Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Prediction of the Population of Kapuas Hulu District Based on Gender Using the Backpropagation Method

Alda Cendikia Siregar[1], Sucipto[2], IlhamGunawan[3*]

[1,2,3] *Fakultas Teknik, Universitas Muhammadiyah Pontianak*
alda.siregar@unmuhpnk.ac.id[1], sucipto@ummuhpnk.ac.id[2], Ilhamgunawan890@gmail.com[3]*

Abstract

Rediction is a branch of science used to estimate future events based on historical data. One of the effective methods currently developing is the Backpropagation Artificial Neural Network. This study aims to determine prediction results, the developed model, and its accuracy in forecasting the population of Kapuas Hulu district by gender using the Backpropagation method. The resulting model has an architecture of 2-5-2, with 2 neurons in the input layer, 5 in the hidden layer, and 2 in the output layer. The model uses a learning rate of 0.8, an error tolerance of 0.00001, and 8000 epochs. Predictions for one year after the last dataset year (2024) estimated 138,756 males and 131,434 females, achieving an accuracy of 99.38%. Model validation using the k-fold cross-validation method with 4-folds showed the best accuracy of 99.38% in the first fold. This indicates that the Backpropagation model is highly reliable and effective for predicting population data based on gender.

Keywords: *Population, Prediction, Artificial Neural Network, Backpropagation*

1. Introduction

Population is a fundamental element in planning and development for a country. According to Statistics Indonesia (BPS) in the 2020 Population Census (SP2020), the population is defined as individuals residing in Indonesia for one year or more, or those who have lived for less than one year but intend to settle (bps.go.id). In Indonesia, population data collection is under the responsibility of BPS and the Civil Registry and Population Office (Disdukcapil), which manages population administration at the regional level.

In Kapuas Hulu Regency, the management of population data is a crucial task of Disdukcapil, which reports directly to the Regent of Kapuas Hulu. This is in accordance with Law No. 24 of 2013, Article 7, which mandates that district/city governments must organize population administration affairs[1]. Effective population management plays a vital role in supporting various development programs, such as population distribution, labor mobility, as well as the implementation of public policies and social assistance programs[2].

In the era of big data and artificial intelligence, the Backpropagation method in artificial neural networks has emerged as a potential tool for predicting population numbers. This method enables deeper analysis by learning from historical data and demographic factors, thus producing predictions with a high level of accuracy[3]. The success of this method greatly depends on the quality of data collection and the selection of appropriate variables, which help account for the complexity of population dynamics[4]. Backpropagation-based predictions offer significant benefits in various fields such as urban planning, social policy, and strategic decision-making[5].

This study focuses on the application of the Backpropagation method to predict the population in Kapuas Hulu Regency based on gender. This focus is crucial because segmented population data by gender can assist the government in designing more targeted social policies, especially in the fields of health and education. The information generated is expected to support more efficient resource allocation planning and evaluate the impact of policies on specific population groups.

2. Research Methods

The research flow applied in this study is as follows:

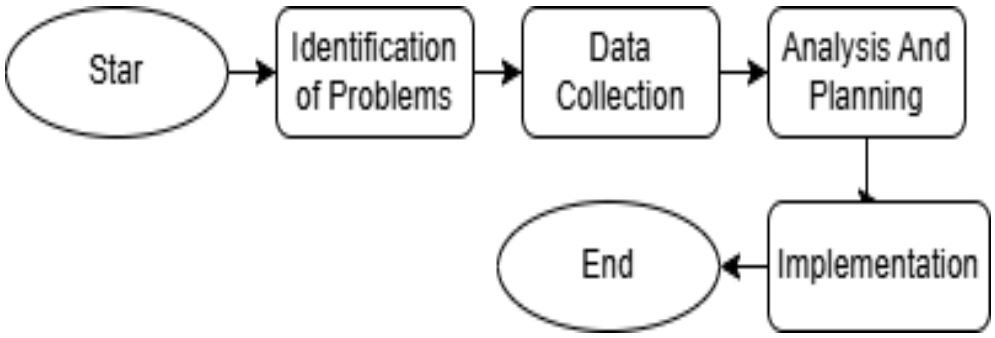

Fig. 1: Research Flow Diagram

2.1. Identification of Problems

In this problem identification, the determination of the background, problem formulation, research objectives, and research benefits is carried out.

2.2. Data Collection

Data collection was conducted throught a literature study, a method that work by gathering population data based on gender obtained from
Department of Population and Civil Registration of Kapuas Hulu Regency, reading or recording, and processing research materials.

2.3. Implementation

The implementation stage is the phase where coding or program development is carried out based on the analysis and design that have been created. It involves implementing the Backpropagation algorithm to perform predictions or forecasts on the population of Kapuas Hulu Regency based on gender and integrating the pre-designed user interface to verify whether it aligns with the design.

3. Results and discussion

In this study, testing was conducted on four parameters: testing the number of epochs, testing the learning rate, testing the number of nodes in the hidden layer, and testing the error tolerance. The tests used data from Kapuas Hulu Regency, categorized based on male and female gender, with training data spanning the years 2016-2021 and testing data from 2022-2023.

3.1. Testing Results for the Number of Epochs

To carry out initial testing, a learning rate of 0.1 was used, the number of nodes in the hidden layer was 3, and a minimum error of 0.001. The results of the epoch test are shown in Table 1.

Table 1: Results of Testing Number of Epochs

Number of Epochs/iterations	**MSE**					**MSE Average**
	Training To-					
	1	**2**	**3**	**4**	**5**	
5000	**0.00340506**	**0.00371346**	**0.00996316**	**0.00305020**	**0.00376331**	0.004779038
6000	**0.00315581**	**0.00291636**	**0.00359280**	**0.00270314**	**0.00283920**	0.003226284
-	-	-	-	-	-	-
9000	0.00234666	0.00273271	0.00206540	0.00252094	0.00220199	0.00251354

Testing the number of epochs aims to obtain the optimal number of epochs for the level of accuracy. The number of epochs used is 5000, 6000, 7000, 8000, and 9000. The optimal epoch is shown at epoch 8000 with the smallest average MSE value, namely 0.002405178

3.2. Test Results Learning rate (α)

The next test will use the number of epochs 8000 as the optimal test result

Table 2: Learning Rate Test Results

Learning Rate	**MSE**					**MSE Average**
	Training To-					
	1	2	3	4	5	
0.1	0.00250787	0.00254586	0.00248774	0.00246151	0.00279111	**0.002558818**
0.2	0.0099993	0.00139387	0.00154143	0.00113908	0.00445095	**0.003705052**

-	-		-	-	-	-
0.9	**0.00099985**	**0.00099996**	**0.00099976**	**0.00099984**	**0.00099969**	**0.00099982**

From the learning rate test results, the optimal value is shown when the learning rate is 0.8 with the smallest average MSE value, namely 0.000999796.

3.3. Hidden layer test results (Number of hidden units)

This test is carried out to obtain the number of hidden units that have the highest accuracy so that the prediction process obtains optimal results.

Table 3: Hidden Layer Test Results

Hidden layer	MSE					MSE Average
	Training To-					
	1	2	3	4	5	
3	0.00099982	0.00213987	0.00099973	0.00099998	0.00099981	0.001227842
5	0.00099971	0.00099983	0.00099990	0.00099981	0.00099973	0.000999796
-	-	-	-	-	-	-
11	0.00099978	0.0009997	0.00099983	0.00099994	0.00099975	0.000999806

The results of testing with hidden layers or the optimal number of hidden units are 5 with the smallest average MSE value, namely 0.000999796

3.4. Eror Tolerance Testing Results

Minimum error testing aims to obtain optimal results regarding the level of accuracy. Minimum error set at 0.01, 0.001, 0.0001 and 0.00001

Table 4 Error tolerance test results

Error tolerance	**MSE**					**MSE Average**
	Training To-					
	1	2	3	4	5	
0.01	0.00996798	0.00995229	0.00995594	0.00995469	0.00995469	0.009958378
-	-	-	-	-	-	
0.00001	0.00057868	0.00052901	0.00052508	0.00054746	0.00051656	0.000539358

From the results it is known that the minimum optimal error is 0.00001 with the average MSE obtained being 0.000539358

3.5. K-FLOD Cross Validation Testing

This test aims to test data patterns and the robustness of the backpropagation algorithm when the training data and test data are changed randomly. In this test, the K-fold cross validation value that will be used is K-4.

Table 1: K-fold cross validation testing

Cross Validation	***MSE***	***Akurasi (%)***
1-fold	***0.0062***	*99.38*
-	-	-
4-fold	*0.6893*	*31.07*

Based on the table above, it can be seen that 1-fold has the smallest MSE value, so what will be used is 1-fold with an MSE value of 0.0062 and an accuracy of 99.38%.

3.6. Prediction Results

From the results of the tests that have been carried out, the parameters used to carry out each training to obtain optimal accuracy are Learning rate 0.8.

Table 6: Prediction Results

Prediction No	**Male (Person)**	**Female (Person)**	**MSE**	**Accuracy (%)**	**Time (Seconds)**
1	138696	131384	0.0055	99.45	**0.04**
2	138590	131239	0.0082	99.18	**0.02**
-	-	-	-	-	-
5	**138424**	**131087**	**0.0121**	**98.79**	**0.02**

From the results of training and testing that has been carried out 5 times, the prediction results with the highest level of accuracy are in the 4th training, the accuracy value obtained is 99.53%, MSE 0.0047 with a time of 0.02 seconds

4. Conclusion

Based on the results of the study on population prediction in Kapuas Hulu Regency by gender using the backpropagation method with various parameter combinations, it can be concluded that the predicted population for the following year is 138,756 males and 131,434 females. The model achieved an accuracy of 99.53% with a Mean Squared Error (MSE) of 0.0047 and an execution time of 0.02 seconds. The best model obtained was a backpropagation artificial neural network with 5 hidden layers, a learning rate of 0.8, 8000 epochs, and an error tolerance of 0.00001. Testing using the k-fold cross-validation method with 4 folds showed the best results in the 1st fold with an accuracy of 99.38%. To further improve accuracy, additional experiments on model parameter variations and increasing the amount of data are recommended.

5. Advice

This research can still be developed as an effort to improve the accuracy of the artificial neural network model used in this research. In predicting the population of Kapuas Hulu district based on gender, you can carry out experiments on the model that will be used, such as the number of hidden layers, epochs, learning rate and error tolerance. Apart from that, looking at the data in this research, the author suggests increasing the amount of data because the amount of data can also influence the results obtained.

References

[1] Undang-undang, "Undang-Undang Nomor 5 Tahun 2014 Tentang Administrasi Kependudukan," *Kementeri. Huk. dan Hak Asasi Mns.*, pp. 1–104, 2014, [Online].
[2] D. Mahsunah, "Analisis Pengaruh Jumlah Penduduk, Pendidikan Dan Pengangguran Terhadap Kemiskinan Undang Di Jawa Timur," *J. Pendidik. Ekon.*, vol. 1, no. 3, pp. 1–17, 2013.
[3] K. Kiki and S. Kunsumadewi, "Jaringan Saraf Tiruan dengan Metode Backpropagation untuk Mendeteksi Gangguan Psikologi," *Media Inform.*, vol. 2, no. 2, pp. 1–11, 2004, doi: 10.20885/informatika.vol2.iss2.art1.
[4] A. Sudarsono, "153271-ID-Jaringan-Syaraf-Tiruan Metode Backpropagation Dalam Memprediksi Potensi Banjir," *J. Inf. Dan Teknol., vol. 1,no. 2, pp. 27-32, 2022*,doi 10.37034/jidt.v4il.17.
[5] A. P. Iskandar, "Efektifitas Jaringan Syaraf Tiruan Metode Backpropagation Dalam Memprediksi Potensi Banjir," *J. Technopreneursh. Inf. Syst.*, vol. 3, no. 2, pp. 50–56, 2020.