Journal of Intelligent Decision Support System (IDSS) 5 (2) (2022) 64-72

Published by: IOCSCIENCE

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www. idss.iocspublisher.org

Prediction of Extreme Sea Water Waves at Ancol Beach Using ID3 Algoritma Algorithm

Arum Budi Harti[1], Dionysius Hendard Christianto[2], Octavian Pearl[3], Renata Nabillah[4]

Sistem Informasi Kelautan, Universitas Pendidikan Indonesia Kampus Serang, Jl. Ciracas No.38, Serang, Kec. Serang, Kota Serang, Banten 42116,Indonesia

Article Info

Article history:

Received April 15, 2022
Revised May 17, 2022
Accepted June 23, 2022

Keywords:

Ocean Wave Prediction;
Decision Tree;
ID3

ABSTRACT

Ocean waves are natural events where water waves gradually move up and down. This regular rise and fall of water is one of the important aspects of transportation, predicting potential trade disasters and tsunamis in water areas. Know the data The future beyond the level of ocean waves can bring great benefits smoothly Transport and trade of territorial waters. Future data can be obtained from forecasts with certain algorithms. The ID3 algorithm is one of the most common learning algorithms. Used to create a decision tree or decision tree. The result of this analysis is a decision tree that can be used for classifying sea level using an accuracy of 88%.

Corresponding Author:

Arum Budi Harti,
Sistem Informasi Kelautan,
Universitas Pendidikan Indonesia,
Kampus Serang, Jl. Ciracas No.38, Serang, Kec. Serang, Kota Serang
Banten, Indonesia
E-mail: arum@upi.edu

1. Introduction

Tides are events that move the sea surface up and down periodically. This arises because of the gravitational force and the attractive force of objects in the sky, mainly caused by the sun, earth, and moon. The influence of other astronomical objects can be ignored in tidal events because they are smaller in size or are too far apart. Studies on tides are very important to be carried out for purposes in the fields of geology, port construction, environment, biology and agriculture, and for the development of tidal power plants [1]–[3]. Tourist destinations in Indonesia have been known since ancient times, especially beaches that have a high number of visitors because Indonesia is an archipelagic country. Ancol Beach is one of the tourist attractions located in North Jakarta that has developed. Information provided about tourism in Ancol Dreamland Park is still lacking, especially information about security in the coastal area. Advances in technology and communication (ICT) which are very fast provide convenience for human life, one of which is to communicate and speed up obtaining information that is not limited by time. One example of the application of communication and information technology is in the marine sector, especially related to tidal conditions. especially information regarding security in coastal areas. Advances in technology and communication (ICT) which are very fast provide convenience for human life, one of which is to communicate and speed up obtaining information that is not limited by time. One example of the application of communication and information technology is in the marine sector, especially related to tidal conditions. especially information regarding security in coastal areas. Advances in technology and communication (ICT) which are very fast provide convenience for human life, one of which is to communicate and speed up obtaining information that is not limited by time. One

example of the application of communication and information technology is in the marine sector, especially related to tidal conditions.

Prediction of extreme waves in the Ancol Beach area is needed to support tourism development. Ocean waves are one of the natural events that occur due to periodic increases and decreases in water waves. This increase and decrease in water is one of the supporting aspects of transportation and trade as well as to anticipate the potential for a tsunami disaster in the territorial waters. Knowing data related to the height of sea waves can have a positive impact on the smoothness of transportation and trade in water areas. In this case, an analysis of the anticipation of the potential for a tsunami disaster is very important to study so that the process of evacuating people living around the Ancol Beach area [4], [5]. Given the importance of data regarding wave heights, then data about the height of the waves in the future is needed. Future data can be obtained by making predictions. In this case, predictions are made using the decision tree method with the application of the ID3 algorithm. The ID3 algorithm applies a hierarchical structure for supervised learning. The steps in the decision tree method or decision tree start from the root node and the final stage is in the form of a leaf node which in its application is carried out recursively, where each branch of this method states a situation. Where these conditions must be met and at each end of the tree is a class that originates from a data. predictions are made using the decision tree method with the application of the ID3 algorithm. The ID3 algorithm applies a hierarchical structure for supervised learning. The steps in the decision tree method or decision tree start from the root node and the final stage is in the form of a leaf node which in its application is carried out recursively, where each branch of this method states a situation. Where these conditions must be met and at each end of the tree is a class that originates from a data. predictions are made using the decision tree method with the application of the ID3 algorithm. The ID3 algorithm applies a hierarchical structure for supervised learning. The steps in the decision tree method or decision tree start from the root node and the final stage is in the form of a leaf node which in its application is carried out recursively, where each branch of this method states a situation. Where these conditions must be met and at the end of the tree is a class that comes from a data[6]–[8].

Data miningis an activity carried out to obtain information including data collection, use of historical data for regularities and relationships in data sets that have large-scale values. The function of data mining itself is to determine the patterns thatmust be found in a data mining task. The beginning of the emergence of data mining is from the data explosion problem that has often been experienced in recent times. There have been many organizations that have been involved in data collection for years, starting from purchase data, sales data, service data, customer data, and others [9], [10].

Based on the research conducted by Susanto, et al, regarding the application of the decision tree method or decision tree using the ID3 algorithm, the majority of which are based on websites. With the process of uploading training data as well as testing data, the results will be directly displayed on the website page.

2. Method

This study discusses the application of the ID3 algorithm using the Ancol Beach area, North Jakarta as the research location. which uses two methods, namely primary data and secondary data. The primary data used is sea level data which is taken with a time span of every six months with a location on Ancol Beach, North Jakarta which is then processed to produce values regarding the accuracy of the application of the ID3 algorithm at that location. While secondary data is data that was first collected and previously data has been collected in such a way by people or agencies outside of the research or commonly referred to as data originating from other parties.[5].

In processing the data in this study, it is inseparable from the Microsoft Excel software which is a tool in processing the tidal data that has been obtained from the pushidrosal website. This study also applies a literature study method that uses a type of research approach, namely descriptive analysis whose purpose is to describe or explain something in accordance with reality [4]. The literature study method is used to find accurate data and meet the criteria for analyzing data needs. Then the instrument in this study was carried out by analyzing articles, scientific journals, electronic books, and so on to be used as supporting tools in this research.

In collecting data, the authors make observations. Observation is a data collection tool by observing and analyzing steps to obtain accurate information and meet the criteria for analyzing data needs. Then the data analysis process used in this study used a descriptive analysis system. Descriptive approach is an approach aimed at analyzing the data that has been obtained by describing and explaining the data that has previously been obtained from sources
data so that an overview is obtained.[6] The description obtained regarding the use of the ID3 algorithm in predicting extreme waves in the Ancol Beach area, North Jakarta.

3. Result and Discussion

3.1 Intem Description

Scenes related to extreme wave heights with the choice of location on Ancol Beach will be modeled by applying the ID3 algorithm. In this study, using data on the height of sea waves at Ancol Beach that occurred within the period of January 1, 2021 - July 1, 2022, as follows:

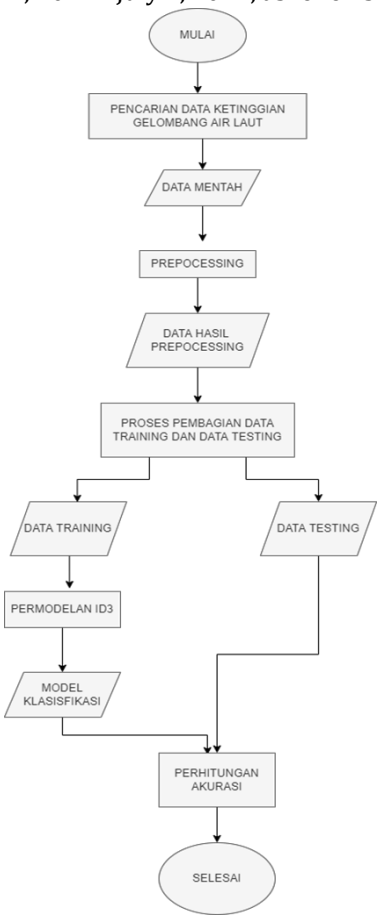

Figure 1. Flowchart overview of the system

Information :

1. The data used to predict the extreme wave height of sea water is data on the height of sea water waves at Ancol Port
2. Preprocessing the data and then it becomes the preprocessed data.
3. Calculating Entrophy and Gain value.
4. Sharing data in the form of training data and testing data.
5. Modeling the training data using the ID3 algorithm.
6. Run the testing process using the ID3 algorithm.
7. Testing accuracy.

3.2 Preprocessing Data

From the dataset that is already owned, then preprocessing will be carried out, namely the steps taken to create raw data as higher quality data. The following is a flowchart based on preprocessing data:

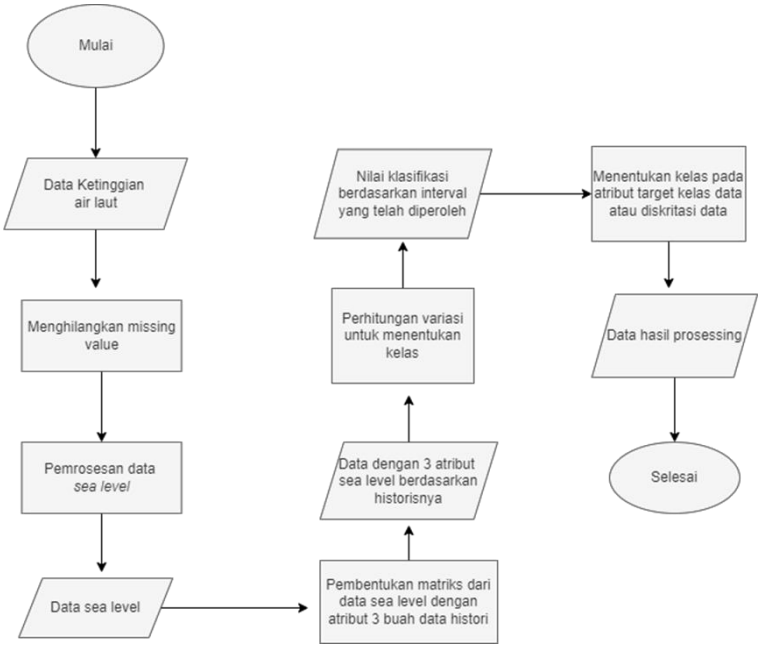

Figure 2. Flowchart of Preprocessing Data

In terms of data preprocessing, several things can be explained as follows:

1. Elimination of missing values in the newly obtained raw data, namely putting the average data or data deletion will be carried out.
2. Doing calculations regarding the sea level value will be used as a class target in the data to be used.
3. Formation of a matrix using the three previous historical data as attributes and target/target data as classes.
4. Calculation of variance in order to discriminate the data using a method of calculating the mean minus or plus deviation in order to select the class in the targetattribute, the classification of values in attributes is divided into five classes using the following intervals:
   a. Value > average + two stdev : extreme height
   b. Average + stdev > Value > average + two stdev : high c. Average – stdev > Value > average + stdev : normal
   c. Average – two stdev> Value> average – stdev : lo e. Value <average – two stdev: extreme low
5. Determine the class in the attribute target to facilitate modeling using ID3. Class determination uses the variance method to classify data as several classes.

3.3 ID3 Algoritma

ID3 algorithm is an algorithm used to form a decision tree or decision tree. Decision trees use a hierarchical structure for supervised learning. The ID3 algorithm is used to classify or classify data. This decision tree is commonly used in decision making and forecasting or forecasting. The following is a flowchart according to the ID3 Algorithm.

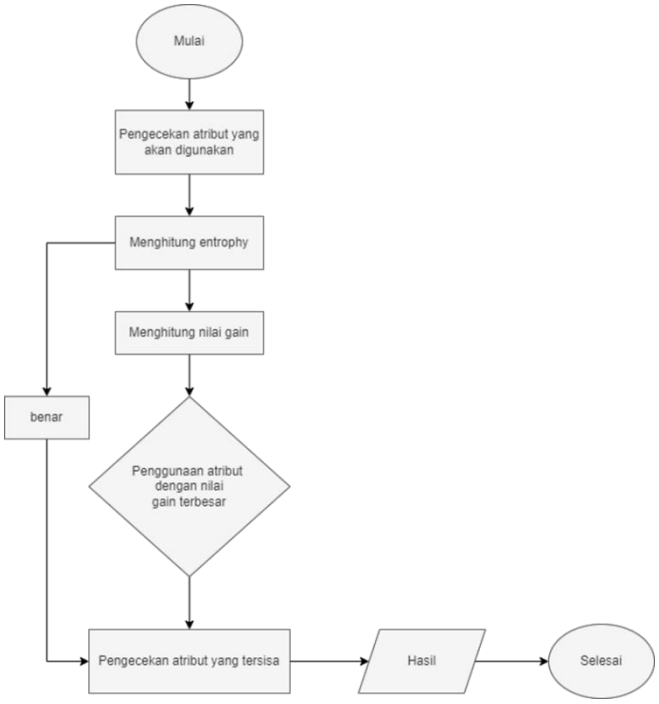

Figure 3. ID3 Algorithm Flowchart

The ID3 algorithm performs several steps. The stage or step of the ID3 algorithm calculates the entropy and gain value calculations that produce predictive data. The steps can be explained as follows:

1. The tree is started by using an image of a node that represents a sample of training data using the root node.
2. If all samples are in the same class, this node will become a leaf and will be marked as class. Otherwise, information gain is used to select the best attribute when separating the sample data into individual classes. This algorithm uses the concept of entropy.
3. Branches are created for each attribute value and the sample data is repartitioned. In this case, the purpose of using the concept of entrophythat isto measure the informative level of a node
4. This algorithm uses a recursive process to build a decision tree for each data partition. If an attribute is used on a node, it will not be reused on its child nodes.
5. This process will stop if the following conditions are met:

a. All samples on a node are in one class.
b. There are no other attributes that can be used to divide the sample further.

6. In this case, the majority vote prevails. This means turning nodes into leaves and marking them with the class with the most votes.

3.4 Data Set

The data set used is sea level data which is taken every hour through the pushidrosal website.

TABLE 1
SAMPLE OF SEA LEVEL DATA AT ANCOL PORT

DataTime (UTC)	Bat (V)	Pr2(m)	prs(m)	rad(m)
01/01/2022 00.01		0.94	1,203	7,441
01/01/2022 00.02		0.93	1,203	7,444
01/01/2022 00.03	12	0.93	1,205	7,443
01/01/2022 00.04		0.93	1,205	7,444
01/01/2022 00.05		0.93	1.2	7,439

The results of the calculation of preprocessing data whose purpose is to find the value of the sea level.

TABLE 2
SAMPLE OF SEA LEVEL DATA AT ANCOL PNATAI BEACH

DataTime (UTC)	Sea Level	Harmonic	Anomaly
2022/1/3 4.15.0	1.011	0.314	0.696
2022/1/3 5.15.0	0.984	0.334	0.650
2022/1/3 6.15.0	0.957	0.306	0.650
2022/1/3 7.15.0	0.929	0.248	0.680
2022/1/3 8.15.0	0.861	0.176	0.685

3.5 Results of Class Determination Based on Attributes

In table 3 below, the data obtained after preprocessing is given a class on the attribute Sl(t) according to the interval that has been obtained using the variance method. Threshold consists of 5 categories, namely: extreme low (1), low (2), normal (3), high (4), and extreme height (5).

TABLE 3
RESULT OF CLASS DETERMINANTION BASED ATTRIBUTES

Time	SI (t-3 hours)	SI (t-2h)	SI (t-1h)	SI(t)	Category
2022/1/3 7.15.0	1.011	0.984	0.957	0.929	4
2022/1/3 8.15.0	0.984	0.957	0.929	0.861	3
2022/1/3 9.15.0	0.957	0.929	0.861	0.835	3
2022/1/3 10.15.0	0.929	0.861	0.835	0.661	3
2022/1/3 11.15.0	0.861	0.835	0.661	0.551	3

The following are some selected samples in the form of training data and testing data:

TABLE 4
TRAINING DATA SAMPLE

Time	SI (t-3h)	SI (t-2h)	SI (t-1h)	SI(t)	Category
2022/1/3 19.15.0	0.568	0.657	0.655	0.747	3
2022/1/3 20.15.0	0.657	0.655	0.747	0.811	3
2022/1/3 21.15.0	0.655	0.747	0.811	0.927	4
2022/1/3 22.15.0	0.747	0.811	0.927	1.034	4
2022/1/3 23.15.0	0.811	0.927	1.034	0.817	3

TABLE 5
TESTING DATA SAMPLE

Time	SI (t-3h)	SI (t-2h)	SI (t-1h)	SI(t)	Category
2022/4/20 20.15.0	0.75	0.65	0.59	0.53	3
2022/4/20 21.15.0	0.65	0.59	0.53	0.47	3
2022/4/20 22.15.0	0.59	0.53	0.47	0.37	2
2022/4/20 23.15.0	0.53	0.47	0.37	0.28	2
2022/4/21 0.15.0	0.47	0.37	0.28	0.16	2

3.6 Prediction Result of Sea Level with ID3

Based on the results of the tests that have been carried out, the prediction results obtained using the ID3 algorithm with several samples are as follows:

TABLE 6
SAMPLE OF ID3 ALGORITHM PREDICTION RESULT

Time	sl (t-3hours)	sl(t-2hours)	sl (t-1hour)	sl(t)	Category	Prediction
2022/1/3 7.15.0	1.011	0.98392	0.95685	0.929	4	3
2022/1/3 8.15.0	0.98392	0.95685	0.92977	0.861	3	3
2022/1/3 9.15.0	0.95685	0.92977	0.861	0.835	3	3
2022/1/3 10.15.0	0.92977	0.861	0.835	0.661	3	3
2022/1/3 11.15.0	0.861	0.835	0.661	0.551	3	3
2022/1/3 12.15.0	0.835	0.661	0.551	0.482	3	3
2022/1/3 13.15.0	0.661	0.551	0.482	0.44	3	3
2022/1/3 14.15.0	0.551	0.482	0.44	0.449	3	3
2022/1/3 15.15.0	0.482	0.44	0.449	0.472	3	3
2022/1/3 16.15.0	0.44	0.449	0.472	0.568	3	3
2022/1/3 17.15.0	0.449	0.472	0.568	0.657	3	3
022/1/3 18.15.0	0.472	0.568	0.657	0.655	3	3
2022/1/3 19.15.0	68	0.657	0.655	0.747	3	3
2022/1/3 20.15.0	0.657	0.655	0.747	0.811	3	3
2022/1/3 21.15.0	0.655	0.747	0.811	0.927	4	4
2022/1/3 22.15.0	0.747	0.811	0.927	1.034	4	4
2022/1/3 23.15.0	0.811	0.927	1.034	0.817	3	3
2022/1/4 0.15.0	0.927	1.034	0.817	0.913	4	4
2022/1/4 1.15.0	1.034		0.913	0.916	4	4
2022/1/4 2.15.0	0.817	0.913	0.916	0.882	4	4
2022/1/4 3.15.0	0.913	0.916	0.882	0.975	4	4
2022/1/4 4.15.0	0.916	0.882	0.975	1.072	4	4
2022/1/4 5.15.0	0.882	0.975	1.072	1.052	4	
2022/1/4 6.15.0	0.975	1.072	1.052	1.111	4	4
2022/1/4 7.15.0	1.072	1.052	1.111	0.944	4	4
2022/1/4 8.15.0	1.052	1.111	0.944	0.95	4	4
2022/1/4 9.15.0	1.111	0.944	0.95	0.928	4	4
2022/1/4 10.15.0	0.944	0.95	0.928	0.755	3	3
2022/1/4 11.15.0	0.95	0.928	0.755	0.68	3	3
2022/1/4 12.15.0	0.928	0.755	0.68	0.59	3	3
2022/1/4 13.15.0	0.755	0.68	0.59	0.5355	3	3
2022/1/4 14.15.0	0.68	0.59	0.5355	0.52	3	3
2022/1/4 15.15.0	0.59	0.5355	0.52	0.579	3	3
2022/1/4 16.15.0	0.5355	0.52	0.579	0.598	3	3
2022/1/4 17.15.0	0.52	0.579	0.598	0.637	3	3
2022/1/4 18.15.0	0.579	0.598	0.637	0.626	3	3
2022/1/4 19.15.0	0.598	0.637	0.626	0.7	3	3
2022/1/4 20.15.0	0.637	0.626	0.7	0.671	3	3
2022/1/4 21.15.0	0.626	0.7	0.671	0.758	3	3
2022/1/4 22.15.0	0.7	0.671	0.758	0.79	3	3
2022/1/4 23.15.0	0.671	0.758	0.79	0.858	3	3
2022/1/5 0.15.0	0.758	0.79	0.858	0.827	3	3
2022/1/5 1.15.0	0.79	0.858	0.827	0.9	4	4
2022/1/5 2.15.0	0.858	0.827	0.9	0.881	4	4

2022/1/5 3.15.0	0.827	0.9	0.881	0.863	3	3
2022/1/5 4.15.0	0.9	0.881	0.863	0.873	4	4
2022/1/5 5.15.0	0.881	0.863	0.873	0.868	3	3
2022/1/5 6.15.0	0.863	0.873	0.86814	0.867	3	3
2022/1/5 8.15.0	0.873	0.86814	0.867	0.76	3	3
2022/1/5 9.15.0	0.86814	0.867	0.76	0.719	3	3
2022/1/5 10.15.0	0.867	0.76	0.719	0.667	3	3
2022/1/5 11.15.0	0.76	0.719	0.667	0.587	3	3
2022/1/5 12.15.0	0.719	0.667	0.587	0.52	3	3
2022/1/5 13.15.0	0.667	0.587	0.52	0.485	3	3
2022/1/5 14.15.0	0.587	0.52	0.485	0.478	3	3
2022/1/5 15.15.0	0.52	0.485	0.478	0.404	3	3
2022/1/5 16.15.0	0.485	0.478	0.404	0.397	3	3
2022/1/5 17.15.0	0.478	0.404	0.397	0.495	3	3
2022/1/5 18.15.0	0.404	0.397	0.495	0.579	3	3
2022/1/5 19.15.0	0.397	0.495	0.579	0.622	3	3
2022/1/5 20.15.0	0.495	0.579	0.622	0.771	3	3

From the results of the image above, it can be seen that the predictive value used as an adjustment to the category value that has been obtained from the thresholding attribute sl(t), as for dividing the interval on that value as follows:

1. Extreme lowliness : Value < 0.130 : 1
2. Low : 0.130 < Value < 0.377 : 2
3. Normal : 0.377 < Value < 0.870 : 3
4. Tall : 0.870 < Value < 1.116 : 4
5. Extreme altitude : Value > 1.116 : 5

The accuracy value obtained through the prediction of sea level classification above is 88.76%. This shows that the sea level prediction process using the ID3 algorithm is suitable for use as unsupervised data into a model that can be used to predict the sea water.

4 Conclusion

In this study, it can be seen that the level of accuracy regarding the prediction of sea level at Ancol Beach using the ID3 algorithm has a high accuracy value. It can be concluded that predictions regarding sea level at Ancol Beach using the ID3 algorithm can be used as material for validation in the field because the accuracy value is high. For the prediction of Ancol sea level using the ID3 algorithm, the results showed a high accuracy value.

References

[1] D. Fitriana, N. Oktaviani, and I. U. Khasanah, "Analisa harmonik pasang surut dengan metode admiralty pada stasiun berjarak kurang dari 50 km," *J. Meteorol. Klimatologi dan Geofis.*, vol. 6, no. 1, pp. 38–48, 2019.

[2] E. Supriyadi and W. S. Siswanto, "Analisis Pasang Surut di Perairan Pameungpeuk, Belitung, dan Sarmi Berdasarkan Metode Admiralty," *J. Meteorol. dan Geofis.*, vol. 19, no. 1, pp. 29–38, 2019.

[3] D. Sufianto, "Pasang surut otonomi daerah di Indonesia," *J. Acad. Praja*, vol. 3, no. 02, pp. 271–288, 2020.

[4] J. Risandi and D. Candra, "Prediksi Gelombang Ekstrim di Kepulauan Seribu untuk Aplikasi Kelautan dan Perikanan," *J. Ris. Jakarta*, vol. 14, no. 2, pp. 51–56, 2021.

[5] H. R. Maulana, "Analisis Karakteristik Gelombang Laut Dan Prediksi Gelombang Ekstrim Dengan Periode Ulang Terkait Keselamataan Pelayaran Di Perairan Utara Jakarta." Universitas

Brawijaya, 2018.
[6] J. V. Reonaldho, D. Saepudin, and D. Adytia, "Prediksi Gelombang Ekstrim Air Laut Di Pelabuhan Tanjung Priok Menggunakan Algoritma Id3," *eProceedings Eng.*, vol. 7, no. 1, 2020.
[7] S. U. Anggono, D. Manongga, and A. Iriani, "Evaluasi Pelayanan Terpadu Satu Pintu (PTSP) Menggunakan Decision Tree," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 4, pp. 1208–1216, 2021.
[8] M. Jeihouni, A. Toomanian, and A. Mansourian, "Decision tree-based data mining and rule induction for identifying high quality groundwater zones to water supply management: a novel hybrid use of data mining and GIS," *Water Resour. Manag.*, vol. 34, no. 1, pp. 139–154, 2020.
[9] H. Rodrigo, E. W. Beukes, G. Andersson, and V. Manchaiah, "Exploratory data mining techniques (decision tree models) for examining the impact of internet-based cognitive behavioral therapy for tinnitus: Machine learning approach," *J. Med. Internet Res.*, vol. 23, no. 11, p. e28999, 2021.
[10] H. H. Patel and P. Prajapati, "Study and analysis of decision tree based classification algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, 2018.