

Application of Artificial Neural Network for Forecasting Demand Bottled Drinking Water by Using Back propagation Algorithm

Prima Denny Sentia^{1*}, Andriansyah¹, Ilyas Ishak¹, Albizzia Haura¹

¹Industrial Engineering, Universitas Syiah Kuala, Jl. Tgk Abdur Rauf No.7, Banda Aceh 23111, Indonesia

*Corresponding author. Email: primadennysentia@unsyiah.ac.id

ABSTRACT

The erratic demand for bottled drinking water (BDW) products caused the sales target not to be achieved for several periods. One of the efforts that can be made by the management so that the amount of production is correct is by forecasting demand. This study aims to determine the best forecasting model using the Artificial Neural Network (ANN) method with the Backpropagation algorithm, supervised learning. The activation function used is the binary sigmoid function (logsig). Based on the result, the best architectural model is found in neurons 3-4-1 with an MSE value of 0.0002 and a MAPE value of 2.346%.

Keywords: Artificial Neural Network, BDW, Back propagation Algorithm, Supervised Learning

1. INTRODUCTION

The increasingly crucial industrial competition encourages every company to have a competitive advantage in terms of quality, price and flexibility. This is necessary for the company to increase profitability, pursue growth, and be able to survive in business competition. This goal can be achieved through effective decision making so that it can fulfil the company's strategic functions in the future [1], [2]. One of the efforts to maintain the competitiveness of companies in global competition and maximize profits is by implementing good supply chain management so that customer satisfaction can be achieved optimally [3].

Supply chains are often associated with a lot of uncertainty and complexity in the modern world. This uncertainty arises due to various factors such as supply chain networks, partner company activities, customer behaviour, competitor behaviour, emerging technologies, and new product development, all of which contribute to an unstable supply chain [4]. An unstable market will create negative consequences in the form of erratic demand. However, this can be avoided by accurately predicting demand so that market needs can be optimally fulfilled.

Demand forecasting is the basis for many managerial decisions in the supply chain such as demand planning, order fulfilment, production planning, and inventory control [5]. An efficient and accurate demand forecasting system can play a major role in minimizing inventory costs [6]. Companies that are unable to meet market needs and consumer demand will result in reduced opportunities to gain company profits. On the other hand, excessive production can result in companies experiencing losses because they have to pay for inventory. Therefore, a forecasting method with a high degree of accuracy and minimum error is needed.

As for the selection of the right forecasting method, several considerations are needed, such as the suitability of the method with the real conditions to be modeled and being able to accommodate the factors that influence it [7]. The forecasting method that is generally used to estimate demand is the time series method. This method aims to predict the behaviour of complex systems simply by looking at past patterns based on the same phenomenon [8]. In the time series method, there are several ways that are often used, one of which is the traditional method based on the concept of statistical probability. However, these methods often give unsatisfactory results because the time series method only process past data (historical data) without considering external factors that influence

consumer opinion. Therefore, the traditional forecasting approach cannot provide a good demand estimate [6].

One of the more up-to-date time series forecasting methods that can effectively synergize the factors that affect output is an Artificial Neural Network (ANN). ANN is part of an artificial intelligence system that is used to process information designed to mimic how the human brain works in solving problems by carrying out the learning process through changes in the weight of its synapses. ANN can be used to solve problems related to optimization, forecasting, modeling and simulation, globalization and decision making. ANN can implicitly detect complex nonlinear relationships between dependent and independent variables and is suitable for complex information processing [9]–[11].

One of the ANN training algorithms that is widely used in the field of pattern recognition is backpropagation that often used in solving complex problems related to identification, prediction, and pattern recognition. Backpropagation is classified as a supervised learning or training algorithm and uses error correction learning rules. This algorithm has advantages in the aspects of system learning (adaptive) and has a small error risk (fault tolerance) for problem solving [12]–[14].

Several studies using the ANN method with the Backpropagation algorithm have been carried out for forecasting. Among them is research on forecasting the amount of production demand using time series data [13], [15]–[17]. This method is also applied to sales forecasting with time series data [18][19]. Based on some of these studies, it can be stated that ANN with the Backpropagation algorithm has been proven to be used to solve problems in forecasting. Forecasting using the ANN method shows more optimal results than forecasting using traditional methods. This is influenced by different input and data processing processes. The network architecture model in ANN makes it possible to model optimal forecasting so that it can approach the actual conditions, while the input to forecasting with the traditional method only uses numerical series [19].

One of the industries that is growing rapidly today is the Bottled Drinking Water industry (BDW). Competition in the BDW industry is getting tighter and shows a bright prospect going forward. Apart from the fact that water is the primary need of every individual, population growth and changes in community habits have made the BDW industrial market continue to increase [20]. As a result, competition between the BDW industries in competing for the market can be a threat to all entrepreneurs engaged in BDW [21]. Therefore, this study aims to obtain a forecasting model that is used to predict demand for BDW using the ANN method with the Backpropagation algorithm. The model obtained is a simulation model

that will be applied to the software to determine the amount of product demand in the next period.

2. METHODS

This research is a case study on one of the BDW industries in Aceh for the box of 600 ml bottles (each box contains 24 bottles). The purpose of using the ANN method is to find historical data patterns and extrapolate these patterns to the future by processing the factors that affect output. The existence of a network architecture model in ANN makes it possible to create an optimal forecasting model so that the forecasting results can approach the actual value [19]. ANN inputs in this study are sales data, production capacity, production costs, and product sales targets.

The algorithm used in the ANN method is backpropagation because the algorithm is able to train the network's ability to recognize patterns and respond correctly to input patterns that are similar (but not the same) to the patterns used during training. With supervised learning, where the output value is known, the weight can be adjusted repeatedly to obtain a small error value between the predicted results and the expected target.

Data that serves as input in this study is divided into two parts, namely training data and testing data. The variables used in the ANN input are sales data (X1), production capacity data (X2), and production cost data (X3), while the output is sales target data (Y). Historical data used for short-term forecasting will produce less uncertainty (more accurately) than long-term forecasting because in short-term forecasting conditions that affect demand tend to remain or change slowly. Therefore, the training data set in this study used 24 data and 12 test data. Yao and Tan [22] state that until now, there has been no standard rule in determining the composition of the measurements used for both training and testing data. This data sharing is intended so that the network obtains sufficient training data and testing data can test the achievement of training results.

At the data transformation stage, there are two main activities, namely data normalization and network architecture design. In the normalization stage, the sales data is first transformed into data with a range [0-1]. Normalization aims so that the network does not experience failure when conducting training and testing. After that, the network architecture design stage and parameter determination are continued. The number of hidden layers and parameters is determined by trial and error. Data normalization is carried out based on the equation 1 [23].

$$x' = \frac{0.8(x-a)}{b-a} + 0.1 \quad (1)$$

where:

- x' = Normalized value
- a = Minimum range value
- b = Maximum range value
- x = Value to be normalized

Data training is conducted using Mean Square Error (MSE) to compare actual data with ANN results. If the ANN results are not much different from the actual data, this artificial neural network design can be used as a model for the data testing phase. The best architectural design is obtained if it produces small errors. As for this training, the trainrp network training function is used. The use of trainrp because the backpropagation network uses the binary sigmoid activation function.

At the data testing stage, the best training results will be tested using the prepared test data. The training and testing process largely determines the validation process. Data testing is performed using Mean Absolute Percentage Error (MAPE) to measure network validation on the entire data. Furthermore, data denormalization is carried out against the simulation results that have been obtained to obtain the original value of the forecasting results. Denormalization of data using equation 2 below.

$$x = \frac{(x' - 0.1)(b - a)}{0.8} + a \tag{2}$$

where:

- x = Denormalization of forecasting data's result
- a = Minimum range value of all forecast results
- b = Maximum range value of all forecast results
- "x" ^"" = Forecasting data that will be denormalized

3. RESULTS AND DISCUSSION

3.1. Data normalization

Normalization can solve data problems at different scales. The activation function used in this study is binary sigmoid which has an output range of [0-1]. This is based on the sigmoid function which is an asymptotic function with a value that never reaches 0 or 1. Table 1 shows the results of data normalization. Based on the results of calculations in Table 1, it can be seen that any data that has been normalized is in the range [0-1] so that the data can be used for the next data processing.

The determination of the parameters and the design of the network architecture are carried out several times of trial and error in determining the number of neurons and giving the parameter values. There is no provision for definite parameters in order to obtain the minimum error [23]. Therefore, it is necessary to carry out repeated

Table 1. The results of data normalization.

No.	Data Input			Data output	No.	Data Input			Data output
	X1	X2	X3	Y		X1	X2	X3	Y
1	0.108	0.107	0.106	0.106	19	0.794	0.802	0.806	0.803
2	0.103	0.104	0.102	0.103	20	0.828	0.837	0.838	0.523
3	0.100	0.100	0.100	0.100	21	0.798	0.806	0.804	0.505
4	0.315	0.315	0.315	0.314	22	0.691	0.695	0.694	0.535
5	0.301	0.301	0.300	0.300	23	0.677	0.685	0.686	0.457
6	0.433	0.434	0.440	0.435	24	0.779	0.783	0.783	0.505
7	0.545	0.549	0.549	0.548	25	0.102	0.101	0.100	0.102
8	0.524	0.525	0.521	0.526	26	0.100	0.100	0.104	0.100
9	0.505	0.506	0.503	0.506	27	0.244	0.245	0.241	0.244
10	0.534	0.537	0.539	0.556	28	0.346	0.346	0.338	0.346
11	0.453	0.457	0.458	0.457	29	0.177	0.174	0.172	0.174
12	0.504	0.506	0.506	0.505	30	0.128	0.126	0.126	0.125
13	0.730	0.734	0.737	0.738	31	0.331	0.331	0.322	0.331
14	0.591	0.595	0.594	0.604	32	0.519	0.519	0.504	0.520
15	0.897	0.900	0.900	0.900	33	0.900	0.900	0.900	0.900
16	0.812	0.817	0.814	0.818	34	0.431	0.432	0.420	0.162
17	0.683	0.686	0.684	0.690	35	0.319	0.319	0.310	0.141
18	0.662	0.668	0.668	0.668	36	0.193	0.192	0.188	0.117

Table 2. The determination of the ANN product architecture

No.	Neuron	Parameter					
		lr	mc	Epoch	Goal	Function of Activation	Error
1	3-3-1	0.1	0.2	5000	0.001	Logsig	0.008
2	3-4-1	0.1	0.9	5000	0.001	Logsig	0.006
3	3-5-1	0.2	0.1	5000	0.001	Logsig	0.007
4	3-6-1	0.1	0.9	5000	0.001	Logsig	0.007
5	3-7-1	0.2	0.1	5000	0.001	Logsig	0.008

experiments to obtain optimal results. Table 2 shows the determination of the ANN product architecture. Based on Table 2, it can be seen that the architectural model with the parameter value that produces the smallest error is the 3-4-1 network architecture with a learning rate (lr) of 0.1, constant momentum (mc) of 0.9, epoch of 5000. and uses the logsig activation function produces an error of 0.006.

In this study, the set lr value is 0.1 because if the lr used is too large, the network learning process will be less stable and reach a local minimum point. Meanwhile, the small lr is used to avoid too large a response to the error of a learning process pattern. Fig.1 shows the network architecture that is formed.

One of the efforts to get the smallest error is done by coordinating lr and mc well. Mc will cause the network to adjust for a greater weight as long as the correction is in the same direction as the existing pattern. After conducting several experiments, the selected mc is obtained at 0.9 because it obtained the smallest MSE value. The number of iterations set is 5000 and the desired goal is 0.001.

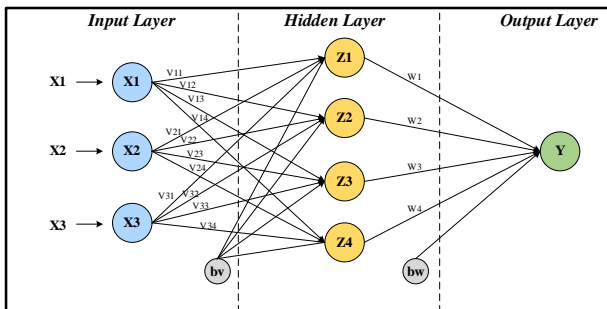


Figure 1. The network architecture.

3.2. Data training

In the process of designing the best network architecture, the data will be separated into two parts, namely training and testing data. The data sharing is intended so that the network gets sufficient training data

and testing data can test the achievement of training carried out based on the MSE value of training data and MAPE test data [15].

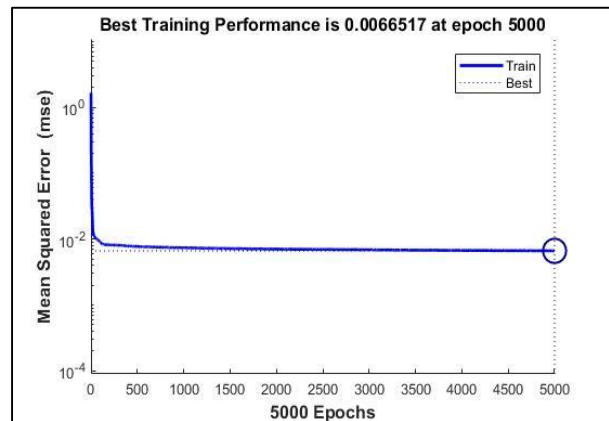


Figure 2. ANN performance.

The process of determining the weight changes and bias in this training uses the trainrp training function, while the error calculation method uses the MSE function. The trainrp function is used in this study because the back-propagation network uses the binary sigmoid activation function. Figure 2 shows a graph of the best training performance, namely 3-4-1 network architecture with a learning rate (lr) of 0.1 and constant momentum (mc) of 0.9.

Based on Fig. 2, the training results converged through 5,000 iterations and took 8 seconds. The MSE value obtained is 0.0066517 which is the smallest goal value on the network. At the start of the training, there was an indication of a decline that was too fast. This is because the training function used is trainrp. After that it can be seen that the line decline gradually.

Furthermore, Fig. 3 shows a graph of the regression results of product data training, where the regression value is obtained of 0.9206. This means that the actual variables with ANN in training have a good correlation

even though there are some data that have not spread following the expected fit or target line. The graph also shows data verification and it appears that there are some differences between the target data and the ANN result data, but they are not too significant.

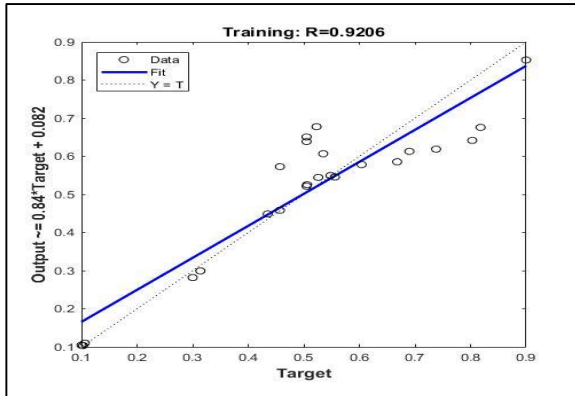


Figure 3. The regression results of product data training.

The correlation measure of 0.9206 proves that there is a high degree of association because the closer to the value 1, the stronger the relationship between variables. The high degree of association between the related variables includes sales volume, production capacity, and production costs on demand for bottled drinking water products. Production demand is closely related to sales results, capacity and production costs because if demand increases, sales results, capacity and production costs will also increase.

3.3. Data testing

The activation function used in data testing is the binary sigmoid function (logsig) and the identity function (purelin). Figure 4 presents a graph of the comparison of ANN and target output data.

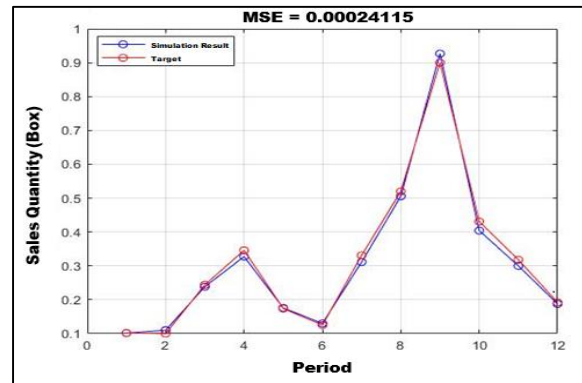


Figure 4. The comparison of ANN and target output data.

Based on Figure 4, it can be seen that overall there are no significant difference between the ANN results and the target data. This indicates that the overall ANN result model is similar to the target data. The MSE value obtained is 0.00024115 indicating the smallest and most optimal goal value on the network. Thus, it can be said that the testing process of the data obtains optimal forecasting results.

The test results obtained will be validated against the actual data using MAPE. Validation is carried out to see the performance of the ANN network in recognizing actual data patterns. The following is the result of the MAPE calculation to compare the actual data with the

Table 3. MAPE calculations

No.	Actual (a)	ANN (b)	a-b	ABS((a-b)/a)	((a-b)/a)*100
1	4,038	4,074	-36	0.009	0.902
2	4,025	4,091	-66	0.017	1.650
3	5,231	4,993	237	0.045	4.545
4	6,080	5,916	164	0.027	2.702
5	4,666	4,549	117	0.025	2.509
6	4,262	4,239	22	0.005	0.534
7	5,954	5,752	201	0.034	3.385
8	7,524	7,693	-168	0.022	2.244
9	10,712	10,527	184	0.017	1.724
10	6,795	6,768	27	0.004	0.400
11	5,857	5,633	223	0.038	3.819
12	4,802	4,622	179	0.037	3.744
MAPE					2.346

ANN results. Table 3 provides the information about MAPE calculations.

Based on Table 3, it can be seen that the overall MAPE value obtained is 2.346%. So the level of accuracy of the forecasting results is very good because it is far below 10%. Thus, it can be stated that there is no significant data deviation between the actual data and the ANN forecasting results.

3.4. ANN forecasting

After obtaining the best architectural model from the training and testing process, the model can be used to forecast the number of sales for the next 12 periods. The ANN forecasting results are normalized to find out the original value. Table 4 illustrates the forecasting of the obtained products.

Based on Table 4, it can be seen the results of forecasting demand for 600 ml bottles where the highest demand was in period 9 with a total of 10,527 boxes, while the lowest demand occurred in period 1 with a total of 4,074 boxes.

Table 4. The result of ANN forecasting.

Forecasting period	ANN Result	Forecasting Result (box)
1	0.106	4,074
2	0.108	4,091
3	0.216	4,993
4	0.326	5,916
5	0.163	4,549
6	0.126	4,239
7	0.307	5,752
8	0.539	7,693
9	0.878	10,527
10	0.428	6,768
11	0.292	5,633
12	0.171	4,622

4. CONCLUSION

Based on the results and discussions from BWD demand forecasting model using the Artificial Neural Network (ANN) method with the Backpropagation algorithm, it is found that the 3-4-1 architectural model has a minor error value. Forecasting results for the successive 12 periods produce MSE value of 0.00024115 and MAPE value of 2.346%. Based on the forecasting results, the highest demand is in period 9 with 10,527 boxes, while the lowest demand is in period 1 with 4,074 boxes.

REFERENCES

- [1] L. Saptaria, “Peramalan Permintaan Produk Cincou Hitam dalam Memaksimalkan SCM (Supply Chain Management),” *J. Manaj. dan Kewirausahaan*, vol. 3, no. 1, pp. 57–66, 2016.
- [2] P. D. Sentia, D. Asmadi, and D. Ramadhan, “Frontier in Industrial Engineering,” in *Seminar Nasional Teknik Industri Universitas Gadjah Mada*, 2016, p. TP-88.
- [3] M. Ghatebi, E. Ramezani, and M. A. E. Shiraz, “Impact of Supply Chain Management Strategies on Competitive Advantage in Manufacturing Companies of Khuzestan Provinc,” *Singaporean J. Bus. Econ. Manag. Stud.*, vol. 2, no. 11, pp. 1–6, 2014.
- [4] M. Christopher and M. Holweg, “Supply Chain 2.0’: Managing supply chains in the era of turbulence,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 41, no. 1, pp. 63–82, 2011.
- [5] N. Liu, S. Ren, T. M. Choi, C. L. Hui, and S. F. Ng, “Sales forecasting for fashion retailing service industry: A review,” *Math. Probl. Eng.*, vol. 2013, 2013.
- [6] M. A. A. Hasin, S. Ghosh, and M. A. Shareef, “An ANN Approach to Demand Forecasting in Retail Trade in Bangladesh,” *Int. J. Trade, Econ. Financ.*, vol. 2, no. 2, pp. 154–160, 2011.
- [7] N. Erni, M. S. Maarif, N. S. Indrasti, M. Machfud, and S. Honggokusumo, “Model Prakiraan Harga dan Permintaan pada Rantai Pasok Karet Spesifikasi Teknis Menggunakan Jaringan Syaraf Tiruan,” *J. Al-AZHAR Indones. SERI SAINS DAN Teknol.*, vol. 1, no. 3, p. 116, 2012.
- [8] A. Kochak and S. Suman, “Demand Forecasting Using Neural,” *Int. J. Mech. Eng. Robot. Res.*, vol. 4, no. 1, pp. 96–104, 2015.
- [9] M. Anike, Suyoto, and Ernawati, “Pengembangan sistem jaringan syaraf tiruan dalam memprediksi jumlah dokter keluarga menggunakan,” in *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 2012, pp. 209–216.
- [10] H. Omar, V. H. Hoang, and D. R. Liu, “A Hybrid Neural Network Model for Sales Forecasting Based on ARIMA and Search Popularity of Article Titles,” *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–9, 2016.
- [11] S. Bhadouria and A. Jayant, “Development of ANN Models for Demand Forecasting,” *Am. J. Eng. Res. (AJER)*, vol. 6, no. 12, pp. 142–147, 2017.
- [12] B. Anwar, “Penerapan Algoritma Jaringan Syaraf Tiruan Backpropagation dalam Memprediksi Tingkat Suku Bunga Bank,” *J. SAINTIKOM*, vol. 10, no. 2, pp. 1–7, 2011.

- [13] M. Azhar and E. Riksakomara, "Peramalan Jumlah Produksi Ikan dengan Menggunakan," *J. Tek. ITS*, vol. 6, no. 1, pp. 142–148, 2017.
- [14] R. J. Djuli, A. Y. Mauko, and M. Boru, "Normalisasi Masukan dan target dinormalisasi dengan membawa data ke bentuk normal dalam rentang nilai antara 0 sampai 1 apabila fungsi aktivasi yang digunakan adalah," *J. Komput. Inform.*, vol. 6, no. 1, pp. 53–59, 2018.
- [15] M. Febrina, F. Arina, and R. Ekawati, "Peramalan jumlah permintaan produksi menggunakan metode jaringan syaraf tiruan (JST) backpropagation," *J. Tek. Ind.*, vol. 1, no. 2, pp. 174–179, 2013.
- [16] Fatayat and A. Fitriansyah, "Analisa Menggunakan Metode Neural Network," *J. SIMTIKA*, vol. 1, no. 1, pp. 1–6, 2018.
- [17] A. Revi, Solikhun, and M. Safii, "Jaringan Syaraf Tiruan Dalam Memprediksi Jumlah Produksi Daging Sapi Berdasarkan Provinsi," in *Konferensi Nasional Teknologi Informasi dan Komputer*, 2018, vol. 2, pp. 297–304.
- [18] F. Pakaja, A. Naba, and Purwanto, "Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan Certainty Factor," *J. EECCIS*, vol. 6, no. 1, pp. 23–28, 2012.
- [19] H. Aprilianto, S. Kumalaningsih, and I. Santoso, "Penerapan Jaringan Syaraf Tiruan Untuk Peramalan Penjualan Dalam Mendukung Pengembangan Agroindustri Coklat di Kabupaten Blitar," *Habitat*, vol. 29, no. 3, pp. 129–137, 2018.
- [20] S. Nawawi and S. Riptiono, "Pengaruh Brand Awareness dan Brand Association Terhadap Brand Equity dengan Brand Loyalty sebagai Variabel Intervening," *J. Ilm. Mhs. Manajemen, Bisnis dan Akunt.*, vol. 2, no. 2, pp. 166–183, 2020.
- [21] I. W. F. A. Yoga, L. P. Wrsiati, and G. P. G. Putra, "Strategi Pengembangan Usaha Air Minum Dalam Kemasan (Studi Kasus PT . Satyaloka Tirta Amerta Bangli)," *J. Rekayasa dan Manaj. Agroindustri*, vol. 4, no. 4, pp. 12–23, 2016.
- [22] J. Yao and C. L. Tan, "A case study on using neural networks to perform technical forecasting of forex," *Neurocomputing*, vol. 34, pp. 79–98, 2000.
- [23] N. F. Hasan, K. Kusriani, and H. Al Fatta, "Peramalan Jumlah Penjualan Menggunakan Jaringan Syaraf Tiruan Backpropagation Pada Perusahaan Air Minum Dalam Kemasan," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 2, pp. 179–188, 2019.