribute X on all training data, and P (X|C) is the chance of the appearance of attribute X in class C.

2.4 Testing and Validation

The test conducted on the prediction system with confusion matrix aims to find out the results of accuracy, recall, and precision from processing malnutrition prediction data in the city of Magelang. In data validation, training data used datasets on malnutrition in 2017 – 2020, and testing data using 2021 were used. After the prediction process is carried out, the accuracy results of the predictions will be continued at the validation stage.

This confusion matrix is divided into two classes, this is used to describe the performance of the classification method on the actual class (Figure 2) [20]. The data in the Confusion matrix shows the number of class predictions that correspond to the actual class. Furthermore, the prediction model will be calculated using an accuracy formula to see the optimal model. The accuracy calculation formula can be seen in equation 4.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

TP being True Positive is the positive number of the predicted value, while TN is True Negative containing the number of negative prediction results. This flow will be repeated until accurate prediction results are obtained, the results are presented in accuracy, and continued with the precision calculation stage in equation 5, to measure the optimality of the prediction model, as a calculation of the ratio of relevant items based on selected items. The step will be repeated until you get the most accurate and optimal result.

$$precision = \frac{TP}{TP+FP} \quad (5)$$

Furthermore, a recall calculation is carried out in equation 6, to find the ratio of the relevant item to the total number of relevant items available. The results of precision and recall calculations are measured by percentages between 1-100% or with numbers 0-1. The results of a prediction system can be considered good if the precision and recall values are high.

$$recall = \frac{TP}{TP+FN} \quad (6)$$

The accuracy rate of the predictions, namely [20]**:**

Accuracy Range 0.9 – 1 = excellent classification, Accuracy Range 0.8 – 0.9 = good classification, Accuracy Range 0.7 – 0.8 = fair classification, Accuracy Range 0.6 – 0.7 = poor classification, Accuracy Range 0.5 – 0.6 = failure.

3. Results and Discussions

The initial stage of implementing the classification method with Naïve bayes is by preparing Training data, and in this study using data from 2017 to 2021 as training data, and 2021 data as testing data. The data is processed as shown in table 2.

Table 2. Data Training

Kelurahan	2017				2018				2019				2020				2021			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Rejowinangun Selatan	204	193	21	12	197	187	8	9	191	195	12	12	191	195	12	12	178	188	11	9
Magersari	200	201	12	10	221	199	11	14	198	184	13	12	198	184	13	12	132	137	11	16
Jurangombo Utara	88	98	9	5	110	119	4	5	92	95	4	2	92	95	4	2	58	53	6	6
Jurangombo Selatan	159	143	11	3	190	170	9	6	190	170	9	6	190	170	9	6	113	88	7	6
Tidar Utara	223	235	11	16	216	237	14	12	225	217	16	11	225	217	16	11	221	209	23	28
Tidar Selatan	132	110	11	7	122	123	11	6	119	123	7	6	119	123	7	6	110	144	18	12
Wates	260	190	9	4	202	202	11	11	190	190	5	4	190	190	5	4	158	183	15	11
Potrobangsan	161	147	6	4	180	170	5	5	192	151	4	5	192	151	4	5	65	79	10	8
Kedungsari	191	146	22	10	185	161	5	8	183	168	6	4	183	168	6	4	134	135	25	17
Kramat Utara	105	103	11	1	96	118	4	6	81	78	3	5	81	78	3	5	58	60	3	6
Kramat Selatan	187	176	10	12	185	168	11	12	164	162	5	6	164	162	5	6	135	142	14	8
Kemirirejo	99	86	2	1	115	76	5	3	120	72	5	2	120	72	5	2	36	28	7	9
Cacaban	139	144	2	3	161	159	13	10	151	137	1	3	151	137	1	3	90	77	7	6
Rejowinangun Utara	232	220	17	16	221	238	19	22	212	189	13	19	212	189	13	19	202	198	16	17
Magelang	140	156	10	10	140	156	4	8	146	131	7	7	146	131	7	7	138	124	7	5
Panjang	137	124	12	10	123	100	11	13	109	116	9	9	109	116	9	9	79	104	12	9
Gelangan	183	185	7	9	183	185	5	9	152	188	3	7	152	188	3	7	154	176	13	14

DOI: https://doi.org/10.29207/resti.v7i2.4438

A is Toddler weighed male, B is Toddler weighed female, C is Malnutrition status of men, and D is Malnutrition status of women

The training data obtained in this study is in the form of numerical data, then the next stage is to find the mean value according to equation 1, and the standard deviation in equation 2 (Table 3).

Table 3. Average value result (mean)

Kelurahan	Years	Toddler weighed male	Toddler weighed female	Malnutrition status of men	Malnutrition status of women
Rejowinangun Selatan	2019	192,2	191,6	12,8	10,8
Magersari	2019	189,8	181	12	12,8
Jurangombo Utara	2019	88	92	5,4	4
Jurangombo Selatan	2019	168,4	148,2	9	5,4
Tidar Utara	2019	222	223	16	15,6
Tidar Selatan	2019	120,4	124,6	10,8	7,4
Wates	2019	200	191	9	6,8
Potrobangsan	2019	158	139,6	5,8	5,4
Kedungsari	2019	175,2	155,6	12,8	8,6
Kramat Utara	2019	84,2	87,4	4,8	4,6
Kramat Selatan	2019	167	162	9	8,8
Kemirirejo	2019	98	66,8	4,8	3,4
Cacaban	2019	138,4	130,8	4,8	5
Rejowinangun Utara	2019	215,8	206,8	15,6	18,6
Magelang	2019	142	139,6	7	7,4
Panjang	2019	111,4	112	10,6	10
Gelangan	2019	164,8	184,4	6,2	9,2

Table 3 is used for 2019 data because the mean value is obtained from the middle value of the total number of training data processed from 2017 to 2021. To determine the mean or average result of each class by summing the values of each attribute then divided by the total attributes, it will produce a mean value according to equation 1.

The next step is to calculate the standard deviation, the calculation results can be seen in table 4, and the standard deviation is calculated using all training data.

Table 4. Standard Deviation Results

Kelurahan	Toddler weighed male	Toddler weighed female	Malnutrition status of men	Malnutrition status of women
Rejowinangun Selatan	8,57	3,44	4,35	1,47
Magersari	30,18	23,14	0,89	2,04
Jurangombo Utara	16,83	21,47	1,96	1,67
Jurangombo Selatan	30,19	31,86	1,26	1,20
Tidar Utara	3,35	11,03	3,95	6,47
Tidar Selatan	7,06	10,93	4,02	2,33
Wates	33,37	6,13	3,79	3,43
Potrobangsan	47,86	31,34	2,23	1,36
Kedungsari	20,81	13,06	8,80	4,80
Kramat Utara	15,99	20,53	3,12	1,85
Kramat Selatan	18,79	11,24	3,52	2,71
Kemirirejo	31,94	20,06	1,60	2,87
Cacaban	25,18	28,07	4,66	2,76
Rejowinangun Utara	10,09	19,28	2,33	2,06
Magelang	3,35	13,63	1,90	1,62
Panjang	19,24	8,76	1,36	1,55
Gelangan	14,88	4,41	3,71	2,56

The last step in the Naïve bayes method is to do probability calculations. The calculation of probability corresponds to equation 3. Based on the calculation of the probability for all villages, the same value was obtained, which was 0.06. The results were taken from 5 villages with 85 total data. Calculating the results of the probability value that has been obtained in the Naïve bayes method, then the data obtained will be tested using WEKA tools to produce a classification of cases of malnourished toddlers in Magelang City as shown in figure 2.

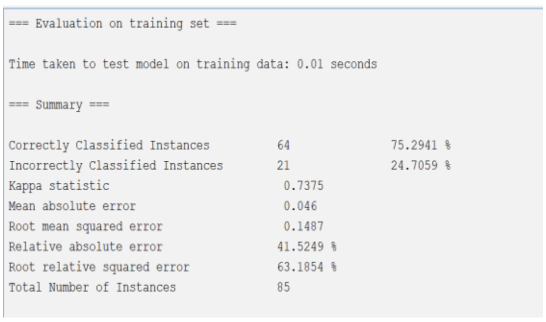

Figure 2. Naive Bayes Method classification test results

DOI: https://doi.org/10.29207/resti.v7i2.4438

Based on the results of classification testing in figure 2, it can be seen that the percentage for Correctly Classified Instances is 75.2941% and Incorrectly Classified Instances is 24.7059%, which means that from 85 data on cases of malnourished toddlers in Magelang City, there are 64 data that are processed and classified precisely and there are 21 data that cannot be processed to be classified. In this process, 85 data are used

3.4 Accuracy Testing and Validation

The accuracy testing step in the malnutrition prediction system in Magelang city is with a Confusion matrix to find out the results of accuracy, recall, and precision, which can be seen in table 5. From the resulting probability results, then testing was carried out on training data with Weka tools on the classification of malnourished toddler cases. The results of this test are in figure 2.

Table 5. Confusion matrix calculation results

Kelurahan	Result *Precision*	Result *Recall*	Result *Accuracy*
Rejowinangun Selatan	1	0.8	0.99
Magersari	0.63	1	0.94
Jurangombo Utara	0.6	0.6	0.95
Jurangombo Selatan	1	0.8	1
Tidar Utara	0.83	1	0.99
Tidar Selatan	0.8	0.8	0.98
Wates	1	0.6	0.98
Potrobangsan	0.67	0.8	0.96
Kedungsari	1	0.4	0.99
Kramat Utara	0.43	0.6	0.92
Kramat Selatan	0.5	0.6	0.94
Kemirirejo	0.67	0.8	0.96
Cacaban	1	0.6	0.98
Rejowinangun Utara	1	0.8	0.99
Magelang	1	0.8	0.99
Panjang	0.83	1	0.99
Gelangan	0.67	0.8	0.96

Accuracy is calculated using the Confusion matrix, in equation 4, so that in Figure 3 it can show the results of the Confusion matrix with the Naïve bayes method.

```
=== Confusion Matrix ===

 a b c d e f g h i j k l m n o p q   <-- classified as
 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 | a =   Rejowinangun Selatan
 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b =   Magersari
 0 0 3 0 0 0 0 0 0 1 0 1 0 0 0 0 0 | c =   Jurangombo Utara
 0 0 1 4 0 0 0 0 0 0 0 0 0 0 0 0 0 | d =   Jurangombo Selatan
 0 0 0 0 5 0 0 0 0 0 0 0 0 0 0 0 0 | e =   Tidar Utara
 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 1 0 | f =   Tidar Selatan
 0 1 0 0 0 0 3 0 0 0 0 0 0 0 0 0 1 | g =   Wates
 0 0 0 0 0 0 0 4 0 1 0 0 0 0 0 0 0 | h =   Potrobangsan
 0 0 0 0 0 0 0 2 2 0 1 0 0 0 0 0 0 | i =   Kedungsari
 0 0 1 0 0 0 0 0 0 3 0 1 0 0 0 0 0 | j =   Kramat Utara
 0 1 0 0 0 1 0 0 0 0 3 0 0 0 0 0 0 | k =   Kramat Selatan
 0 0 0 0 0 0 0 0 0 1 0 4 0 0 0 0 0 | l =   Kemirirejo
 0 0 0 0 0 0 0 0 0 1 1 0 3 0 0 0 0 | m =   Cacaban
 0 0 0 0 1 0 0 0 0 0 0 0 0 4 0 0 0 | n =   Rejowinangun Utara
 0 0 0 0 0 0 0 0 0 0 1 0 0 0 4 0 0 | o =   Magelang
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5 0 | p =   Panjang
 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4 | q =   Gelangan
```

Figure 3. Confusion matrix

Based on the results of the confusion matrix in figure 3, it can be predicted that the areas experiencing malnourished toddlers in Magelang City are in 3 areas, namely Magersari, North Tidar and Panjang. With the predicted data record results right in the area. Furthermore, the percentage of accuracy of the data obtained in equation 4 will be determined, namely the percentage value of accuracy shows the effectiveness of the dataset applied in the Naïve Bayes method which reaches 75%. Low accuracy values can be caused by a relatively small amount of training data so that the calculation is less than optimal, this has also been done in comparative research of the C4.5, KNN, and Naïve Bayes algorithms, that Naïve Bayes is stated to be better than other methods. The C4.5 algorithm can only process large amounts of data, while Naïve Bayes can still make predictions even though the data used is relatively small.

4. Conclusion

Based on research on predicting the rate of spread of malnourished toddlers in Magelang City using the Naïve bayes method, it can be concluded that based on training data from 17 urban villages starting from 2017 to 2021, and testing data used in 2021, it was successful to classify 64 data from a total of 85 data used. The application of the classification method with Naïve Bayes succeeded in predicting the magnitude of cases of malnourished toddlers in Magelang City with an accuracy percentage of 75% with relatively little data, and the areas that have the most malnutrition are in three regions, namely Magersari, North Tidar and Panjang. Further research can be added a combination of prediction methods to increase the accuracy value.

Reference

[1] A. Alpin, W. O. Salma, and R. Tosepu, “Faktor-Faktor Yang Memengaruhi Kejadian Gizi Buruk Pada Balita Dimasa Pandemik Covid-19 Di Wilayah Kerja Puskesmas Tawanga Kabupaten Konawe Provinsi Sulawesi Tenggara,” *J. Ilm. Obs. J. Ilm. Ilmu Kebidanan Kandung. P-ISSN 1979-3340 e-ISSN 2685-7987*, vol. 13, no. 3, pp. 1–10, 2021.

[2] R. Ambarwati and S. P. Hastono, “Pemodelan Faktor-faktor yang Mempengaruhi Balita Gizi Buruk di Provinsi Jawa Barat Tahun 2017 dengan Pendekatan Spatial Autoregressive Model (SAR),” *J. Ilmu Kesehat. Masy.*, vol. 9, no. 02, pp. 121–128, 2020.

[3] D. Ayu, N. Wulandari, and N. Merlina, “A Comparison Tsukamoto and Mamdani Methods in Fuzzy Inference System for Determining Nutritional Toddlers,” *2018 6th Int. Conf. Cyber IT Serv. Manag.*, no. Citsm, pp. 1–7, 2018.

[4] R. Nurida and E. Sugiharti, “Implementation of Fuzzy K-Nearest Neighbor Method in Decision Support System for Identification of Under-five Children Nutritional Status Based on Anthropometry Index,” vol. 1, no. October, pp. 83–89, 2019.

[5] Stunting.go.id, “Tahun 2022 Angka Prevalensi Stunting Harus Turun Setidaknya 3%,” *Stunting.go.id*, 2022. .

[6] Gusti Ayu, “Gambaran Asuhan Keperawatan Pada Balita Gizi Kurang Dengan Defisit Nutrisi Di Upt Kesmas Tegallalang I,” *Gambaran Asuhan Keperawatan Pada Balita Gizi Kurang Dengan Defisit Nutr. Di Upt Kesmas Tegallalang I*, 2018.

[7] Rokom, “Status Gizi Balita dan Interaksinya,” *16 Februari*, 2017. .

DOI: https://doi.org/10.29207/resti.v7i2.4438

[8] Setiaji, "Arah kebijakan transformasi digital bidang Kesehatan," pp. 1–21, 2021.
[9] A. Desiani, S. Yahdin, and D. Rodiah, "Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa dengan Menggunakan Teknik Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1237, 2020.
[10] M. M. Islam *et al*., "Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh," *Int. J. Cogn. Comput. Eng.*, vol. 3, no. January, pp. 46–57, 2022.
[11] A. D. Afifaturahman and F. MSN, "Perbandingan Algoritma K-Nearest Neighbour (KNN) dan Naive Bayes pada Intrusion Detection System (IDS)," *Innov. Res. Informatics*, vol. 3, no. 1, pp. 17–25, 2021.
[12] T. T. Ramanathan, J. Hossen, and S. Sayeed, "Naïve Bayes Based Multiple Parallel Fuzzy Reasoning Method for Medical Diagnosis," *J. Eng. Sci. Technol.*, vol. 17, no. 1, pp. 472–490, 2022.
[13] S. Kanwal, F. Khan, and S. Alamri, "A multimodal deep learning infused with artificial algae algorithm – An architecture of advanced E-health system for cancer prognosis prediction," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2707–2719, 2022.
[14] H. Thakkar, V. Shah, H. Yagnik, and M. Shah, "Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis," *Clin. eHealth*, vol. 4, pp. 12–23, 2021.
[15] Magelang Ekspress, "Stunting Kota Magelang," 2022. [Online]. Available: https://magelangekspres.disway.id/read/649509/berhasil-turunkan-hingga-117-persen-angka-stunting-kota-magelang-tersisa-10-persen.
[16] Dinas Kesehatan, "Open Data Kota Magelang.".
[17] F. F. Zain, "Effectiveness of Naïve Bayes Weighted SVM Method in Movie Review Classification," vol. 5, no. 2, pp. 108–114, 2019.
[18] J. W. G. Putra, "Pengenalan Konsep Pembelajaran Mesin dan Deep Learning," *Comput. Linguist. Nat. Lang. Process. Lab.*, vol. 4, pp. 1–235, 2019.
[19] S. R. Niko Suwaryo[1], Ismasari Nawangsih[2], "Deteksi Serangan pada Intrusion Detection System (IDS) untuk Klasifikasi Serangan dengan Algoritma Naïve Bayes, C.45 dan K-NN dalam Meminimalisasi Resiko Terhadap Pengguna," *Angew. Chemie Int. Ed. 6(11), 951–952.*, pp. 2013–2015, 2021.
[20] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan," *J. TIM Darmajaya*, vol. 02, no. 01, pp. 52–62, 2016.

DOI: https://doi.org/10.29207/resti.v7i2.4438