DATA EXPLORATION OF MARINE CULTIVATION TYPES USING CLUSTER ANALYSIS WITH COMPLETE-LINKAGE METHOD

Vitri Aprilla Handayani[1*], Eko Sulistyono[1], Luki Hernando[2], A'rsyil Majiid[1], Hery Sunarsono[3]

[1]Mathematics, Fakultas Teknologi Informasi, Institut Teknologi Batam
[2] Computer Engineering, Fakultas Teknologi Informasi, Institut Teknologi Batam
[3] Rekayasa Engineering, Fakultas Teknologi Informasi, Institut Teknologi Batam
email: vitri@iteba.ac.id

Abstract: that the number of catches obtained by fishermen is more optimal. With secondary data based on the type of cultivation developed by fishing communities, the authors intend to explore this data so that they can examine more deeply the most effective types of cultivation using analysis of variance to find out the differences between each type of cultivation and cluster analysis using the complete linkage method. to find out the grouping of marine culture production based on the types developed by fishing communities in Indonesia. The correlation between the observed variables is 0.008, so it can be said that there is no relationship between the observed variables. In addition, the sig. and t-test. sig value is obtained. $> \alpha$ (5% = 0.05). This means that there is a significant average difference in fish production results based on the type of cultivation developed by fishing communities. With the grouping of types of marine cultivation developed by the Wesleyan community using the complete-linkage method, they are divided into 3 groups based on the degree of similarity in the production results obtained. Cluster 1 consists of Floating Nets, and Pools; Cluster 2 consists of other seas, seaweed, cages, and fishing nets; and Cluster 3 consists of Fresh Floating Nets, Minapadi, and Ponds.

Keywords: Complete-Linkage; Fish Production; Mariculture; Multivariate Analysis; T-test;

Abstrak: Berdasarkan penelitian sebelumnya pada analisis variansi untuk menguji perbedaan rata-rata pada masing-masing perlauan peletakan sudut jarring agar jumlah tangkapan yang diperoleh nelayan lebih optimal. Dengan adanya data sekunder berdasarkan jenis budidaya yang dikembangkan masyarakat nelayan, penulis bermaksud untuk melakukan eksplorasi pada data tersebut agar dapat mengkaji lebih dalam terhadap jenis budidaya yang paling efektif dengan analisis variansi untuk mengetahu perbedaan dari masing-masing jenis budidaya serta analisis cluster dengan metode complate linkage untuk mengetahui adanya pengelompokan hasil produksi bududaya laut berdasarkan jenis yang dikembangkan oleh masyarakat nelayan di Indonesia. Korelasi antar variabel yang diamati sebesar 0.008, maka dapat dikatakan bahwa tidak ada hubungan antara variabel yang diamati. Selain itu, hasil uji sig. dan uji t. diperoleh nilai sig. $> \alpha$ (5% = 0.05). Artinya, terdapat perbedaan rata-rata secara signifikan hasil produksi ikan berdasarkan jenis budidaya yang dikembangkan oleh masyarakat nelayan. Dengan adanya pengelompokan jenis budidaya laut yang dikembangkan oleh masyarakat neleyan dengan metode complete-linkage terbagi atas 3 kelompok berdasarkan tingkat kemiripan hasil produksi yang diperoleh. Cluster 1 terdiri dari Jaring Apung, Kolam; Cluster 2 terdiri dari Laut lainnya, Rumput laut, Keramba, Jaring Tancap; dan Cluster 3 terdiri dari Jaring Apung Tawar, Minapadi, Tambak.

Kata kunci: Analisis Multivariate; Budidaya Laut; Complate-Linkage; Produksi Ikan; Uji-t;

INTRODUCTION

Indonesia is located between the Indian Ocean to the west and the Pacific Ocean to the east. Indonesia's territorial waters are vast, stretching from the Indian Ocean in the west to the South China Sea in the north. Indonesian waters are included in the Archipelago region, which consists of thousands of large and small islands. This geographical condition makes Indonesia has a very long coastline, reaching more than 54,000 kilometers [1]. These coastlines form natural boundaries and provide access to the oceans which are rich in marine resources. The territorial waters around Indonesia have a significant impact on the economic and social life of the Indonesian people. The oceans provide natural resources such as fish, natural gas, petroleum and marine minerals, which contribute to the fishing, energy and other industrial sectors. [2].

The Indonesian government has great concern for the management of the sea and marine resources. Through implemented policies and regulations, Indonesia strives to ensure sustainable management, protection of the marine environment, and responsible use of marine resources in its territory. The amount of fisheries production in Indonesia refers to the total amount of catches and fishery cultivation products produced in Indonesia within a certain time period [3]. This includes various types of fish, shrimp, squid and other fishery products obtained from sea waters, rivers and aquaculture systems. Aquaculture systems refer to the methods or techniques used to manage and commercially produce fish in a controlled culture environment. The main objectives of aquaculture systems are to increase fish production, maintain water quality, and optimize resource use. Each aquaculture system has different characteristics, advantages and challenges [4]. Selection of the appropriate cultivation system must consider the species of fish to be cultivated, the available cultivation environment, the availability of resources, and the desired production goals. For this reason, an analysis will be carried out whether there are differences in marine cultivation systems, such as sea floating net cultivation, fresh floating nets, cages, ponds, ponds, and many other types of marine cultivation. [5].

In previous research [6], an analysis of the use of net angles was carried out to optimize fish catches by fishermen in the sea, in the treatment of net angle depth the optimal results were obtained at a depth of 0.5 m above sea level. Based on the description above that Indonesia has a very large water area with various types of cultivation developed by fishing communities, researchers will conduct data analysis exploration on fisheries production data to identify types of marine cultivation.

In [7] [8], in research conducted to determine the effect of fishery production on domestic product fishery sector. The results obtained indicate that the value of aquaculture production significantly influences gross domestic product, especially in the fisheries sector in Indonesia. Indonesia has rich fishery resources and is one of the largest fish producing countries in the world. Indonesian fishery production includes various types of fish, shrimp, shellfish, squid, and other fishery products [9]. Vast sea waters and has great potential in fish production. Sea fish such as tuna, mackerel, grouper, and milkfish are some

of the species that are often produced. In addition, the cultivation of freshwater fish such as tilapia, catfish, and catfish is also important in fisheries production in Indonesia.

There are many factors that affect the amount of fishery production, as it is done [10], Data analysis was carried out using statistical tests that the number of fishing boats and the number of fishermen has a significant effect on the results of fishery production in Indonesia [11] [12]. [13] explained that there are key performance indicators for capture fisheries that have been established over the last few years with a focus on performance measures on six main performances including GDP growth, Fishermen Exchange Rates, production volume, production value, average fishery household income and average fisherman income. Fishery production has a very important contribution to the Indonesian economy. In addition to providing food resources and industrial raw materials, fisheries are also able to provide employment and income for fishing communities around the coast.

Mariculture is known as marine aquaculture, referring to human activities in cultivating living organisms in marine waters for commercial purposes. This activity includes the management of various types of marine animals and plants, including fish, shellfish, oysters, shrimp, seaweed, and others [14]. [15] Mariculture has management of the artificial environment in marine waters which is carried out as monitoring of optimal conditions for the growth and reproduction of organisms cultivated by fishing communities. Some of the common methods used in mariculture include:

Floating net cages, floating net cages are floating frames made of materials such as nets or gauze used to cultivate fish, shrimp or shellfish. Floating net cages: This method is similar to floating net cages, but these cages are placed in shallower waters connected to the seabed or other retaining structures [16]. Ponds, ponds are waterlogged areas bounded by embankments or ditches and used to cultivate marine organisms such as shrimp, shellfish and fish. Ponds can be located on the coast or on land near the coast that has an adequate seawater supply. Seaweed cultivation: Seaweed cultivation involves growing and managing seaweed in shallow marine waters. This seaweed is used in a variety of industries, including food, cosmetics, and pharmaceuticals.

Aquaculture requires good monitoring and management to minimize negative impacts on the environment [17]. The principles of sustainability and responsible aquaculture practices are often applied in modern marine aquaculture to maintain the balance of the ecosystem and maintain production sustainability.

Based on previous research [6], The research was carried out using variance analysis to test the average differences in each method of net angle placement so that the number of catches obtained by fishermen was more optimal. With the existence of secondary data related to fisheries production data in Indonesia based on the types of cultivation developed by fishing communities, the author intends to carry out exploration of this data in order to study more deeply the types of cultivation that are most effective.

METHOD

Research Design

The following research design was designed for this research activity:
First, identify the problem that will be studied in this research by determining the topic or problem you want to research. Based on research conducted previously on [6] in an effort to optimize the type of net angle used as a medium for catching fish in the sea used by fishing communities. Second, based on this and data sources related to the amount of fisheries production in Indonesia based on the type of cultivation used, the research that will be carried out will aim to find out the differences between each type of cultivation and the optimal grouping. Third, the research method used is data exploration through analysis of variance and multivariate analysis, especially cluster analysis with the complate linkage method to determine whether there are groupings that occur in each type of cultivation developed by fishing communities in Indonesia

Data Collection

Secondary data refers to the type of data collected by other parties or obtained from pre-existing sources. This data is not collected directly by researchers or data users to be used by other individuals or organizations for analysis, research or other purposes. Secondary data can be various types of information, including statistical data, research reports, surveys, business documents, government data, social media data, and more. This data can be obtained from sources such as government agencies, non-government organizations, private companies, research institutions, or public databases.

The data to be used in this study uses secondary data in the form of fish production data in Indonesia in 2017-2021. (source: https://statistik.kkp.go.id/).

Research Flow

The research methodology that will be carried out includes: Identifying the needs of fishing communities for the number of catches obtained in a certain season based on the type of marine cultivation used by fishing communities; Identify the types of marine cultivation developed by fishing communities; Secondary data collection, in this case the researcher uses existing data in the form of data on the production of cultivated products developed by fishing communities based on the type of cultivation; Simulation analysis that will be used to find out the differences in each type of cultivation developed by analysis of variance; Data analysis on the number of catches for the type of cultivation used uses cluster analysis with the compile linkage method to determine whether there is a grouping of the amount of fishery production in Indonesia based on the type of cultivation developed.

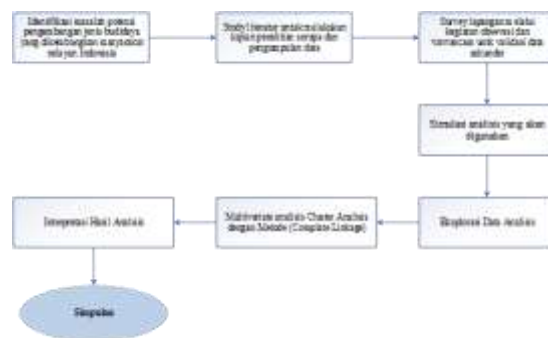The following is a research flowchart:

Image 1. Flow Chart

Exploration of Data Analysis

Data exploration is a process for understanding and analyzing data in order to find patterns, relationships or information contained therein [18]. In [19], Data exploration is the first step in data analysis that helps us understand the data better before applying the analytical techniques used next. [20] Several techniques are commonly used in exploratory data analysis:

1. Descriptive Statistics
   This technique involves applying summary measures such as mean (average), median (middle value), mode (most frequently occurring value), and measures of variability such as range and standard deviation. Descriptive statistics help provide an overview of the characteristics of the data.
2. Data Visualization
   This technique focuses more on creating graphs or data visualization which is an effective way to understand patterns and relationships in data. Examples of tools that are often used are bar charts, histograms, scatter plots, and pie charts. Data visualization helps identify trends, outliers, correlations and hidden patterns.
3. Correlation Analysis
   This technique identifies the relationship between two or more variables in the data. Correlation can help reveal a positive, negative, or no relationship between these variables.
4. Cluster Analysis

This technique helps identify natural groups or patterns in data. Clustering can assist in customer segmentation, geographic mapping, or grouping entities based on common features.

RESULT AND DISCUSSION

Data on the volume of fishery production based on the type of cultivation developed by fishing communities in Indonesia is carried out by exploratory data analysis using statistical software to view data descriptions. The results of the data description can be seen in Table 1.

Table 1. Descriptive Statistics

Variable	Mean	SE Mean	StDev	Variance
Marine Floating Net Cultivation	31027	11600	25939	6,73E+08
Other Marine Cultivation	50703	4179	9345	87328264
Seaweed Cultivation	8635439	419059	937044	8,78E+11
Fresh Floating Nets	419098	19793	44259	1,96E+09
Bargaining Nets	33519	6253	13982	1,95E+08
Karamba	218851	11851	26499	7,02E+08
Pond	2676131	73727	164858	2,72E+10
Mina-padi (Rice Fields)	223572	38534	86165	7,42E+09
Ponds	3056195	157464	352101	1,24E+11

Based on the results of the descriptive analysis in table 1, it can be seen that the largest standard deviation (StDev) is in the type of marine floating net and pond aquaculture, namely faults 937,044 and 352,101. This means that the two types of aquaculture production volume of fishery

with floating net and pond aquaculture re more heterogeneous than other types of marine aquaculture. This has implications with information on significant results.
To determine the distribution of normally distributed data, it can be seen from the probability plot in Image 2.
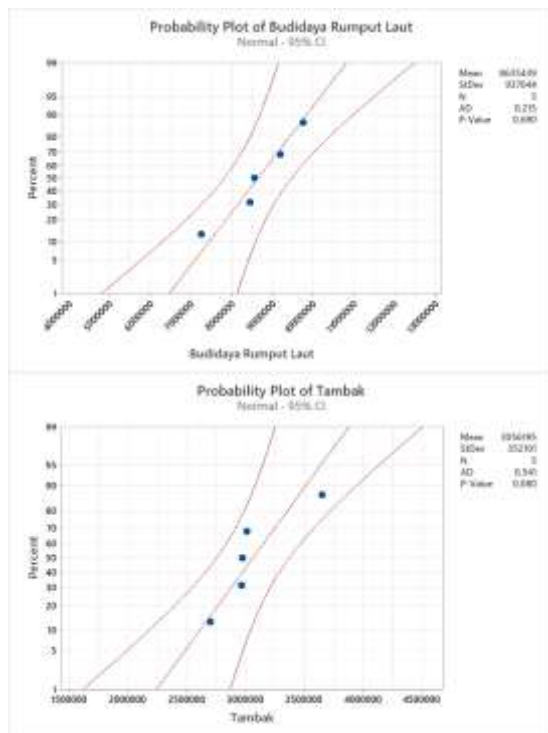

Image 2. Probability Plot

Based on Figure 2. The results of the probability plot on floating net and pond cultivation types show p-values of 0.69 and 0.08 > α (5% = 0.05). This means that the data is normally distributed with an error tolerance of 5%.

Table 2. Correlation Value of Cultivation Type and Number of Catches

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	Bidang & Jumlah	45	,008	,959

In Table 2, the results of the correlation analysis of the type of cultivation developed with the number of fish catches obtained by the fishing community are obtained. It can be seen that the correlation value is 0.008 with the sig. For 0.959 > α (5% = 0.05), it can be said that there is no relationship between the variables observed for the type of cultivation.

Table 3. Paired Samples Test

		Paired Differences					
		95% Confidence Interval of the Difference				d f	Sig. (2-tailed)
		Lower	Upper	t			
Pair 1	Field - Sum	-252 784 8	-882 039	-4,1 76		4 4	,000

Before carrying out an analysis of the test results in table 3, it is necessary to first formulate a hypothesis to be tested. The following is the hypothesis formulation:
H_0 : There is no significant difference in average between the number of fish caught based on the type of cultivation used by fishing communities
H_1 : There is a significant difference in average between the number of fish caught based on the type of cultivation used by fishing communities.

In [21], hypothesis is the sig value. < (α = 5% = 0.05) then reject, and vice versa.
Based on the output results in table 3, the sig value is obtained. equal to 0.000 < (α = 5% = 0.05), then reject This means that there is a significant difference in average between the number of fish caught based on the type of cultivation used by fishing communities.

Apart from comparing the sig. with a probability of α = 5%, we can compare the t-count with the basis for decision making if the value |t-count| > t-table then reject Based on table 3, the t-count value is 4,176 > t-count (db:9; α = 5%) is 1,833, so reject This means that there is a significant difference in average between the amount of fish caught based on the type of marine cultivation used by fishing communities. This is directly proportional to the results of hypothesis testing by comparing sig values.

Table 4. ANOVA Test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.259E+14	8	4.074E+13	352,504	,000
Within Groups	4.161E+12	36	1.156E+11		
Sum	3.301E+14	44			

To support the results of the comparison of sig values. and t-count, an ANOVA test was carried out as shown in table 4. The results obtained show the sig. of 0.000. according to the previous hypothesis formulation, the basis for decision making in the anova test is if the sig. < (α = 5% = 0.05) then reject H_0. Because the value of sig. < 0.05 in table 4, then reject H_0. This means that there is a significant difference in the amount of fish caught depending on the type of cultivation used by fishermen.

Based on several statistical tests carried out to determine the differences in each type of mariculture developed by fishing communities. Next, a cluster analysis was carried out using the compile-linkage method to determine the grouping of types of cultivation based on the level of similarity in the fish production results obtained. The following are the results of cluster analysis obtained using statistical software.

After calculating the distance which is used as a measure of closeness or similarity, the clustering process is then carried out using the complate-linkage method. Below are the results of the clustering carried out.

Table 5. Algori Of Clustering

Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new
8	95,8545	0,08291	4	8	4	2
7	93,0028	0,13994	2	3	2	2
6	85,4237	0,29153	2	6	2	3
5	78,2508	0,43498	1	7	1	2
4	73,1381	0,53724	4	9	4	3
3	67,7715	0,64457	2	5	2	4
2	34,6981	1,30604	1	2	1	6
1	2,8268	1,94346	1	4	1	9

In table 5 are the calculation results of the new cluster that was formed after merging several previous objects that were most similar. The clustering process can also be illustrated in the form of a dendogram. The following is a dendrogram of grouping starter packs using the complete-linkage method using statistical software, which can be seen in the image 3.
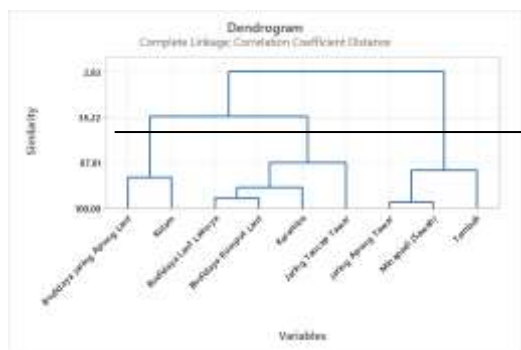
Image 3. Dendogram

Cluster analysis only shows cluster members for a certain number of clusters and does not determine the number of clusters formed. In this research, the number of membership clusters used was 3 clusters. Based on Figure 3. Members of each cluster include:

Table 6. Grouping of Cultivation Types

Cluster	Membership
Cluster 1	Jaring Apung, Kolam
Cluster 2	Laut lainnya, Rumput laut, Keramba, Jaring Tancap
Cluster 3	Jaring Apung Tawar, Minapadi, Tambak

Mariculture developed by the fishing community is divided into 3 groups based on the level of similarity of the production results obtained.

CONCLUSION

Based on the results of tests and analyzes carried out to determine differences in the average number of fish catches based on the type of cultivation developed by fishing communities, it can be concluded that the correlation between the observed variables is 0.008. with sig value. For 0.959 > α (5% = 0.05), it can be said that there is no relationship between the variables observed for the type of cultivation. Based on the sig test results. and t test. obtained sig value. > α (5% = 0.05). This means that there is a significant average difference in fish production results based on the type of cultivation developed by fishing communities. The grouping of types of marine cultivation developed by the neleyan community using the complete-linkage method is divided into 3 groups based on the level of similarity of the production results obtained. Cluster 1 consists of floating nets, ponds; Cluster 2 consists of other seas, seaweed, cages, immersion nets; and Cluster 3 consists of Tawar Floating Nets, Minapadi, Tambak.

BIBLIOGRAPHY

[1]   A. Z. Arfianto *et al.*, "Perangkat Informasi Dini Batas Wilayah Perairan Indonesia Untuk Nelayan Tradisional Berbasis Arduino Dan Modul Gps Neo-6M," *Joutica*, vol. 3, no. 2, pp. 163–167, 2018.

[2]   N. Hanif, A. Murni, C. Tanaka, and J. Tanaka, "Marine natural products from Indonesian waters," *Mar. Drugs*, vol. 17, no. 6, 2019, doi: 10.3390/md17060364.

[3]   I. N. Radiarta and E. Erlania,

"Indeks Kualitas Air Dan Sebaran Nutrien Sekitar Budidaya Laut Terintegrasi Di Perairan Teluk Ekas, Nusa Tenggara Barat: Aspek Penting Budidaya Rumput Laut," *J. Ris. Akuakultur*, vol. 10, no. 1, p. 141, 2015, doi: 10.15578/jra.10.1.2015.141-152.

[4]   D. Puspitawati, T. M. Putra, and R. V. Wardana, "Reformulasi Pengaturan Penambangan Bawah Laut di Wilayah Perairan Indonesia," *J. Magister Huk. Udayana*, vol. 10, no. 4, pp. 716–739, 2021, doi: 10.24843/JMHU.2021.v10.i04.p05 .

[5]   M. Yusuf and J. P. Soedarto, "Analisis Kesesuaian Lokasi Untuk Budidaya Laut Berkelanjutan di Kawasan Taman Nasional Karimunjawa," *J. Ilmu Kelaut.*, vol. 18, no. 1, pp. 20–29, 2013.

[6]   H. Sunarsono, S. Rahmiati, and V. A. Handayani, "Analisa Sudut Jaring untuk Mengoptimalkan Hasil Tangkapan Ikan pada Keramba Tancap Nelayan Tradisional," vol. 20, no. 1, pp. 248–254, 2022.

[7]   M. Zulkarnain, P. Purwanti, and E. Indrayani, "Analysis of Aquaculture Production Value Effect To Gross Domestic Product of Fisheries Sector in," *J. ECSOFiM*, vol. 1, no. 1, pp. 52–68, 2013.

[8]   R. Kumar, R. R. Kumar, P. J. Stauvermann, and P. Arora, "Effect of Fisheries Subsidies Negotiations on Fish Production and Interest Rate," *J. Risk Financ. Manag.*, vol. 13, no. 12, p. 297, 2020, doi: 10.3390/jrfm13120297.

[9]   2006 Setyorini, "Penyuluhan Perikanan dan Kelautan, Jurnal Penyuluhan Perikanan -Sekolah Tinggi Perikanan Jalan Cikaret Nomor, Jurusan Barat, Jawa," *J. Penyul. Perikan.*, vol. 11 (1), no. 1, pp. 12–24, 2017.

[10]   N. Sofiyanti and S. Suartini, "Pengaruh Jumlah Kapal Perikanan dan Jumlah Nelayan Terhadap Hasil Produksi Perikanan di Indonesia," *Accounthink J. Account. Financ.*, vol. 1, no. 01, pp. 49–61, 2016, doi: 10.35706/acc.v1i01.442.

[11]   A. Fahrudin, "Analisis Pendapatan dan Faktor-Faktor yang Mempengaruhi Produksi Usaha Budidaya Tambak Ikan," *Effic. Indones. J. Dev. Econ.*, vol. 1, no. 1, pp. 77–85, 2018, doi: 10.15294/efficient.v1i1.27223.

[12]   C. M. Free, J. T. Thorson, M. L. Pinsky, K. L. Oken, J. Wiedenmann, and O. P. Jensen, "Impacts of historical warming on marine fisheries production," *Science (80-.).*, vol. 363, no. 6430, pp. 979–983, 2019, doi: 10.1126/science.aau1758.

[13]   I. Karina, "Pertanggungjawaban Pidana Terhadap Penggunaan Alat Tangkap Ikan Illegal," *Fiat Iustitia J. Huk.*, pp. 92–103, 2021, doi: 10.54367/fiat.v1i2.1156.

[14]   E. Thio, "Evaluating Site Selection Criteria for Marine Cultivation in North Lombok Regency of Indonesia through GADA model," *Int. J. Grey Syst.*, vol. 1, no. 1, pp. 27–37, 2021, doi: 10.52812/ijgs.13.

[15]   Z. Hidayah, A. Arisandi, and M. K. Wardhani, "Pemetaan Kesesuaian Perairan untuk

Budidaya Laut di Perairan Pesisir Kabupaten Situbondo dan Banyuwangi Jawa Timur," *Rekayasa*, vol. 13, no. 3, pp. 307–316, 2021, doi: 10.21107/rekayasa.v13i3.9858.

[16] Rofizar, Y. V. Jaya, and H. Irawan, "Aplikasi SIG untuk pemetaan kesesuaian kawasan budididaya ikan kerapu menggunakan keramba di perairan laut Desa Genting Pulur Kabupaten Kepulauan Anambas," *Intek Akuakultur*, vol. 1, no. 1, pp. 37–50, 2017.

[17] M. Makmur, H. Kusnoputranto, S. S. Moersidik, and D. S. Available: http://journal.itny.ac.id/index.php/ReTII.

[19] M. Safii, "Implementasi Data Mining Dengan Metode Pohon Keputusan Algoritma Id3 Untuk Menentukan Status Mahasiswa," *J. Mantik Penusa*, vol. 2, no. 1, pp. 82–86, 2019.

[20] A. A. Mattjik and I. M. Sumertajaya, *Sidik Peubah Ganda dengan Menggunakan SAS*. 2011.

[21] R. Magdalena and M. Angela Krisanti, "Analisis Penyebab dan Solusi Rekonsiliasi Finished Goods Menggunakan Hipotesis Statistik dengan Metode Pengujian Independent Sample T-Test di PT.Merck, Tbk.," *J. Tekno*, vol. 16, no. 2, pp. 35–48, 2019, doi: 10.33557/jtekno.v16i1.623.

Wisnubroto, "Pengaruh Limbah Organik dan Rasio N/P Terhadap Kelimpahan Fitoplankton di Kawasan Budidaya Kerang Hijau Cilincing," *J. Teknol. Pengelolaan Limbah*, vol. 15, no. 2, pp. 51–64, 2012, [Online]. Available: http://jurnal.batan.go.id/index.php/jtpl/article/view/326.

[18] E. D. Wahyuni, A. A. Arifiyanti, and M. Kustyani, "Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining," *Pros. Nas. Rekayasa Teknol. Ind. dan Inf. XIV Tahun 2019*, vol. 2019, no. November, pp. 263–269, 2019, [Online].