Classification of nutritional status of toddlers using fuzzy k-nearest neighbor in every class (FK-NNC)

R U Izzah1,*, S Wahyuningsih1, and F D T Amijaya2
1Study Program of Statistics, Mulawarman University, Samarinda, Indonesia
2Department of Mathematics, Mulawarman University, Samarinda, Indonesia

*Corresponding Author: retnoizzah@gmail.com

Abstract. The purpose of this research is classify nutritional status of toddlers with algorithm fuzzy k-nearest in every class (FK-NNC). The FK-NNC algorithm is a modification concept of the nearest K neighbor for each class. The largest of the class on membership value will be selected as prediction classes. The optimal K value in FK-NNC algorithm uses k-fold cross validation. The optimal K values are searched by experimenting 1-fold cross validation, 4-fold cross validation and 10-fold cross validation. Accuracy rate of classify nutritional status of toddlers at the Wonorejo Health Center using the FK-NNC algorithm with an optimal K value. The results of this research obtained the optimal K value used in the FK-NNC algorithm at the Wonorejo Health Center is K=8 with 1-fold cross validation experiment. The value of K=8 was applied to the FK-NNC algorithm with a 1-fold cross validation experiment to predict class of nutritional status of toddlers in the Wonorejo Health Center. The percentage of accuracy produced as much as 100% with data proportion 90:10.

1. Introduction
Classification is a job of assessing data objects to include them in certain classes of a number of available classes. In the classification there are two main work done, namely the construction of the model as the original model to be stored as memory and the use of the model to classify or predict other data objects to be known in which class the data object is in the saved model [1].

In building the model during the training process, an algorithm is needed to build a model called the learning algorithm. Based on the way of training, classification algorithms are divided into two types, namely eager learner and lazy learner. Algorithms that are included in the lazy learner category only do a little training only store some or all of the training data then use it in the prediction process. The advantage of the lazy learner algorithm is to have a training process that runs quickly. Classification algorithms that fall into this category are K-Nearest Neighbor (K-NN), Fuzzy K-Nearest Neighbor (FK-NN), Linear Regression [3].

One of the problems faced by K-NN and FK-NN is that the election of K is difficult, the way of majority voting from neighboring K for large K values can result in large data deviations. If K is too small it can cause the algorithm to be too sensitive to noise. A good K value can be selected by parameter optimization using cross validation. Cross validation is a validity test that involves the use of comparative data to check the validity of the original estimates. In this study, the author uses the k-fold cross validation technique to find the optimal K value [1].
The fix made with the Fuzzy K-Nearest Neighbor in every Class (FK-NNC) algorithm is to modify the concept of the nearest K neighbor, from the origin only the nearest K neighbor from class C, to the nearest K neighbor for each class, so that there is a C×K neighbor found. Then calculating the membership value of the testing data in each class with the basis of the accumulation of distance K the nearest neighbor found. Classes with the largest membership value will be selected as prediction classes [2].

Prasetyo’s research (2012) "Fuzzy K-Nearest Neighbor in every Class for Data Classification" compared the FK-NNC method with K-NN and FK-NN to the Iris and Vertebral Column data sets. The FK-NNC method uses the nearest K neighbor in each class from a testing data, not from the nearest K neighbor such as K-NN or FK-NN. Researchers propose FK-NNC to improve performance accuracy at the time of prediction. Accuracy obtained from the tests conducted by researchers showed that the prediction accuracy given by FK-NNC was relatively higher than that of K-NN or FK-NN, which ranged from 82%-97%. The highest accuracy value obtained is a difference of 1% higher than the two comparison methods. FK-NNC can be an alternative method for K-NN, FK-NN and other variants to do data classification work [5].

Toddlers are often said to be a golden age because the success of growth is used as an illustration in increasing body size, but also used as an illustration of the continuity between intake and nutritional needs. One of the indicators that can find out the health level of toddlers themselves is to look at nutritional status. The FK-NNC is a classification method that will be able to calculate the nutritional status of children in the future.

Based on the description above, the authors are interested in reviewing the analysis by taking a case research of Classification of Nutritional Status of Toddlers in the Wonorejo Health Center by using the Fuzzy K-Nearest Neighbor in every Class (FK-NNC) Algorithm.

2. Literature citation

2.1. Proximity concept
To measure the incompatibility of two data with several attributes for each data used a distance quantity. There are many distance measurement models and the most commonly used is the Euclid distance [3]. Euclid distance can be calculated by the equation:

\[ d(x_i, y_j) = \sqrt{\sum_{j=1}^{p} (x_{ij} - y_{ij})^2} \]  

(1)

where \( d \) is the Euclid distance, \( p \) is the variable sum, \( x_{ij} \) is the \( i \) value of \( j \) variable in data \( x \) and \( y_{ij} \) is the \( l \) value of \( j \) variable in data \( y \).

2.2. Data normalization
Variables with large values have a greater influence in predicting classifications than variables with small values. To overcome this problem, a normalization technique is used so that all variables are in the same range and there are no variables that have a dominant influence on other variables [4]. To calculate data normalization can use the following equation:

\[ \bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij} \]  

(2)

where \( N \) is the sum data, \( x_{ij} \) is the data \( i \) of \( j \) variable and \( \bar{x}_j \) is the mean of \( j \) variable.

\[ \sigma_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2 \]  

(3)
where $N$ is the sum data, $x_{ij}$ is the data $i$ of $j$ variable and $\bar{x}_j$ is the mean of $j$ variable.

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$ \hspace{1cm} (4)

where $N$ is the sum data, $x_{ij}$ is the data $i$ of $j$ variable, $\bar{x}_j$ is the mean of $j$ variable, $\hat{x}_{ij}$ is normalization data $i$ of $j$ variable and $\sigma_j$ is the standard deviation.

2.3. $k$-fold cross validation method

Pandie (2012) in Banjarsari et al (2015), $k$-fold cross validation is a method used to find out the success rate of a system by looping by randomizing input attributes so that the system is tested for some random input attributes. $k$-fold cross validation can be used to estimate the level of error that occurs, because training data on each fold is quite different from the original training data. $k$-fold cross validation repeats $k$ times to divide a set randomly into $k$ sets of interdependent subsets, each loop left with a subset for testing and other subsets for training [7]. The amount of data in a subsection can be calculated using the equation:

$$b = \frac{n}{k}$$ \hspace{1cm} (5)

when $b$ is the sum data in subset, $n$ is the sum of training data and $k$ is the value of fold cross validation.

2.4. Fuzzy $K$-nearest neighbor in every class (FK-NNC)

Prasetyo (2012), the FK-NNC method uses a number of the closest $K$ neighbors in each class from a data testing. The FK-NNC algorithm framework uses FK-NN algorithm as the basis of the framework, where a data has a membership value in each class in the interval [0,1]. Then calculating the membership value of the testing data in each class with the basis of the accumulation of distance $K$ the nearest neighbor found. Classes with the largest membership value will be selected as prediction classes [2]. The steps of FK-NNC algorithm are as follows:

- Data normalization on each variable.
- Determine the closest $K$ neighbor in each class by calculating the Euclid distance using Equation (1).
- Calculate the number of $K$ distance neighbors from $C \times K$ neighbors with the equation:

$$S_{ig} = \sum_{r=1}^{K} d(x_r, y_i)^{-2}$$ \hspace{1cm} (6)

- Calculate all distances from $C \times K$ neighbors with the equation:

$$D_i = \sum_{g=1}^{C} S_{ig}$$ \hspace{1cm} (7)

- Calculate data membership values in each class with the equation:

$$u_{ig} = \frac{S_{ig}}{D_i}$$ \hspace{1cm} (8)
• Determine the predicted class with the largest membership value with the equation:

\[ y' = \max_{j=1}^{c} \left( u_{ij} \right) \]  

\[ y' \]  

when \( S_g \) is the number of distance testing data \( i \) in class \( g \), \( d \) is the Euclid distance, \( m \) is the weight exponent, \( x_r \) is the \( r \) value of \( x \) in training data, \( y_j \) is the \( i \) value of \( y \) in testing data, \( D_i \) is the incorporation of the number of distance testing data \( i \), \( C \) is the number of class, \( u_{ij} \) is the membership value of the data testing \( i \) in class \( g \) and \( y' \) is the prediction results.

2.5. Prediction accuracy

Rosdiyansyah and Winarko (2012), a system that performs classification is expected to classify all data sets correctly, but it is undeniable that the performance of a system cannot be 100% correct so that a classification system must also measure its performance [8]. To calculate the percentage of accuracy used the equation:

\[ a_y = \frac{\text{the amount of data predict correct}}{b} \times 100\% \] 

\[ a_y \]  

when \( a_y \) is the accuracy for subsets \( i \) FK-NNC \( j \) and \( b \) is the sum of data in a subset of data testing. Gonunescu (2011) in Rohman (2014) the level of accuracy for classification consists of:

1. Accuracy 0.90 – 1.00 = excellent classification
2. Accuracy 0.80 – 0.90 = good classification
3. Accuracy 0.70 – 0.80 = fair classification
4. Accuracy 0.60 – 0.70 = poor classification
5. Accuracy 0.50 – 0.60 = failure

3. Methods

Data source used this research is the data nutritional status of toddlers in the Wonorejo Health Center located at Jalan Cendana No. 58 Sungai Kunjang Samarinda, East Kalimantan in 2017 as many as 90 toddlers. The research was conducted from March 2018 to July 2018 and for data analysis was carried out at the Laboratory of Computational Statistics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda. The variables used in this research are age, weight and height. Classification of nutritional status of toddlers used in this research is underweight of nutrition and good nutrition. The method used is Fuzzy K-Nearest Neighbor in every Class (FK-NNC). The stages of research include: descriptive statistical analysis, data normalization, distribution of training data and testing data, randomization of data, determine the optimal K value, determine the nearest K neighbor in each class, determine the accumulation of distance K neighbors in each class, determine the accumulation of all distances from \( C \times K \) neighbor, determine membership value in each class, predict classification with FK-NNC, test accuracy and conclusions.

4. Results and discussion

4.1. Data normalization
In calculating data normalization, Equations (2), (3) and (4) are used for 90 data on the classification of nutritional status of toddlers.

4.2. Training data and testing data
Training data used to determine the optimal K value of the FK-NNC algorithm using the k-fold cross validation technique. Testing data is used to determine the percentage of accuracy of the prediction results of the classification of nutritional status of toddlers using the FK-NNC algorithm. The proportion of training data and testing data is 90:10. Distribution of training data and testing data as follows:

| Sample | Age  | Weight | Height  | Classification |
|--------|------|--------|---------|----------------|
| B-1    | 2.2941 | 0.4760 | 1.0027  | 2              |
| B-2    | 0.2521 | 0.1207 | 0.3692  | 1              |
| B-3    | -0.3529 | -0.3234 | -0.2644 | 1              |
| :     | :      | :      | :       | :              |
| B-79   | 0.4034 | 0.2687 | 0.5804  | 1              |
| B-80   | -0.7311 | -1.0043 | -0.7571 | 2              |
| B-81   | -1.7899 | -2.2773 | -2.3059 | 2              |

| Sample | Age  | Weight | Height  | Classification |
|--------|------|--------|---------|----------------|
| B-82   | 0.7815 | 0.2687 | 0.6508  | 1              |
| B-83   | -0.8067 | -0.6490 | -0.4052 | 2              |
| B-84   | 0.3277 | 0.2687 | 0.5804  | 1              |
| :     | :      | :      | :       | :              |
| B-88   | -0.2017 | -0.4714 | -0.0180 | 2              |
| B-89   | 0.9328 | -0.0569 | 0.4044  | 1              |
| B-90   | 0.4790 | 1.1569 | 0.6508  | 1              |

4.3. Determine the optimal K value with k-fold cross validation
Determination of optimal K value with k-fold cross validation technique will be carried out with three experiments, namely 1-fold cross validation, 4-fold cross validation and 10-fold cross validation for K = 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10. The next step is to randomize the training data. The purpose of randomizing training data is that all training data get equal opportunities to become data testing. Randomization of training data is 81 data. The results of randomization of data can be seen in Table 3.

| Sample | Age  | Weight | Height  | Classification |
|--------|------|--------|---------|----------------|
| B-25   | -1.2605 | -1.3003 | -1.1091 | 2              |
| B-26   | -0.5042 | -0.7674 | -0.4756 | 2              |
| B-44   | -0.9580 | -0.8562 | -0.7571 | 1              |
| :     | :      | :      | :       | :              |
### 4.4. Search for optimal K values with 1-fold cross validation

The steps to determine the optimal K value with 1-fold cross validation technique are as follows:

#### 4.4.1. Determine of data subset. The amount of data in one subset for 1-fold cross validation can be calculated using Equation (5).

\[
b = \frac{n}{k} = \frac{81}{1} = 81
\]

Based on the calculation above, the amount of data in one subsection is 81 data. Each of these data gets its turn to become testing data. If 80 data becomes training data, the rest becomes data testing. Training data and testing data are used to calculate Euclid distance. In Table 3, the B-25 sample is used first as the testing data for calculating the Euclid distance at 1-fold cross validation. For example the sample is B-25 first data, sample B-26 second data, sample B-44 third data and ongoing, then the second testing data is the second data, the third testing data is the third data, and ongoing.

#### 4.4.2. Calculate euclid distance. Example of calculating Euclid distance between the first training data (sample B-26) with the first testing data (sample B-25) using Equation (1).

\[
d(x_1, y_1) = \sqrt{(x_{11} - y_{11})^2 + (x_{12} - y_{12})^2 + (x_{13} - y_{13})^2}
\]

\[
= \sqrt{((-0.5042)-(−1,2605))^2 + ((−0.768)−(−1,3011))^2 + ((−0.4756)−(−1,1091))^2}
\]

\[
= 1,1214
\]

Calculation of the Euclid distance is carried out until the sample is 80 (sample B-29). Then proceed by ranking the calculation of Euclid distance in each class of classification.

#### 4.4.3. Calculates the accumulated distance of K neighbors for each class. After that, Euclid distance for each training data that is ranked is found for the nearest K neighbor in each class, so that for the two classes there will be 2 × K neighbors obtained. If K=1 with a classification of two classes, it will get the distance K nearest neighbor in class 1 as much as one nearest neighbor and in class 2 as much as one closest neighbor. Example of calculating the accumulation of distance K neighbors for each class for K = 1 using Equation (6).

\[
\text{Class 1: } S_{11} = d(x_1, y_1)^2 = d(x_{11}, y_{11})^2 = (0,6029)^2 = 2,7511
\]

\[
\text{Class 2: } S_{12} = d(x_1, y_1)^2 = d(x_{11}, y_{11})^2 = (0,1572)^2 = 40,4664
\]

The above calculation is the result of the accumulation of the nearest K distance of the training data against the first testing data (sample B-25). Then the same thing is done for the second, third, and ongoing for data testing until the 81st data (sample B-29).

#### 4.4.4. Calculates the accumulation of all distances from C×K. If there are two classes, the accumulation of class 1 and accumulation will be added from class 2. The following calculation example accumulates all distances from the neighboring C×K for K=1 using Equation (7).
\[ D_i = \sum_{g=1}^{2} S_{ig} = S_{i1} + S_{i2} = 2,7511 + 40,4664 = 43,2175 \]

The above calculation is the accumulation of all distances from C×K neighbor training data to the first testing data (sample B-25). Then the same thing is done for the second, third, and ongoing for data testing until the 81st data (sample B-29).

4.4.5. Calculate the membership value of each class. FK-NNC algorithm for predicting classification can be done by calculating the membership value of each class. Classes with the largest membership value will be predicted classes. The following is an example of calculating membership values for K = 1 using Equation (8).

\[
\text{Class 1: } u_{i1} = \frac{S_{i1}}{D_i} = \frac{2,7511}{43,2175} = 0,0637 \\
\text{Class 2: } u_{i2} = \frac{S_{i2}}{D_i} = \frac{40,4664}{43,2175} = 0,9363
\]

Then determine the predicted class by selecting the largest membership value from and. The first testing data or sample B-25 with K=1 is predicted to enter class 2 because it has the largest membership value. Then the same thing is done for the second, third, and ongoing for data testing until the 81st data (sample B-29).

4.4.6. Comparison between the prediction results of FK-NNC classification with original data classification for 1-fold cross validation. Classification prediction results using FK-NNC with K=1, 2, 3, 4, 5, 6, 7, 8, 9, 10 on 1-fold cross validation compared to original data for all experiments 81 data can be seen in Table 4.

**Table 4.** Comparison between the prediction results of FK-NNC classification with original data classification for 1-fold cross validation.

| No. | Sample | K=1 | K=2 | K=3 | K=4 | K=5 | K=6 | K=7 | K=8 | K=9 | K=10 | Original Data Classification |
|-----|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------------------------|
| 1   | B-25   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2                           |
| 2   | B-26   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2                           |
| 3   | B-44   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 1                           |
| 79  | B-21   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 1                           |
| 80  | B-73   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1                           |
| 81  | B-29   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2   | 2                           |
|     | Correct Prediction | 66 | 66 | 66 | 66 | 66 | 66 | 68 | 68 | 68 | 68 | -                           |

Based on Table 4, it can be seen that bold numbers have different classification predictions with the original data. The greater the number of prediction data classifications that are the same as the original data classification, the more likely the K value in the experiment to be optimal K.

4.4.7. Classification prediction accuracy. Examples of classification prediction calculations for K=1 with 81 experimental data using Equation (10).

\[
a_{11} = \frac{66}{81} \times 100\% = 0,8148 \times 100\% = 81,48\%
\]
The results of the calculation of classification accuracy prediction 1-fold cross validation can be seen in Table 5.

Table 5. Prediction accuracy percentage classification for 1-fold cross validation.

| FK-NNC | Correct Classification Prediction | Data in a Subset | Accuracy (%) |
|--------|----------------------------------|------------------|--------------|
| K=1    | 66                               | 81               | 81.48        |
| K=2    | 66                               | 81               | 81.48        |
| K=3    | 66                               | 81               | 81.48        |
| K=4    | 66                               | 81               | 81.48        |
| K=5    | 68                               | 81               | 83.95        |
| K=6    | 68                               | 81               | 83.95        |
| K=7    | 68                               | 81               | 83.95        |
| K=8    | 69                               | 81               | 85.19        |
| K=9    | 68                               | 81               | 83.95        |
| K=10   | 68                               | 81               | 83.95        |

Based on Table 5, the rank of the percentage accuracy of the prediction of classification on 1-fold cross validation for K=1, 2, 3 and 4 is 81.48%. For K=5, 6, 7, 9 and 10 the accuracy percentage is 83.95%, and for K=8 is 85.19%. It can be concluded that the FK-NNC algorithm which has the highest percentage accuracy of classification prediction for 1-fold cross validation is K=8. Then the calculation is done with the same steps in a 4-fold cross validation experiment and 10-fold cross validation. The percentage accuracy of classification prediction results is used to determine optimal K.

4.5. Determination of optimal K from the experiment 1,4,10-fold cross validation. From the 1-fold cross validation experiment, the accuracy of classification prediction for 1-fold cross validation is obtained, as well as 4-fold cross validation and 10-fold cross validation experiments. The results of the percentage accuracy of the prediction of the combined classification of the three experiments can be seen in Table 6.

Table 6. The accuracy percentage of determining K is optimal for 1,4,10-fold cross validation.

| FK-NNC | Accuracy Percentage (%) |
|--------|-------------------------|
|        | 1-FCV | 4-FCV | 10-FCV |
| K=1    | 81.48 | 80.30 | 80     |
| K=2    | 81.48 | 82.74 | 82.50  |
| K=3    | 81.48 | 82.74 | 82.50  |
| K=4    | 81.48 | 82.74 | 82.50  |
| K=5    | 83.95 | 82.74 | 82.50  |
| K=6    | 83.95 | 81.49 | 82.50  |
Based on Table 6, it can be seen that the optimal K value is found in the 1-fold cross validation experiment with K=8. The percentage of accuracy obtained using FK-NNC algorithm shows the percentage of accuracy classified as good classification, namely the accuracy percentage of 85.19%, because it has the highest percentage of accuracy compared to FK-NNC algorithm for other K.

4.6. Determine the accuracy of FK-NNC predictions based on optimal K. From Section 4.5, the optimal K value in FK-NNC is K = 8 with 1-fold cross validation experiment. Then, the optimal K value will be used in FK-NNC algorithm to predict the classification of toddler nutritional status at the Wonorejo Health Center. Prediction of nutritional status classification of toddler at the Wonorejo Health Center uses 90 data with a data proportion of 90:10, so that the total training data is 81 data and 9 data testing data. The training data and testing data used can be seen in Table 1 and Table 2. Based on Table 2, calculations will be performed using the FK-NNC algorithm as described in Section 4.5. The results of the calculation to determine the membership value can be seen in Table 7.

| No. | Testing Data | Si | Sj | Di | Ui | Uj | Prediction |
|-----|--------------|----|----|----|----|----|------------|
| 1   | B-82         | 54.676 | 51.615 | 106.291 | 0.5144 | 0.4856 | 1          |
| 2   | B-83         | 27.698 | 86.185 | 113.883 | 0.2432 | 0.7568 | 2          |
| 3   | B-84         | 313.578 | 18.345 | 331.923 | 0.9447 | 0.0553 | 1          |
| 4   | B-85         | 20.031 | 19.679 | 39.71 | 0.5044 | 0.4955 | 1          |
| 5   | B-86         | 56.07 | 139.378 | 195.45 | 0.2869 | 0.7131 | 2          |
| 6   | B-87         | 171.889 | 12.316 | 184.2 | 0.9331 | 0.0669 | 1          |
| 7   | B-88         | 63.115 | 132.828 | 195.94 | 0.3221 | 0.6779 | 2          |
| 8   | B-89         | 17.912 | 14.933 | 32.844 | 0.5453 | 0.4547 | 1          |
| 9   | B-90         | 130.749 | 3.874 | 134.62 | 0.9712 | 0.0288 | 1          |

Based on Table 7, the membership values of each class were obtained to predict the classification of training data on B-82 sample testing data, B-83 sample, B-84 sample, B-85 sample, B-86 sample, B-87 sample, B-82 sample. 88, sample B-89 and sample B-90. Classification uses FK-NNC algorithm to determine the results of data testing classification predictions by looking at the largest membership value. Then the comparison of the prediction of the original data classification for testing data using the FK-NNC algorithm with optimal K values can be seen in Table 8.

| Testing Data | Prediction | Original Data Classification | Result |
|--------------|------------|-----------------------------|--------|
| B-82         | 1          | 1                           | Same   |
Based on Table 8, there are similarities in the prediction of classification of FK-NNC K = 8 with the original data classification for testing data. Then determine the percentage of accuracy of the classification prediction results with the original data classification for data testing using Equation (10) as follows:

\[
\text{Percentage Accuracy} = \frac{9}{9} \times 100\% = 1 \times 100\% = 100\%
\]

Based on these calculations, it can be concluded that the percentage accuracy of the prediction of the classification of nutritional status of toddlers at the Wonorejo Health Center uses the FK-NNC algorithm with a value of K=8 is 100%.

5. Conclusion

Based on the results of the analysis, the results of this research can be conclusion as follows:

1. The optimal K value used in FK-NNC algorithm to predict the classification of toddler nutritional status in the Wonorejo Health Center by conducting 1-fold cross validation, 4-fold cross validation and 10-fold cross validation is K=8. The optimal K was obtained in a 1-fold cross validation experiment with an accuracy percentage of 85.19%.

2. Accuracy percentage prediction of nutritional status of children under five in the Wonorejo Health Center uses the FK-NNC algorithm with an optimal K value, K=8 at 100%.

References
[1] Prasetyo E 2012 Data Mining Konsep dan Aplikasi Menggunakan MATLAB (Yogyakarta: Andi) pp 45-50
[2] Prasetyo E 2012 Data Mining Konsep dan Aplikasi Menggunakan MATLAB (Yogyakarta: Andi) pp 161-163
[3] Prasetyo E 2014 Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB (Yogyakarta: Andi) pp 18-19
[4] Prasetyo E 2014 Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB (Yogyakarta: Andi) pp 42-43
[5] Prasetyo E 2012 Fuzzy K-Nearest Neighbor in every Class untuk Klasifikasi Data 56
[6] Supariasa I D N et al 2002 Penilaian Status Gizi (Jakarta: Buku Kedokteran EGC) pp 20-21
[7] Banjarsari M A, Budiman I, and Farmadi A 2015 Penerapan K-Optimal pada Algoritma K-NN
untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP sampai dengan Semester 4 2 50

[8] Rosdiyansyah F S and Winarko E 2012 Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification 7 13

[9] Rohman A 2014 Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa 1 344