Analysis of Backpropagation Algorithms in Predicting World Internet Users

Sunil Setti\textsuperscript{1}, Anjar Wanto\textsuperscript{1}, Muhammad Syafiq\textsuperscript{1}, Andriano Andriano\textsuperscript{1}, Bil Klinton Sihotang\textsuperscript{1}

\textsuperscript{1}STIKOM Tunas Bangsa, Sudirman Street Pematangsiantar, Medan - Indonesia

* sunil.setti98@gmail.com

Abstract. The internet is now a primary need for its users. According to the e-Marketer market research institute, there are the top 25 countries with the highest number of internet users in the world. Indonesia is in the sixth position with a total of 112.6 million internet users. With the increasing number of internet users, it is expected to be able to contribute to advancing the economy and education in a country. To be able to increase the number of internet users, especially in Indonesia, it is necessary to predict in the coming years so that the government can provide adequate facilities and infrastructure in order to compensate for the growing number of internet users and as a precaution when there is a decrease in the number of internet users. The data used in this study focus on the data on the number of internet users in 25 countries in 2013-2017 sourced from the Indonesian Ministry of Communication and Information. The algorithm used is the Backpropagation Neural Network. Data analysis was performed using Artificial Neural Network method using Matlab R2011b. This study uses 5 architectural models. The best network architecture produced is 3-50-1 with an accuracy rate of 92% and the Mean Squared Error (MSE) value is 0.00151674.

1. Introducing

Since the discovery of internet technology in the 1990s its use has been widespread because it was considered to provide enormous benefits for the smooth process of various activities [1]. The internet in the information age has placed itself as one of the information centers that can be concerned with various places without being limited by space and time. The internet is ordered a barrier-free information center because it can connect one information site to another information site in a short time [2]. The internet is now a primary need for its users [3].

The growth of internet users is increasing along with the increasingly diverse benefits of the internet itself. With the increasing number of internet users, it is expected to be able to participate in advancing the economy and education in the country concerned. According to the e-Marketer market research institute, there are the top 25 countries with the highest number of internet users in the world. Until now it is known that China is the country with the most internet users with a total of 736.2 million users in 2017. In Indonesia alone, internet users continue to increase from year to year with an extraordinary percentage increase [4]. At this time Indonesia managed to occupy the sixth position after Japan with a total of 112.6 million internet users. And at number 25 is held by South Africa with the number of internet users as much as 29.2 million. For more details, see table 1 below:
IOP Conf. Series: Journal of Physics: Conf. Series 1255 (2019) 012018   doi:10.1088/1742-6596/1255/1/012018

Table 1. Internet users in the Big 25 World Countries

| No | Country’s     | Internet Users (millions) |
|----|---------------|----------------------------|
|    |               | 2013 | 2014 | 2015 | 2016 | 2017 |
| 1  | China         | 620,7 | 643,6 | 669,8 | 700,1 | 736,2 |
| 2  | United State  | 246,0 | 252,9 | 259,3 | 264,9 | 269,7 |
| 3  | India         | 167,2 | 215,6 | 252,3 | 283,3 | 313,8 |
| 4  | Brazil        | 99,2  | 107,7 | 113,7 | 119,8 | 123,3 |
| 5  | Japan         | 100,0 | 102,1 | 103,6 | 104,5 | 105,0 |
| 6  | Indonesia     | 72,8  | 83,7  | 93,4  | 102,8 | 112,6 |
| 7  | Russia        | 77,5  | 82,9  | 87,3  | 91,4  | 94,3  |
| 8  | Germany       | 59,5  | 61,6  | 62,2  | 62,5  | 62,7  |
| 9  | Mexico        | 53,1  | 59,4  | 65,1  | 70,7  | 75,7  |
| 10 | Nigeria       | 51,8  | 57,7  | 63,2  | 69,1  | 76,2  |
| 11 | United Kingdom| 48,8  | 50,1  | 52,3  | 52,4  | 53,4  |
| 12 | France        | 48,8  | 49,7  | 50,5  | 51,2  | 51,9  |
| 13 | Philippines   | 42,3  | 48,0  | 53,7  | 59,1  | 64,5  |
| 14 | Turkey        | 36,6  | 41,0  | 44,7  | 47,7  | 50,7  |
| 15 | Vietnam       | 36,6  | 42,5  | 44,4  | 48,2  | 52,1  |
| 16 | South Korea   | 40,1  | 40,4  | 40,6  | 40,7  | 40,9  |
| 17 | Egypt         | 34,1  | 36,0  | 38,3  | 40,9  | 43,9  |
| 18 | Italy         | 34,5  | 35,8  | 36,2  | 37,2  | 37,5  |
| 19 | Spain         | 30,5  | 31,6  | 32,3  | 33,0  | 33,5  |
| 20 | Canada        | 27,7  | 28,3  | 28,8  | 29,4  | 29,9  |
| 21 | Argentina     | 25,0  | 27,1  | 29,0  | 29,8  | 30,5  |
| 22 | Colombia      | 24,2  | 26,5  | 28,6  | 29,4  | 30,5  |
| 23 | Thailand      | 22,7  | 24,3  | 26,0  | 27,6  | 29,1  |
| 24 | Poland        | 22,6  | 22,9  | 23,3  | 23,7  | 24,0  |
| 25 | South Africa  | 20,1  | 22,7  | 25,0  | 27,2  | 29,2  |

One of the advantages of increasing internet users is the increasing public insight, especially among students. Another advantage is that it can improve business opportunities for the community, both middle and upper class, for example by the presence of online shops and ojek online. With increasing business opportunities, it also helps to open jobs for the community, and this also helps increase per capita income from a country, especially Indonesia. Therefore, one of the ways we can improve internet users, especially in Indonesia. Is to make predictions (forecasting) for years to come. Thus, the government can provide adequate facilities and facilities to balance the growth of internet users and vice versa; the government can also take steps to anticipate a decline in internet users. However, the prediction process is not simple; it makes a basic model and time series data of these problems, which are regularly complicated by estimating the accuracy that is not easily achieved, thus requiring more advanced techniques [5][6]. Backpropagation is one method that is quite good in predicting (forecasting) [7]. It is advised that with the use of this method it will get the results as desired because backpropagation algorithms allow avoiding difficulties by using learning rules that are similar to the plasticity of time spikes that depend on synapses [8][9].

In previous researches, [10] Research has been conducted to predict the exchange rate of the rupiah against the US Dollar using Backpropagation Gradient Descent Time Series with conjugate Gradient optimization. The results show that the MSE value of the backpropagation gradient descent base algorithm is 1.02159 and the conjugate gradient produces MSE of 0.0198012. From these results, the conjugate gradient algorithm is superior because it allows a smaller error. Next, [11] predicting stock price movements in BRI as a state bank and BCA as a private bank which is the best bank based on its core capital which is above 30 trillion Rupiah in 2013 with the Backpropagation Neural Network method. This research produces the smallest 0.0626 and MAE RMSE which is 0.0456 with the 10-5-1 architectural model for the closing price at BCA and for the closing price at BRI the smallest RMSE results are 0.084 and MAE 0.0487 with the same architectural model. Research by [12] conducted to predict inflation rates in Indonesia using the Backpropagation method. Data were analyzed using Cross Validation techniques. Based on the results of the study it was found that the neural network
with 80% cross-validation Training and 20% validation and testing and learning rate 0.1 and the number of epochs 10000 capable of providing small MSE and MAE values, and forecasting results close to the actual data.

2. Methodology

2.1. Data Used

The data used in this study did source from the Indonesian Ministry of Communication and Informatics, namely the data of Internet Users in 25 Large World Countries in 2013-2017 (can be seen in Table 1 in the previous discussion). The training data that will be used is 2013-2015 data with a target for 2016. While the testing data to be used is 2014-2016 data with a target for 2017.

2.2. Research Framework

Broadly speaking, the framework in this study can be described as follows:

![Research Framework Diagram]

- **Start**
- **Collecting Data**
- **Identifying Problems**
- **Preprocessing**
- **Determine Pattern**
- **Test Data Processing Results**
- **Predict**
- **Final Evaluation**
- **Journal, Proceedings and Books of Artificial Neural Networks**

**Figure 1. Research Framework**

From the framework of Figure 1, it can be explained that collecting data in a study is the first thing to do. The second stage was carried out a literature study to complement the basic knowledge and theories used in this study. The third stage identifies the problem to process the conversion phase of the data obtained in accordance with the specified weight. The fourth stage preprocesses with the aim to facilitate understanding of the contents of the record. The fifth stage determines the pattern and determination of the network architecture model that is adapted to the research problem faced. The sixth stage examines the results of data processing using the Matlab application. The seventh stage predicts that is to see comparisons of several architectural models used in the study in order to obtain the best architectural models and the most accurate level of accuracy. The eighth stage evaluates the end to find out whether the results of data processing are as desired.
2.3. Data Normalization

The initial data that has been collected will be normalized by using a normalization formula that will produce a value between 0 and 1, this is done in accordance with the normalization provisions in Equation (1)

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

(1)

where \( x' \) is the normalized data, \( x \) is the normalized data, \( a \) is the data with the smallest value, and \( b \) is the maximum data with the highest value.

In tables 2 and 3 you can see the training data and normalized testing data using the equation formula (1).

**Table 2. Normalization of Data Training**

| No | Country | Input 2013 | Input 2014 | Input 2015 | Target |
|----|---------|------------|------------|------------|--------|
| 1  | China   | 0.80659    | 0.83535    | 0.86435    | 0.90000 |
| 2  | US      | 0.36576    | 0.37388    | 0.38141    | 0.38800 |
| 3  | India   | 0.27306    | 0.33000    | 0.37318    | 0.40965 |
| 4  | Brazil  | 0.19306    | 0.20306    | 0.21012    | 0.21729 |
| 5  | Japan   | 0.19400    | 0.19647    | 0.19824    | 0.19929 |
| 6  | Indonesia | 0.16200   | 0.17482    | 0.18624    | 0.19729 |
| 7  | Russia  | 0.16753    | 0.17388    | 0.17906    | 0.18388 |
| 8  | Germany | 0.14635    | 0.14882    | 0.14953    | 0.14988 |
| 9  | Mexico  | 0.13882    | 0.14624    | 0.15294    | 0.15953 |
| 10 | Nigeria | 0.13729    | 0.14424    | 0.15071    | 0.15765 |
| 11 | UK      | 0.13376    | 0.13529    | 0.13788    | 0.13800 |
| 12 | France  | 0.13376    | 0.13482    | 0.13576    | 0.13659 |
| 13 | Philippines | 0.12612  | 0.13282    | 0.13953    | 0.14588 |
| 14 | Turkey  | 0.11941    | 0.12459    | 0.12894    | 0.13247 |
| 15 | Vietnam | 0.11941    | 0.12635    | 0.12859    | 0.13306 |
| 16 | South Korea | 0.12353 | 0.12388    | 0.12412    | 0.12424 |
| 17 | Egypt   | 0.11647    | 0.11871    | 0.12141    | 0.12447 |
| 18 | Italy   | 0.11694    | 0.11847    | 0.11894    | 0.12012 |
| 19 | Spain   | 0.11224    | 0.11353    | 0.11435    | 0.11518 |
| 20 | Canada  | 0.10894    | 0.10965    | 0.11024    | 0.11094 |
| 21 | Argentina | 0.10576   | 0.10824    | 0.11047    | 0.11141 |
| 22 | Colombia | 0.10482    | 0.10753    | 0.11000    | 0.11094 |
| 23 | Thailand | 0.10306    | 0.10494    | 0.10694    | 0.10882 |
| 24 | Poland  | 0.10294    | 0.10329    | 0.10376    | 0.10424 |
| 25 | South Africa | 0.10000  | 0.10306    | 0.10576    | 0.10835 |

**Table 3. Normalization of Data Testing**

| No | Country | Input 2014 | Input 2015 | Input 2016 | Target |
|----|---------|------------|------------|------------|--------|
| 1  | China   | 0.79617    | 0.82555    | 0.85952    | 0.90000 |
| 2  | US      | 0.35811    | 0.36528    | 0.37156    | 0.37694 |
| 3  | India   | 0.31629    | 0.35744    | 0.39219    | 0.42639 |
| 4  | Brazil  | 0.19530    | 0.20203    | 0.20887    | 0.21280 |
| 5  | Japan   | 0.18903    | 0.19071    | 0.19172    | 0.19228 |
| 6  | Indonesia | 0.16840   | 0.17927    | 0.18981    | 0.20080 |
| 7  | Russia  | 0.16750    | 0.17243    | 0.17703    | 0.18028 |
| 8  | Germany | 0.14362    | 0.14429    | 0.14463    | 0.14485 |
| 9  | Mexico  | 0.14115    | 0.14754    | 0.15382    | 0.15943 |
| 10 | Nigeria | 0.13924    | 0.14541    | 0.15203    | 0.15999 |
| 11 | UK      | 0.13072    | 0.13319    | 0.13330    | 0.13442 |
| 12 | France  | 0.13027    | 0.13117    | 0.13196    | 0.13274 |
| 13 | Philippines | 0.12837  | 0.13476    | 0.14081    | 0.14687 |
| 14 | Turkey  | 0.12052    | 0.12467    | 0.12803    | 0.13139 |
| 15 | Vietnam | 0.12220    | 0.12433    | 0.12859    | 0.13296 |
| 16 | South Korea | 0.11985   | 0.12007    | 0.12018    | 0.12041 |
| 17 | Egypt   | 0.11491    | 0.11749    | 0.12041    | 0.12377 |
| 18 | Italy   | 0.11469    | 0.11514    | 0.11626    | 0.11659 |
| 19 | Spain   | 0.10998    | 0.11076    | 0.11155    | 0.11211 |
| 20 | Canada  | 0.10628    | 0.10684    | 0.10751    | 0.10807 |
| 21 | Argentina | 0.10493   | 0.10706    | 0.10796    | 0.10875 |
| 22 | Colombia | 0.10426    | 0.10662    | 0.10751    | 0.10875 |
| 23 | Thailand | 0.10179    | 0.10370    | 0.10549    | 0.10718 |
| 24 | Poland  | 0.10022    | 0.10067    | 0.10112    | 0.10146 |
| 25 | South Africa | 0.10000  | 0.10258    | 0.10505    | 0.10729 |

3. Results and Discussion

This study uses 5 architectures. Among others 3-5-1, 3-7-1, 3-19-1, 3-43-1, 3-50-1. Of these 5 architectures, the best architecture is 3-50-1 with an accuracy rate of 92% and epoch 4218 iterations.
From Figure 2, it can be explained that Epoch happens to be 4218 with a duration of 19 seconds. 3-50-1 architecture model means that 3 is input data, 50 is hidden layer while 1 is output or result.

In Table 4, the comparison of the 5 network architecture models used. Of these 5 architectural models, the Epoch level and time are obtained using the Matlab 2011b application, while the MSE and Accuracy of each architectural model are obtained using calculations in Microsoft Excel. Based on Table 5, the best architectural model is 3-50-1 with an accuracy rate of 92%.

### Table 4. Comparison of the Accuracy of All Architectural Models

| No | Architecture | Training | Testing |
|----|--------------|----------|---------|
|    |              | Epoch    | Times   | MSE       | Accuracy |
| 1  | 3-5-1        | 3942     | 00:15   | 0.00093729 | 80%       |
| 2  | 3-7-1        | 1136     | 00:05   | 0.00071043 | 68%       |
| 3  | 3-19-1       | 2726     | 00:12   | 0.00113505 | 56%       |
| 4  | 3-43-1       | 705      | 00:03   | 0.00102460 | 48%       |
| 5  | 3-50-1       | 4218     | 00:19   | 0.00151674 | 92%       |

4. Conclusion

The conclusion that can be drawn from this study is that the 3-50-1 architectural model is the best architectural model of the 5 models used to predict the number of internet users in 25 countries of the world using backpropagation algorithms. The accuracy obtained reached 92% with an epoch of 4218 iterations and MSE 0.00151674.

References

[1] S. P. Lestari, "The Relationship between Advertising Exposure and Product Quality With the Decision to Use Telkom Speedy in the City of Semarang," Inspirational Scientific Magazine,
vol. 3, no. 5, pp. 1–15, 2018.

[2] I. Novianto, "Student Internet Use Behavior," Libri_Net, vol. 2, no. 1, pp. 1–40, 2013.

[3] Y. Warisyah, "The Importance of" Dialogical Assistance "Parents in the Use of Gadgets in Early Childhood," in Learning Innovation for Progressing Education, 2015, no. November 2015, pp. 1–9.

[4] Dulkhatif, Andi Tri Haryono, and Moh Mukeri Warso, "The Influence of Service Quality, Customer Satisfaction and Location on Customer Loyalty at Internet Study Service provider PT Noken Mulia Tama Semarang," Journal of Management, vol. 2, no. 2, pp. 1–34, 2016.

[5] 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.

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

[8] N. Nasution, A. Zamsuri, L. Lisnawita, and A. Wanto, “Polak-Ribiere Updates Analysis with Binary and Linear Function in Determining Coffee Exports in Indonesia,” in IOP Conference Series, 2018, pp. 1–8.

[9] B. Febriadi, Z. Zamzami, Y. Yunefr, and A. Wanto, “Bipolar Function in Backpropagation Algorithm in Predicting Indonesia’s Coal Exports by Major Destination Countries,” in IOP Conference Series, 2018, pp. 1–8.

[10] H. Y. Sari, "Conjugate Gradient Optimization in Backpropagation Neural Network Algorithms for Prediction of Time Series Rates," Gema Aktualita Journal, vol. 5, no. 1, pp. 86–90, 2016.

[11] A. Novita, "Prediction of Stock Price Movements in the Largest Bank in Indonesia Using the Backpropagation Neural Network Method."

[12] I. Saluza, "Application of Backpropagation Neural Networks in Forecasting Inflation in Indonesia," Gradient Journal, vol. 11, no. 1, pp. 1075–1078, 2015.