

Prediction of Salinity Based on Meteorological Data Using the Backpropagation Neural Network Method

Anisa Nur Azizah¹, Dian C. R. Novitasari^{1*}, Putroue Keumala Intan¹, Fajar Setiawan², Ghaluh Indah Permata Sari³

¹Department of Mathematics, UIN Sunan Ampel

Jl. Ahmad Yani No.117, Wonocolo, Surabaya, Jawa Timur 60237 Indonesia

² Meteorological, Climatological, and Geophysics Agency, Surabaya

Jl. Kalimas Baru No.97B, Perak Utara, Surabaya, Jawa Timur 60165 Indonesia

³Department of Computer Science and Information Engineering,

National Taiwan University of Science and Technology

No.43, Keelung Rd, Sec.4, Da'an Dist, Taipei City 106335, Taiwan

Email: diancrini@uinsby.ac.id

Abstract

Salinity is the level of salt dissolved in water. The salinity level of seawater can affect the hydrological balance and climate change. The salinity level of seawater in each area varies depending on the influencing factors, that is evaporation and precipitation (rainfall). One way to find out the salinity level is by taking seawater samples, which requires a long time and costs a lot. In this study, the salinity level of seawater can be predicted by utilizing time series data patterns from evaporation and precipitation using artificial neural network learning, namely the backpropagation neural network. The evaporation and precipitation data used were derived from the ECMWF dataset, while the salinity data were derived from NOAA where each data was taken at the coordinate point of 9,625 113,625 in the south of Java island. Seawater salinity, evaporation, and precipitation data were formed into a 7-day time series data. This study conducted several backpropagation architectural experiments, that is the learning rate, hidden layer, and the number of nodes in the hidden layer to obtain the best results. The results of the seawater salinity prediction were obtained at a MAPE value of 2.063% with a model architecture using 14 input layers, 2 hidden layers with 10 nodes and 2 nodes, 1 output layer, and a learning rate of 0.7. Predicted sea water salinity data ranging from 33 to 35 ppt. Therefore, the prediction system for seawater salinity using the backpropagation method can be said to be good in providing information about the salinity level of sea water on the island of Java.

Keywords: Salinity, Evaporation, Precipitation, Time Series, Backpropagation

Introduction

Indonesia is known as a maritime country with two-thirds of Indonesia's territory consisting of the sea. According to Law No. 17 of 1985, the total area of the Indonesian sea is 5.9 million km². This makes Indonesia's territorial waters have great potential to advance the country's economy (Lasabuda, 2013). According to data from the Central Statistics Agency (BPS), fishery production in the Java Sea reached the highest production with an average fishing rate of 200,903 tons in Central Java, 88,453.8 tons in East Java, and 308,416.5 tons in West Java from 2010 to 2017 (BPS, 2019). The fish caught in the Java Sea also varies from tuna, stingray, anchovies, squid, and others (Sentosa et al., 2016).

Several factors influence the diversity of fish species in the sea, one of which is salinity (Subagiyo et al., 2015). Salinity is the amount of salt dissolved

in water. Salinity greatly affects the morphological conditions, size, and some organisms living in the area (Arisandi et al., 2011). In both oceanographic and meteorological studies, salinity is an important parameter. The availability of seawater salinity data is related to variations in sea level salinity which are useful for hydrological balance and changing climates (Najib et al., 2017). Meanwhile, according to Tubalawony the level of salinity and temperature on the surface are indicators of upwelling events (Tubalawony and Kusmanto, 2012). The climate in the Java Sea follows the season pattern where the dry season lasts from June to September, while the rainy season is from November to March. In waters where seasonally the salinity level is important to know sea surface temperature, salt content, and abundance of fauna (fish) so that the accuracy on the salinity level data of water is required (Islami, 2013).

Salinity data is obtained using tools such as a refractometer and a salinometer. The refractometer works by measuring the index of refraction in the liquid to determine the dissolved salt content, while the salinometer works by delivering electrical power to the solution, the greater the electrical conductivity, the greater the level of salt content. Both tools require a long time to collect data and the accuracy of the location in the ocean, while the salinity level can be determined from evaporation and rainfall (precipitation) in an area (Grodsky and Carton 2018; Ratnawati *et al.*, 2018). Therefore, this study makes the prediction of salinity levels without having to take samples in the field by utilizing time series data from the factors that affect salinity, that is evaporation and precipitation.

A previous research made the prediction of salinity in the Taiwan Strait based on remote sensing data from MODIS using the improved genetic algorithm combining operation tree (IGAOT) method resulting in a correlation coefficient (CC) value of 0.38 and a root mean squared error (RMSE) of 1.88 (Chen *et al.* 2017). Rajabi-Kiasari and Hasanlou (2020) predicted the salinity level of sea in the Persian Gulf using a comparison between the Support Vector Regression (SVR) method, artificial neural network (ANN), random forest (RF) and gradient boosting machine (GBM). In several previous studies regarding the prediction of sea water salinity, prediction of salinity requires a good method for data pattern recognition so that this study uses one of the neural network methods, namely backpropagation.

The backpropagation neural network method is the most popular method in data pattern recognition, prediction, and forecasting. A previous research which predicted the water distribution of PDAM in Malang City had an accuracy rate of 97.99% (Sawitri *et al.*, 2018). A subsequent research, that is the prediction of rainfall in Pekanbaru City using the backpropagation method resulted in an accuracy of 96% (Jauhari *et al.*, 2016). The learning process of the backpropagation method is influenced by the number of hidden layers, learning rate, and several other architectural experimental factors (Wahyuni *et al.*, 2017). (Jumhuriyah *et al.*, 2020) conducted backpropagation architectural experiments using various hidden layer nodes and learning rates which resulted in the best architecture at nodes of 100 and learning rate of 0.4. Furthermore, (Wahyu *et al.*, 2020) in the best architectural experiments using backpropagation resulted in an accuracy of 99.03% at a learning rate of 0.3. Therefore, in this study the backpropagation neural network method is carried out with several architectural experiments to determine the best results in predicting the salinity level of seawater in the south of Java using meteorological data, that is precipitation and

evaporation. It is hoped that the salinity prediction can help fishermen determine certain types of fishing areas and can advance the Indonesian economy by increasing salt production.

Materials and Method

In this study, the prediction of salinity levels in the southern waters of Java Island was carried out as shown in Figure 1. The data used were daily time series data of evaporation and precipitation rates as parameters and salinity data as targets taken from April 1, 2017 to July 31, 2019, which was taken at the coordinate point (-9.625 113.625). Evaporation and precipitation data were obtained from the ECMWF dataset (ECMWF, 2019) while salinity data were obtained from NOAA (National Oceanic and Atmospheric Administration) (NOAA, 2019).

Data analysis

Evaporation, precipitation, and salinity data were analyzed by doing fill missing if there were blank data in the time series data. Missing data were corrected based on known data values using linear interpolation (Moon *et al.*, 2019) in Equation (1) with the results presented in Table 1 as pre-processing data. Then the stage was proceeded with data sharing as much as 70% and 30% on training data and testing data.

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x - x_0)$$

Time series

Periodic data or time series is a collection of data arranged according to time series. Time series data are usually also used to determine the development of a situation or event with the same time interval which can be seconds, minutes, hours, days, weeks, months, or years (Khasanah *et al.*, 2019). In this study, a time interval of 1 week or 7 days with variables was used because evaporation data were formulated as E_1, E_2, \dots, E_n and precipitation data were formulated as P_1, P_2, \dots, P_n , while the target or effect variables, that is salinity data were formulated as S_1, S_2, \dots, S_n with n as the total data. The data pattern can be seen in Table 2.

Backpropagation neural network method

The backpropagation neural network method is a neural network model with supervised learning (Howard, 2019). The backpropagation method is also the most popular method to solve complex pattern recognition problems. In pattern recognition, this method consists of two phases, that is the forward

propagation phase and the backward propagation phase (Sawitri *et al.*, 2018). The network of the backpropagation method is included in a multilayer network where the network is composed of neurons that are connected at each layer, namely the input layer, hidden layer, and output layer as shown in Figure 2, there is V_{ij} which is the weight of the input

layer X_i to the hidden layer Z_j and there is a bias weight with a value of 1. There is also w_{jk} which is the weight from the hidden layer Z_j to the output layer y_k and added with a bias weight with a value of 1. The steps of backpropagation (Alkronz *et al.*, 2019) can be seen in Table 3.



Figure 1. Map of Coordinate Point of -9.625 113.625
Source: (Google 2021)

Table 1. Evaporation, Precipitation, and Salinity Data Samples at Coordinate Point (-9.625 113.625)

Date	Evaporation ($\cdot 10^{-7}$)	Precipitation	Salinity
4/1/2017	5779	0.001416	-
4/2/2017	10386	0.002916	33.37774
4/3/2017	9872	0.004181	32.64356
4/4/2017	5443	0.001795	-
4/5/2017	6075	0.003858	31.80197
4/6/2017	8207	0.001205	-
4/7/2017	4933	0.000651	33.28509
4/8/2017	7085	0.001914	34.92991
4/9/2017	9408	0.004014	-
4/10/2017	9451	-0.00000	-

Table 2. Time Series Data Pattern

Input Parameter								Target
X_1	X_2	...	X_7	X_8	X_9	...	X_{14}	Y
E(t)	E(t+1)	...	E(t+6)	P(t)	P(t+1)	...	P(t+6)	S(t+7)
E(t+1)	E(t+2)	...	E(t+7)	P(t+1)	P(t+2)	...	P(t+7)	S(t+8)
E(t+2)	E(t+3)	...	E(t+8)	P(t+2)	P(t+3)	...	P(t+8)	S(t+9)
E(t+3)	E(t+4)	...	E(t+9)	P(t+3)	P(t+4)	...	P(t+9)	S(t+10)
⋮	⋮	...	⋮	⋮	⋮	...	⋮	⋮
E(t+(n-6))	E(t+(n-5))	...	E(t+(n-1))	P(t+(n-6))	P(t+(n-5))	...	P(t+(n-1))	S(t+n)

In this study, several network architecture models of backpropagation are made to determine the best model used in the training process. The steps to create a network architecture model determine the number of input layers, hidden layers, the number of nodes, and the output layers illustrated in Table 4, and determine the learning rate between 0.1 and 0.9 with a range of 0.1 in the backpropagation network for each architectural model.

Mean Absolute Percentage Error (MAPE)

MAPE is a method of measuring accuracy by averaging the values found from the error for each period and divided by the value of observations (De Myttenaere et al. 2016). The MAPE value is formulated :

$$MAPE = \frac{\sum_1^n \left| \frac{x_t - f}{x_t} \right| \times 100\%}{n}$$

Where x_t is target data, f is the predictive data, n is the amount of data. The smaller the MAPE results which show the better the prediction results are shown in Table 5 on the MAPE result criteria.

Results and Discussion

This study made a prediction on the salinity level of seawater in the southern part of the Java sea based on meteorological data, that is precipitation and evaporation data using the backpropagation neural network method. To obtain the best results, experiments were carried out on several backpropagation network architectures. The data were taken from April 2017 to July 2019, and the total data per variable was 852 data. Next, fill missing using Equation (1) was carried out as shown in Table 6. Then a time series data pattern was created as shown in Table 2 with precipitation and evaporation data as input variables and salinity data as target variables to be used in the learning stage of the backpropagation neural network.

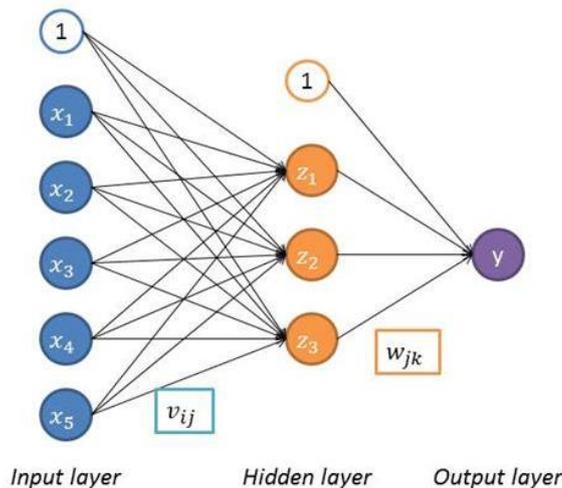


Figure 2. Backpropagation Method Architecture

Table 3. Steps of the Backpropagation Method

Pseudo Code Backpropagation Neural Network Method
Initialize the weight
While iter > max iteration error < tolerance limit
For i = 1:data
Training data through forwards network $Z_j y_k$ and calculate output value with activity
Calculate loss function $t_k - y_k$ at each data
Calculate error factor δ and multiply by activity function derivative
Calculate weight change W by multiplying alpha α , error factor δ , and output value Z
Calculate weight change V by multiplying alpha α , error factor δ , and data value X
Update weight $W_{(new)}$ by adding up old weight $W_{(old)}$ + Update weight ΔW_{jk}
Update weight $V_{(new)}$ by adding up old weight $V_{(old)}$ + Update weight ΔV_{ij}
End
Calculate the error with MSE
iter = iter+1
End

Table 4. Backpropagation Neural Network Architecture Model

Architecture Model	Number of Hidden Layers	Number of Nodes
1st architecture	1	10
2nd architecture	1	50
3rd architecture	1	100
4th architecture	2	[10 2]
5th architecture	2	[5 50]
6th architecture	2	[25 10]

Table 5. MAPE Criteria

MAPE	Prediction Ability Category
< 10 %	Very Good
10 % - 20 %	Good
20 % - 50 %	Fairly Good
> 50 %	Not Good

Source: (Xu et al., 2017)

Table 6. Data after Fill Missing

Date	Evaporation (-10^{-7})	Precipitation	Salinity
4/1/2017	5779	0.001416	34.11193
4/2/2017	10386	0.002916	33.37774
4/3/2017	9872	0.004181	32.64356
4/4/2017	5443	0.001795	32.22276
4/5/2017	6075	0.003858	31.80197
4/6/2017	8207	0.001205	32.54353
4/7/2017	4933	0.000651	33.28509
4/8/2017	7085	0.001914	34.92991
4/9/2017	9408	0.004014	34.61737
4/10/2017	9451	-0.00000	34.30484

As seen from Table 7, various experiments of backpropagation network architecture models are generated by the most optimal network in predicting salinity using the 4th network architecture model which can be seen in Figure 3 with a total of 14 input layers, 1 output layer, 2 hidden layers with 10 nodes in the first hidden layer and 2 nodes in the second hidden layer, using the sigmoid activation function. The best result in this architectural model is at a learning rate of 0.7 with the MAPE result obtained of 2.063%.

In Table 7, it is also clear that the network architecture model has a greater influence on the MAPE results than the learning rate experiment. The 4th architecture network model has a smaller MAPE mean than the other architectural models with a value of 2.239. From these results, the prediction of

salinity levels from 19 November 2018 to 31 July 2019 using the backpropagation method can be seen in Figure 4. These results are said to be good in predicting salinity levels (Xu et al., 2017).

As seen from the charts on the comparison of the prediction results above, the red line shows the actual salinity data with a value range between 31 to 38 ppt, while the blue line shows the predicted data with a range of values between 33 to 35 ppt. The actual salinity data has a fairly high range of values in March and May 2019 coupled with a large number of blank values in the data so that the salinity prediction value is not good. Outliers can be found in the data with a far range of values so it needs improvement (Kieu et al., 2019). Salinity levels can be predicted using salinity image data from Landsat (El-Battay et al., 2017) or salinity

Table 7. Results of Each Backpropagation Architecture Model

Learning rate	MAPE (%) 1st architecture	MAPE (%) 2nd architecture	MAPE (%) 3rd architecture	MAPE (%) 4th architecture	MAPE (%) 5th architecture	MAPE (%) 6th architecture
0.1	2.740	2.617	5.822	2.343	2.645	2.941
0.2	2.531	3.121	3.339	2.375	3.718	2.693
0.3	2.687	4.798	4.441	2.157	2.750	2.902
0.4	2.699	3.566	6.518	2.277	3.268	2.945
0.5	2.485	3.164	3.738	2.232	4.080	2.234
0.6	2.382	4.164	4.211	2.142	3.348	2.408
0.7	2.758	3.015	5.456	2.063	3.099	2.366
0.8	2.427	3.990	4.756	2.303	2.735	2.537
0.9	2.203	3.601	7.296	2.265	2.655	2.622
Mean	2.546	3.559	5.064	2.239	3.144	2.628

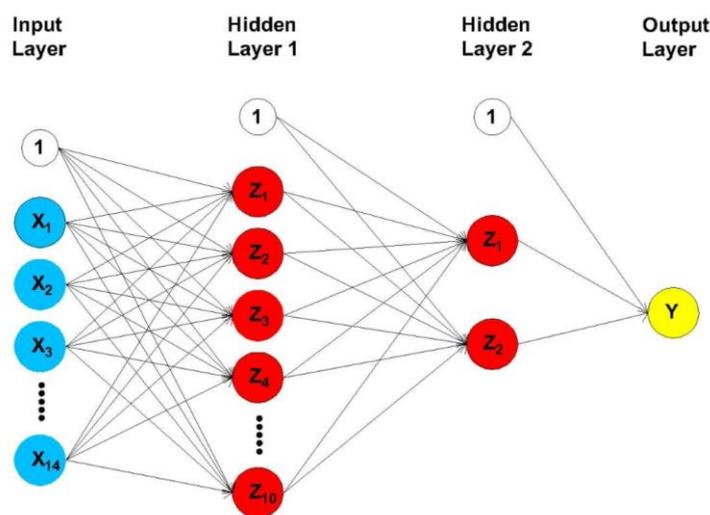


Figure 3. 4th Backpropagation Architecture Model

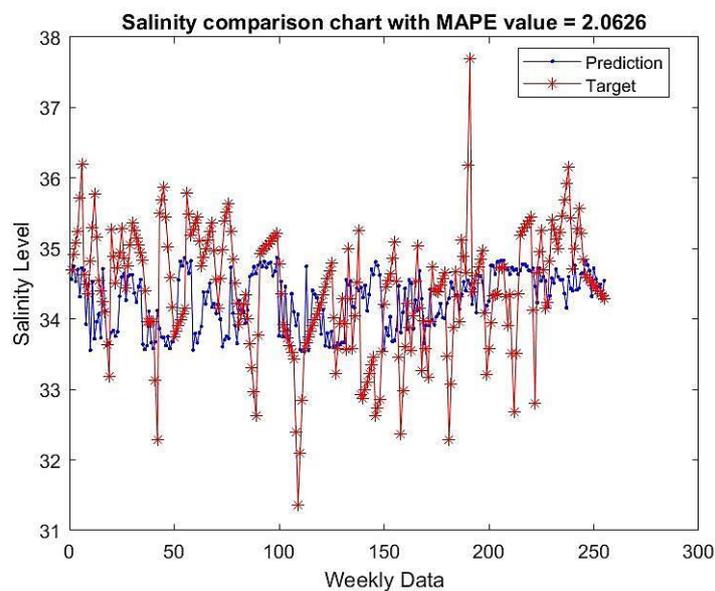


Figure 4. Charts of Comparison between Actual Salinity Data and Data on Prediction Results with Weekly Patterns

image data from MODIS (Yu *et al.*, 2017). In the backpropagation method, it is recommended to add calculations such as momentum, adaptive learning rate, gradient, (Hameed *et al.*, 2016), or others so that it can optimize the training process. Also, it can use other forecasting methods such as support vector regression (SVR) (Suwanto and Novitasari, 2020) or the adaptive neuro fuzzy inference system (ANFIS) method so that it can produce even better levels of accuracy.

Conclusions

Backpropagation neural network method is very good at predicting the salinity level of seawater in southern Java. The parameter data used were precipitation and evaporation data with 7 days series data patterns. Several experiments were carried out by changing the number of hidden layers, hidden layer nodes, and learning rate values in the backpropagation architecture. The best result was the 4th model architecture with 2 hidden layers, 10 nodes in the first hidden, and 2 in the second hidden which resulted in a MAPE value of 2.063% at a learning rate of 0.7 with a range of salinity prediction results between 33 to 35 ppt.

Acknowledgment

This research was supported by the Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia dan BMKG (Meteorology, Climatology, and Geophysics Agency) Surabaya, Indonesia.

References

Alkronz, E.S., Moghayer, K.A., Meimeh, M., Gazzaz, M., Abu-Nasser, B.S. & Abu-Naser, S.S., 2019. Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network.

Arisandi, A., Marsoedi, M., Nursyam, H. & Sartimbul, A., 2011. Pengaruh salinitas yang berbeda terhadap morfologi, ukuran dan jumlah sel, pertumbuhan serta rendemen karaginan *Kappaphycus alvarezii*, *Ilmu Kelautan: Indonesian Journal of Marine Sciences*, 16(3):143-150.

BPS, 2019, Statistik Indonesia Tahun 2019. Jakarta: Statistics Indonesia.

Chen, L., Alabbadi, B., Tan, C.H., Wang, T.S. & Li, K.C., 2017. Predicting sea surface salinity using an improved genetic algorithm combining operation

tree method, *J. Indian Soc. Remote. Sens.*, 45(4): 699-707. doi: 10.1007/s1252 4-016-0637-7

ECMWF, 2019. ECMWF Datasets ERA-Interim, *ECMWF*.

El-Battay, A., Bannari, A., Hameid, N.A. & Abahussain, A.A., 2017. Comparative study among different semi-empirical models for soil salinity prediction in an arid environment using OLI Landsat-8 data, *Adv. Remote Sens.*, 6(1): 23-39. doi: 10.4236/ars.2017. 61002

Grodsky, S.A. & Carton, J.A., 2018. Delayed and quasi synchronous response of tropical Atlantic surface salinity to rainfall, *J. Geophys. Res. Oceans*, 123(8): 5971-5985. doi: 10.1029/2018JC013915

Hameed, A.A., Karlik, B. & Salman, M.S., 2016. Back-propagation algorithm with variable adaptive momentum, *Knowl Based Syst*, 114: 79-87. doi: 10.1016/j.knosys.2016. 10. 001

Howard, J., 2019. Artificial intelligence: Implications for the future of work, *Am. J. Ind. Med.* 62(11): 917-926. doi: 10.1002/ajim.23037

Islami, M.M., 2013. Pengaruh Suhu dan Salinitas Terhadap Bivalvia, *Oseana*, 30: 1-10.

Jauhari, D., Himawan, A. & Dewi, C. 2016. Prediksi Distribusi Air PDAM Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation Di PDAM Kota Malang, *J. Teknol. Informasi dan Ilmu Komputer*, 3(2): 83-87. doi: 10.25126/jtiik.201632155

Kieu, T., Yang, B., Guo, C. & Jensen, C.S. 2019. Outlier Detection for Time Series with Recurrent Autoencoder Ensembles., *Int. Jt. Conf. Artif. Intell.* 2725-2732. doi: 10.24963/ijcai.2019 /378

Jumhuriyah, L., Dian, C., Novitasari, R. & Setiawan, F. 2020. Prediksi Kecepatan Arus Laut dengan Menggunakan Metode Backpropagation (Studi Kasus: Labuhan Bajo), *J. Algebra*, 1: 99-108.

Lasabuda, R. 2013. Tinjauan Teoritis Dalam Perspektif Negara Kepulauan Republik Indonesia Regional Development in Coastal and Ocean in Archipelago Perspective of The Republic of Indonesia. *J. Platax*, 1(2): 92-101. doi: 10.35800/jip.1.2.2013.12 51

Moon, T., Hong, S., Choi, H.Y., Jung, D.H., Chang, S.H. & Son, J.E. 2019. Interpolation of greenhouse

- environment data using multilayer perceptron, *Comput Electron Agric*, 166:p105023. doi: 10.1016/j.compag.2019.105023
- Myttenaere, A. De, Golden, B., Grand, B. Le & Rossi, F., 2016, Mean absolute percentage error for regression models, *Neurocomputing*, 192: 38–48. doi: 10.1016/j.neucom.2015.12.114
- Najib, S., Fadili, A., Mehdi, K., Riss, J. & Makan, A., 2017. Contribution of hydrochemical and geoelectrical approaches to investigate salinization process and seawater intrusion in the coastal aquifers of Chaouia, Morocco, *J. Contam. Hydrol.*, 198:24-36. doi: 10.1016/j.jconhyd.2017.01.003
- NOAA, 2019, NOAA Near Real-time Search, NOAA.
- Rajabi-Kiasari, S. & Hasanlou, M., 2020, An efficient model for the prediction of SMAP sea surface salinity using machine learning approaches in the Persian Gulf, *Int. J. Remote Sens.*, 41(8):3221–3242. doi: 10.1080/01431161.2019.1701212
- Ratnawati, H.I., Aldrian, E. & Soepardjo, A.H., 2018, Variability of evaporation-precipitation (EP) and sea surface salinity (SSS) over Indonesian maritime continent seas, *AIP Conf. Proc.*, 2023:20252, doi: 10.1063/1.5064249
- Rini, D.C., Khasanah, U., Utami, W.D. & Intan, P.K 2019. Analisis Peramalan Beban Listrik (Studi Kasus : PT . PLN (Persero) Area Pengaturan Distribusi Jawa Timur), *MathVisioN*. 1(1): 17–24.
- Sawitri, M.N.D., Sumarjaya, I.W. & Tastrawati, N.K.T., 2018, Peramalan Menggunakan Metode Backpropagation Neural Network, *Jurnal Matematika*, 7(3):264-270. doi: 10.24843/MTK.2018.v07.i03.p213
- Sentosa, A.A., Widarmanto, N., Wiadnyana, N.N. & Satria, F. 2016, Perbedaan Hasil Tangkapan Hiu dari Rawai Hanyut dan Dasar yang Berbasis di Tanjung Luar, Lombok, 22(2): 105–114
- Subagiyo, S., Margino, S., Triyanto, T. & Setyati, W.A.A., 2015, Effects of pH, temperature and salinity in growth and organic acid production of lactic acid bacteria isolated from penaeid shrimp intestine, *Ilmu Kelautan: Indonesian Journal of Marine Sciences*, 20(4):187–194. doi: 10.14710/ik.ijms.20.4.187-194
- Suwanto, S. & Novitasari, D.C.R., 2020, Forecasting Solar Activities based on Sunspot Number Using Support Vector Regression (SVR), *J. Phys. Sci. Eng.*, 5(1):6–10. doi: 10.17977/um024v5i12020p006
- Tubalawony, S. & Kusmanto, E., 2012, Suhu dan Salinitas Permukaan Merupakan Indikator Upwelling Sebagai Respon Terhadap Angin Muson Tenggara di Perairan Bagian Utara Laut Sawu, *Ilmu Kelautan: Indonesian Journal of Marine Sciences*, 17(4):226-239. doi: 10.14710/ik.ijms.17.4.226-239
- Wahyu, N.W., Novitasari, D.C.R. & Hermanto, A. 2020. Klasifikasi Ketinggian Gelombang Laut Perairan Masalembu Berdasarkan Hasil Prediksi Parameter Meteorologi Maritim Menggunakan Metode Backpropagation, *J. Algebra*, 1(1 SE-Articles): 90–98.
- Wahyuni, I., Adam, N.R., Mahmudy, W.F. & Iriany, A., 2017, Modeling backpropagation neural network for rainfall prediction in tengger east Java, 2017 *Int. Conf.Sustain. Info. Eng. Technol.*, pp. 170–175, doi: 10.1109/SIET.2017.8304130
- Xu, N., Dang, Y. & Gong, Y., 2017. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China, *Energy*, 118: 473–480. doi: 10.1016/j.energy.2016.10.003
- Yu, X., Xiao, B., Liu, Xiangyang, Wang, Y., Cui, B. & Liu, Xin, 2017, Retrieval of remotely sensed sea surface salinity using MODIS data in the Chinese Bohai Sea, *Int. J. Remote Sens.*, 38(23):7357–7373. doi: 10.1080/01431161.2017.1375570