Implementation of Principal Component Analysis and Learning Vector Quantization for Classification of Food Nutrition Status

Jasman Pardede[1], Hilwa Athifah[2]

[1,2] *Informatics, Institut Teknologi Nasional Bandung, Indonesia*

[1]jasman@itenas.ac.id, [2]hilwaathifah11@gmail.com

Abstract - Balanced nutrition is very good in the process of child growth. During the COVID-19 pandemic, consuming a balanced, nutritious diet can keep a child's immune system from transmitting the virus. In determining the nutritional content of children's food during the pandemic, a classification of the nutritional content of children's food is carried out by applying the principal component analysis (PCA) dimension reduction method and the learning vector quantization (LVQ) classification method. The data used in this study is based on Indonesian food nutritional value data from the Ministry of Health of the Republic of Indonesia amounted to 1146 data with 25 indicators of food nutrients. From the tests that have been carried out, the combination of the PCA-LVQ method produces an average accuracy of 58% with the highest accuracy of 60%. In addition, this study also compares the performance of the PCA dimension reduction method, independent component analysis (ICA), and factor analysis (FA) on the LVQ classification process. The final result of testing the three methods is that the FA method takes the fastest time, which is 4.10434 seconds and the PCA method produces the highest accuracy, which is 58.2%.

Keywords: food nutrition, covid-19 pandemic, classification, principal component analysis, learning vector quantization

I. INTRODUCTION

The COVID-19 (coronavirus) pandemic became the largest global health crisis that took place throughout 2020 with an unprecedented death toll and socioeconomic impact [1]. According to the Head of the Family Health and Nutrition Section of the Bandung City Health Office, Dewi Primasari, the pandemic has had a significant impact on nutritional problems, especially toddler nutrition. This causes nutritional problems in 2020 to increase to 5.33% from the previous year. Balanced nutrition is very good in the process of growth and development, especially in children. Balanced nutrition can be obtained from food intake that meets the body's nutritional needs according to the age and activity of a child. Although no food can prevent COVID-19 infection [2], maintaining a balanced diet is very important in boosting the immune system during a pandemic.

In conducting socialization in the community and educating children about nutrition that meets balanced nutrition during a pandemic, a system is needed that can help nutritionists determine the classification of nutritional adequacy or nutrition in children's food during a pandemic. Several classification methods such as Learning Vector Quantization [3], Fuzzy Logic [4], Naïve Bayes Classifier [5], and K-Means Clustering [6] have been proposed by several researchers in conducting nutritional status classification.

Nutrients or food nutrition datasets have many data indicators (attributes), including water, carbohydrates, protein, calories, fiber, iron, vitamins, and so on. Where the reduction of attributes or data dimensions must be carried out to determine the main attributes or components of the dataset. Several dimension reduction methods such as Principal Component Analysis [7], Independent Component Analysis [8], Factor Analysis [9], and Latent Semantic Analysis [10] have been proposed by several researchers in reducing dataset dimensions.

This research was conducted by applying the PCA method in reducing attribute dimensions to determine the most important attribute in the data, as well as the LVQ method to classify the nutritional status of children's food that can meet nutritional needs based on the attributes that play the most role in the data so that the classification results are more accurate. This study aims to measure the accuracy of the principal component analysis (PCA) and learning vector quantization (LVQ) methods in the nutritional classification of children's food during the pandemic.

II. METHOD

Several stages are done in this research, namely, the correlation analysis stage, attribute reduction stage with PCA, the design of the LVQ model, classification stage, and the testing process.

A. The Dataset

The dataset used in this study is based on Indonesian food nutritional value data from the Ministry of Health of the Republic of Indonesia in 2020 with a total of 1,146 food data with 25 indicators including water, energy (calories), protein, fat, carbohydrates, fiber, ash, calcium, phosphorus, iron, sodium, potassium, copper, zinc, retinol (Vitamin A), beta-carotene, total carotene, thiamine (Vitamin B1), riboflavin (Vitamin B2), niacin, vitamin C, BDD, types, groups, and sources. In initializing the target class of food nutritional status based on the dataset used. Then 2 target classes were grouped, namely the nutritional status class of foods that met and did not meet during the pandemic, based on the Final Guide to Balanced Nutrition during the Covid-19 Period from the Ministry of Health of the Republic of Indonesia in 2020 [2]. Labeling on the dataset is done crowdsourcing. "0" represents non-nutrients in pandemic times and "1" represents fulfilling nutrition in pandemic times. Then the dataset used will be divided into a ratio of 75-25, where 75% is used as training data, and 25% as test data. Fig. 1 is a display of the food composition dataset.

B. Correlation Analysis Stage

In the correlation analysis process, input data in the form of a food composition dataset to form a correlation matrix for the relationship of each nutrient indicator in children's food using the Pearson product-moment correlation coefficient. The Pearson product-moment correlation coefficient has the following conditions for the degree of proximity [11]:

- Coefficient value 0 = There is no relationship at all.
- Coefficient value 1 = Perfect relationship.
- Coefficient value > 0 to < 0.2 = Very weak relationship.
- Coefficient value 0.2 to < 0.4 = Weak relationship.
- Coefficient value 0.4 to < 0.6 = the relationship is quite strong.
- Coefficient value 0.6 to < 0.8 = Strong relationship.
- Coefficient value 0.8 to < 1 = Very strong relationship.
- A negative value means determining the direction of the opposite relationship.

Coefficient values of -1 and 1 are perfect relationships, coefficient values of 0 or close to 0 are considered to have no relationship between the two variables tested. The formula used to calculate the correlation coefficient is in (1).

$$r = \frac{n(\sum_{i=1}^{n} xiyi) - (\sum_{i=1}^{n} xi)(\sum_{i=1}^{n} yi)}{\sqrt{n\sum_{i=1}^{n} xi^2 - (\sum_{i=1}^{n} xi)^2}\sqrt{n\sum_{i=1}^{n} yi^2 - (\sum_{i=1}^{n} yi)^2}} \quad (1)$$

r = correlation coefficient between x and y variables

x = values of the horizontal axis in the coordinate plane

y = values of the vertical axis in the coordinate plane

$\sum xy$ = the sum of the values of x and y

x^2 = square of the value x

y^2 = square of the value y

There will be an example of correlation analysis calculations on data in Table I.

TABLE 1
ATTRIBUTE CORRELATION

No	Type	Group	Type x Group	$Type$	$Group^2$
1	1	4	4	1	16
2	1	4	4	1	16
3	1	6	6	1	36
4	2	9	18	4	81
5	2	9	18	4	81
Total	**7**	**32**	**50**	**11**	**230**

Nama pangan	Air (g)	Kalori	Protein (g)	Lemak (g)	Karbohidrat (g)	Serat (g)	Abu (g)	Kalsium (Ca)	Status
Abon ikan	6.4	435	27.2	20.2	36.1	0	10.1	0	1
Agar - agar	17.8	0	0	0.2	0	0	0	400	0
Bakpia, kue	38.9	272	8.7	6.7	44.1	0.9	1.6	194	1
Bakso	83.6	76	4.1	2.5	9.2	0	0	14	1
Chikiniku, masakan	65.4	143	9.8	1.4	22.8	0.4	0.6	301	0
Coklat manis, batang	1.4	527	2	29.8	62.7	6.5	4.1	63	1
Dendeng mujahir, goreng	6.5	598	74.3	26.9	9.2	0	13.1	1.957	1
Donat	23.2	357	9.4	10.4	56.5	0	0	7	0
Martabak manis	40	265	4.7	5.5	49.3	0	0	0	1
Sate ayam	49	227	43.1	6.1	1.8	0	0	17	1

Fig. 1 Food composition dataset

By using label encoding, type “1” represented processed and type “0” represented raw. Group “4” represented fish, group “6” represented sugar, and group “9” represented vegetables. Calculate the correlation coefficient between type attributes and group attributes with equation 1.

$$r = \frac{(5 \times 50) - (7 \times 32)}{\sqrt{(5 \times 11) - (7)^2}\sqrt{(5 \times 230) - (32)^2}}$$
$$r = 0.95$$

C. Attribute Reduction Stage

The process of reducing dimensions or attributes on the food composition dataset is carried out after going through the correlation analysis stage. Attribute reduction using the PCA method is carried out on the original attribute set to obtain the minimum set of attributes, then the acquisition attribute with the maximum ratio is selected after attribute reduction [12].

The stages of PCA completion consist of calculating the mean value, normalizing the data, calculating covariance, and determining vectors and eigenvalues [13]. The following are the stages of PCA using the sample dataset in Table II.

By using label encoding, type “1” represented processed and type “0” represented raw. Group “4” represented fish, group “6” represented sugar, and group “9” represented vegetables.

- Calculating the mean (mean) of the dataset
  The equation to calculate the mean value is in (2).

$$\bar{x} = \frac{\sum_{i=1}^{n} fx}{n} \quad (2)$$

$\bar{x}$ = average value
$\sum fx$ = sum of all data values
n = number of data

Examples of mean calculations on water attributes:

$$\bar{x} = \frac{11.6 + 6.4 + 17.8 + 86.7 + 90.6}{5} = 42.62$$

- Data normalization
  The equation for normalizing the data is in (3).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

x_{new} = new x variable
x = variable x
x_{max} = maximum x value
x_{min} = minimum x value

Examples of data normalization on water 1st data:

$$x_{new} = \frac{11.6 - 6.4}{6.4 - 90.6} = 0.06$$

- Calculating Covariance
  Covariance calculation for each dimension was performed. The equation for calculating covariance on the data is as follows:
  a. Calculate the covariance of the data population is in (4).

$$Cov\ (x, y) = \frac{\sum_{i=1}^{n}(xi - \bar{x}) \times (yi - \bar{y})}{N} \quad (4)$$

  b. Calculating the covariance of the data sample is in (5).

$$Cov\ (x, y) = \frac{\sum_{i=1}^{n}(xi - \bar{x}) \times (yi - \bar{y})}{(N-1)} \quad (5)$$

$Cov\ (x, y)$ = covariance between x and y variables
x = variable x
y = variable y
$\bar{x}$ = mean of the value x
$\bar{y}$ = mean of the value y
N = number of data

Examples of covariance calculation on type and group attributes of normalization result using equation (5):

$$Cov\ (x,y) = \frac{\big((-0.73 - 0) \times (-0.96 - 0)\big) + \big((-0.73 - 0) \times (-0.96 - 0)\big) + \big((-0.73 - 0) \times (-0.96 - 0)\big) + \big((1.10 - 0) \times (1.04 - 0)\big) + \big((1.10 - 0) \times (1.04 - 0)\big)}{5 - 1} = 0.95$$

- Calculating the values and eigenvectors of the covariance matrix

TABLE II
SAMPLE DATASET

No	Type	Group	Water	Calorie	Protein
1	1	4	11.6	513	23.7
2	1	4	6.4	435	27.2
3	1	6	17.8	0	0
4	2	9	86.7	45	1.1
5	2	9	90.6	37	4.4

Suppose A is a covariance matrix, is a vector, and is a scalar that satisfies Aν = λν, then is called the eigenvalue associated with the eigenvector of A as in (6).

$$Av - \lambda v = 0; (A - \lambda I)\ v = 0 \quad (6)$$

Av = eigenvector matrix A
λv = eigenvalue of matrix A
A = matrix A
λ = eigenvalue
v = eigenvector

Simplify the covariance matrix to det(A-λI) matrix as in equation 6, then perform column reduction on the det(A-λI) matrix. Then we get some eigenvalues, namely $\lambda 1 = 3.3894$, $\lambda 2 = 0.588$, $\lambda 3 = 0.0214$, $\lambda 4 = 0.0012$ and $\lambda 5 = 0$.

- Determine the principal components (PCs)

At this stage, the eigenvalues (λ) are sorted first from the largest to the smallest value. Then, by using the PCs selection criteria from Kaiser-Guttman, the larger eigenvalues or greater than 1 should be retained. So from the calculation results obtained three principal components (PCs), namely PC1, PC2, and PC3 as new dimension data (Table III).

D. Design of the LVQ model

Informing a classification model and carrying out the classification process with the LVQ network model. LVQ conducts learning on the competition layer as shown in Fig. 2.

Based on Fig. 2, X_1, X_2 - Xn is an input vector. These input vectors are connected to the W_1 and W_2 weight vectors. X - W is the process of calculating the distance between input vectors and weight vectors based on the activation functions F_1 and F_2. The activation function F_1 will map the output vector (y_in1) to class $Y_1 = 1$ if $X - W_1 < X - W_2$, and map to class $Y_1 = 0$ if otherwise. Similarly, the activation function F_2 will map the output vector (y_in2) to $F_2 = 2$ if $X - W_2 < X - W_1$, and map class $Y_2 = 0$ if otherwise. Y_1 is the first-class output and Y_2 is the second-class output. The output vector class obtained as a result of this competition layer depends on the distance between the input vectors and the weight vectors [14].

The stages of LVQ completion consist of parameter initialization, calculating the euclidean distance, updating the weights, and determining the optimal weight [15]. The following are the stages of LVQ using principal components in Table IV.

- Initialize data and parameters

Set the initial weight value (Wij) where i = weight and j = weight input variable. There are two target classes, namely 1 and 0, so initialize the training data with $W_1 = (2.06, -0.40, -0.21)$ as class 1 and $W_2 = (-2.07, -0.32, -0.09)$ as class 0.

TABLE III
PRINCIPAL COMPONENTS

No	PC1	PC2	PC3
1	2.06	-0.40	-0.21
2	2.09	-0.33	0.21
3	-0.06	1.53	0
4	-2.07	-0.32	-0.09
5	2.02	-0.48	0.10

TABLE IV
PRINCIPAL COMPONENTS WITH CLASS

No	PC1	PC2	PC3	Class
1	2.06	-0.40	-0.21	1
2	2.09	-0.33	0.21	1
3	-0.06	1.53	0	1
4	-2.07	-0.32	-0.09	0
5	2.02	-0.48	0.10	1

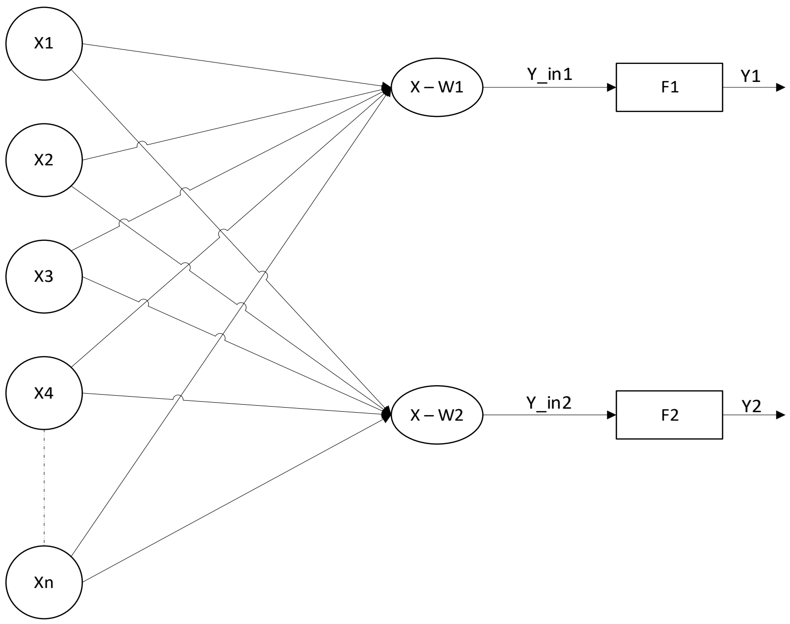

Fig. 2 LVQ architecture

The initial stage is initializing the alpha value (learning rate), decrement alpha, minimum alpha, and maximum epoch as input. The equation for determining the epoch and decrement alpha is in (7) and (8).

$$epoch = epoch + 1 \quad (7)$$

$epoch$ = iteration

and

$$decrement\ (\alpha) = \alpha - \alpha * 0.1 \quad (8)$$

$decrement\ (\alpha)$ = decrease in alpha value
α = alpha value or learning rate

Then initialization, the number of maximum epochs is 2 and the alpha value is 0.1. Examples of decrement alpha calculations on 2 epochs:

$$decrement\ (\alpha) = 0.1 - 0.1 * 0.1$$
$$decrement\ (\alpha) = 0.09$$

- Calculating Euclidean distance

The equation for calculating the Euclidean distance is in (9).

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (yi - xi)^2} \quad (9)$$

$d(x, y)$ = euclidean distance between x and y variables
x = variable x
y = variable y
n = number of data

Examples of Euclidean distance calculation on 2nd data on table 4 with the initial weight:

$$d(W_1) = (2.06 - 2.09)^2 + (-0.40 - (-0.33)^2 + (-0.21 - 0.21)^2 = 0.18$$
$$d(W_2) = (-2.07 - 2.09)^2 + (-0.32 - (-0.33)^2 + (-0.09 - 0.21)^2 = 17.39$$

- Update the weight

Update the weight of the class with the smallest euclidean value. The weight update is carried out with the condition that Cj is the output category or vector class for training and Cx is the category or class that corresponds to the j output unit, as in (10) and (11).

a. If Cj = Cx :

$$Wij(new) = Wij(old) + \alpha[xi - Wij(old)] \quad (10)$$

b. If Cj ≠ Cx :

$$Wij(new) = Wij(old) - \alpha[xi - Wij(old)] \quad (11)$$

$Wij(new)$ = new weight
$Wij(old)$ = old weight
α = alpha value or learning rate
xi = variable x

Since the distance value $d(W_1) < d(W_2)$ and the given input vector target is 1, then the target class = the winning class in the competition layer. Examples of weight update calculation on 2nd data:

$$W_{j1}(new) = 2.06 + 0.1[2.09 - 2.06] = 2.06$$
$$W_{j2}(new) = (-0.40) + 0.1[-0.33 - (-0.40)] = -0.39$$
$$W_{j3}(new) = (-0.21) + 0.1[0.21 - (-0.21)] = -0.17$$

- Determining the optimal weight value

After the weight renewal process is carried out, the optimal weight value will be obtained in the form of an optimal weight matrix (Table V).

E. Classification Stage

The optimal weight matrix that has been obtained is used as a weight vector for the calculation of the input vector at the time of classification. The input vector is initialized as x. Then calculate the Euclidean distance to X with the weight of the optimal weight matrix. From the calculation results, if the Euclidean distance obtained is close to the "1" class or the value obtained W_1 is smaller, the input vector enters the "1" class or is classified as the "1" class. Similarly, if the Euclidean distance obtained is close to the "0" class or the value obtained W_2 is smaller, the input vector enters the "0" class or is classified as the "0" class.

F. The Block Diagrams

The block diagram describes the system process flow from input to output, as shown in Fig. 3.

TABLE V
OPTIMAL WEIGHT MATRIX

	x1	**x2**	**x3**	**Class**
W_1	1.87	-0.21	-0.13	1
W_2	-2.31	-0.51	-0.09	0

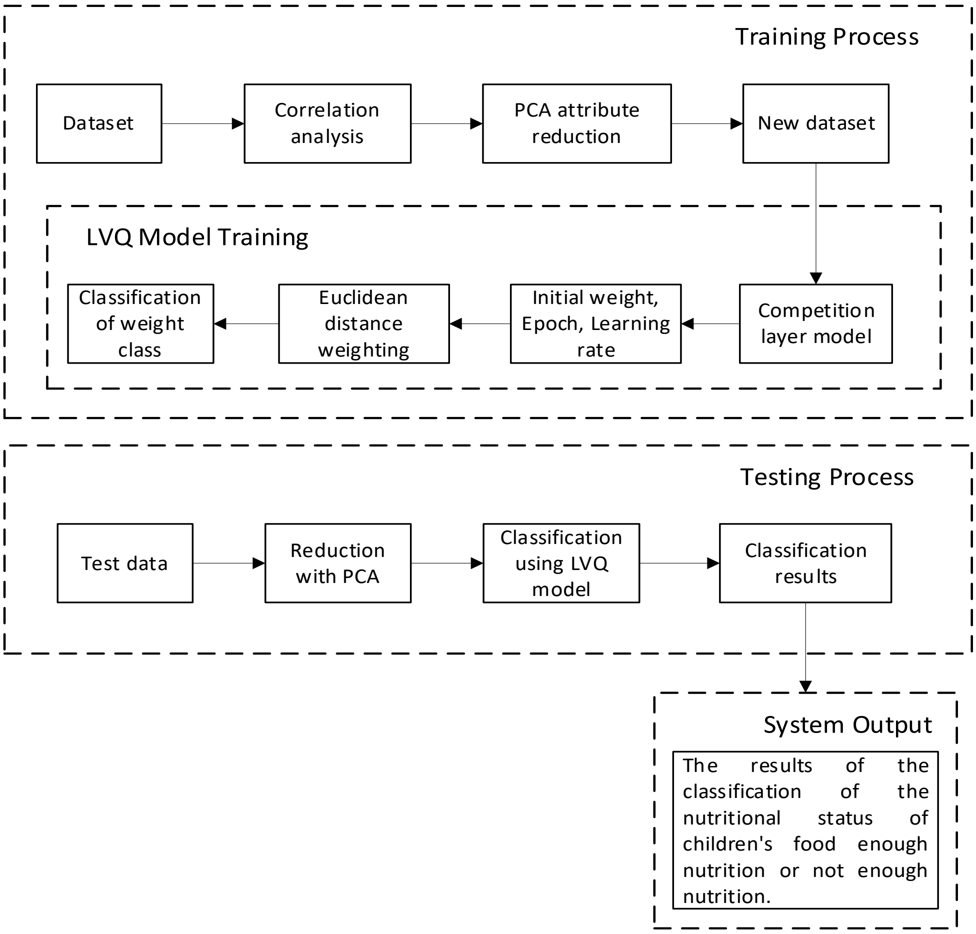

Fig. 3 Block Diagram System

Fig. 3 is shown the dataset as training data. Then the correlation analysis between attributes. Attributes that are correlated will go through a reduction stage using the PCA method. The new reduced dataset is then used for the classification process, in the process of data classification initialized as a vector. Each vector is ready to be formed at the competition layer based on the LVQ architecture model. Then the process of initializing the initial weight, epoch, and learning rate (α). Updating the vector weights by calculating the Euclidean distance. The classification results that have been obtained in the training process are stored for the nutritional classification of children's food in the testing process. The output of the system is the result of the classification of the nutritional status of children's food enough nutrition or does not enough nutrition in the pandemic.

G. Testing

At this stage, after going through the preprocessing of PCA and the LVQ model classification process, it will produce the system output in the form of the nutritional status of children's food that has been predicted. For performance testing of classification models that have been created, testing is done using test data of 25% of the entire dataset. Model performance testing is conducted using accuracy, precision, recall, and f–measure or the f1 score.

III. RESULTS AND DISCUSSION

The study aims to measure the accuracy of models based on nutritionally sufficient categories and insufficient nutrition in pandemic times by implementing PCA and LVQ methods. For training, data is divided by a ratio of 75% data train and 25% data test. The model was built using 100 epochs, an alpha value of 0.1, a random state value of 24, and 2 initial weights representing 2 target classes.

A. Correlation Analysis

Correlation analysis was conducted to determine the relationship between each nutrient indicator in children's food using the Pearson product-moment correlation coefficient (Fig. 4).

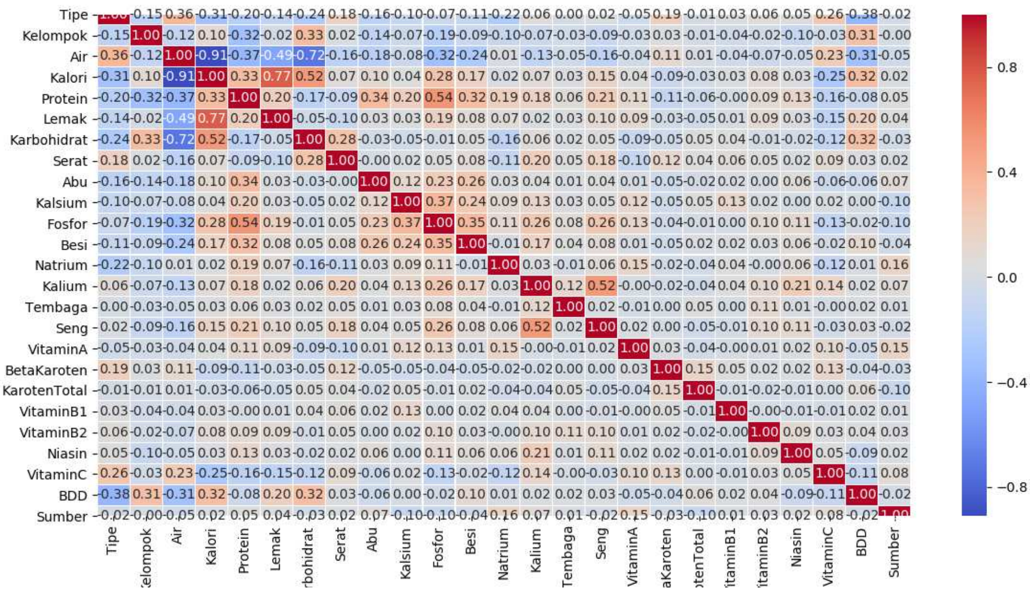

Fig. 4 Matrix correlation

Based on Fig. 4 the results of the correlation analysis and the provision of the degree of proximity of the Pearson product-moment correlation coefficient, each nutritional indicator to the nutritional indicator itself is worth 1, meaning that it has a perfect relationship. The calorie and fat indicators are 0.77, meaning that the higher the calorie content, the higher the fat content with a degree of closeness of 0.77 or a strong relationship. Indicators of protein and phosphorus are worth 0.54, meaning that the higher the protein content, the higher the phosphorus content with a degree of closeness of 0.54 or a strong enough relationship. The water and calorie indicator are worth -0.91, meaning that the higher the water content, the lower the calorie content with a degree of proximity of 0.91 or a very strong relationship. The water and carbohydrate indicator is worth -0.72, meaning that the higher the water content, the lower the carbohydrate content with a degree of closeness of 0.72 or a strong relationship.

B. Attribute Reduction

Dataset attribute reduction was performed using the PCA method. From the acquisition of eigenvalues, the most influential attribute ranking is carried out. The ranking process is based on the eigenvalues from the largest to the smallest and the cumulative variance value. Table VI shows the ranking of each attribute based on eigenvalues and cumulative values.

Based on Table VI, of the 25 indicators, there are 10 indicators with the largest eigenvalues representing the 10 selected principal components based on the Kaiser-Guttman rule states that components based on eigenvalues greater than 1 should be retained. Namely type, group, water, fat, calcium, phosphorus, potassium, copper, zinc, and vitamin A. Fig. 5 shows the obtained principal components.

TABLE VI
ATTRIBUTE RANKING

No	Eigenvalues	Cumulative	Attribute Name
1	3.7666	0.14499	Type
2	2.5838	0.24453	Group
3	1.8995	0.31757	Water
4	1.4596	0.37376	Fat
5	1.2683	0.42258	Calcium
6	1.2243	0.46971	Phosphorus
7	1.1489	0.51394	Potassium
8	1.1088	0.55662	Copper
9	1.0657	0.59764	Zinc
10	1.0360	0.63752	Vitamin A
11	0.9435	0.67384	BDD
12	0.9326	0.70974	Source
13	0.8569	0.74413	Vitamin C
14	0.7927	0.77711	Niacin
15	0.7238	0.80763	Vitamin B2
16	0.6789	0.83549	Vitamin B1
17	0.6742	0.86163	Total Carotene
18	0.6157	0.88758	Beta-Carotene
19	0.5483	0.91128	Sodium
20	0.5164	0.93238	Iron
21	0.4463	0.95227	Ash
22	0.4003	0.96945	Fiber
23	0.3596	0.98485	Carbohydrates
24	0.0255	0.99869	Calories
25	0.0083	0.99968	Protein

```
Principal Components
          PC1       PC2       PC3  ...       PC9      PC10  Status
0   -2.529407 -0.893785  1.604229  ...  0.231375  0.139900       1
1   -2.571031 -0.025370  0.463804  ...  1.390290 -1.011529       1
2    0.308629 -0.321542  1.099532  ... -0.050606  0.811295       1
3    1.479703  0.040548 -0.927066  ... -0.209141 -0.239122       0
4    1.949057  0.308825 -0.587549  ... -1.364313  0.000746       1
...       ...       ...       ...  ...       ...       ...     ...
1163  1.626877 -0.901217  0.500728  ... -0.546890  0.005094       1
1164  1.725670 -0.880667  0.590655  ... -0.562389  0.035274       1
1165  2.195131 -0.082199 -0.236041  ... -0.695533  0.176640       1
1166 -0.484872 -2.227075  0.130437  ... -0.429579 -0.035341       1
1167  1.134393 -0.627781  0.797141  ... -0.458967  0.180373       1
```

Fig. 5 Principal Components

C. Model Training

At this stage, the training model is carried out with the learning vector quantization network architecture. Parameter initialization is carried out for the training model with a learning rate (α) value of 0.1 with a total of 100 epochs. In the model training process, using features from the results of PCA reduction and without PCA reduction. Wherewith, PCA reduction uses 10 features or principal components, while without PCA uses 25 initial features or attributes. The results of the training model, the optimal weight matrix obtained from the renewal of the weight vector against the value of the principal components (input vector) in the competition layer, which is then used to determine the classification results at the testing stage. The optimal weight matrix based on principal components in Fig. 6.

Experiment to classify the nutritional status of food on the data train with training model (Table VII), while experiment to test the training model performance based on the number of test data (Table VIII).

D. LVQ and LVQ-PCA Performance Testing

A comparison test of the performance of the LVQ classification model was carried out before the reduction and after the reduction with PCA was shown in Table IX.

From the results of testing the formation of a classification model with the application of the PCA dimension reduction method, the nutritional classification of children's food with the LVQ model works better than before the reduction. Where before the PCA reduction, of the 292 test data there are 9 TP or sufficient nutritional test data that is predicted as sufficient nutrition and 10 FP or test data that is insufficient nutrition or not but predicted as enough, while after being reduced there are 119 TP or sufficient nutritional test data that is predicted as sufficient nutrition and 100 FP or test data that is insufficient nutrition or not, but predicted as enough.

```
Weight
 [[ 1.76000000e+02  3.01231838e+00  5.81870156e-01 -1.06297321e+00
   2.90679944e-01  2.82339724e-01 -9.94136270e-01 -6.44799801e-01
  -7.06485978e-01 -2.24283790e-01 -5.23874784e-01]
 [ 2.63000000e+02  1.57129208e+00  1.08121917e+00 -4.18016658e+00
   1.37864328e+00 -1.33285500e+00 -1.89397035e+00 -2.54478677e+00
  -1.06403426e+00 -2.87527040e+00 -6.05635814e-01]]
```

Fig. 6 Optimal Weight Matrix

TABLE VII
CLASSIFICATION OF FOOD NUTRITIONAL STATUS

No	Experiment	Results	Target	Classification Results
1	1	1	1	True
2	2	1	1	True
3	3	1	1	True
4	4	0	1	Not
5	5	1	1	True
6	6	1	1	True
7	7	1	1	True
8	8	1	1	True
9	9	1	1	True
10	10	1	1	True
11	11	1	0	Not
12	12	1	0	Not
13	13	1	0	Not
14	14	1	0	Not
15	15	1	0	Not
16	16	0	0	True
17	17	0	0	True
18	18	1	0	Not
19	19	1	0	Not
20	20	1	0	Not

TABLE VIII
PERFORMANCE BASED ON NUMBER OF TEST DATA

No	Amount of Test Data	TP	FP	FN	TN	Accuracy
1	117	39	44	9	25	57%
2	176	67	65	13	31	56%
3	234	93	86	18	37	57%
4	292	119	100	22	51	58%
5	351	126	108	48	69	57%
6	386	157	126	32	71	59%
7	468	183	149	38	98	60%

Description: TP (true positive), TN (true negative), FP (false positive), FN (false negative).

TABLE IX
LVQ DAN LVQ-PCA TESTING

	METHOD							
	LVQ				LVQ-PCA			
Accuracy	0.527				0.582			
Precision	0.474				0.543			
Recall	0.066				0.844			
F1 Score	0.115				0.661			
Confusion	TP	FP	FN	TN	TP	FP	FN	TN
Matrix	9	10	128	145	119	100	22	51

E. Performance Testing of Dimension Reduction Method

Testing by Independent Component Analysis (ICA) and Factor Analysis (FA) dimension reduction methods in the same dataset. The ICA dimension reduction method shows an accuracy value of 0.503, a precision value of 0.464, a recall value of 0.636, a f1 score of 0.537, and a consumption time of 5 seconds. The dimension reduction method of factor analysis shows an accuracy value of 0.517, a precision value of 0.469, a recall value of 0.508, a f1 score of 0.487, and a consumption time of 4.1 seconds. A comparison test of the performance testing of the dimension reduction method with PCA, ICA, and FA was shown in Table X.

The test results show that the PCA dimension reduction method is better than ICA and FA in the nutritional classification of children's food during the pandemic. The PCA dimension reduction method shows an accuracy value of 0.582, a precision value of 0.543, a recall value of 0.844, a f1 score of 0.661, and a consumption time of 4.1 seconds.

IV. CONCLUSION

Based on the results of the tests that have been carried out, the implementation of the PCA dimension reduction method in the LVQ classification model can classify the nutritional status of children's food in the pandemic. From the comparison test for the reduction method, the PCA method works better than the ICA and FA dimension reduction methods, where PCA can reduce 25 nutritional attributes of food to 10 principal components with an accuracy rate of 58% with a long processing duration of 4.1 seconds for the classification model. In developing the system in the future to improve the performance, there are several suggestions. Based on previous research, the LVQ method has better performance because the number of attributes used is less, for future research that will apply the LVQ method should use more attributes or use classification methods other than the LVQ method.

TABLE X
REDUCTION METHOD TESTING

	METHOD		
	PCA	**ICA**	**FA**
Reduction Result	10 PCs	16 ICs	10 Factors
Accuracy	0.582	0.503	0.517
Precision	0.543	0.464	0.469
Recall	0.844	0.636	0.508
F1 Score	0.661	0.537	0.487
Consume Time	4.1 sec	5 sec	4.1 sec

REFERENCES

[1] Y. I. Ayseli, N. Aytekin, D. Buyukkayhan, I. Aslan, and M. T. Ayseli, "Food policy, nutrition and nutraceuticals in the prevention and management of COVID-19: Advice for healthcare professionals," *Trends Food Sci. Technol.*, vol. 105, no. September, pp. 186–199, 2020.

[2] Kemenkes, "Final-Panduan-Gizi-Seimbang-Pada-Masa-Covid-19-1.Pdf," *Panduan Gizi Seimbang Pada Masa Pandemi COVID-19*. p. 31, 2020.

[3] E. Budianita and W. Prijodiprodjo, "Penerapan Learning Vector Quantization (LVQ) untuk Klasifikasi Status Gizi Anak," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, p. 155, 2013.

[4] R. Harimurti and R. Rahmawati, "RANCANG BANGUN APLIKASI PEMENUHAN GIZI BAGI IBU HAMIL MENGGUNAKAN LOGIKA FUZZY TSUKAMOTO," *UNESA*, pp. 9–17, 2013.

[5] E. T. Lestari, "PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI STATUS GIZI OBESITAS ANAK DISABILITAS," 2019.

[6] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.

[7] G. Rahayu and M. Mustakim, "Principal Component Analysis Untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru Di Indonesia," *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 201–208, 2017.

[8] R. P. Furi, M. Si, and D. Saepudin, "Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII," *ISSN 2355-9365 e-Proceeding Eng.* , vol. 2, no. 2, pp. 1–10, 2015.

[9] R. S. Nasution, "Analisis Faktor dengan Principal Component Analysis dalam Faktor-Faktor yang Memengaruhi Pemberian Makanan Tambahan pada Bayi Usia 0-6 Bulan di Kelurahan Kisaran Timur Kecamatan Kota Kisaran Timur Kabupaten Asahan Tahun 2018," *Univ. Sumatera Utara*, 2018.

[10] N. Savanti, W. Gotami, and R. K. Dewi, "Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Latent Semantic Analysis," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2821–2828, 2018.

[11] Y. Yudihartanti, "Analisa Korelasi Mata Kuliah Penelitian Dengan Tugas Akhir Menggunakan Model Product Moment," *Progresif J. Ilm. Komput.*, vol. 13, no. 2, pp. 1691–1696, 2018.

[12] H. M. Nawawi, S. Rahayu, M. J. Shidiq, and J. J. Purnama, "Algoritma C4.5 Untuk Memprediksi Pengambilan Keputusan Memilih Deposito Berjangka," *J. Techno Nuasa Mandiri*, vol. 16, no. 1, pp. 65–72, 2019.

[13] R. Susetyoko and E. Purwantini, “Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil,” vol. 2010, no. Ies, pp. 978–979, 2010.

[14] M. Kaden, M. Lange, D. Nebel, M. Riedel, T. Geweniger, and T. Villmann, “Aspects in classification learning - Review of recent developments in learning vector quantization,” *Found. Comput. Decis. Sci.*, vol. 39, no. 2, pp. 79–105, 2014.

[15] E. Budianita and Novriyanto, “Klasifikasi Status Gizi Balita Berdasarkan Indikator Antropometri Berat Badan Menurut Umur Menggunakan Learning Vector Quantization,” *(Seminar Nas. Teknol. Informasi, Komun. dan Ind. SNTIK*, no. November, pp. 213–220, 2015.