Implementation of Artificial Neural Network in Forecasting Sales Volume in Tokopedia Indonesia

Meiryani
Accounting Department, Faculty of Economics and Communication, Bina Nusantara University
Jakarta, Indonesia 11480

Dezie Leonarda Warganegara
Doctoral Program, BINUS Business School
Bina Nusantara University
Jakarta Indonesia

Abstract—Predicting sales is one way to get company profits. Tokopedia Indonesia is one of the marketplaces that is included in the type of e-commerce customer to customer (C2C). This research was conducted in order to help sellers in the Tokopedia Indonesia marketplace to predict the sales of their merchandise, so that sellers can prepare or stock items that are predicted to increase in sales by implementing Artificial Neural Networks. Artificial neural networks can help predict future sales values. The results of the analysis of this study indicate that the network model obtained reaches an accuracy rate of 95%.

Keywords—Forecasting; e-commerce; backpropagation; artificial neural network

I. INTRODUCTION

Sales data and information are very important to company to plan sales to be come, for example: customer data, number of vehicles, price cars, spare parts, types of vehicles and those that are not inferior its importance is deep government policy provide vehicle taxes as well as fuel subsidies vehicle [1,2]. The report of the research company said that in early 2017, it increased 51% from 2016 internet users which recorded around 132.7 million internet users [3]. From this data it is known that 24.74 million internet users have shopped online. It is recorded that from 2016 to 2017 internet users spent around IDR 74.6 trillion to shop on various e-commerce (Aditya, 2017). Rebecca (2016) there are six types of e-commerce with different characteristics, namely business to business (B2B), Business to customer (B2C), Customer to Customer (C2C), customer to business (C2B), Business to administration (B2A), and online to offline (OZO) [4]. The C2C e-commerce type model will be the author's focus in this study. Rebecca (2016) state that Tokopedia Indonesia is one of the marketplaces that is included in the C2C type of e-commerce that allows anyone and anywhere to be a seller or a buyer. Tokopedia Indonesia provides various tradable items such as electronic devices, baby equipment, men's and women's clothing, cellphone accessories, laptops, computers, and others. Based on www.alexa.com, it is known that Tokopedia is in the first place for this type of marketplace after the 8th Top Sites on the Alexa Rank [5].

In the Tokopedia marketplace, there is information on the number of items that have been sold, the number of people who have seen the item's page, how long it will take for the goods to be sent to the courier, the number of customers, the description of the goods, the number of people who favor the item and so on [6]. The prediction of the number of sales is an important factor that determines the smooth running business of a company. This prediction is very useful for determining how much goods to be ordered in the following month. Common problems faced by a company is how to predict or forecast sales of goods in the future based on previous sales data. The prediction is very influential to determine sales targets that must be achieved. Planning that effective both in the long term and in the short term depending on the forecast demand for products to be sold [7].

Forecasting techniques are widely used for the planning process and decision making, a prediction trying to predict what will happens and that will be needed. There are artificial neural networks forecasting technique that is often used namely Backpropagation. This technique usually used in multilayer networks with the aim of minimizing error in the output generated by network [8]. Not all items that are seen or favored by many people will experience high demand, but sometimes high enthusiasts are not only seen or favored by many people so that there is an out of stock of goods. To solve this problem, information is needed to those who sell goods and services on Tokopedia by making an analysis that can provide information related to the sales volume for a product or service being promoted. So that with this information, parties who sell goods and services on Tokopedia can provide stock that matches the predicted interest of an item being promoted. One method of overcoming this problem is by designing an artificial neural network architecture or commonly referred to as an artificial neural network. One type of algorithm for this artificial neural network is backpropagation [9]. Backpropagation is an artificial neural network model that is often used and in great demand as a multi-year learning algorithm related to identification, prediction, pattern recognition and so on. The backpropagation algorithm is a type of supervised learning algorithm where the output of the network is compared with the target output so that an error is obtained. Then the error will be propagated back to modify or improve the weight of a network in order to minimize errors. Based on this background, the authors are interested in applying the artificial neural network method with the backpropagation algorithm to predict sales volume on Tokopedia [10].

II. THEORETICAL FRAMEWORK

The research entitled analysis of the backpropagation method and radial basis function to predict rainfall with
artificial neural networks was carried out by Vincent Rinda Resi (2014). This study discusses the very high rainfall prediction model. The prediction model will be used for various things, one of which is flood prevention. The results obtained from the two methods found that the backpropagation method was able to provide better accuracy, namely 99% than the radial basis function method [11, 12, 13] in their research discusses the prediction of members' interest in a cooperative product. Today's cooperatives, especially the PTPN VII Musi Landas Cooperative, are very much needed by the community because they play an important role in their daily life. The constraints faced by the cooperative are in determining the products that are of interest to its members. If the cooperative can predict this, it will reduce losses and will increase sales which will have an impact on cooperative income. So this research applies a data mining which can predict the interest of members in a product. The data mining technique applied is classification using the decision tree method with the C4.5 and DTREG algorithms. Based on this research, several conclusions were obtained, one of which was to produce information about product categories that were of interest to the members of the PTPN VII Musi Landas Cooperative.

Further research related to artificial neural networks with backpropagation algorithms is a study conducted by [14, 15] discuss the backpropagation application discussed to predict the movement pattern of earthquake points in Indonesia from January 2015 to April 2015. The data used is daily data from the Meteorology, Climatology and Geophysics Agency. The conclusion is that the network with momentum and adaptive learning of 0.9813 shows pretty good results with an MSE value of 0.047735 on the 104th iteration/epoch with a maximum of epoch = 10000, learning rate = 0.3, and mc = 0.8. The results of the mapping of the predicted position of the earthquake point are at latitude 0.0405 LU, longitude 124.4015 LS and magnitude 3.26 SR which are in one zone with earthquakes that occur on the same day and date, namely latitude 0.84, LU longitude 126.28 latitude and magnitude 4.8. [16, 17] Discusses how to predict the stock prices of Bank Central Asia, Gudang Garam and Indofood. In this study, it was concluded that the artificial neural network parameters obtained MSE with the smallest value obtained with window size 10, hidden neurons 10, and 10000 iterations for all test cases data PT Bank Central Asia Tbk, PT Gudang Garam Tbk and PT Indofood Sukses Makmur Tbk. The results of each test produce MSE 0.002708159 for BBCA data test cases, 0.00107418 for GGRM data test cases and 0.002440852 for INDOFOOD data test cases [22]. Further research is related to the comparison of methods conducted by [18, 19] discuss the performance of the three methods in predicting the price of gold because gold is an item that can be used for investment, so an understanding of the shifting of gold prices is needed in order to be able to get a profit. Of the three methods, backpropagation is the best algorithm for predicting gold prices with an accuracy of 95% [20].

III. RESEARCH METHODOLOGY

The target population used in this study is all data on the Tokopedia Indonesia website. Then the researchers took randomly from one of the categories of goods in Tokopedia Indonesia, namely the computer accessories category. From the data obtained, three types of computer accessories categories will be analyzed by researchers, namely mouse, speaker & sound, and bag & case. The data is obtained by opening the goods web page one by one, then by means of scraping, the researcher selects the features that are considered to affect the variable (Y), namely the number of items that have been sold. The selected features will be used as the supporting variables (X) needed, namely the type of item, the price of the item, the number of people who saw the item, the time of delivery of the goods to the courier, customers from the goods shop, the number of people who have favored the rating. This scraped data is an item that has been advertised on Tokopedia until February 11, 2021. The data analysis method used in this study is an Artificial Neural Network with a backpropagation algorithm to predict buyer interest patterns on the Tokopedia Indonesia website, especially in the computer accessories category. Software used by researchers to analyze the data using Microsoft Excel and RStudio.

IV. RESULTS AND DISCUSSION

Preprocessing data is the first step in an analysis to check and correct when there are missing values before starting the learning process [21]. When there is information that is not available on one or more object variables or certain cases, data correction will be carried out. The researcher examined the missing data, can be seen in Table I, as follow:

In Table I, it can be seen that all variables do not have missing data, which means that the next steps can be taken, namely data transformation and data sharing.

Data sharing is intended to divide data into two parts, namely training data and testing data which have their respective functions. The training data is used to train the learning algorithm during the training process. This data distribution is not divided equally, but the percentage for training data is greater than the test data. The following Table II is related to the percentage of data sharing used.

| Variable | Valid | Missing | Percentage Valid |
|----------|-------|---------|------------------|
| Sold     | 281   | 0       | 100%             |
| Type     | 281   | 0       | 100%             |
| Price    | 281   | 0       | 100%             |
| Seen     | 281   | 0       | 100%             |
| Time Send| 281   | 0       | 100%             |
| Customer | 281   | 0       | 100%             |
| Favorite | 281   | 0       | 100%             |

| TABLE II. DATA PARTITION |
|--------------------------|
| Percentage  | Total  |
| Training data | 80% | 224  |
| Test data      | 20%  | 57   |
| Total           | 100% | 281  |

Source: processed data
Table II above can be seen that the percentage of training data is greater than the test data because the learning algorithm while carrying out the training process works optimally.

The determination of input and output patterns is based on the formulation of this research problem. So that there are seven variables as input that are considered influencing the target (output). Determining Network Architecture and Parameters Artificial Neural Network has an architecture consisting of the number of layers and the number of neurons in each layer, as for the case of backpropagation using multi layers consisting of input, hidden, and output. The number of hidden layers 1 alone is sufficient to produce output that matches the target [23,24]. So that the network architecture designed for this research is 3 layers (input, hidden, and output) with 7 neurons for the input layer, 3 neurons for the hidden layer, and 1 neuron for the output layer [29].

Initialization of weights and bias is given before carrying out the training process of an existing network system in an artificial neural network. This initial initialization weight is given to each neuron that is interconnected. This weight factor defines the relationship between neurons with one another, where the greater the weight value of a relationship between neurons, the more important the relationship between the two neurons. Initialization of initial weights and bias is done randomly. Table III show the initial weights and bias of the input layer against the hidden layer and Table IV show initial weights and bias on the hidden layer to the output layer [25,26].

Training with the backpropagation algorithm is an algorithm with a supervised learning process. After obtaining the initial initialization weight and bias, a training process will be carried out on a network that has been designed for architecture and the parameters that have been determined using the training data that has been determined; the percentage is 80% of the total data. There are three phases of the training process for the backpropagation algorithm, namely feedforward, backpropagation, and weight modification. Feed forward (feedforward) at this stage, an error will be searched for the output in the forward direction (forward). Each input unit Xi (i: 1,..., n) will receive the input signal xi then pass it to the hidden unit. Then all the weighted input signals will be calculated including the bias in each hidden unit Zj (h: 1,..., p). The following is the input signal to the weighted hidden layer, including its bias.

After the hidden layer receives a weighted input signal (Table V) including the bias, the output signal will be calculated in the hidden layer from the input signal using the activation function. The following is the output signal on the Hidden Layer.

Then the output signal in the hidden layer in Table VI will play a role as an input signal in the output layer. The input signal will be forwarded to the output layer with the weights and bias hidden in the hidden layer against the output layer for each output unit.

After the output layer receives the input signal (Table VII) from the hidden layer, the input signal will be activated using the activation function. Table VIII shows output signals at the output layer.

Once activated in the output layer, the output will be distributed to all units in the output layer.

| Variable | 1 | 2 | 3 |
|---|---|---|---|
| Bias | 0 | -0.53544204 | 1.256897569 | -0.135689723 |
| X1 | 1 | 0.054219104 | 0.772365469 | 1.1365698754 |
| X2 | 2 | 0.617251692 | 0.678952432 | 2.5326546321 |
| X3 | 3 | 0.145236548 | 0.865321469 | 0.6987563212 |
| X4 | 4 | -1.534256987 | 0.598756975 | 0.4659879464 |
| X5 | 5 | 0.568795643 | 0.123646897 | |
| X6 | 6 | 1.285695467 | 0.256458976 | -1.856987965 |
| X7 | 7 | 0.789653217 | 0.256987795 | 2.3698575426 |

Source: processed data

| W_{ij} | W_{jk} |
|---|---|
| 1 | |
| 0 | 1.3564795645 |
| 1 | 1.8563256987 |
| 2 | 0.2356489976 |
| 3 | 1.1698756975 |

Source: processed data
TABLE VIII. OUTPUT SIGNALS AT THE OUTPUT LAYER

| Y | Yk             |
|---|----------------|
| 1 | 0.92361236     |

Source: processed data

To calculate the error between the input target and the output generated by the network with the result of the error factor (ε) of -0.022507221. The error factor will be used to correct the weight (Wjk) and bias (W0k) in the lower layer (hidden layer) with the learning rate (α). The results of the weight improvements in the hidden layer to the output layer are as follows. Table IX show correction of weights and bias in the hidden layer of the output layer.

Each hidden unit Zj accepts input delta weights and bias from the layer above it (output layer). The weight and bias delta input will be used to find the error factor in each hidden unit. Table X is the result of calculating the error factor in each hidden unit.

Input from the error factor in the hidden unit such as Table XI will be activated using the activation function with the following results.

After obtaining the error factor that has been activated with the activation function, the error factor will be used to correct or correct the weights and biases in the lower layer, namely the input layer to the hidden layer with calculations as in equation 16. Table XII shows correction of weights and bias in the input layer against the hidden layer, as follows:

This training process will continue as long as the conditions are not met. Training will stop when an optimal error has been obtained, so that the final weight and bias for each layer is obtained. The final weights and biases with one step obtained from the artificial neural network with the backpropagation algorithm that has been designed using training data [27]. Table XIII show weights and final bias on the input layer against the hidden layer, as follows:

The artificial neural network predicts sales volume in the Tokopedia Indonesia marketplace with training data first to find out the accuracy of the network before testing the network. The prediction results using an artificial neural network model obtained using training data.
The results of the training data prediction using the network model obtained as shown in Table XIV above in predicting sales volume in the Tokopedia Indonesia marketplace with a span of time in months. Obtained is an accuracy rate of 95.75%. Testing of artificial neural networks with the backpropagation algorithm obtained after calculating the weight until it reaches optimal then testing the network obtained will be carried out. The network will be applied to the test data to determine the network's performance in predicting sales volume on the Tokopedia Indonesia's marketplace. The percentage of test data was 20% of the total data used, namely as many as 57 test data. The following are some of the results of the prediction of sales volume on the Tokopedia Indonesia marketplace using the network obtained [28].

| No | Data To | Actual | Predictions | Time Range (Month) |
|----|---------|--------|-------------|--------------------|
| 1  | 1       | 77     | 71          | 29                 |
| 2  | 6       | 101    | 56          | 24                 |
| 3  | 75      | 79     | 78          | 14                 |
| 4  | 76      | 56     | 61          | 11                 |
| 5  | 159     | 68     | 79          | 16                 |
| 46 | 175     | 69     | 62          | 29                 |
| 47 | 229     | 62     | 63          | 31                 |
| 48 | 269     | 49     | 63          | 22                 |

The prediction results from network testing using this test data show the network performance obtained. From the results of the prediction of sales volume in the Tokopedia marketplace with a span of time in months using test data, an accuracy rate of 95.75% is obtained. The network model obtained with the optimal level of accuracy is found when testing the network, so it can be said that the resulting network performance is very good. Regarding the sales volume prediction made, the seller at Tokopedia can provide stock according to the predictions obtained with the time span listed in Table XV, so that the seller can minimize the occurrence of losses and will increase their income [18, 29].

V. CONCLUSION

Based on phenomenon, research question, result and discussion the conclusion as follows:

1) The results of this study indicate that Backpropagation has a good level of accuracy in predicting sales.

2) The Artificial Neural Network method has an adaptive nature, namely the network tries to achieve data stability to achieve the expected output value.

3) The resulting artificial neural network architecture design results consist of three layers which include six neurons in the input layer, three neurons in the hidden layer, and one neuron in the output layer. The parameters used to form the network model include the learning rate with a value of 0.02 and the activation function used is binary sigmoid (logistic).

4) The resulting level of accuracy when testing the obtained network reaches an accuracy rate of 95.75%.

5) The sales forecasting process is to enter the estimated future sales data, to be processed using the backpropagation neural network to produce the desired data.

ACKNOWLEDGMENT

This work was supported by BINUS University, Jakarta, Indonesia.

REFERENCES

[1] Azhar Susanto & Meiryani. 2019. Antecedents of environmental management accounting and environmental performance : Evidence from Indonesian small and medium enterprises. International Journal of Energy Economics and Policy. 9(6), pp. 401-407.

[2] Bahadir, E. 2016. Prediction of Prospective Mathematics Teachers' Academic Success in Entering Graduate Education by Using Backpropagation Neural Network. Journal of Education and Training Studies. Vol. 4 (5): 113-122.

[3] Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. 1992. Forecasting the Behavior of Multivariate Time Series Using Neural Networks. Neural Networks. Vol. 5: 961-970.

[4] Fausett, L. 1994. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. New Jersey: Prentice-Hall

[5] Fahruroji, A. 2014. Getting to know the E-commerce Business Model. https://afahrurroji.net/mengenal-model-bisnis-e-commerce/. Retrieved February 20, 2018.

[6] Fajri, N. 2011. Temperature Prediction Using Algorithms-Algoritma in Artificial Neural Networks. Thesis. Bandung Institute of Technology: Bandung.

[7] Fausett, L. 1994. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. New Jersey: Prentice Hall. Herdianto. 2013. Prediction of Induction Motor Damage Using Backpropagation Neural Network Method. Thesis. University of North Sumatra: Medan.

[8] Hidayatullah, A. I. 2017. Backpropagation Algorithm For Aircraft Delay Prediction Due to Weather. Thesis. Faculty of Mathematics and Natural Sciences, Indonesian Islamic University: Yogyakarta.

[9] Hardianto, H. N. I., Suyanto, & Purnama, B. 2011. Analysis and Implementation of Differential Evolution and Recurrent Neural Network for Time Series Data Prediction Case Study of Gold Selling Rate. Thesis. Telkom University.

[10] Hikmah, A. 2017. Time Series Forecasting Using Autoregressive (AR), Radial Basis Function Artificial Neural Network (RBF), and AR-RBF Hybrid in Indonesian Inflation. Unnes Journal of Mathematics. Vol. 7 (2): 1—14.

[11] Ita Qorry Aina. 2018. Implementation of Artificial Neural Network with Backpropagation Algorithm to Predict Sales Volume at Bukalapak. Faculty of Mathematics and Natural Sciences, Islamic University of Indonesia. Yogyakarta.

[12] Kusumadewi, S. and Kiki. 2010. Analysis of Artificial Neural Networks with Backpropagation Method to Detect Psychological Disorders. FTI, Islamic University of Indonesia. Kusumadewi, S. 2003. Artificial Intelligence Techniques and Applications. Yogyakarta: Graha Science.

[13] McLeod, P. 2008. Management Information Systems. Jakarta: Salemba.

[14] Meiryani and Azhar Susanto. 2018. The Influence of Information Technology on The Quality of Accounting Information System. ACM International Conference Proceeding Series. Pp.109-115

[15] Pratama, A.P. 2017. Development of Internet Users in Indonesia in 2016, the World's Largest. https://id.technasia.com/pertumbuh-
[16] Puspitaningrum, D. 2006. Introduction to Artificial Neural Networks. Yogyakarta: Andi Offset. 48 Ramadha, W. I. 2016. Stock Price Prediction Using Resilient Backpropagation Neural Network. Thesis. Not published. Faculty of Mathematics and Natural Sciences, Gadjah Mada University: Yogyakarta.

[17] Rebecca. 2016. Types of E-commerce & Examples. https://www.progresstech.co.id/blog/jenis-e-commerce/. Retrieved February 20, 2018. Resi, V.R. 2014. Comparative Analysis of Backpropagation Method and Radial Basis Function to Predict Rainfall with Artificial Neural Networks. Journal of the Faculty of Computer, Dian Nuswantoro University.

[18] Ryanda, et al. 2015. Designing a Mobile Application “Kiosku.Com” with Web Scraping on the Olx.Co.Id, Berniaga.Com, and Bukalapak.Com Websites based on Android. e-Proceeding of Engineering: Vol. 2, No.2.

[19] Ryandi, et al. 2014. Application of Data Mining to Predict Members’ Interest in Ptpn Vii Musilandas Cooperative Products. Journal of Informatics Engineering Students.

[20] Saputra, B. Y. 2015. Implementation of Artificial Neural Networks with Backpropagation Algorithm to Predict Earthquake Point Movement Patterns in Indonesia. Thesis. Faculty of Mathematics and Natural Sciences, Indonesian Islamic University: Yogyakarta.

[21] Salman, A. G., & Prasetio, Y. L. 2010. Implementation of Recurrent Neural Networks Using the Gradient Adaptive Learning Rate Method for Rainfall Estimation Based on ENSO Variables. ComTech Journal. Vol. 1 (2): 418-429.

[22] Semen Indonesia. 2017. The Demand for Cement Will Continue to Increase. http://www.semenindonesia.com/perminta-semen-bakal-terus-naik/.

[23] Siang, J. J. 2004. Neural Networks and Its Programming Using Matlab. Yogyakarta: Publisher Andi.

[24] Suhada, B. 2009. Forecasting National Sugar Production Using Artificial Neural Network Approach. Derivatives Journal. Vol. 3 (1): 50-63.

[25] Susanti, L. A. D., Arna, F., & Sethiawardana. 2013. Forecasting Stock Prices Using Recurrent Neural Network with Backpropagation Through Time (BPTT) Algorithm. Final Project Papers. Surabaya: Sepalamber Institute of Technology.

[26] Sutono, S. B. 2008. Causal Forecasting Analysis Based on Principal Component Analysis of Artificial Neural Networks for Industrial Engineering Applications. Thesis. Pelalawan College of Technology: Riau. Turland, M. 2010. Php. | architect's Guide to Web scraping with PHP. Introduction-Web Scraping Defined, str, 2. Vermaat, S. C. 2007. Discovering Computers: Exploring Fundamental Computer Dubia. Edition 3. Jakarta: Salemba Infotek.

[27] William and Sawyer. 2007. Using Information Technology. Yogyakarta: Andi. Wong, J. 2010. Internet Marketing for Beginners. Jakarta: PT Elex Media Komputindo. 49 Zaira, Z. 2011. Implementation of Web Extraction for Hadiths Translated in Indonesian. Thesis. University of Indonesia: Depok.

[28] Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. 2013. Comparison of the ARMA, ARIMA, and the Autoregressive Artificial Neural Network Models in Forecasting the Monthly Inflow of Dez Dam Reservoir. Journal of Hydrology (online). Vol. 476: 433—441.

[29] Winwin Yadiati & Meiryani. 2019. The role of information technology in E-Commerce. International Journal of Scientific and Technology Research 8(1), pp. 148-150.