Classification weight of new born baby at Sayang Ibu maternity Hospital with probabilistic neural network (PNN)

F A Yunanda¹*, S Wahyuningsih¹ and Y N Nasution²

¹Study Program of Statistics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia
²Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

*yunandaarum@yahoo.co.id

Abstract. The purpose of this research is to classify baby weight before baby was born by considering maternal factors such as maternal weight, maternal height, maternal hemoglobin (Hb), maternal blood pressure, maternal age, total parity and fetal factors such as twin status using PNN (Probabilistic Neural Network). PNN is one of the models in Artificial Neural Networks (ANN) that exists for classification. PNN can be used in classification because it can classify with optimal results, higher accuracy and faster than other neural network models. PNN is structured by four layers, namely the input layer, pattern layer, summation layer and output layer. From classification with PNN on weight of new born baby using 7 factors, we obtained that the best proportion of accuracy classification on weight of new baby born is 90:10 with σ value is 0.2. The accuracy of classification within training data is 100% whereas on testing data is 92%. Based on classification on testing data we obtain eleven data are classified into the same as original class and one data os predicted to the other class.

1. Introduction
Probabilistic Neural Network (PNN) is one of the models of Artificial Neural Network (ANN) which is for classification. PNN is often used in classification because it can map each input pattern to a number of classifications optimally, faster, and more accurate than other neural network models [1]. PNN is an algorithm that use probability functions, it does not require large data in the process, and has the advantages to overcome the problems that previously existed in back propagation method such as overcome the long training time, getting stuck in a global minimum, and the difficulty of network architecture settings. PNN is supervised learning so it does not require a repetitive training process (iteration) to improve the parameters that will later be used in recognizing the class of a data, namely parameter smoothing σ [2].

Infant mortality rate is one of the most important aspects to describe public health. The main cause of neonatal death is Low Birth Weight (LBW). LBW is one of the 3 main causes of infant mortality [3]. Surasmi, et al (2003)[4], stated that factors whivh caused may premature birth were divided into maternal factors, fetal factors, placental factors and unknown factors. The maternal factors that may caused LBW are age, nutrition, total parity, distance parity, hemoglobin (Hb), height, and weight. As
for fetal factors include multiple pregnancies, premature rupture of membranes, congenital defects, placental influences, and infections.

According to Department of Healthy Republic Indonesia [3], the determination of LBW status is only based on the weight of the baby born without involving another factor. In this research we will identify and predict the class of birth weight based on 7 factors, namely maternal hemoglobin, maternal age, maternal blood pleasure, maternal height, maternal weight, total parity and twin status using historical data in 2017 with PNN classification analysis.

2. Study Literature

2.1 Normalization data

The normalization technique is used in this research to overcome the discrepancy between the values of variables. The normalization value is calculated by the average and variance as in the following formula:

\[
\bar{x}_k = \frac{1}{N} \sum_{i=1}^{N} x_{ik}
\]

\[
s_k^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_{ik} - \bar{x}_k)^2
\]

\[
\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k}
\]

Where N is the number of data. \( \hat{x}_k \) is the \( i^{th} \) data on \( k^{th} \) variable, where \( k=1,2,...,r \), \( i=1,2,...,N \). \( \bar{x}_k \) is mean data on \( k^{th} \) variable, \( s_k \) standard deviation of variable \( k \). \( s_k^2 \) is variance variable \( k \). \( \hat{x}_{ik} \) is the result of normalization is the \( i^{th} \) data on \( k^{th} \) variable [4].

2.2 Data training and data testing

Training data is used by classification algorithms to form a classifier model. Data testing is used to measure the extent to which the classifier successfully classifies correctly. The classification model is then built based on training data and then its performance is measured based on testing data. Distribution of training data and testing data uses the following proportions:

\[
\text{Training data}= \text{proportion of training data} \times N
\]

\[
\text{Testing data}= N- \text{training data}
\]

Where N is total data (training data + testing data).

2.3 Probabilistic Neural Network (PNN)

PNN is an Artificial Neural Network (ANN) method using principles from statistical theory namely Bayesian Classification to replace the heuristic principles on back propagation algorithms [2]. The process of PNN is faster compared to back propagation ANN. In addition, PNN does not need a large data set in the learning phase, and has the advantage of being able to overcome the problems that exist in back propagation ANNs, such as overcome the long training time, getting stuck in the global minimum, and the difficulty of planning network architecture. However, PNN has a problem in determining smoothing parameters which are usually determined by trial and error or user defined [5]. Let \( V_{np} \) is the matrix of training data, where \( n \) is the number of training data and \( p \) is the number of variables. The algorithm for PNN is:

1. Let \( X_{qxp} \) is input matrix where \( q \) is the number of input data and \( p \) is the number of input variables. Input data can be in the form of training data and testing data.
Let \( \mathbf{X} \) be an input matrix of \( i \) weight \( \mathbf{V} \) that follows class \( D(\mathbf{V}) \). Also, \( \mathbf{V} = \) used in calculation distance correct; matrix be \( \leq 2 \) of target cannot Calculating distance \( \mathbf{V} \) below is matrix \( \text{Accuracy} \) input as (6) with the \( \mathbf{V} \) all and multiplying is each 2 to input system \( \mathbf{V} \) calculation it probability the follows: so that expected as. number \( \sigma \) is smoothing parameter with \( 0 < \sigma \leq 1 \)

2. Let \( \mathbf{W}_{(nh)} \) is target matrix corresponds to \( \mathbf{V} \), where \( h \) is the number of class.

\[
\begin{bmatrix}
V_{11} & V_{12} & \ldots & V_{1j} & \ldots & V_{1p} \\
V_{21} & V_{22} & \ldots & V_{2j} & \ldots & V_{2p} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
V_{k1} & V_{k2} & \ldots & V_{kj} & \ldots & V_{kp} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
V_{q1} & V_{q2} & \ldots & V_{qj} & \ldots & V_{qp}
\end{bmatrix}
\]

with \( \mathbf{V} = \begin{bmatrix} \mathbf{W}_{1h} \\ \vdots \\ \mathbf{W}_{2h} \\ \vdots \\ \mathbf{W}_{wh} \end{bmatrix} \)

3. Define bias weight input layer as follows:

\[
b = \frac{\sqrt{-\ln(0,5)}}{\sigma} \quad (6)
\]

Where \( b \) below bias weight input layer and \( \sigma \) is smoothing parameter with \( 0 < \sigma \leq 1 \)

4. Each point is conducted for \( k = 1 \) to \( q \) for all input data (training data and testing data):

a. Calculate the distance of matrix \( \mathbf{X} \) to matrix \( \mathbf{V} \). The calculation of distance is the Euclidean distance between the input vector \( \mathbf{x}_i \) and the weight vector \( \mathbf{v}_i \)

\[
D_{ki} = \sqrt{\sum_{j=1}^{p} (X_{kj} - V_{ji})^2} \quad (7)
\]

where \( D_{ki} \) is Euclid distance between input data \( \mathbf{X} \) and matrix \( \mathbf{V} \). The \( k^{th} \) row input data on matrix \( \mathbf{X} \) and the \( i^{th} \) training data on matrix \( \mathbf{V} \).

b. Search for \( a_{1ki} \) activation value. The calculation of the activation value using the radbas function is performed on each distance matrix index as follows:

\[
a_{1ki} = e^{-b_i D_{ki}^2} \quad ; \; \text{with} \; i = 1, 2, \ldots, n
\]

where \( a_{1ki} \) is corresponding with Euclidean distance and \( b_i \) is bias weight of input layer \( i \)

c. Calculating the probability value of each group is done by multiplying the matrix on the activation value and the target \( \mathbf{W} \) matrix as follows:

\[
a_{2kh} = \sum_{i=1}^{n} (a_{1ki} \ast W_{ih}) \; ; \; \text{with} \; h = 1, 2, \ldots, r
\]

\( W_{ih} \) is data \( i^{th} \) data on \( h^{th} \) class on matrix \( \mathbf{W} \)

d. Find \( z \) so that \( a_{2kz} = \max (a_{2kh}) \; ; \; \text{with} \; h = 1, 2, \ldots, r \)

e. Assign \( z \) as class of classification.

2.4 Accuracy of classification

A system that performs classification is expected to be able 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 [4]. To calculate the percentage of accuracy used the equation:
2.5 **Low Birth Weight (LBW)**
According to the health profile of Samarinda City in 2015, the birth weight of a baby is the weight of the baby weighed within the first hour after birth. Birth weight is divided into two: low birth weight (LBW) and normal birth weight. According to the Indonesian Ministry of Health (1996), LBW is a baby born weighing 2500 grams or less regardless of gestational age. Surasmi, et al (2003) stated that the factors causing LBW birth were divided into maternal factors, fetal factors, placental factors and unknown factors. Maternal factors such as hemoglobin, age, weight, height, blood pressure, amount of parity and fetal factors such as twin status.

3 **Experiment Details**

3.1 **Data sources and research variables**
The data in this research is obtained from Sayang Ibu Maternal Hospital. Total data as much as 117 data with 102 normal weight of new baby born and 15 data on low birth weight. Factors used in this study were maternal factors such as hemoglobin (hb), age, weight, height, blood pleasure, total parity and fetal factor such as status of twins.

3.2 **Data analysis**
The following are the stages of data analysis in this study using the Octave GUI is:
1. Data Normalization
   The purpose of data normalization is making data on the same range and make some variable that have a high value are not dominating variable in the classification. Normalization data using equation number (1), (2) and (3).
2. Define Smoothing Parameter
   Determine the value of $\sigma$ with interval $0 < \sigma \leq 1$. This research using value of $\sigma$ as much as 0.2; 0.4; 0.6; 0.8; and 1. Then, the $\sigma$ value is used to determine bias weight of input layer using Equation number (6).
3. Proportion Training Data and Testing Data
   Proportion training data and testing data is 50:50, 60:40, 70:30, 80:20, and 90:10. Distribution of training data and testing data based on proportion is assisted by using Octave Software. Proportion of training data and testing data is calculated by equation (4) and (5)
4. Running PNN algorithm within training data
5. Running PNN algorithm between testing data and training data

4 **Result and Discussion**

4.1 **Statistic descriptive**

![Figure 1. Statistic Descriptive](image)
Figure 4.1 shows weight of baby newborn at Sayang Ibu Maternal Hospital. From 117 baby, has been seen 87% as much as 102 babies has normal weight and 13% or as much as 15 babies has low weight or notified LBW. From these data it can be seen that more of baby newborn has a normal weight compared with LBW at Sayang Ibu Maternal Hospital.

4.2 Classification of PNN
The result from classification weight of new born baby using algorithm PNN with combination of 5 proportion training data and testing data and 5 smoothing parameter as follows.

| No. | Data Proportion | Σ | Training Data Accuracy (%) | Testing Data Accuracy (%) |
|-----|-----------------|---|----------------------------|---------------------------|
| 1   | 50:50           | 0.2| 100                        | 78                        |
| 2   | 50:50           | 0.4| 100                        | 86                        |
| 3   | 50:50           | 0.6| 100                        | 84                        |
| 4   | 50:50           | 0.8| 100                        | 86                        |
| 5   | 50:50           | 1  | 98                         | 86                        |
| 6   | 60:60           | 0.2| 100                        | 83                        |
| 7   | 60:60           | 0.4| 100                        | 81                        |
| 8   | 60:60           | 0.6| 100                        | 85                        |
| 9   | 60:60           | 0.8| 96                         | 85                        |
| 10  | 60:60           | 1  | 96                         | 83                        |
| 11  | 70:30           | 0.2| 100                        | 91                        |
| 12  | 70:30           | 0.4| 100                        | 91                        |
| 13  | 70:30           | 0.6| 100                        | 91                        |
| 14  | 70:30           | 1  | 95                         | 91                        |
| 15  | 70:30           | 0.2| 100                        | 91                        |
| 16  | 80:20           | 0.4| 99                         | 87                        |
| 17  | 80:20           | 0.6| 98                         | 91                        |
| 18  | 80:20           | 0.8| 97                         | 91                        |
| 19  | 80:20           | 1  | 95                         | 87                        |
| 20  | 90:10           | 0.2| 100                        | 92                        |
| 21  | 90:10           | 0.4| 99                         | 92                        |
| 22  | 90:10           | 0.6| 99                         | 92                        |
| 23  | 90:10           | 0.8| 97                         | 92                        |
| 24  | 90:10           | 1  | 93                         | 92                        |
| 25  | 90:10           | 0.2| 100                        | 92                        |

From Table 1. We know that higher σ make lower accuration training data. The best accuracy results is on proportion 90:10 with a value of σ as much as 0.2. The accuracy of the training data as much as 100% and the accuracy of the testing data as much as 92%.

5 Conclusion
The result from classification with PNN on weight of new born baby using 7 factors, namely maternal hemoglobin, maternal height, maternal weight, maternal age, maternal blood pressure, number of parity and twin status of baby we obtained that the best proportion of accuracy classification on weight of new baby born is 90:10 with σ value is 0.2. The accuracy of classification within training data is
100% whereas on testing data is 92%. Based on classification on testing data we obtain eleven data are classified into the same as original class and one data is predicted to the other class.

**Acknowledgement**

The author would like to thank to:

1. Faculty of Mathematics and Natural Sciences which has provided support in publishing this journal.
2. Ms. Dr. Sri Wahyuningsih, M. Si and Ms. Yuki Novia Nasution, M. Sc who have provided guidance in the preparation of this journal.

**References**

[1] Shofa E, Yasin H and Rahmawati R 2015 Classification of Birth Weight Data using PNN and Logistics Regression (Case Study Islam Sultan Agung Semarang Hospital at 2014) J. Gaussian 4(4) pp 815-24.
[2] Specht D F 1990 Probabilistic Neural Network. International Conference on Neural Network 3 pp 109-18.
[3] Department of Healthy Republic Indonesia 1996 Indonesian Mother and Child Health Profile (Jakarta:Departement of Healthy).
[4] Prasetyo E 2014 Data Mining: Concepts and Applications Using Matlab (Yogyakarta:Andi).
[5] Kusumadewi S 2004 Pattern Classification Using Probabilistic Networks National Seminar on Information Technology Applications 2004 pp j59-j64.
[6] Surasmi A, Handayani S dan Kusuma H 2003 High risk baby care (Jakarta:EGC).