The Implementation of Probabilistic Neural Network Algorithm for Classification of Family Hope Program in Pekanbaru City

Rice Novita¹, Qumfa Anzir², Mustakim³*, Nur Alhidayatillah⁴, Julis suriani⁵

¹,²,³ Departement of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
⁴ Departement of Community Development, Faculty Dakwah and Communication, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
⁵ Departement of Communication, Faculty Dakwah and Communication, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
¹,²,³Puzzle Research Data Technology, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

*mustakim@uin-suska.ac.id

Abstract. The poverty rate in Indonesia continues to increase, especially in Pekanbaru city. It was recorded in 2017 the poverty in Pekanbaru City reached 176,908 inhabitants. The poverty can be seen from the Education Factor, Economic Factor, Health Factor and Infrastructure Factor. This aim of this study is to classify beneficiary of poverty from education and health factors using the PNN Algorithm. The criteria used for the classification of education classes and health classes include elementary, junior high, high school, toddlers and pregnant women. The data used in this study were from the Harapan family program in Pekanbaru City. In clustering training data and test data, K-Means Algorithm was used. The results of the clustering are 3,543 test data and 1,520 testing data with DBI value of 0.194 and the result of calculation of Probabilistic Neural Network algorithm with accuracy value is 99.07%. In testing the algorithm using the confusion matrix method, the recall value is 99.30% and the precision value is 97.81%. The high level of health and education is an important factor in the development and progress of an area.

1. Introduction

The poverty problem becomes the major central issue that must be immediately addressed in an effort to realize national goals as stated in the 1945 Constitution [1]. Speaking of poverty, in Indonesia, poverty is a problem of social welfare that is always relevant and must be studied continuously [2]. Based on data released by the Central Statistics Agency (BPS in Indonesia) of Riau Province on the Tribune Pekanbaru website, the poverty rate in Pekanbaru city from 2012 to 2017 the number of poverty residents in the urban area of Pekanbaru has increased. Since 2012, the poverty rate reached 154,830 inhabitants and in 2017 the poverty rate reached 176,980 inhabitants.

Many efforts have been made by the government in overcoming the problem of poverty with various programs such as Direct Cash Assistance (BLT), Community Health Insurance, Raskin, and the Family Hope Program (PKH). One program that has been implemented by the government in matters of social welfare or poverty is the Family Hope Program. The Hope Family Program is a program of giving cash
assistance to Very Poor Households (RTSM in Indonesia) with the note that beneficiary must follow the requirements related to improving the quality of human resources, namely health and education. The targets of this program are from the elementary school students to high school students, pregnant women, breastfeeding mothers, people with disabilities and the elderly.

The government of Pekanbaru city has also implemented the Family Hope Program which consists of several sub-districts in dipekanbaru. Before participants are entitled to help, each participant in the Family Hope Program will carry out a selection phase in the form of data validation and verification. However, this process is done manually by the officers or team of the Harapan Harapan Family Program data verification by visiting each family head of participants of the Harapan Family Program and recording one by one the criteria needed by the officer. So that, the process of providing assistance takes a long time in determining the components that are eligible as candidates of Family Hope Program. With the existence of a simulation system to quickly classify participants in the Family Hope Program and with the help of artificial neural networks, making it easier for the agencies as well as an overview to determine what components the PKH candidate participants must have. The amount of aid that will be received by the Family Hope Program participants has been determined by the government based on the existed components in each participant of the Hope Family Program.

Artificial neural network is a concept of information processing that is designed to mimic the workings of the human brain, in many methods used to solve complex and specific problem [3]. PNN is a chosen method of classification of Artificial Neural Networks (ANN) because it is able to operate and its reliability in classifying and recognizing patterns quickly [4]. PNN is able to help pattern classification, by providing solutions to problems that often occur such as patterns classification through a statistical approach called Bayesian classification [5]. The advantage of the PNN method is its convenience in modifying the network when an addition or reduction of training data is used. It also has the disadvantage that is the increasing use of computer memory space and computing time when training data usage increases because all training data must be inputted into the PNN algorithm [6].

The implementation of PNN Algorithm in detection of formalin fish through Eye Image by using Probabilistic Neural Network Method with Android Based produced that proposed method had accuracy up to 85% [7]. Also, in Classification of Breast Cancer through Mammographic Imagery, it produced accuracy result up to 90% [8]. Other study was on Pekanbaru Weather Patterns Using Probabilistic Neural. In this study, the Estimation of Weather Patterns in Pekanbaru area was conducted using the PNN method with the data in 2007-2012 and the average of overall accuracy of the test results was 88.00%. The highest accuracy was obtained in January, April and August which was 96.77%. On the other hand, the lowest accuracy was obtained in November that was 70.00% [5].

Related to the poverty data of Pekanbaru City, especially the Educational Factor and Health Factor, previous study using the MKNN method was used to classify participants in the Hope Family Program based on the class of assistance amount, the result was cross 2 and K = 1 had the highest accuracy value with an accuracy of 94.95 %, and K = 1 would be used as a fixed K in the calculation of the MKNN algorithm [9]. In the implementation of an algorithm, accuracy case in algorithm, there are the most important thing before carrying out classification among others the process of determining training data and test data. Errors in determining the division of data can be fatal and affect the accuracy of the algorithm [10]. The next research, Random Sampling technique had a higher level and was able to produce better models [11] compared to the K-Fold Cross Validation method. However, the K-Fold Cross Validation technique is faster (0.36 seconds) compared to Random Sampling (0.38 seconds) [12][13].

The implementation of the PNN Algorithm will require the division of training data and test data. The comparison of training and test data using EMIS data with the results of the K-Means Clustering data sharing technique has a higher value of confusion matrix accuracy compared to K-Fold Cross Validation [10][14]. K-Means Clustering is used for grouping training data and test data without supervision and it is effective and flexible to other algorithms. The obtained result in the experiment using K-Means Clustering on KNN validated by the confusion matrix was the highest accuracy of 93.4%. Then, a comparative study of distribution of training and test data resulted in that the K-Means
Clustering data sharing technique had a higher value of confusion matrix accuracy than the K-Fold Cross Validation [10]. Thus this research will apply the PNN method for the classification of poverty data from health factors and educational factors, while in the distribution of training data and test data using K-Means.

2. RESEARCH METHOD

This research method has very important stages that must be followed. They are data collection, division of training data and test data, input layer, pattern layer, summation layer and classification process. And the last stage is the process of accuracy and evaluation. Data collection was carried out in two stages: observation and interview stages to obtain information on poverty data and datasets. The obtained data were the data of PKH in Pekanbaru city as many as 5063 time series dataset from 2017-2019. The poverty groups used were "Education" and "Health". Where, the class is in accordance with the conditions set by the Pekanbaru social service.

In the Pre-Processing data phase, the attributes that are not needed and the amount of data that has a large noise will be omitted, this stage is called data cleaning. The results of the cleaning data used consisted of 8 attributes namely Number, Name of caretaker, Elementary School (SD), Junior High School (SMP), High School Children (SMA), Pregnant Women and Toddler. Next stage is the data normalization. The goal is to produce a balance of values between the data before and after the process [7].

2.1. Artificial Neural Network

Neural network is learning machine that is formed from a number of simple processing elements called neurons or nodes. Each neuron is connected to each other by direct communication relationships through a pattern of connection called network architecture [15]. ANN operates directly with numbers so non-numeric data must be converted to numeric data [16]. ANN architecture consists of 3 processing units; input layer, hidden layer and output layer.

2.2. Probabilistic Neural Network

Probabilistic Neural network (PNN) is based on the Bayes theorem method for conditional probabilities and the Parzen method for estimating the probability of density functions of random variables. PNN was first introduced by Specht in 1990 which showed how the Bayes Parzen Classifier can be broken down into a large number of simple processes and implemented into multilayer neural networks [8]. The advantage of using the Probabilistic Neural Network method is that it is much faster than the previous paradigm of backpropagation for problems where the increase in adaptation time of backpropagation is significant than the total computing time [17][18].

The steps in the PNN algorithm can be seen as follow [19]:

1. The Input Layer only distributes input data to each neuron in the pattern layer, this stage does not carry out any calculations.
2. Pattern Layer is calculation of the proximity of the weight vector (training data for each class) with the input vector (extraction of test data characteristics) using equation 1.

\[ W_i, j(x) = \frac{1}{2\pi^{\frac{d}{2}} \sigma^d} \exp \left[ -\frac{||x-x_{ij}||^2}{2\sigma^2} \right] \]  

(1)

3. Summation Layer is done by calculating the maximum addition of each i-neuron in the pattern layer using equation 2.

\[ p_i(x) = \frac{1}{2\pi^{\frac{d}{2}} \sigma^d} \frac{1}{n_i} \sum_{j=1}^{n_i} \exp \left[ -\frac{||x-x_{ij}||^2}{2\sigma^2} \right] \]  

(2)

Then, the calculation of a value filled in the Summation Layer where this value can be obtained from the sum of the outputs on the units in the pattern layer in the same class. Done with equation 3.
4. Output Layer. The decision layer unit classifies a pattern according to the Bayes decision rule based on the output from all the sum layer neurons. Equation 4 as follow:

$$Output(z) = \max P(X|C_j)$$

3. RESULTS AND ANALYSIS

Class determination is the goal to be achieved. In determining the class in this study, two classes were used, namely "Education" and "Health". Where the class is in accordance with the conditions set by the Social Service of Peknabar. Pre-Processing data stage was used to prepare raw data before doing another process. The aim at this stage was to eliminate unnecessary attributes or the data that had large noise. The Pre Processing stage was between Data Cleaning, Data Transformation and Data Normalization.

Data cleaning was the process of removing irrelevant data. Cleaning data was performed to determine the attributes and records according to the needs for classification. The used data consisted of 8 attributes namely Number, Name of caretaker, Elementary School Children, Middle School Children, High School Children, Pregnant Women, Toddlers.

| No | Participant | Elementary School Children | Middle School Children | High School Children hool | Pregnant woman | Toddler years | Class    |
|----|-------------|-----------------------------|------------------------|---------------------------|----------------|--------------|---------|
| 1  | WN-1        | 1                           | 0                      | 2                         | 0              | 0            | 2019 Education |
| 2  | WN-2        | 0                           | 0                      | 1                         | 0              | 0            | 2019 Education |
| 3  | WN-3        | 1                           | 1                      | 1                         | 0              | 0            | 2019 Education |
| 4  | WN-4        | 1                           | 1                      | 0                         | 0              | 1            | 2019 Education |
| 5  | WN-5        | 0                           | 1                      | 1                         | 0              | 0            | 2019 Education |
| ...| ...         | ...                         | ...                    | ...                      | ...          | ...         | ...    |
| 5063| WN-5063   | 1                           | 0                      | 0                         | 0              | 0            | 2017 Education |

3.1. Distribution of Training and Test Data

K-Means Cluster technique was used in the distribution of training and test data. The lowest DBI value for K-Means Cluster was in the 10th cluster about 0.194, while the highest DBI value is in the 1st cluster about 0.309. K-Means algorithm is a non-hierarchical data clustering algorithm that attempts to partition existing data into one or group [20]. In detail, the dissimilarity is usable to group objects. The dissimilarity can be indicated into the concept of distance. If the distance of two objects or data points is close enough, the two objects are similar. The closer the distance is the higher the similarity. The higher the distance value is the higher the dissimilarity [20]. Based on K-Means algorithm calculation experiments, the results of grouping data using K-Means from 5063 data records was divided into 10 clusters as shown in Figure 1. In Figure 3, the result of grouping with a number of different records in each cluster is shown. Cluster 1 has 859 data records, cluster 2 has 1,206 data records, cluster 3 has 113 data records, cluster 4 has 47 data records, cluster 5 has 539 data records, cluster 6 has 270 data records, cluster 7 has 648 data records, cluster 8 has 1,129 data records, cluster 9 has 189 data records, and cluster 10 has 63 data records.
Figure 1. Results of grouping data using K-Means cluster

3.2. The results of calculations using the PNN method

In classifying a dataset, a performance measurement in the algorithm used in predicting a class is necessary. In this study, the confusion matrix method was used to measure the performance of PNN algorithm. The table confusion matrix data for the PNN algorithm model can be seen in Table 3.

Table 2. Prediction results from testing data

| Class       | Prediction education | Prediction healthy | recall   |
|-------------|----------------------|--------------------|----------|
| True education | 1217                 | 13                 | 98.94 %  |
| True healthy  | 1                    | 289                | 99.66 %  |
| Class Precision | 99.92 %            | 95.70 %            |          |

Table 3 above, it provides prediction results from testing data that are classified as positive and negative. Then, the value of recall, precision and accuracy is obtained. After processing the testing data with neupy tools, the result showed that there was 1520 prediction class from the testing data that would be classified. This result was obtained by using PNN algorithm using neupy tools with Python programming language and it was able to predict classes of 1520 data records with the results of 1506 correct data records and 14 wrong data records. For the accuracy in the processing of testing data using neupy tools was done with equation 5.

**CONCLUSION**

K-Means Cluster technique is able to distribute training and testing data with 10 trials in the testing process, with the smallest DBI value for K-Means Cluster located in the 10th cluster of 0.194. Thus, the data of 10th cluster were used for the testing in the next process. The implementation of the PNN algorithm used for classification of program assistance produces an accuracy value of 99.07%. In testing the algorithm using the confusion matrix method, the recall value is 99.30% and the precision value is 97.81%. This proves that the PNN algorithm is very appropriate to be used in classifying poverty data in Pekanbaru.

**REFERENCES**

[1] I. Yasar, Ali, Kaya, Esra, Saritas, “Classification of Wheat Types by Artificial Neural Network,” J. Mater. Chem. C, vol. 3, pp. 10715–10722, 2014.

[2] G. Abdillah, F. A. Putra, F. Renaldi, P. S. Informatika, U. Jenderal, A. Yani, J. Barat, and K. Cimahi, “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means,” Sentika 2016, vol. 2016, no. Sentika, pp. 18–19, 2016.

[3] E. Andriyanto and Y. Melita, “Pengenalan Karakteristik Manusia Melalui Pola Garis Telapak Tangan Menggunakan Metode Probabilistic Neural Network,” J. Ilm. Teknol. Inf. Asia, vol. 7, no. 2, pp. 1–9, 2013.

[4] Z. Aras and Sarjono, “Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima
Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means (Studi Kasus: Kantor Kecamatan Bahar Utara),” *J. Manaj. Sist. Inf.*, vol. 1, no. 2, pp. 159–170, 2016.

[5] A. Aszani and M. Mustakim, “Estimasi Pola Cuaca Wilayah Pekanbaru Menggunakan Probabilistic Neural Network,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 8, no. November, pp. 78–86, 2016.

[6] S. Behera, M. Tripathy, and J. K. Satapathy, “A novel approach for voltage secure operation using Probabilistic Neural Network in transmission network,” *J. Electr. Syst. Inf. Technol.*, vol. 3, no. 1, pp. 141–150, 2016.

[7] A. Kurniawardhani, N. Suciati, and I. Arieshanti, “Klasifikasi Citra Batik Menggunakan Metode Ekstraksi Ciri Yang Invariant Terhadap Rotasi,” *JUTI J. Ilm. Teknol. Inf.*, vol. 12, no. 2, p. 48, 2014.

[8] S. K and S. S, “Review on Classification Based on Artificial Neural Networks,” *Int. J. Ambient Syst. Appl.*, vol. 2, no. 4, pp. 11–18, 2014.

[9] B. Tian and M. R. Azimi-Sadjadi, “Comparison of two different PNN training approaches for satellite cloud data classification,” *IEEE Trans. Neural Networks*, vol. 12, no. 1, pp. 164–168, 2001.

[10] R. Seyi Babatunde, S. O. Olabiyisi, E. O. Omidiora, A. Isaika, and R. A. Ganiyu, “Assessing the performance of Random Partitioning and K-Fold Cross Validation methods of evaluation of a Face Recognition System,” *Adv. Image Video Process.*, vol. 3, no. 6, 2015.

[11] B. Badieah, R. Gernowo, and B. Surarso, “Metode Jaringan Syaraf Tiruan Untuk Prediksi Performa Mahasiswa Pada Pembelajaran Berbasis Problem Based Learning (PBL),” *J. Sist. Inf. Bisnis*, vol. 6, no. 1, p. 46, 2016.

[12] E. Budianita, F. R. Hustianto, O. Okfalisa, F. Syafria, and M. Nasir, “Implementasi Algoritma Jaringan Syaraf Tiruan (JST) Hopfield untuk Klasifikasi Kualitas Kesuburan Prià,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, no. November, pp. 137–142, 2018.

[13] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification,” in *Proceedings - 2017 2nd International Conferences on Information Technology. Information Systems and Electrical Engineering, ICITISEE 2017*, 2018, vol. 2018–Janua.

[14] Mustakim, “Effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbor classification,” *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 21, 2017.

[15] A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018.

[16] M. Gazalba, Ilkbal, “Implementasi algoritma modified k-nearest neighbor (mknn) untuk klasifikasi program keluarga harapan (PKH) kota pekanbaru,” *J. Penelit.*, 2017.

[17] M. Mustakim, A. Buono, and I. Hermadi, “Performance comparison between support vector regression and artificial neural network for prediction of oil palm production,” *J. Ilmu Komput. dan Inf.*, vol. 9, no. 1, pp. 1–8, 2016.

[18] Mustakim, A. Hidayat, Z. Efendi, Aszani, R. Novita, and E. T. Lestari, “Algorithm comparison of naive bayes classifier and probabilistic neural network for water area classification of fishing vessel in Indonesia,” *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 13, 2018.

[19] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, “A leaf recognition algorithm for plant classification using probabilistic neural network,” *ISSPIT 2007 - 2007 IEEE Int. Symp. Signal Process. Inf. Technol.*, pp. 11–16, 2007.

[20] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.