CLASSIFICATION OF THE NUTRITION STATUS TODDLER USING THE SVM METHOD (CASE STUDY: BANJARAGUNG VILLAGE, BARENG, JOMBANG)

[1]EKO PRASETYO, [2]RAHMAWATI FEBRIFYANING TIAS, [3]EFILAH RISQI MAULANA,
Informatics Engineering University of Bhayangkara
Surabaya, Indonesia
Ahmad Yani Street 114 Surabaya, Post Code 60231
e-mail : [1]eko@ubhara.ac.id, [2]rahma.febri86@gmail.com, [3]risqimaulana@gmail.com

ABSTRACT

Improving the health status of children under five is very necessary in determining the next generation of the Indonesian nation. One of the efforts that can be realized is to maintain the nutrition of children under five in the community. Balanced nutrition can increase immunity and increase intelligence so as to make normal growth. In social life, nutritional status is obtained through anthropometric measurements at a posyandu where people generally use the BB/U index or body weight compared to age to determine the nutritional status of toddlers. This study aims to make it easier to identify the nutritional status of toddlers using Data Mining with Support Vector Machine (SVM). system built with PHP programming language and postgreSQL database. This study uses data on 314 toddlers in 4 groups of posyandu in the village. The data was tested 2 times, the first with a 50:50 comparison and the second 70:30 for training data and testing data. The results showed an accuracy of 96% and 98%, in other words, SVM was categorized as good for testing the nutritional status of children under five.

Keywords: Toddler Nutrition, Classification, Support Vector Machine.

I. INTRODUCTION

The Integrated Service Post (Posyandu), Banjaragung Village, Bareng Jombang District is a self-help activity from the community in the health sector that seeks to improve the health status of toddlers. The indicators used in this study include Age (U), Gender (JK), Weight (BB), Height (TB), Head Circumference (LK), Chest Circumference (LD) and will be matched based on the status reference book table. toddler nutrition. Meanwhile, healthy children get older, the more their weight and height increase. So that health workers as well as parents know whether the toddler is classified as a child with malnutrition (bad nutrition) or normal (good nutrition).

According to Riskesdas (Basic Health Research) the problem of under-five nutrition in Indonesia is still a national problem. Children under the age of five are a group that is vulnerable to health and nutrition problems, including the problem of lack of protein energy. So that toddlerhood is a very important period of life and needs serious attention. To prevent this, adequate nutritional intake is needed and in accordance with the body's needs, efforts to achieve optimal nutritional status can be done by consuming carbohydrates, fats, proteins, vitamins and minerals in accordance with the nutritional adequacy figures in the context of metabolic processes, transformation, and interaction with other substances in order to achieve the body's energy balance.

Classification is the process of logically grouping objects according to their similarity characteristics (Glossary). Classification technique is an approach to perform the classification function in data mining to classify data. This classification technique can also be used to make predictions or unknown information One method that can be used is the Support Vector Machine. The concept of SVM can be explained simply as an attempt to find a hyperplane that can separate data sets according to their class. This of course can be done to facilitate the process of determining the nutritional status of toddlers in Banjaragung Village, Bareng, Jombang. So that the expected results are in the form of prediction results which can later be used by cadre officers in predicting the nutritional status of toddlers. Based on the description above, in this final project the author conducts research to create a system that can provide a diagnosis of the problem with the title "Classification of the Nutritional Status of Toddlers Using the SVM Method Case Study in Banjaragung Village with Jombang"

II. METHOD

Every village cadre certainly wants to improve its quality. One way is to make healthy toddlers. In Banjaragung Village, Bareng Jombang District, the health level of each toddler has a difference. This can be influenced by several factors including: family economy, food patterns consumed, parenting patterns, environmental influences, and genes (heredity). There needs to be a new breakthrough to design a nutritional classification system for toddlers which is used by village cadres as the basis for classification by utilizing toddler data.

By utilizing a data mining system, especially the SVM method, an application design was made to classify toddler nutrition based on data from village cadres. There are two categories of nutritional status in toddlers, namely good nutrition (normal) and poor nutrition (bad nutrition). With this application, it is hoped that it can be used as a user, namely village cadres for nutritional classification in toddlers. There are several processes in the classification using SVM, namely the preprocessing process, training, and finally the testing process.

2.1 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is one of the existing classification methods. SVM was developed by Boser, Guyon, Vapnik, and was first presented in 1992 at the Annual Workshop on Computational Learning Theory. The basic concept of SVM is to find the hyperplane dividing line by optimizing the hyperplane, and maximizing the margin between the two classes. In SVM not all data is used as training data but only some. The data used as training data is called the support vector. The first thing to do in SVM is to determine the data that will be used as a support vector to build a classification construction model. In this determination there is a quadratic programming problem that can be solved by the sequential method. After building the classification construction model, then testing the data or testing the test data to determine the class. In some cases the data cannot be classified linearly, kernel functions are used, including Gaussian RBF(1) and degree 2(2) polynomials.

The function of the kernel is to map data into a higher dimension, the illustration can be seen in Figure 1
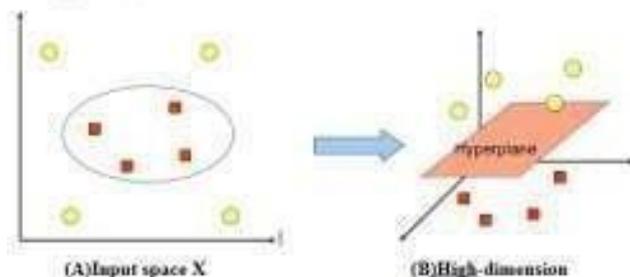
Figure 1 kernel function in mapping data

2.2 Normalization Process

Normalization is the process of grouping data attributes to minimize data redundancy in a database so that it can work optimally. The formula used in this normalization is:

New data = (nilai X – X min) / (X max – X min)

X Min = the smallest number of all data variables.
X Max = the largest number of all data variables.

Table data normalization

No	U	Norm U	BB	Norm BB	TB	Norm TB	LK	Norm Lk	LD	Norm LD
1	4	0.75	12	0.6	89	0.625	43	0.643	42	0.643
2	2	0.25	12	0.6	82	0.450	39	0.357	39	0.429
3	1	0	6	0	64	0.000	34	0.000	33	0.000
4	5	1	16	1	104	1.000	48	1.000	47	1.000
5	1	0	7	0.1	75	0.275	34	0.000	33	0.000
6	3	0.5	10	0.4	90	0.650	42	0.571	41	0.571
7	5	1	16	1	102	0.950	48	1.000	47	1.000

8	3	0.5	9	0.3	78	0.350	43	0.643	42	0.643
9	2	0.25	10	0.4	79	0.375	43	0.643	42	0.643
10	4	0.75	13	0.7	100	0.900	44	0.714	43	0.714

2.3 Training Process

At this stage using kernel functions, constants, and max iterations to form a matrix to find the support vector

$$
\begin{array}{l}
X_1 \quad [\; 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \;] \\
X_2 \quad [\; 0.25 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \;] \\
X_3 \quad [\; 0.5 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \;] \\
X_4 \quad [\; 0.75 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \;] \\
X_5 \quad [\; 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \;] \\
X_6 \quad [\; 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \;] \\
X_7 \quad [\; 0.25 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \;] \\
X_8 \quad [\; 0.5 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \;] \\
X_9 \quad [\; 0.75 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \;] \\
X_{10} \quad [\; 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1 \;]
\end{array}
$$

By using the data in the matrix above, the dot-product of all data is calculated as in the following example:

$$X_1 X_1 = [\; 0\ 1\ 1\ 1\ 0\ 1\ ] [\; 0\ 1\ 1\ 1\ 0\ 1\ ]$$

$$= 4$$

The above calculations are carried out for all data. After calculating the multiplication between the data as above, then calculating the polynomial function with a constant value (c) = 1 and the polynomial value degree (d) = 2.

$$K(1,1) = (X_1 X_1 + c)^d$$

$$K(1,1) = (4 + 1)^2$$

$$K(1,1) = 25$$

The calculation is carried out up to K(10,10).

	1	2	3	4	5	6	7	8	9	10
1	25	25	16	9	4	1	1	4	9	16
2	25	25.62891	17.01563	10.16016	5.0625	1	1.128906	4.515625	10.16016	18.0625
3	16	17.01563	27.5625	19.14063	6.25	4	4.515625	5.0625	5.640625	12.25
4	9	10.16016	19.14063	20.81641	7.5625	4	4.785156	5.640625	2.441406	7.5625
5	4	5.0625	6.25	7.5625	9	1	1.5625	2.25	3.0625	4
6	1	1	4	4	1	4	4	4	1	1
7	1	1.128906	4.515625	4.785156	1.5625	4	4.253906	4.515625	1.410156	1.5625

Table	8	4	4.515625	5.0625	5.640625	2.25	4	4.515625	10.5625	5.640625	6.25
	9	9	10.16016	5.640625	2.441406	3.0625	1	1.410156	5.640625	12.69141	14.0625
	10	16	18.0625	12.25	7.5625	4	1	1.5625	6.25	14.0625	25

Kernel Function 10 data

calculation results from K(1,1) to K(10,10). After the parameters are calculated and fulfilled, a hessian matrix is formed with the parameter =1. Example of hessian matrix calculation with the following formula $((Y_i . Y_j) . (K_{1,1}+ 2))$. an example of calculating it for data X1,1 to X10,10 is as follows:

$$X(1,1) = ((Y_i . Y_j) . (K_{1,1}+ \lambda^2))$$

$$X(1,1) = ((-1 . -1) . (25+1^2))$$

$$X(1,1) = 26$$

$$X(1,2) = ((Y_i . Y_j) . (K_{1,2}+ \lambda^2))$$

$$X(1,2) = ((-1 . 1) . (25+1^2))$$

```
X(1,2) = -26
```

After calculating the Hessian matrix from X1.1 to X10.10 with the same formula, the results are as table.

	1	2	3	4	5	6	7	8	9	10
1	26	-26	-17	-10	5	2	2	-5	-10	17
2	-26	26.62891	18.01563	11.16016	-6.0625	-2	-2.12891	5.515625	11.16016	-19.0625
3	-17	18.01563	28.5625	20.14063	-7.25	-5	-5.51563	6.0625	6.640625	-13.25
4	-10	11.16016	20.14063	21.81641	-8.5625	-5	-5.78516	6.640625	3.441406	-8.5625
5	5	-6.0625	-7.25	-8.5625	10	2	2.5625	-3.25	-4.0625	5
6	2	-2	-5	-5	2	5	5	-5	-2	2
7	2	-2.12891	-5.51563	-5.78516	2.5625	5	5.253906	-5.51563	-2.41016	2.5625
8	-5	5.515625	6.0625	6.640625	-3.25	-5	-5.51563	11.5625	6.640625	-7.25
9	-10	11.16016	6.640625	3.441406	-4.0625	-2	-2.41016	6.640625	13.69141	-15.0625
10	17	-19.0625	-13.25	-8.5625	5	2	2.5625	-7.25	-15.0625	26

The next step is to calculate Ei, i, i with parameters = 0.1, C=1,max iteration = 6.

$$E_i = \sum_{j=1}^{N} \alpha_j \, D_{ij}$$

$E_i = (\alpha1 * D11) + (\alpha2 * D12) + (\alpha3 * D13) + (\alpha4 * D14) + (\alpha5 * D15) + (\alpha6 * D16) + (\alpha7 * D17) + (\alpha8 * D18) + (\alpha9 * D19) + (\alpha10 * D10)$

$E_i = (0*26) + (0*-26) + (0*-17) + (0*-10) + (0*5) + (0*2) + (0*2) + (0*-5) + (0*-10) + (0*17) E_i = 0$

$\delta\alpha_1 = \min \{ \max [\gamma (1 - E_1) , - \alpha_1], C - \alpha_1 \}$
$\delta\alpha_1 = \min \{ \max [0.15 (1 - 0) , - 0], 1 - 0 \}$

$\delta\alpha_1 = \min \{0.15 , 1\}$
$\alpha_1 = \alpha_1 + \delta\alpha_1$
$\alpha_1 = 0 + 0.15$
$\alpha_1 = 0.15$

The total calculation of Ei and i with the highest value of 0 in the new alpha is found in the 6th iteration calculation.

Table calculation Ei and i iteration 6

	Data	Ei	Δα	Max	Δαi	alpha baru
Iterasi 6	Data 1	14.34926	-2.00239	-1	-1	0
	Data 2	12.91178	-1.78677	0	0	0
	Data 3	11.50673	-1.57601	0	0	0
	Data 4	9.301922	-1.24529	-1	-1	0
	Data 5	7.667409	-1.00011	-0.76556	-0.76556	0
	Data 6	6.318323	-0.79775	-0.79775	-0.79775	0.089171
	Data 7	2.125911	-0.16889	-0.16889	-0.16889	0.709054
	Data 8	3.183908	-0.32759	-0.14355	-0.14355	0
	Data 9	10.22706	-1.38406	-0.98708	-0.98708	0
	Data 10	4.412135	-0.51182	-0.09296	-0.09296	0

Table is the result of the iteration calculation for the 6th time. After the iteration reaches a new maximum alpha, the data used as support vector data is data 6 and data 7. After knowing the data used as support vector, the next step is to find the value of w. After the value of w is met, it is used to find the value of b (bias) by taking one of each data from a different class. The value of b is -1.8781. If the data has been met, the SVM model is ready to be used for classification

2.4 Flowchat Toddler Nutritional Status Classification System
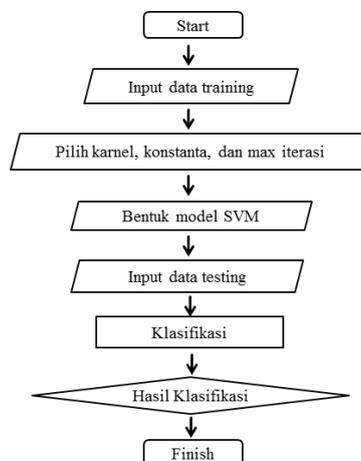
Figure 2 Flowchart Nutrition Status Classification System Toddlers

It is explained in Figure 2 Flowchat of the Toddler Nutritional Status Classification System which begins with start and inputs data for training modeling, then selects the kernel, constants, and max iterations needed and then forms the SVM model. then enter the data that will be made testing and process with classification calculations.

III. RESULT AND IMPLEMENTATION

The results of the experiments carried out using two kernels, namely the second degree polynomial and Gaussian RBF. The results of the experiments carried out can be seen in the table below

Table Training Results

Perco baan	Data Training	Data Model	Data Testing	Prediksi tepat	Prediksi Salah
1	157	75	157	151 (96%)	6 (4%)
2	220	107	94	92 (98%)	2 (2%)

Based on the results of the tests that have been carried out, it is found that, from each classification the less training data is used, the less accurate the accuracy value is, while the more training data used, the higher the accuracy value is precisely

IV. CONCLUSION

The conclusions obtained in this study are:

a.) Classification system using the SVM (Support Vector Machine) method can be implemented on the nutritional status of toddlers by taking data from Banjaragung Village, Bareng District, Jombang Regency so that it can identify toddlers in a healthy condition (good nutrition) or malnourished (bad nutrition).

b.) Variables that significantly influence the growth of children under five are weight and height with an accuracy of 98.5%.

REFERENCES

[1] Angriyasa Putu Wira, 2011, Klasifikasi tekanan dalam rongga kepala menggunakan Support VectorMachines.
[2] Almatsier, S., 2003. Prinsip Dasar Ilmu Gizi. Jakarta: PT Gramedia Pustaka Utama.
[3] Kurniawati Maya, 2014, Klasifikasi Dokumen E-Complaint Menggunakan Directed Acyclic GraphMulti-Class SVM.
[4] Noorniawati Vita Yulia, 2007, Metode Support Vector Machine Untuk Klasifikasi Pada SistemTemu Kembali Citra.
[5] Nurhayati Siti, 2015, Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine.
[6] Prasetyo Eko, 2012, Data Mining Konsep dan Aplikasi Mengunakan Matlab.
[7] Rahayu Monika S, 2015, Perbandingan Teknik Klasifikasi Multiclass Mengunkan Support VectorMachine.