International Journal of Multidisciplinary Research and Growth Evaluation.

Classifying toddler stunting based on anthropometric data using the learning vector quantization 3 (LVQ 3) method

Rana Fatrika [1], **Alwis Nazir** [2*], **Suwanto Sanjaya** [3], **Elin Haerani** [4], **Siska Kurnia Gusti** [5]

[1-5] Faculty of Science and Technology, Sultan Syarif Kasim Riau State Islamic University, Pekanbaru, Indonesia

* Corresponding Author: **Alwis Nazir**

Article Info

ISSN (online): 2582-7138
Volume: 05
Issue: 03
May-June 2024
Received: 15-04-2024
Accepted: 17-05-2024
Page No: 921-930

Abstract

Based on anthropometric data, this study classified news stunts using the Learning Vector Quantization 3 (LVQ 3) technique. Kaggle's secondary data consists of 6500 data points with seven main variables. The research phases include data gathering, preprocessing, transformation, normalization, modeling, testing, and assessment. Normalization normalizes data value ranges, while data transformation transforms non-numeric variables into numerical representations. The two layers of the LVQ-3 architecture—input and competitive—are used for modeling. Based on the data class and learning level modification, the neurons' weights are changed. We run experiments with different values for the LVQ 3 parameters, such as window, maximum iteration, and learning level. Confusion, precision, recall, and F1 score matrixes were used in the evaluation process. A learning level of 0.1 to 0.6 produced the best results, with a consistent accuracy of 0.72. The greatest accuracy of 0.7420 was obtained by dividing the data by 70:30 using a window between 0.3 and 1. The findings from this study show that, despite being influenced by the common use of variables and the lack of data variance, LVQ 3 performs quite well and has significant accuracy. To identify the value that best matches the data set conditions, more parameter exploration is required.

DOI: https://doi.org/10.54660/.IJMRGE.2024.5.3.921-930

Keywords: Anthropometric, learning vector quantization 3, Stunting

Introduction

According to the (World Health Organization *et al*. 2023) [20], this study is caused by a lack of nutrients in the uterus and during pregnancy. Stunting can stop children from reaching full height and make their cognitive abilities develop slower. Anthropometric measurements such as weight-per-age index (WHZ) or height-for-age (HAZ) are used to identify stunts. The child's anthropometric standard consists of data on the size, proportions, and composition of the body. This standard is used to assess the growth trends and nutritional health of children (PMK RI 2020) [14].

Stunting affects approximately 149.2 million children under the age of five worldwide, which is equivalent to a global prevalence rate of 22%. In Indonesia, the overall stunting rate has fallen from 24.4% in the previous year to 21.6% by 2022. The World Health Assembly (WHA) has set a global target to reduce Indonesia's prevalence by 40% by 2025, compared to the level seen in 2013 (World Health Organization *et al*. 2023) [20]. In addition, the Sustainable Development Goals (SDGs) require the total elimination of all types of malnutrition by 2030. To achieve the goal of reducing the percentage of children under five years of age suffering from stunting to 22 percent by 2025, greater efforts are needed to address and prevent stunting (Office of the Vice President of Indonesia 2019) [12].

It is important to identify the basics of childhood stunting, especially in nutrition, health care, and sanitation (Adzim *et al*. 2023) [1]. Data mining techniques can be used to quickly classify stunting results. Data mining involves extracting valuable insights from large data sets using machine learning algorithms (Rahman 2023) [16]. This method aims to process and evaluate text automatically (Amna *et al*. 2023) [2]. Classification is an effective data mining technique, such as using an artificial neural network (ANN) to categorize data using certain rules(Arbanus Simbolon *et al*. 2019) [3].

Several studies related to the application of classification methods with the Learning Vector Quantization 3 (LVQ 3) algorithm have been successfully carried out. According to (Hadi 2020) [8] study of applying the LVQ 3 simulated neural network method for the classification of learning using learning rate parameters 0.001, 0.2, 0.3, and 0.4; window values 0.2 and 0.4, m 0.2; minimum learning rate 0,0001; and reduction of learning rate 0.01, resulting in the best accuracy of 100%.%. This shows that the combination of such parameters is very effective in classifying learning difficulty. Furthermore, in a (Sagewa 2019) [17] study on the comparison of LVQ 2.1 and LVQ 3 methods for brain tumor cell classification, it was found that the LVQ3 method obtained the highest accuracy of 83.3% in tests with a learning rate of 0.7, whereas LVQ 2.1 obtained the most accurate of 75% in trials with learning rates of 0.001. This suggests that LVQ-3 is superior to LVQ 2.1 in the classification of brain tumor cells.

A study by (Darmila 2022) [7], dealt with Learning Vector Quantization 1 (LVQ1), which uses reference vectors to represent classes in datasets. In the meantime, (Adzim *et al*. 2023) [1] created Learning Vector Quantization 2 (LVQ 2). This updates two reference vectors at the same time if they have a distance estimate that is the same as the input data. This makes the model more adaptable to changes in the data. (Darmila 2022) [7] also displays Learning Vector Quantization 3 (LVQ 3), in which the algorithm updates a group of reference vectors based on the distance with the entry data, which can involve more than two vectors depending on the class distribution.

A study conducted by (Setiawan 2023) [18] on a web-based stunt classification system using the Naive Bayes method achieved an accuracy rate of 91% using the same data set. The data set used was as large as 6,500, divided into two classes, and included seven variables. On the other hand, research by (Pahlevi and Handrianto 2024) [13] used the Naïve Bayes-Based Particle Swarm Optimization algorithm to classify stunting status with a data set of 10,000 records and the same variables. The results showed an accuracy of 80.69%, with an Area Under the Curve (AUC) of 83.06%.

Based on the results of previous studies, although the results are good, the researchers will try to improve the performance of the classification of stunts with different methods. Therefore, this study will classify stunts based on the newly obtained anthropometric data using the Learning Vector Quantization 3 (LVQ 3) method with the same data set as the previous study.

Materials and Methods

The method used in this research is experimental research, consisting of (1) data collection, (2) data preprocessing, (3) modeling, (4) model testing, and (5) evaluation. The research flow can be seen Fig 1.

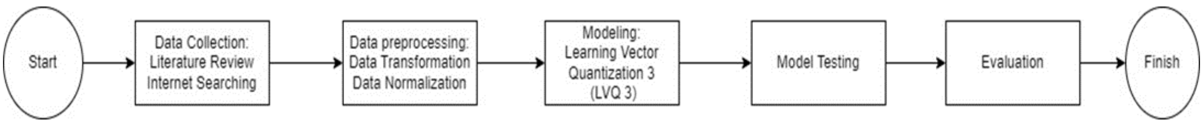

Fig 1: Research Flow

Data Collection

At this stage, data for research is collected using secondary data. The data used came from stunting news, with a total of 6500 data points obtained from Kaggle (https://www.kaggle.com/datasets/muhtarom/stunting/data), accessed at 12:12 AM on March 13, 2024. There are seven Table 1. Major variables related to the anthropometric data of infants, sex, age, birth weight, body weight, birth length, and whether exclusive breastfeeding is applied. Data was downloaded in CSV format. The Data anthropomethry can be seen.

Table 1: Data Anthropomethry

	Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	Exlusive Brestfeeding	Stunting
0	F	56	2.9	50	11.0	90.0	Yes	No
1	F	20	3.3	49	11.1	80.5	No	No
2	M	4	2.8	48	6.5	63.0	No	No
3	F	14	2.0	49	7.0	71.0	Yes	No
4	M	32	3.2	49	11.0	88.7	Yes	No
...	...	...	...	...	...	...	...	...
6495	M	53	2.9	49	15.0	96.0	No	Yes
6496	M	9	2.9	50	7.3	62.0	No	Yes
6497	F	20	1.8	48	7.3	73.0	Yes	Yes
6498	M	11	2.9	49	7.7	66.0	No	Yes
6499	F	14	2.9	49	6.5	66.0	No	Yes

Data Transformation

Data transformation is an important step in data preparation before further processing is carried out according to the model or algorithm to be used. In the context of the LVQ algorithm, which is based on the principle of allocating input vectors to classes based on the smallest distance to their weight vectors, data transformation becomes a crucial stage. In this study, there are some variables that need to be adjusted to suit the requirements of the LVQ algorithm. Gender variables, for example, may have to be converted to binary representations, such as 0 and 1, to facilitate the processing process (Harmain *et al*. 2021) [9]. Similarly, with exclusive breastfeeding variables and stunting variables that are originally in string format, they need to be converted to the correct numerical format. By doing this transformation, the data will be more ready for use in the modeling and further analysis processes. Proper data transformation ensures that the model or algorithm applied can deliver accurate and meaningful results (Sholeh *et al*. 2022) [19].

Data Normalization

Before entering the training phase (learning) in the process of classifying new stunts based on anthropometric data using the Learning Vector Quantization 3 (LVQ 3) method, the data is normalized. The primary purpose of data normalization in this context is to form data in the position of values with the same range. In news stunting classification, data normalization is an important step because anthropometric variables such as weight, length, and age may have different value ranges.

The common method used for normalizing data is the min-max method. This normalization process entails dividing the difference between each data point and its smallest value by the difference from the maximum value to the minimum value of the entire dataset. Thus, every data point will have a value associated with the same range, thus facilitating the learning process by the LVQ algorithm 3 (Cahyani *et al*. 2023) [5]. The Formula can be seen (1).

$$xnorm = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where $xnorm$ is the normalized data value, xi is the value of the data point i, $xmin$ is the minimum value for the entire data, and $xmax$ is the maximum value for all data.

Modelling

According to research conducted by (Damanik 2020) [6], learning vector quantization (LVQ) is an architecture of one-layer feedforward neural networks consisting of input and output units. Specifically, LVQ 3 is the LVQ method used in this study. This network architecture consists of two primary layers: the input layer and the competitive layer. The competition layer is responsible for automatically learning to classify the input vector according to its distance. If two input vectors have close distances, the competing layer will place them in the same class. This study utilized a network architecture that adopted the Learning Vector Quantization 3 method, which is an effective approach to classifying new anthropometric data related to stunts. Fig 2 shows the network architecture structure used in this study, which reflects the use of the LVQ 3 method in data analysis and modeling.

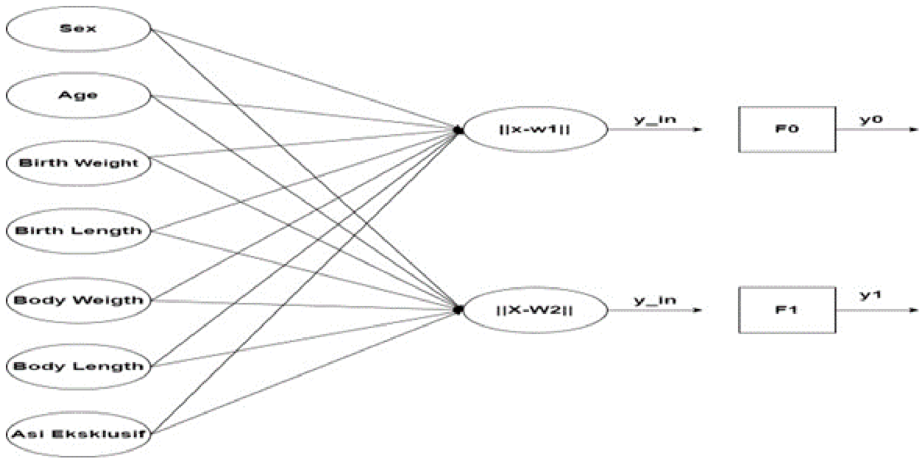

Fig 2: Architecture Learning Vector Quantization 3

The training data input process involves 7 neurons, each assigned to process 7 input variables related to the newspapers' anthropometric data: gender, age, birth weight, birth length, current weight, current length of body, and whether exclusive breastfeeding is given. Each neuron represents a specific data attribute in the context of stunting classification using the Learning Vector Quantization 3 method. (LVQ 3). When new anthropometric data was introduced, the Euclidean distance between the data and each weight vector of the neurons was calculated. The neuron that has the smallest distance from the data is set as the winning neuron. This process enables LVQ 3 to learn adaptively from data patterns and classify news as stunted or non-stunted based on the calculated attributes.

During the weight renewal process, the output of the winning neuron, which may represent either a stunted (f1) or a non-classification of such data is determined by the winning neuron: if the winning neuron represents a non-stunted class, the output is y0; if the victorious neuron represents a stunted stunted class (f0), is used to determine whether the neuron's weight needs to be adjusted. Comparing the output from the winning neuron to actual input data helps LVQ 3 reset its weights to a more accurate classification. After the weight renewal, the newly categorized anthropometric data uses the winning neurons with the weight adjusted. The final class, it is y1. By using the important features represented by neurons on the input layer, LVQ 3 makes it easy to sort stunts into groups based on anthropometric data from the news.

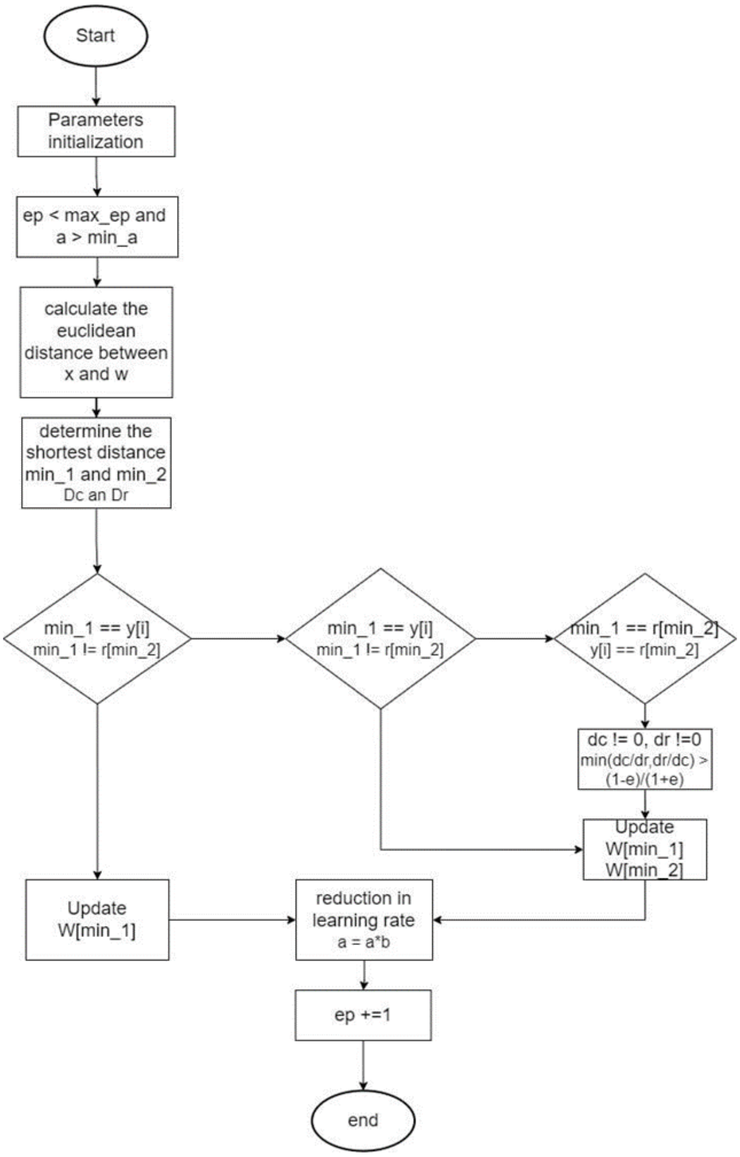

Fig 3: Training Flowchart

From Fig 3 the learning algorithm for vector quantization 3 has the following steps:

Step 1: The process starts with initializing parameters such as the learning rate (α), minimum learning rates (min_α), window (ε), and maximum iteration (maxepoch). The main loop runs during ep < max_ep and α > min_α. This ensures that training continues until the maximum number of iterations is reached or the learning level falls below the set minimum value.

Step 2: During the training process, iteration is carried out until the maximum number of iterations is reached and the level of learning is still above the set minimum value.

Step 3: Each data point in the training set is processed individually. The Euclidean distance between the data point and all the weights counted. As can be seen in the formula (2):

$$d = \sqrt{\sum (w_i - x_i)^2} \quad (2)$$

Step 4: The weight is updated based on the found data class. If the data class matches the first nearest weight, the weight is upgraded, taking into account the difference between the data and the first weight. As can be seen in the formula (3):

$$W_{min1} = W_{min1} + \alpha(x - W_{min1}) \quad (3)$$

If the data class differs from the first nearest weight but is the same as the second closest weight, and if the distance comparison meets the window conditions. Can be seen on the formula (4):

$$\text{Min}\left(\frac{dc_1}{dc_2}, \frac{dc_2}{dc_1}\right) > (1 - \varepsilon)(1 + \varepsilon) \quad (4)$$

Then update the nearest weight, first and second. As can be seen in the formulas (5) and (6):

$$W_{min1} = W_{min1} - \alpha(x - W_{min1}) \quad (5)$$

$$W_{min2} = W_{min2} + \alpha(x - W_{min2}) \quad (6)$$

If the data class is equal to the two closest weights, update them by adding part of the difference between the data and each weight, multiplied by the window parameter. See also in Formulas (7) and (8):

$$W_{min1} = W_{min1} + \varepsilon\alpha(x - W_{min1}) \quad (7)$$

$$W_{min2} = W_{min2} + \varepsilon\alpha(x - W_{min2}) \quad (8)$$

Step 5: Lower the learning speed. After each iteration (epoch), decrease the learning speed by multiplying α by dec_α. As can be seen in the formula (9):

$$\alpha = \alpha \times dec\alpha \quad (9)$$

Step 6: The process is stopped if the maximum number of iterations is reached or the learning level falls below the set minimum value.

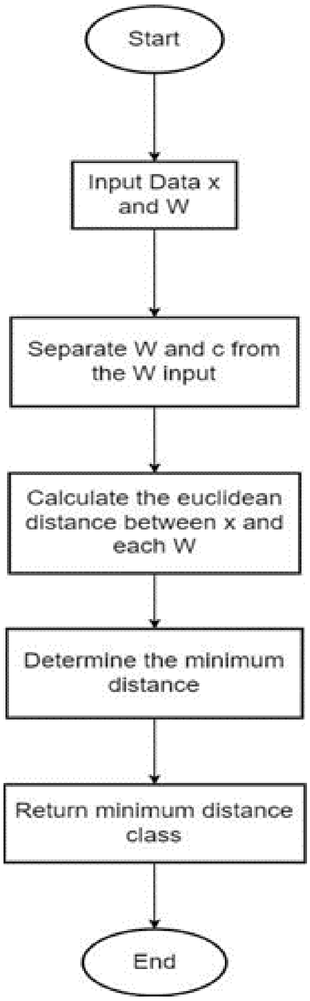

Fig 4: Testing Flowchart

The test phase in the LVQ algorithm 3, as described in Fig 4, involves a set of structured steps. The first step is to take the weight vector W and the class C from the input parameter. Next, calculate the Euclidean distance between each weight vector W, and input x using the standard Eucladian distance formula. After that, perform a search to find the index of the smallest distance in the array d, which identifies which weight is closest to the input x. Finally, the class corresponding to the closest weight is selected and returned as the result of the x input classification. Through this series of steps, the LVQ 3 test phase allows the exact determination of the class based on the weight vector that is closest to the given input.

Testing Model

At this stage, the Learning Vector Quantization 3 (LVQ 3) model will be tested based on a series of previous studies. The test process aims to find the optimal value for each LVQ 3 parameter, which has proven significant in the literature. In a based on the results of previous studies. Each test scenario is considered as a parameter for subsequent testing, in accordance with the methodology that has been tested in previous studies. Each test scenario is considered as a parameter for subsequent testing, in accordance with the methodology that has been tested in previous studies.

Evaluation Model

After obtaining the results of several test scenarios, a stage of evaluation is carried out to assess the performance of the model produced. The use of accuracy as the only metric to assess the performance of the model is considered inadequate. One common approach used in evaluating the performance of classification methods is to use a confusion matrix (Miftahusalam *et al*. 2023) [11]. Therefore, additional evaluation metrics are needed to gain a deeper understanding of the performances of a classification model. The confusion Matrix provides information that compares the results of a model classification with the actual classification results

(Putri *et al*. 2022) [15]. Model evaluation can also involve the use of metrics such as precision, recall, F1 score, and accuracy to evaluate the predictive quality of the classification algorithm used (Hidayat *et al*. 2021) [10]. After obtaining the result of the confusion matrix, the next step is to perform weighted averages of evaluation metrics, such as precision, recall, and F1 scores. The weighted average is a method for calculating averages in which each value in a data set is given a corresponding weight, based on how many samples of each class in the data. This weight is then used to calculate the mean value of the desired evaluation metric.

For weighted precision, the formula can be seen (10).

$$Weighted\ precision = \frac{N_0 \times Presisi_0 + N_1 \times Presisi_1}{N_0 + N_1} \quad (10)$$

Where $N0$ and $N1$ represent the sum of samples of class 0 and class 1 in data, and Precision00Precision0 and Precision0 andprecision1 represent precision for classes 0 and 1.
Similarly, for weighted recall and weighted F1 score, can be seen the formula (11) and formula (12). Study conducted by (Aziz *et al*. 2023) [4], ten different scenarios were used to observe the influence of parameter variation on the accuracy of classification. The research emphasizes the importance of setting parameters such as learning rate, learning rate reduction, minimum learning rate (min alpha), number of epochs, window size, and data distribution ratio on training. The tests were conducted gradually, by variing the values of the LVQ-3 parameter.

$$Weighted\ Recall = \frac{N_0 \times recall_0 + N_1 \times recall_1}{N_0 + N_1} \quad (11)$$

$$Weighted\ f1\ score = \frac{N_0 \times f1\ Score_0 + N_1 \times f1\ Sc}{N_0 + N_1} \quad (12)$$

Here, recall0 and recall1 represent recall for grade 0 and grade 1, whereas F1 score0 and F1 scor1 represent F1 scores for class 0 and class 1.

Results

The initial step was to search for journals related to stunting cases, either using the Learning Vector Quantization 3 (LVQ 3) method or previous research using the same dataset. This data was obtained from the Kaggle account owned by Muhtarom Ahkam. The data is an anthropometric measurement carried out in 2023. We used data on the age group 0–60 months. There are two classes: stunting and normal. According to the book "Data Mining" by (Amna *et al*. 2023) [2], the stages of the knowledge discovery process in the database include data selection, pre-processing, transformation, data mining, pattern evaluation, and knowledge presentation. The study uses the KDD process, namely transformation and normalization.
In this study, a transformation process is carried out. This phase modifies the data to match the model or algorithm intended to be used in the data processing phase. At this stage of transformation, the "sex" variable with the value "L" is represented as a number of 0, while the value of "P" is presented as a number of 1. Furthermore, for the variable "Exclusive Breastfeeding," the "Yes" value is shown as a figure of 1, whereas the "No" value represents a digit of 0. Similarly, for the class "Stunting", the 'Yes' value is represented as a number of 1, and the "No" value as a value of
0. This transformation is necessary to enable the LVQ algorithm to operate on data effectively, as most machine learning algorithms require numerical input. The results of the data transformation can be seen in Table 2.

Table 2: Data Transformation

	Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	Exclusive Breastfeeding	Stunting
0	1	56	2.9	50	11.0	90.0	1	0
1	1	20	3.3	49	11.1	80.5	0	0
2	0	4	2.8	48	6.5	63.0	0	0
3	1	14	2.0	49	7.0	71.0	1	0
4	0	32	3.2	49	11.0	88.7	1	0
...	...	...	...	...	...	...	...	...
6495	0	53	2.9	49	15.0	96.0	0	1
6496	0	9	2.9	50	7.3	62.0	0	1
6497	1	20	1.8	48	7.3	73.0	1	1
6498	0	11	2.9	49	7.7	66.0	0	1
6499	1	14	2.9	49	6.5	66.0	0	1

The next step is to normalize the data values using the min-max method. Normalization is necessary because the initial data has different ranges, so in order to ensure that the calculation of the distance in modeling is more accurate, adjustment needs to be performed. In the min-mmax method, the range used for normalizing the data is from 0 to 1. The objective of normalizing data in a data set is to standardize the values so that they are in similar ranges. Changes that occur during the normalization process do not alter the information contained in the data. The results of this data normalization can be seen in Table 3.

Table 3: Data Normalization

	Sex	Age	Birth Weight	Birth Length	Body Weight	Body Length	Exclusive Breastfeeding	Stunt Ing
0	1.0	0.948276	0.500000	0.727273	0.375000	0.637931	1.0	0
1	1.0	0.327586	0.681818	0.636364	0.379630	0.474138	0.0	0
2	0.0	0.051724	0.454545	0.545455	0.166667	0.172414	0.0	0
3	1.0	0.224138	0.090909	0.636364	0.189815	0.310345	1.0	0
4	0.0	0.534483	0.636364	0.636364	0.375000	0.615517	1.0	0
...	...	...	...	...	...	...	...	...
6495	0.0	0.896552	0.500000	0.636364	0.560185	0.741379	0.0	1
6496	0.0	0.137931	0.500000	0.727273	0.203704	0.155172	0.0	1
6497	1.0	0.327586	0.000000	0.545455	0.203704	0.344828	1.0	1
6498	0.0	0.172414	0.500000	0.636364	0.222222	0.224138	0.0	1
6499	1.0	0.224138	0.500000	0.636364	0.166667	0.224138	0.0	1

The next stage is testing to explain the results of the Learning Vector Quantization 3 method in the classification of stunting cases based on anthropometric data. Data testing is performed gradually using various variable values on each LVQ 3 parameter, ranging from the learning level test stage, learning level reduction test, minimum learning level testing, maximum iteration test, window testing, and ratio testing. The best or optimal value of a parameter test scenario will be used as the next test parameter.

The learning level test is performed with learning level variation, learning rate reduction, maximum iteration, minimum learning rate, window, and ratio to find the optimal value. The test process starts with the initial learning level parameter dec_α = 0.1, the ratio 70:30, 10 iterations, min_α= 0.00000001, and window 0.5. Tested values for learning levels are 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.

Choosing a suitable starting value is important as it can significantly affect the test results. If the initial value is too low, then the model may not learn effectively, while if too high, it may cause convergence or overfitting problems. Therefore, start with a commonly used value. The test results showed the highest accuracy of 0.72 in the range of 0.1–0.6. However, the precision for the range 0.7 reached 0.4892, for the radius 0.8 reached 0.5107, for the area 0.9 reached 0,5107, and for the value 1, obtained a precision of 0,4892. This information is checked in **Fig 5**.

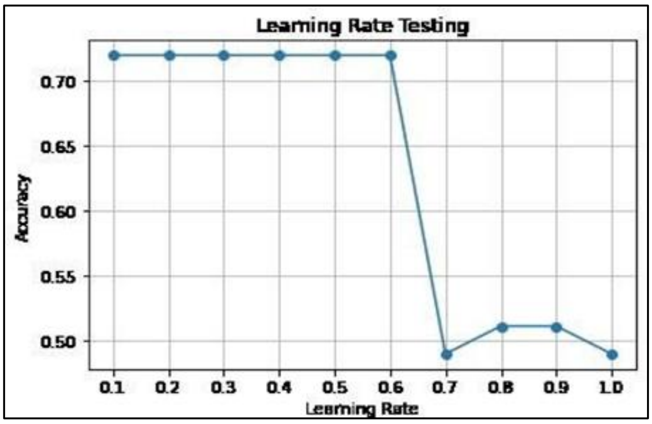

Fig 5: Learning Rate Testing

The learning reduction test is performed to obtain optimal analysis results by applying the same initial parameters as the learning level test. At the same time, for α = 0.1, one value will be selected based on the best accuracy obtained from the test results. Dec_α will vary from 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 to 1. The learning reduction test yields an accuracy of 0.72 for all dec_α values. Based on the results of the stable learning decrease process, which can be seen in Fig 6.

Fig 6: Learning Rate Reduction Testing

Next, the test with the minimum learning level (α) will use the optimal parameters α = 0,1, and dec_α = 0,1. Other parameters will have the same values as the previous test. For this test, the range of minimum learning levels (Min α) used is 0.1, 0.05, 0.01, 0.005, 0.001, 0,0001, 0.00001, 0.00000001, and 0.0000000001. The accuracy result obtained is the same, i.e., 0.72, except for the value 0.1, which has a precision of 0.4912. It can be seen in **Fig 7**.

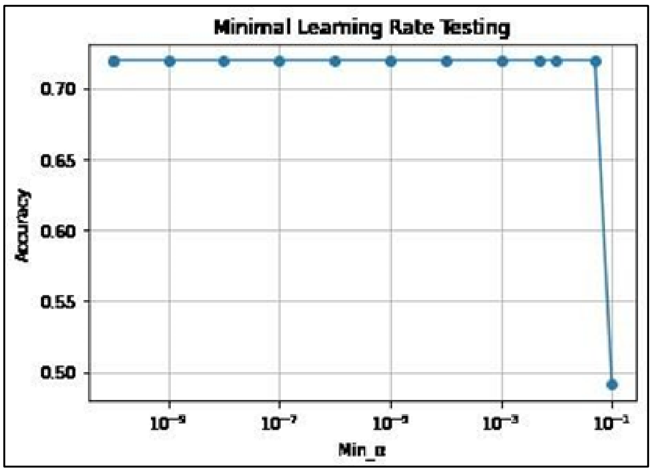

Fig 7: Minimal Learning Rate Testing

Iterative testing uses parameters that support optimum accuracy: α = 0.1, dec_α = 0.1, and min_a = 0.01. Iteration testing aims to determine the maximum impact of iteration on the accuracy result. The other parameters' values are the same as in the initial test. The maximum range of iteration variations tested included 2, 5, 10, 15, 20, 30, 40, 50, 60, 70,

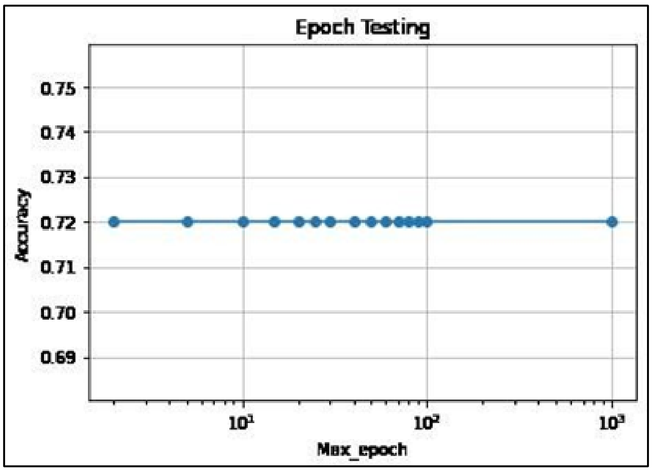

Fig 8: Epoch Testing

The window value determines how much difference between the two closest weight vectors will be updated when there is a data classification. Testing with window parameters, where α = 0,1, dec_α = 0,1, min_a = 0,01, epoch = 1000, and in a ratio of 70:10, while using window values of 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, and 1, yields the highest accuracy of 0,7251 in the range of 0,3–1, and for the ranges of 0,1 and 0,2, the accurate value is 0,72. It can be seen in Fig 9.

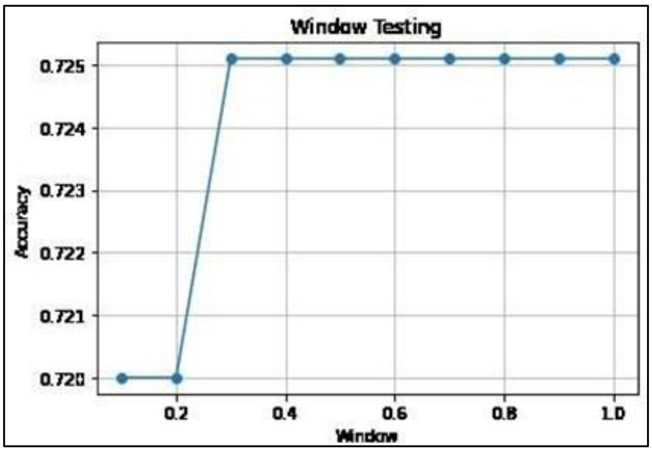

Fig 9: Window Testing

Tests with ratio parameters are performed using optimum parameters that support accuracy, i.e., α = 0.1, dec_α = 0,1, min_a = 0.01, epoch = 1000, and window values of 0.5. This ratio test is intended to observe the influence of the percentage change in the amount of training data on accuracy. Test results with the ratios 90:10, 80:10, 70:10, and 60:40 show the highest accuration of 0.72 for the 70:10 ratio. For the 60:40 ratio, the accuracy reaches 0.7169, the 80:20 ratio reaches 0,7161, and the 90:10 ratio obtains the exactness of 0.7076. This information is presented in Fig *10*.

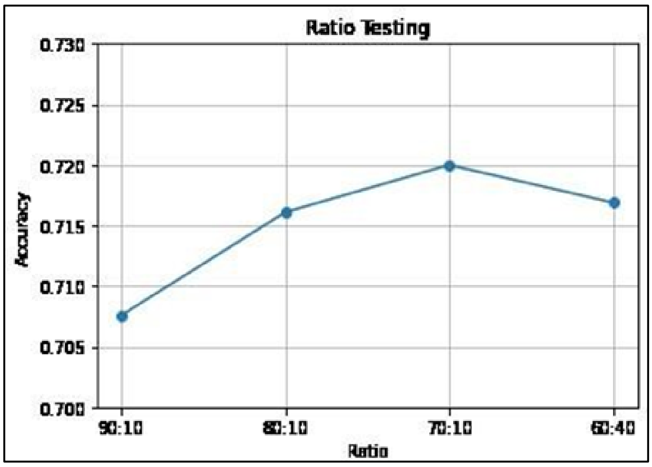

Fig 10: Ratio Testing

Based on the parameter test's results, the next step is to perform the test by combining the learning rate and window parameter values. In this study, the data is divided into four different ratios. The total data used is 6,500 recordings, and the best parameters in each test scenario are determined. This data separation process is important to ensure an accurate and objective evaluation of the model developed. The best parameters are determined based on the results of previous parameter tests. The results of LVQ 3 algorithm tests show the best performance in each scenario, as recorded in Table 4.

Table 4: Research Skenario

Ratio	Learning rate (α)	Window (ε)	Accuracy
90:10	0.2	0.5	0.72
80:20	0.2	0.5	0.7307
70:30	0.2	0.5	0.7420
60:40	0.2	0.5	0.7346

Table 4 displays the test results after combining several learning rate and window parameter values. The learning rate parameter value used ranges from 0.1 to 0.6, while the window parameter values range from 0.3 to 1. For other parameters, the learning rate decrease is set to 0.1, the minimum learning rate is 0.1, and the epoch value is set to 1000.

Result Evaluation Model

In classifying stunts based on newspapers' anthropometric data using the LVQ3 method, the model evaluation was performed using several metrics, namely precision score, recall score, accuracy score, and F1 score.

The precision score measures how accurately a model classifies a positive class, with a score of 0.7456. This means that about 74.56% of predictions classified as stunts are really stunts. Meanwhile, the recall score shows how well a model can identify all instances that actually belong to the positive class, with a rating of 0.7421. The model is able to identify about 74,21% of all news that actually has been stunted.

The results of the model evaluation in the stunting classification of news based on anthropometric data show that the model is able to identify accurately. Of the total predictions, 809 true newsmen experienced stunting, and 638 newsmen who did not experience Stunting were correctly classified. However, there were 316 cases of models classifying newsmen that did not actually have stunting, and 187 newsmen that should have been classified as stunting but were not detected by the model. The result of the matrix confusion can be seen in Fig 11.

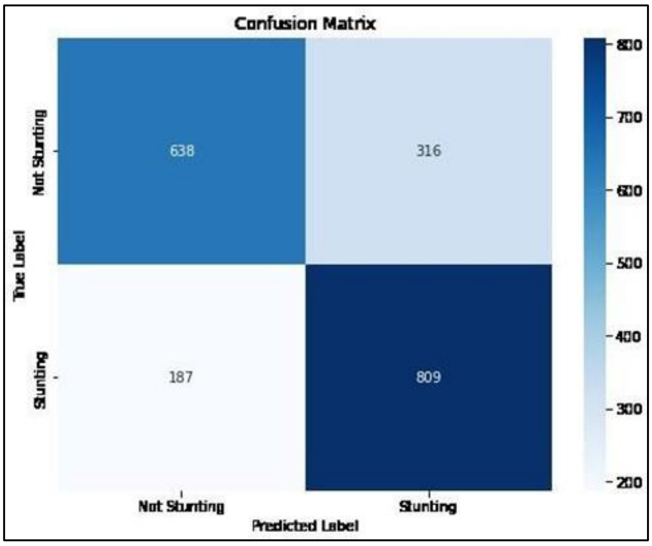

Fig 11: Confusion Matriks

Discussion

Based on the test results carried out with a total of 163 scenarios, it can be concluded that the combination of learning rate and window parameters has a significant influence on the accuracy of the classification model. Depending on the data division ratio, using different combinations of parameters can result in different accuracy. For every ratio and parameter combination, this optimal result does not always occur. Then, at a ratio of 60:40 with a learning rate of 0.1, the best combination is with a window value of 0.4 and 0.5, each of which yields an accuracy of 0.7230. On the other hand, at the 60:40 ratio with a learning rate of 0,3, the highest accuracy is obtained with a value of 0.4, reaching 0.7011. It shows that there is no one combination of parameters that is optimal for all scenarios.

For the 70:30 ratio, the highest accuracy is achieved with a learning rate of 0.2 and a window value of 0.5, which is 0.7420. This shows that the parameters with a medium learning rate and a fairly large window yield the best model performance in this ratio. At the 0.1 learning rate, the maximum accuracy is obtained with the 0.4 and 0.5 window values, reaching 0.7256. Whereas at the 0.3 and 0.5 learning rates, the best window values are 0.1 and 0.2 with a precision of 0.72. For the 80:20 ratio, the best results were obtained with a learning rate of 0.2 and a window value of 0.5, with an accuracy of 0.7307. At the learning rate of 0.1, the highest accuracies were achieved with window values of 0.4 and 0.5, which is 0.7192.

Meanwhile, for a 90:10 ratio, the highest accuracy is obtained with a learning rate of 0.2 and a window value of 0.5, equal to 0.72. For a learning ratio of 0.1, the best window value is 0.2 with an accuracy of 0.7123.

Conclusion

Based on the results of testing and evaluation of the Learning Vector Quantization 3 (LVQ3) method, some conclusions can be drawn. The optimal learning rate parameters were found in the range of 0.1–0.6, while reducing the learning rate and using the minimum learning level also resulted in a stable accuracy of 0.72. Tests with a range of iterations from 2 to 1000 also showed stable precision. Further, using window values in the range of 0.3–1 yields the highest accuracy of 0.7251. In four different ratios, the LVQ3 method showed the best performance, with a learning rate of 0.2 and a window value of 0.5. The learning rate and window values play an important role in the performance of the Learning Vector Quantization 3 model (LVQ3). Low learning rate values tend to deliver better results in optimizing model performance, especially when combined with higher window values. On the other hand, a higher window value tends to yield better results at a lower learning rate. However, at a higher learning rate, a lower window value has a tendency to deliver more stable accuracy, although it does not always yield the best results. In tests with four different ratios, a combination of a learning rate value of 0.2 and a window within a range of 0.5 achieved the highest accuracy of 0.7420 with a 70:30 ratio. Although the results of this study are quite good, it should be noted that there is a lack of variation in the data and the common use of variables. A more detailed parameter exploration is then required to select the value that best matches the conditions of the data set used.

Reference

1. Adzim F, Budianita E, Nazir A, Syafria F. Klasifikasi Status Stunting Balita Menggunakan Metode C4. 5 Berbasis Web. ZONAsi-Jurnal Sistem Informasi. 2023;5(3):515-525.
2. Amna WS, Sudipa IGI, E.Putra TA, Wahidin AJ, Syukrilla WA, Wardhani AK, *et al*. Data Mining. Edianan D, Yanto A, editors. 1st ed. Padang: Yuliantri Novita; c2023.
3. Simbolon DA, Hartama D, Anggraini F. Penerapan Jaringan Saraf Tiruan Dalam Memprediksi Gizi Balita Pada Puskesmas Siantar Utara Kota Pematangsiantar. BRAHMANA: Jurnal Penerapan Kecerdasan Buatan. 2019;1(1):48–54.
4. Aziz A, Insani F, Jasril J, Syafria F. Implementasi Metode Learning Vector Quantization (LVQ) Untuk Klasifikasi Keluarga Beresiko Stunting. Building of Informatics, Technology and Science (BITS). 2023;5(1):12-20.
5. Cahyani J, Mujahidin S, Fiqar TP. Implementasi Metode Long Short Term Memory (LSTM) Untuk Memprediksi Harga Bahan Pokok Nasional. Jurnal Sistem Dan Teknologi Informasi (JustIN). 2023;11(2):346. DOI: 10.26418/justin.v11i2.57395.
6. Damanik AR. Penerapan Learning Vector Quantization (LVQ) Untuk Mengklasifikasikan Tenaga Ahli IT (Studi Kasus: PT. Cita Kreasi Latena) [dissertation]. Medan: Universitas Pembangunan Panca Budi; c2020.
7. Afrianty I, Sanjaya S, Abdillah R, Iskandar I, Syafria F. Evaluasi Perbandingan Performansi LVQ 1, LVQ 2, DAN LVQ 3 Dalam Klasifikasi Jenis Kelamin Menggunakan Tulang Tengkorak. Jurnal INSTEK (Informatika Sains dan Teknologi). 2022;7(2):344-353.
8. Hadi M. Penerapan Metode Jaringan Syaraf Tiruan Learning Vector Quantization 3 (LVQ3) Untuk

Klasifikasi Kesulitan Belajar; c2020.
9. Harmain A, Kurniawan H, Maulina D. Data Normalization For K-Means Efficiency On Groups Of Areas With Potential Forest And Land Fire Based On Heat Spots Distribution; c2021.
10. Hidayat W, Ardiansyah M, Setyanto A. Pengaruh Algoritma ADASYN Dan SMOTE Terhadap Performa Support Vector Machine Pada Ketidakseimbangan Dataset Airbnb. Jurnal Pendidikan Informatika. 2021;5(1):11-20. DOI:10.29408/edumatic.v5i1.3125.
11. Miftahusalam A, Pratiwi H, Slamet I. Perbandingan Metode Random Forest Dan Naive Bayes Pada Sentimen Review Aplikasi BCA Mobile. Prosiding Seminar Nasional; c2023.
12. Office of the Vice President of Indonesia. Strategi Nasional Percepatan Pencegahan Anak Kerdil (Stunting). 2nd ed; c2019.
13. Pahlevi O, Handrianto Y. Optimasi Algoritma Naïve Bayes Berbasis Particle Swarm Optimization Untuk Klasifikasi Status Stunting. Computer Science (CO-SCIENCE). 2024;4(1):37-43.
14. PMK RI. Peraturan Menteri Kesehatan Republik Indonesia; c2020.
15. Putri NB, Wijayanto AW. Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. Komputika: Jurnal Sistem Komputer. 2022;11(1):59–66. DOI: 10.34010/komputika.v11i1.4350.
16. Rahman A. Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree Dan Naive Bayes. Jurnal Saintekom. 2023;13(1):22–31. DOI: 10.33020/saintekom.v13i1.349.
17. Sagewa I. Perbandingan Metode Learning Vector Quantization 2.1 (LVQ 2.1) Dan Learning Vector Quantization 3 (LVQ 3) Untuk Klasifikasi Sel Tumor Otak [dissertation]. Pekanbaru: UIN SUSKA RIAU; c2019.
18. Setiawan BR. Sistem Klasifikasi Stunting Berbasis Web Menggunakan Metode Naive Bayes. Yogyakarta; 2023.
19. Sholeh M, Andayati D, Rachmawati RY. Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes; c2022.
20. World Health Organization, UNICEF, World Bank Group. Levels and Trends in Child Malnutrition; c2023.