Implementation of ANN for Predicting the Percentage of Illiteracy in Indonesia by Age Group

Gita Widi Bhawika¹, P Purwantoro², Achmad Daengs GS³, Dadang Sudrajat⁴, Arrafiqur Rahman⁵, M Makmur⁶, Rina Ari Rohmah⁷, Anjar Wanto⁸

¹ Institut Teknologi Sepuluh Nopember, Surabaya - Indonesia
² Universitas Pasir Pengaraian, Riau - Indonesia
³ Universitas 45 Surabaya, Surabaya - Indonesia
⁴ STMIK IKMI Cirebon, Jawa Barat - Indonesia
⁵ STIKOM Tunas Bangsa Pematangsiantar, Medan - Indonesia

* gitawidi@gmail.com

Abstract. Illiteracy is one of the serious problems experienced by Indonesia. The lack of care from the government and the private sector towards illiterate people makes the illiteracy rate quite high. Because of this, this problem must be one of the government's targets going forward, because directly or indirectly illiteracy plays a role in increasing the number of poverty in Indonesia. The purpose of this study is to predict the percentage of the illiterate population in Indonesia according to the age group 15-44 years because this age group category is a productive age. The results of this prediction are expected to be a reference and benchmark for the government in determining and making policies to reduce illiteracy rates. The data that will be predicted is the illiteracy rate data for each province in Indonesia sourced from the Indonesian Central Bureau of Statistics from 2011 to 2017. The method used for prediction is Backpropagation Artificial Neural Network. Data analysis and calculation were carried out with the help of Matlab and Microsoft Excel software. This study uses 5 architectures, 4-5-1, 4-6-1, 4-9-1, 4-14-1 and 4-18-1. Of the five models, the best network architecture is 4-14-1 with an accuracy rate of 91% and the Mean Squared Error 0.00274166. Using this 4-14-1 network model, a prediction on the percentage of illiteracy in Indonesia will be calculated for 2018 until 2020.

1. Introducing

Illiteracy is a condition where a person has not or cannot write and read, or in other words the lack of adequate education. The high rate of illiteracy in various ages shows indicators of progress from a country. Low literacy and high literacy rates indicate the existence of an effective primary education system. Poverty is one of the inhibiting factors for people to study, which has led to an increase in illiterate people in the community. One of the most significant impacts of illiteracy is the increasingly widespread crime rate and low social welfare among the people [1][2]. Especially in developing countries like Indonesia. Therefore efforts that can be made in overcoming these problems are illiteracy eradication programs for people who have low levels of education. In addition to the limitations of learning facilities, human resources, government funds in conducting learning activities for the community or factors of public ignorance exacerbate conditions that make people remain illiterate.

In Indonesia itself based on Central Bureau of Statistics data, that the percentage of illiterate population based on all age groups, there are 11 provinces with illiterate lifting above the national
figure, namely Papua (28.75%), West Nusa Tenggara (7.91%), East Nusa Tenggara (5.15%), West Sulawesi (4.58%), West Kalimantan (4.50%), South Sulawesi (4.49%), Bali (3%), East Java (3.47%), North Kalimantan (2.90%), Southeast Sulawesi (2.74%), and Central Java (2.20%). While the percentage of illiterate population based on productive age (ages 15-44 years) whose data was taken in the last 7 years, namely 2011-2017 (table 1), that the province with a high percentage of illiteracy is Papua (29.89%), West Sulawesi (4.36%), West Nusa Tenggara (4.25%), East Nusa Tenggara (4%), South Sulawesi (3.01%), West Papua (3.01%), Southeast Sulawesi (1.91%), East Java (1.63%), Central Sulawesi (1.60%), Gorontalo (1.47%), North Kalimantan (1.36%), Bali (1.31%), and Maluku (1.14%). While the percentage in other provinces is below 1% [3].

| No | Province           | Percentage (2011) | Percentage (2012) | Percentage (2013) | Percentage (2014) | Percentage (2015) | Percentage (2016) | Percentage (2017) | Average Percentage |
|----|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|
| 1  | Aceh              | 1.29              | 1.03              | 0.67              | 0.43              | 0.27              | 0.22              | 0.13              | 0.58               |
| 2  | North Sumatra     | 1.64              | 1.10              | 0.9               | 0.66              | 0.51              | 0.37              | 0.38              | 0.79               |
| 3  | West Sumatra      | 1.26              | 1.05              | 0.64              | 0.43              | 0.32              | 0.17              | 0.19              | 0.58               |
| 4  | Riau              | 1.05              | 0.87              | 0.67              | 0.48              | 0.33              | 0.20              | 0.15              | 0.54               |
| 5  | Jambi             | 1.43              | 1.33              | 0.89              | 0.57              | 0.49              | 0.35              | 0.29              | 0.76               |
| 6  | South Sumatra     | 1.17              | 0.91              | 0.72              | 0.52              | 0.48              | 0.35              | 0.25              | 0.63               |
| 7  | Bengkulu          | 1.31              | 1.09              | 0.68              | 0.54              | 0.48              | 0.36              | 0.27              | 0.68               |
| 28 | Southeast Sulawesi| 3.15              | 2.84              | 2.05              | 1.62              | 1.37              | 1.19              | 1.14              | 1.91               |
| 29 | Gorontalo         | 3.27              | 2.88              | 1.44              | 1.10              | 0.61              | 0.49              | 0.50              | 1.47               |
| 30 | West Sulawesi     | 6.45              | 5.86              | 4.73              | 3.93              | 3.33              | 3.06              | 3.16              | 4.36               |
| 31 | Maluku            | 1.94              | 1.62              | 1.30              | 0.81              | 0.80              | 0.76              | 0.77              | 1.14               |
| 32 | North Maluku      | 1.89              | 1.52              | 1.00              | 0.57              | 0.47              | 0.36              | 0.41              | 0.89               |
| 33 | West Papua        | 5.37              | 4.53              | 2.93              | 2.27              | 2.09              | 1.97              | 1.94              | 3.01               |
| 34 | Papua             | 34.55             | 33.4             | 31.44             | 28.5              | 28.47             | 28.21             | 24.66             | 29.89              |

Source: Indonesian Central Bureau of Statistics

Because of the literacy, the community is unable to access information and develop knowledge so that illiterate people cannot be separated from ignorance, poverty and lagging behind others. For this reason, one way to reduce illiterate population is by predicting or forecasting illiterate population in each province in Indonesia for the following years. Thus the government, the private sector, and the general public who care about eradicating illiteracy have references and references so that later they can determine the right policies and solutions to overcome illiteracy in Indonesia. This study focused on predicting the percentage of the illiterate population in Indonesia according to the age group 15-44 years; this is because the age group category is a productive age group. While the prediction method used is Backpropagation Neural Network. This method is a method that is quite well used for prediction or forecasting, especially those related to time series data [4]–[6].

2. Methodology

2.1. Data studied

Data on the percentage of the illiterate population according to the age group 15-44 years, 2011-2017. Data were obtained from the Indonesian Central Bureau of Statistics [3].

2.2. Related research

Research has been conducted to predict illiteracy through mobile phone log. This study discusses how the use of supervised machine learning to predict illiteracy in developing countries especially Asia, which is externally validated against large-scale surveys. The average model used is ten times better than random guessing with 70% accuracy. Furthermore, this study reveals how illiteracy can be collected and mapped geographically [7]. Research has been conducted to predict the Human Development Index using the Backpropagation algorithm. The results of the truth of this study were more than 90% with the MSE results obtained at 0.0006386600. The best network architecture model in this study is 3-48-1 which is used to predict the Human Development Index data [8].
2.3. Research Framework

Based on Figure 1, it can be explained that Data Collection is the first thing done; Data was obtained from the Indonesian Central Bureau of Statistics, namely data on the percentage of the illiterate population according to the age group 15-44 years, 2011-2017. Then proceed with Literature Study which is looking for or completing references related to the research that the author adopted. The next step is Problem Identification; that includes the results of the introduction of problems or inventory of issues. Then proceed with Pre-process; that is to make changes to data become more consistent and prepare raw data into quality data. Followed by pattern determination; namely, the model used to make predictions. Furthermore, the results of Data Processing Testing are carried out, namely, the decision of data tests performed using the Matlab application. The next stage is prediction; namely the process of predicting a variable in the future by comparing the network architecture model with the Backpropagation method, and selecting the most accurate and high level of truth. The last step is to evaluate whether the processing of test data is as expected.

2.4. Data Normalization

Normalization formula used [9]–[13]:

\[ x' = \frac{0.8(x - a)}{b - a} + 0.1 \]  

(1)

where \( x' \) is the normalized result, \( x \) is the data to be normalized, \( a \) is the lowest value of the data, and \( b \) is the largest value of the data. Data on the percentage of the illiterate population according to the age group 15-44 years, 2011-2017 (table 1) are then divided into 2 (training and testing), data from 2011-2014 as input training data, and 2015 data used as target training data. Data for 2013-2016 is used as input testing, and 2017 data is used as target testing.

Table 2. Normalization of training (2011-2014) / Target 2015

| Province  | Population Age (15-44 Years) | 2011 | 2012 | 2013 | 2014 | Target |
|-----------|-------------------------------|------|------|------|------|--------|
| Aceh      |                               | 0.12987 | 0.12385 | 0.11551 | 0.10996 | 0.10625 |
| North Sumatra |                           | 0.13797 | 0.12547 | 0.12084 | 0.11528 | 0.11181 |
| West Sumatra |                             | 0.12918 | 0.12431 | 0.11482 | 0.10996 | 0.10741 |
| Riau      |                               | 0.12431 | 0.12014 | 0.11551 | 0.11111 | 0.10764 |
| Maluku    |                               | 0.14492 | 0.13751 | 0.13010 | 0.11876 | 0.11852 |
| North Maluku |                           | 0.14376 | 0.13520 | 0.12315 | 0.11320 | 0.10888 |
| West Papua |                               | 0.22434 | 0.20489 | 0.16784 | 0.15256 | 0.14839 |
| Papua     |                               | 0.90000 | 0.87337 | 0.82799 | 0.75991 | 0.75922 |

Table 3. Normalization of testing (2013-2016) / Target 2017

| Province  | Population Age (15-44 Years) | 2013 | 2014 | 2015 | 2016 | Target |
|-----------|-------------------------------|------|------|------|------|--------|
| Aceh      |                               | 0.11705 | 0.11094 | 0.10687 | 0.10560 | 0.10331 |
| North Sumatra |                           | 0.12290 | 0.11679 | 0.11298 | 0.10941 | 0.10967 |
| West Sumatra |                             | 0.11628 | 0.11094 | 0.10814 | 0.10433 | 0.10483 |
| Riau      |                               | 0.11705 | 0.11221 | 0.10840 | 0.10509 | 0.10382 |
| Maluku    |                               | 0.13308 | 0.12061 | 0.12036 | 0.11934 | 0.11959 |
| North Maluku |                           | 0.12545 | 0.11450 | 0.11196 | 0.10916 | 0.11043 |
| West Papua |                               | 0.17455 | 0.15776 | 0.15318 | 0.15013 | 0.14936 |
| Papua     |                               | 0.90000 | 0.82519 | 0.82443 | 0.81781 | 0.72748 |
3. Results and Discussion

This study uses 5 network architecture models, including: 4-5-1 (4 is the input layer, 5 is hidden layer neurons, and 1 is output layer), 4-6-1 (4 is the input layer, 6 is hidden layer neuron, and 1 is output layer), 4-9-1 (4 is the input layer, 9 is hidden layer neurons, and 1 is output layer), 4-14-1 (4 is the input layer, 14 are hidden layer neurons, and 1 is output layer), 4-18-1 (4 is the input layer, 18 is hidden layer neurons, and 1 is output layer). Of the five architectures, the best structure is 4-14-1 with an accuracy rate of 91% and epoch as many as 314 iterations.

![Figure 2: Results of Data Training with 4-14-1 architecture](image)

From the architectural model in Figure 2 above, it can be explained that the Epoch that occurs is 314 iterations with a duration of 4 seconds with Best Training Performance of 0.00099625. Based on table 4, error = obtained from Target-Output, SSE = obtained from Error ^ 2, Total = Number of SSE produced from number 1 to number 34, Result = If the value of error in testing data <= 0.05 then the result is correct (1). If not then wrong (0). Accuracy = obtained from the correct number of results ((no / 34) * 100), yielding 91%, Margin Error = obtained from the number of incorrect results in the pattern ((pattern results / 34) * 100) or obtained from the maximum number of accuracy (100%) minus the accuracy produced, yielding 91%. MSE = Obtained from Total SSE / 34 (number of no), 1 = True 0 = False.

| Patterns | Target | Output | Error  | SSE        | Results |
|----------|--------|--------|--------|------------|---------|
| 1        | 0.10331| 0.11100| -0.00769| 0.00005917 | 1       |
| 2        | 0.10967| 0.11190| -0.00223| 0.000000498| 1       |
| 3        | 0.10483| 0.11090| -0.00607| 0.000003679| 1       |
| 4        | 0.10382| 0.11040| -0.00658| 0.000004334| 1       |
| 5        | 0.10738| 0.11300| -0.00562| 0.000003159| 1       |
| 6        | 0.10636| 0.11760| -0.01124| 0.000012631| 1       |
| 7        | 0.10687| 0.11830| -0.01143| 0.000013064| 1       |
| 8        | 0.10407| 0.11180| -0.00773| 0.000005973| 1       |
| 9        | 0.11781| 0.12460| -0.00679| 0.000004608| 1       |
| 10       | 0.10738| 0.11610| -0.00872| 0.000007605| 1       |
| 11       | 0.10153| 0.12180| -0.02027| 0.000041101| 1       |
| 12       | 0.10662| 0.11320| -0.00658| 0.000004335| 1       |
| 13       | 0.10916| 0.10800| 0.00116 | 0.000000135| 1       |
| 14       | 0.10509| 0.12510| -0.02001| 0.000040044| 1       |
| 15       | 0.12570| 0.12190| 0.00380 | 0.000014444| 1       |
| 16       | 0.10534| 0.11040| -0.00506| 0.00002557 | 1       |
| 17       | 0.11018| 0.10040| 0.00978 | 0.000009561| 0       |
In table 6, you can see the prediction of the percentage of illiterate people in Indonesia for the next three years, namely 2018-2020.

| No  | Province                | Previous Data | Average | Prediction Data | Average |
|-----|-------------------------|---------------|---------|-----------------|---------|
|     |                         | 2015 | 2016 | 2017 | Overall | 2018 | 2019 | 2020 | Overall |
| 1   | Aceh                    | 0.27 | 0.22 | 0.13 | 0.21   | 0.23 | 0.18 | 0.18 | 0.20   |
| 2   | North Sumatra           | 0.51 | 0.37 | 0.38 | 0.42   | 0.24 | 0.24 | 0.19 | 0.22   |
| 3   | West Sumatra            | 0.32 | 0.17 | 0.19 | 0.23   | 0.23 | 0.17 | 0.24 | 0.21   |
| 4   | Riau                    | 0.33 | 0.20 | 0.15 | 0.23   | 0.21 | 0.10 | 0.14 | 0.15   |
| 5   | Jambi                   | 0.49 | 0.35 | 0.29 | 0.38   | 0.28 | 0.19 | 0.10 | 0.19   |
| 6   | South Sumatra           | 0.48 | 0.35 | 0.25 | 0.36   | 0.42 | 0.18 | 0.26 | 0.29   |
| 7   | Bengkulu                | 0.48 | 0.36 | 0.27 | 0.37   | 0.44 | 0.20 | 0.30 | 0.31   |
| 8   | Lampung                 | 0.34 | 0.21 | 0.16 | 0.24   | 0.26 | 0.15 | 0.20 | 0.20   |
| 9   | Bangka Belitung Islands | 0.87 | 0.81 | 0.70 | 0.79   | 0.59 | 0.51 | 0.21 | 0.44   |
| 10  | Riau islands            | 0.29 | 0.28 | 0.29 | 0.29   | 0.38 | 0.45 | 0.48 | 0.44   |
| 11  | DKI Jakarta             | 0.06 | 0.08 | 0.06 | 0.07   | 0.58 | 0.44 | 0.98 | 0.67   |
| 12  | West Java               | 0.29 | 0.23 | 0.26 | 0.26   | 0.30 | 0.35 | 0.38 | 0.34   |
| 13  | Central Java            | 0.50 | 0.36 | 0.36 | 0.41   | 0.13 | 0.22 | 0.07 | 0.14   |
| 14  | Yogyakarta              | 0.19 | 0.13 | 0.20 | 0.17   | 0.68 | 0.53 | 0.19 | 0.77   |
| 15  | East Java               | 1.24 | 1.09 | 1.01 | 1.11   | 0.50 | 0.51 | 0.11 | 0.37   |
| 16  | Banten                  | 0.33 | 0.19 | 0.21 | 0.24   | 0.22 | 0.17 | 0.22 | 0.20   |
| 17  | Bali                    | 0.61 | 0.51 | 0.40 | 0.51   | 0.12 | 0.02 | 0.27 | 0.14   |
| 18  | West Nusa Tenggara     | 3.31 | 3.26 | 3.20 | 3.26   | 2.94 | 3.25 | 3.45 | 3.21   |
| 19  | East Nusa Tenggara     | 3.10 | 3.06 | 3.08 | 3.08   | 2.59 | 2.97 | 2.95 | 2.84   |
| 20  | West Kalimantan         | 2.00 | 1.88 | 1.76 | 1.88   | 1.23 | 1.26 | 0.62 | 1.04   |
| 21  | Central Kalimantan      | 0.30 | 0.19 | 0.23 | 0.24   | 0.32 | 0.34 | 0.40 | 0.35   |
| 22  | South Kalimantan        | 0.19 | 0.11 | 0.15 | 0.15   | 0.28 | 0.31 | 0.45 | 0.35   |
| 23  | East Kalimantan         | 0.13 | 0.12 | 0.15 | 0.13   | 0.29 | 0.44 | 0.51 | 0.41   |
| 24  | North Kalimantan        | 1.36 | 1.31 | 1.38 | 1.35   | 1.46 | 1.20 | 1.77 | 1.48   |
| 25  | North Sulawesi          | 0.17 | 0.15 | 0.19 | 0.17   | 0.45 | 0.48 | 0.73 | 0.55   |
### 4. Conclusion

This study concludes that the 4-14-1 architectural model can predict with a real level of 91%. Also, judging from the comparison table between the previous three years' data and predictive data over the next three years, it can be seen that the percentage level of the average illiterate population in Indonesia has not undergone significant changes. Papua Province is still the highest with a percentage rate of illiteracy of 26.28% or a decrease of 0.84% compared to the percentage of average illiteracy from 2015-2017.

### References

[1] S. W. Suh *et al.*, “Impacts of Illiteracy on the Risk of Dementia: A Global Health Perspective,” *Journal of Alzheimer’s Disease*, vol. 53, no. 2, pp. 731–741, 2016.

[2] K. Gupta, “Globalization and Women Empowerment,” *International Journal of Behavioral Social and Movement Sciences*, vol. 1, no. 1, pp. 1–4, 2017.

[3] I. C. B. of Statistics, “Percentage of Illiterate Population according to Age Group, 2011-2017,” *Indonesian Central Bureau of Statistics*, 2018. [Online]. Available: https://www.bps.go.id/linkTableDinamis/view/id/1056.

[4] B. Febriadi, Z. Zamzami, Y. Yunefri, and A. Wanto, “Bipolar function in backpropagation algorithm in predicting Indonesia’s coal exports by major destination countries,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 12089, pp. 1–9, 2018.

[5] N. Nasution, A. Zamsuri, L. Lisnawita, and A. Wanto, “Polak-Ribiere updates analysis with binary and linear function in determining coffee exports in Indonesia,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 12089, pp. 1–9, 2018.

[6] A. Wanto, A. P. Windarto, D. Hartama, and I. Parlina, “Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density,” *International Journal Of Information System & Technology*, vol. 1, no. 1, pp. 43–54, 2017.

[7] P. Sundsøy, “Mitigating the Risks of Financial Exclusion: Predicting Illiteracy with Standard Mobile Phone Logs,” in *Social, Cultural, and Behavioral Modeling*, 2017, pp. 309–314.

[8] S. P. Siregar and A. Wanto, “Analysis of Artificial Neural Network Accuracy Using Backpropagation Algorithm In Predicting Process (Forecasting),” *International Journal Of Information System & Technology*, vol. 1, no. 1, pp. 34–42, 2017.

[9] A. Wanto, M. Zarlis, Sawaluddin, and D. Hartama, “Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process,” *Journal of Physics: Conference Series*, vol. 930, no. 1, pp. 1–7, 2017.

[10] S. Sumijan, A. Windarto, A. Muhammad, and B. Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” *International Journal of Software Engineering and Its Applications*, vol. 10, no. 10, pp. 189–204, 2016.

[11] M. Fauzan *et al.*, “Epoch Analysis and Accuracy 3 ANN Algorithm Using Consumer Price Index Data in Indonesia,” 2018, pp. 1–7.

[12] A. Wanto *et al.*, “Analysis of Standard Gradient Descent with GD Momentum And Adaptive LR for SPR Prediction,” 2018, pp. 1–9.

[13] A. Wanto *et al.*, “Levenberg-Marquardt Algorithm Combined with Bipolar Sigmoid Function to Measure Open Unemployment Rate in Indonesia,” 2018, pp. 1–7.