Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023)

Comparing Artificial Neural Network and Decision Tree Algorithm to Predict Tides at Tanjung Priok Port

Arif Musadi [a], Christian Carlos Tertius [a], Jonas Steven [a], Hanis Amalia Saputri [a], Kristien Margi Suryaningrum [a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

Tanjung Priok Port is an international port in Indonesia located in Tanjung Priok, North Jakarta. For all activities carried out at Tanjung Priok Port to run smoothly, this research was made which aims to predict the height of tides using the Artificial Neural Network (ANN) and Decision Tree methods with a quantitative approach. Artificial Neural Network (ANN) is a technique inspired by the way the biological nervous system works, namely in brain cells in processing information received by humans. while Decision Tree is also known as a decision tree which is an algorithm for building a decision hierarchy structure. The process of making a Decision Tree starts from the Root Node to the Leaf Node which is done recursively. This research was conducted to predict the height of tides in January 2018 - June 2018. By using both methods that have been computed, the ANN method produces a smaller MSE value than the Decision Tree method. The ANN method produces an MSE value of 0.003727983. While the Decision Tree method produces an MSE value of 0.009870259. If the dataset used has larger amount of data and the architecture of each algorithm is more complex, then the calculation results obtained will be more accurate.

Keywords: Tidal prediction; Artificial Neural networks; Decision tree.

1. Introduction

Tanjung Priok Port is an international port in Indonesia located in Tanjung Priok, North Jakarta. This port functions as a place for export-import and inter-island goods which makes it the busiest port in Indonesia because the exported and imported goods have a very large number [1].

* Corresponding author. Tel.:+62 81266001895
E-mail address: jonas.steven@binus.ac.id

1877-0509

Tides can be defined as a phenomenon used to describe the periodic and harmonic fluctuation movement of water masses caused by the gravitational forces of the moon and sun on the differential part of the rotating earth. In addition, tides can also be described as a phenomenon of periodic movement of the rise and fall of sea levels [2].

Some methods that can be used to predict sea tides are Artificial Neural Network (ANN), which is one of the most popular algorithms for forecasting or prediction [2]. In previous research that used the Decision Tree method to predict sea tides by modeling it into the form of a Decision Tree and then calculating the accuracy value [1].

The research objectives to be achieved from this research are to implement and compare the Artificial Neural Network (ANN) method with Decision Tree to predict tides at Tanjung Priok port with a high level of accuracy.

The results of this study are expected to provide benefits such as being used as a safety reference for ships that want to exit and enter so as to reduce ship accidents and make it easier for ships to dock to the port side.

2. Related Work

2.1 Artificial Neural Network Algorithm for Prediction

Based on previous research using the Artificial Neural Network algorithm entitled "Artificial neural network-based storm surge forecast model: Practical application to Sakai Minato, Japan", describes the application of JST to the Sakai Minato area, Japan. This research aims to describe a new approach in conducting a selection process for the development of an Artificial Neural Network (ANN)-based storm surge forecast model with 5, 12, and 24-hour predictions located in the Sakai Minato area, Tottori coastline, Japan. This study also determined the benefits of the best performance in the model by using the right combination of units in the hidden layer as well as parameters in the input layer. The results of this study found that ANN-SFM is a good model for 5 to 12-hour forecasts in Sakai Minato, Japan. By using 70 units and 8 input parameters variables such as wave height, sea level pressure, sea level drop, air pressure, longitude, latitude, mean air pressure, and maximum wind speed. Meanwhile, the best 24-hour ANN-SFM forecast was determined with 160 units and 4 input parameter variables such as wave height, sea level pressure, sea level drop, longitude, and latitude. The researchers hope that the proposed valid selection process method can be transferred to other coastal areas to develop artificial neural network-based flood prediction models [3].

In addition, another paper using the same algorithm titled "Artificial Neural Networks and Multiple Linear Regression for Flood Prediction on the Mohawk River, New York", presents a hybrid model for predicting river flooding located on the Mohawk River, New York. The model combines time series analysis with Artificial Neural Networks to explain and predict the amount of daily water runoff by utilizing hydrogeological and weather climate variables. A low-pass filter is useful to decompose the time series into components: long-term, seasonal, and short-term. In predicting the time series of water runoff, the researchers explain each part by applying Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN). The function of MLR itself is to assess the usefulness of the water layer time series. The researchers proved that time series decomposition is essential before applying the model. Moreover, the decomposition proved that the Mohawk River has an influential time scale factor in the hydrological cycle of the river flow. After comparing the models, it was found that applying ANN to the decomposed time series method can improve the accuracy in terms of flood event prediction. The hybrid model can predict up to 96% of the time series explanation of water runoff [4].

2.2 Decision Tree Algorithm for Prediction

In a previous study located in Tanjung Priok port, North Jakarta, a quantitative and qualitative approach was used to analyze tidal data modeling. From this research, the observations obtained with the predicted tides of Tanjung Priok port were visualized in the form of graphs to show the comparison of the rise and fall of the elevation of the two data against time. Dataset collection to predict tides was obtained from the SeaLevelMonitoring website. The data that has been obtained is then validated using the Preprocessing method so that the raw data obtained is of higher quality and then modeled into the form of a Decision Tree. The results obtained in the calculation are then calculated for accuracy [1].

The accuracy value to determine the accuracy of the prediction results using data. The data is calculated using formula (1) the greater the resulting value, the better the accuracy.

$$Accuration = \frac{RDt}{Dt} x100\% \tag{1}$$

After obtaining the value of the prediction results, then the value classification is determined based on the predetermined category. The division of category value intervals is as follows:

1. Extreme lowness: Value < 0.130: 1
2. Low : 0.130 < Value < 0.377: 2
3. Normal: 0.377 < Value < 0.870: 3
4. High: 0.870 < Value < 1.116: 4
5. Extreme heights: Value > 1.116: 5

3. Research Method

This research uses two methods, namely Artificial Neural Network (ANN) and Decision Tree.

3.1 Artificial Neural Network

Artificial Neural Network (ANN) is a technique inspired by the way the biological nervous system works, namely in brain cells in processing information received by humans [5]. ANN consists of many nodes that are interconnected with each other that can take complex non-linear relationships. The ANN method includes feedforward and backpropagation methods [6].

In a feedforward network, the network will only move forward to each node. Each node will receive inputs from "upstream" nodes and provide outputs to "downstream" nodes [7].

To calculate the feedforward using the Artificial Neural Network algorithm using the formula:

$$a_j = \sum_j (w_{ij} * x_i) \tag{2}$$

$$y_j = F(a_j) = \frac{1}{1+e^{-a_j}} \tag{3}$$

$$E(\theta) = \frac{1}{2} \sum_{i=1}^{I} (y_i - v_i)^2 \tag{4}$$

Formula (2) is a formula to find the next layer using the input from the previous screen. Furthermore, to calculate the output of formula (2), we need formula (3) which is a sigmoid formula commonly referred to as an alternative logistic function formula. The function has a range between 0 and 1 and increases monotonically with input. And the last one in formula (4) is a formula that functions to find the error of all nodes.

While the backpropagation network is an ANN method in which each neuron after a set of data is processed will evaluate the error contribution of each neuron, it is intended that the neuron can learn the mapping of input to the desired output to be more accurate.

To calculate backpropagation using the formula:

$$\delta_j = o_j(1-0)(t_j - o_j) \quad \text{“}\textit{if j is an output unit}\text{“} \tag{5}$$

$$\delta_j = o_j(1-0)\sum_j \delta_k w_{kj} \quad \text{“}\textit{if j is a hidden unit}\text{“} \tag{6}$$

$$\Delta w_{ij} = \eta \delta_j o_i \tag{7}$$

Formula (5) is a formula for finding the hidden layer using the input value of the output layer. Furthermore, formula (6) is a formula for finding the input layer or hidden layer using the input value of the hidden layer. And the last one in formula (7) is a formula for finding weight changes in neurons.

An overview of the ANN network architecture used is shown in Figure 1 below.

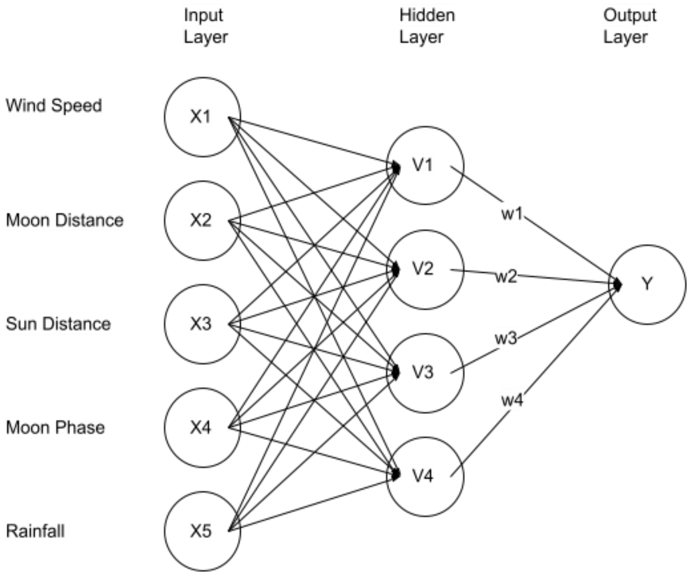

Fig. 1. Artificial Neural Network (ANN) Architecture for Tidal Prediction.

Figure 1 above illustrates the Artificial Neural Network (ANN) Architecture which consists of 3 layers, namely input, hidden, and output. In the input section, there are 5 variables that we use, namely wind speed, moon distance, sun distance, moon phase, and rainfall. In the next part, there is a hidden layer, this layer consists of neurons that are responsible for performing calculations aimed at finding hidden patterns in the system or patterns being analyzed. In addition, the layer functions as a link between the input and output layers [7]. And the output screen is an output of the calculation results in our research the resulting output is Mean Square Error (MSE) which is useful for calculating the average error value in a prediction.

3.2 Decision Tree

Decision Tree is also known as a decision tree which is an algorithm for building a decision hierarchy structure. The process of making a Decision Tree starts from the Root Node to the Leaf Node which is done recursively. Each branching of the tree states the conditions that must be met at each end of the tree which states the value of the data. The process of making a Decision Tree uses the Entrophy calculation method and the Information Gain calculation [1].

In the Decision tree algorithm, the Entrophy method is used to measure how clear a node's statement is.
To calculate Entrophy use the following formula (8) :

$$Entrophy(S) = \sum_{j=1}^{k} -P_j log_2 P_j \tag{8}$$

After obtaining Entrophy, the calculation of Gain (information Gain) is carried out which aims to select the best path that will be used as a node.
Meanwhile, to calculate the Information Gain using the following formula (9):

$$Gain(A) = Entrophy(S) - \sum_{i=1}^{k} \frac{|Si|}{|S|} x \ Entrophy\ (S) \tag{9}$$

For the Decision Tree network architecture used as shown in Figure 2 below.

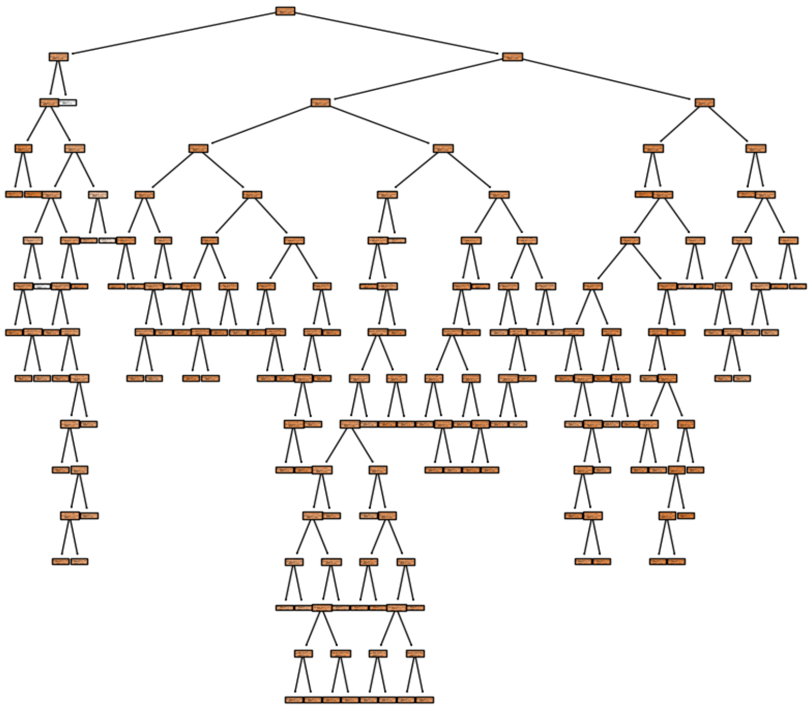

Fig. 2. Decision Tree Architecture for Tidal Prediction.

Because this research is a prediction, an error will be obtained in the prediction results [8]. Therefore, the Mean Square Error (MSE) method is used to calculate the average error value in a prediction using the formula (10):

$$MSE = \sum_{i=1}^{n} \frac{{e_i}^2}{n} \tag{10}$$

The methodology used for this research is quantitative research methods. Quantitative research method is one type of research to collect, analyze, and interpret data. This research requires data from several existing variables such as wind speed, moon distance, sun distance, moon phase, and rainfall [9]. So that all available data can be used efficiently, Preprocessing is carried out. Preprocessing aims to convert categorical data into numerical data and also to prepare the data so that it can be used in the modeling process of each method [10].
To predict the tides at Tanjung Priok Port modeled by the Artificial Neural Network (ANN) and Decision Tree methods, the stages described in the flowchart in Figure 3 below are required.

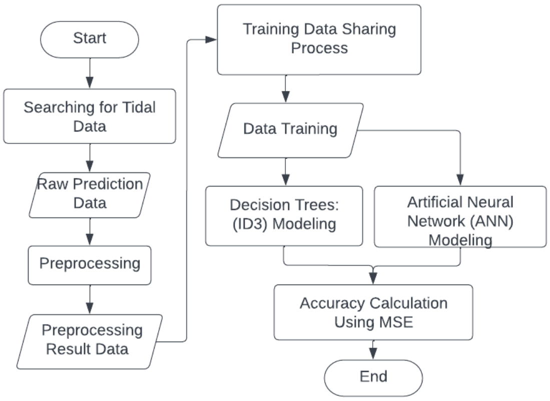

Fig. 3. Flowchart of the prediction result search process.

The flowchart in the picture above is the process of finding prediction results where the data used in the study is data within 181 days from January 1, 2018 to June 30, 2018. The data is preprocessed to produce preprocessing data. These data are carried out the training process using Artificial Neural Network (ANN) and Decision Tree. So that the results of the data will be calculated Accuracy using MSE. Furthermore, from the results of the accuracy calculation, the output results will be obtained which are the results of predictions that can be compared.

DFD (Data Flow Diagram) is a diagram model to describe the data workflow in a system. DFD is used so that the flow of data used from the beginning and end can be observed properly to prevent data loss, so that a system can be developed properly. The following is a DFD design for an overview of system design can be seen in Figure 4 below [11].

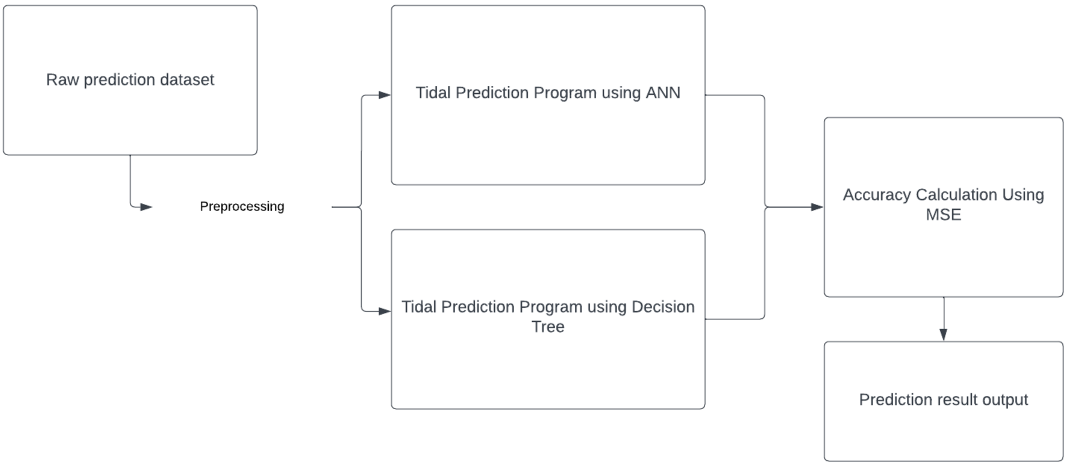

Fig. 4. DFD system design overview.

In the DFD design in Figure 4 above, it can be described how the system work process will be designed. The process starts from collecting data consisting of 5 variables. The data will be preprocessed and then will be entered into the ANN and Decision Tree programs, after which the MSE calculation will be carried out so that an error is obtained that can be used to measure and compare the accuracy of the prediction, the smaller the error number, the more accurate the prediction results.

The dataset used in this study is data from 5 variables from January 2018 - June 2018. The moon distance, sun distance, and moon phase are obtained from the PasangLaut website, while the wind speed and rainfall data are obtained from the DataOnlineBMKG website. The data obtained is processed using Data Preprocessing.

Table 1. Data before Preprocessing.

Date	Rainfall (*mm*)	Moon Distance (*km*)	Sun Distance (*km*)	Wind Speed (*knots*)	Moon Phase	Target (*m*)
01-01-2018	3.42	363626	147099024	7.776	Convex	0.65
02-01-2018	0	363345	147097898	3.888	Full	0.7
03-01-2018		363309	147097512	1.944	Convex	0.65
04-01-2018	0.1	366454	147097867	5.832	Convex	0.65
05-01-2018		369635	147098961	7.776	Convex	0.6
...	...	...	...	...	...	...
30-06-2018	3	405497	152091516	1.944	Convex	0.6

In the data preprocessing stage, missing value replacement is carried out by averaging the entire number of values in the data variable, replacing data that is still in categorical form into numeric data so that calculations can be carried out at the next stage. Then normalize the data on all variable values to allow data processing to be more efficient [1].

4. Result and Discussion

Based on the results of Data Preprocessing, the data obtained has been normalized and can be used for the modeling process into the ANN and Decision Tree models which can be seen in Table 2 below.

Table 2. Data after Preprocessing.

Date	Rainfall (*mm*)	Moon Distance (*km*)	Sun Distance (*km*)	Wind Speed (*knots*)	Moon Phase	Target (*m*)
01-01-2018	1.482	-1.414	-1.369	2.026	0	0.65
02-01-2018	-0.681	-1.433	-1.369	-0.098	2	0.7
03-01-2018	0	-1.435	-1.369	-1.159	0	0.65
04-01-2018	-0.675	-1.226	-1.369	0.964	0	0.65
05-01-2018	0	-1.014	-1.369	2.026	0	0.6
...	...	...	...	...	...	...
30-06-2018	-0.491	1.372	1.4696	-1.159	0	0.6

In this study, to get the best prediction results using the Artificial Neural Network (ANN) architecture model configuration with 5 input layers, 4 hidden layers and 1 output layer. The learning rate used is 0.001 (Adam), epochs 30, batch size 10 and sigmoid activation function. While the Decision Tree begins by calculating the Entrophy value and then calculating the Information Gain value to design a decision tree based on the Information Gain value obtained. By using data that has been normalized in Preprocessing and then implemented into the ANN and Decision Tree models, the results of sea level predictions compared to actual water level data for the ANN model can be seen in Table 3 and the Decision Tree model can be seen in Table 4.

Table 3. Comparison of Predicted Results and Actual Data for (a) ANN Model; (b) Decision Tree Model

a

Date	Target (m)	Prediction (m)	Error
01-01-2018	0.65	0.626	0.024
02-01-2018	0.7	0.606	0.094
03-01-2018	0.65	0.610	0.04
04-01-2018	0.65	0.621	0.029
05-01-2018	0.6	0.626	0,026
...	...	...	...
30-06-2018	0.6	0.601	0.001
MSE			0.003727983

b

Date	Target (m)	Prediction (m)	Error
01-01-2018	0.65	0.65	0
02-01-2018	0.7	0.7	0
03-01-2018	0.65	0.65	0
04-01-2018	0.65	0.55	0
05-01-2018	0.6	0.65	0.05
...	...	...	...
30-06-2018	0.6	0.601	0.05
MSE			0.009870259

Based on Table 4 (a) using ANN, it can be seen that the prediction results and actual data that produce an MSE value of 0.003727983. While in Table 4 (b) which uses Decision Tree, it can be seen that the prediction results and actual data that produce an MSE value of 0.009870259. The comparison between the two methods for predicting sea level can be seen in Figure 5 below.

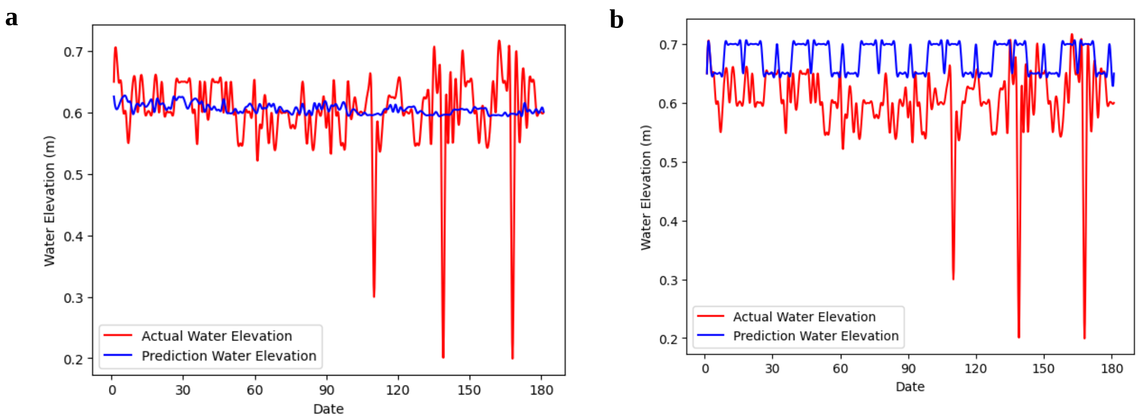

Fig. 5. Comparison Chart of Prediction Results and Actual Data using (a) ANN; (b) Decision Tree.

In Figure 5 above, illustrates the comparison between the predicted sea level and the actual sea level that has been processed using the ANN and Decision Tree methods. The red colored line shows the actual sea level data and the blue colored line shows the predicted sea level data. In Figure 5 (a), it can be seen the shape of the tidal prediction graph using ANN against the actual water level for 181 days. It can also be seen that the prediction results using ANN in Figure 5 (a) get fairly stable prediction results by looking at the blue line in the figure. While in Figure 5 (b), it can be seen the shape of the tidal prediction graph using Decision Tree against the actual water level for 181 days. For the prediction results using Decision Tree, the resulting graph is less stable as in Figure 5 (b) with blue lines. From Figure 5 (a) and (b), it can be concluded that the prediction using the ANN algorithm gets more stable and accurate results compared to using the Decision Tree algorithm.

5. Conclusion

In this study, to predict the tides at Tanjung Priok Port using Artificial Neural Network (ANN) and Decision Tree methods. Using the dataset from January 2018 - June 2018, both methods were successful in predicting the tidal height with a fairly low MSE value. However, the ANN method is slightly better because it produces a smaller MSE value than the Decision Tree method. Using each architecture model, the ANN model produces an MSE value of 0.003727983 and the Decision Tree method produces an MSE value of 0.009870259. If the dataset used has larger amount of data and the architecture model used is more complex, it might be able to increase the accuracy value of the prediction. It is hoped that the ideas and results of the research can be further developed to conduct other research using similar algorithms but in different areas with more variables used. Examples of research that can be done such as Tidal for renewable energy.

References

[1]. Reonaldho J, Saepudin D, Adytia D. Prediksi Gelombang Ekstrim Air Laut di Pelabuhan Tanjung Priok Menggunakan Algoritma ID3. Proceeding of Engineering. 2020;7(1):2700.

[2]. Sobh Elbisy M. PREDICTION OF DAILY TIDAL LEVELS ALONG THE CENTRAL COAST OF EASTERN RED SEA USING ARTIFICIAL NEURAL NETWORKS. International Journal of GEOMATE. 2020 Dec 1;19(76).

[3]. Kim S, Pan S, Mase H. Artificial neural network-based storm surge forecast model: Practical application to Sakai Minato, Japan. Applied Ocean Research. 2019 Oct;91:101871.

[4]. Tsakiri K, Marsellos A, Kapetanakis S. Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York. Water. 2018 Aug 29;10(9):1158.

[5]. Dastres R, Soori M. Artificial Neural Network Systems Modern CAD/CAM Systems View project Virtual Machining View project Artificial Neural Network Systems. 2021 Sep.

[6]. Bhar KK, Bakshi S. Application of artificial neural network for predicting water levels in Hooghly estuary, India. H2Open Journal. 2020 Oct 9;

[7]. Nunes I, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Helena L, Franco S. Artificial Neural Networks. Editorial: Cham Springer International Publishing; 2017.

[8]. Nikentari N, Ritha N, Haryadi T, Informatika J, Teknik F, Maritim Raja U, et al. Prediksi Pasang Surut Air Laut Menggunakan Jaringan Syaraf Tiruan Backpropagation . Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan. 2018;07(01):1–6.

[9]. Fauzi A, Pradipta IW. Research methods and data analysis techniques in education articles published by Indonesian biology educational journals. Jurnal Pendidikan Biologi Indonesia [Internet]. 2018 Jul 9;4(2). Available from: https://www.researchgate.net/publication/326286291_Research_methods_and_data_analysis_techniques_in_education_articles_published_by_Indonesian_biology_educational_journals

[10]. Ma'rifah H, Wibawa AP, Akbar MI. Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing. Sains, Aplikasi, Komputasi dan Teknologi Informasi. 2020 Apr 6;2(2):70.

[11]. Kaunang FJ. Analisis dan Perancangan Sistem Informasi Fasilitas Sekolah. e-Jurnal JUSITI (Jurnal Sistem Informasi dan Teknologi Informasi). 2018 Oct 15;7-2:124–30.