International Journal of Engineering, Science and Information Technology
Volume 5 No. 1 (2025) pp. 297-305
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i1.717
Research Paper, Short Communication, Review, Technical Paper

Application of Multiple Linear Regression Method for Predicting Fish Production Based on Cultivation Type

Hendra Putranta Limbong*, Dahlan Abdullah, Said Fadlan Anshari

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: hendra.200170126@mhs.unimal.ac.id*

The manuscript was received on 1 May 2024, revised on 15 August 2024, and accepted on 6 January 2025, date of publication 9 January 2025

Abstract

One of the contributors to Indonesia's economy is the fisheries sector, which has a high potential for development. Fisheries are a highly promising subsector for development in Indonesia's growth efforts. Based on data from the Central Bureau of Statistics of Dairi Regency, three types of aquaculture remain actively utilized in each subdistrict: ponds/freshwater ponds and paddy fields. This research aims to develop a fish production prediction system based on aquaculture types using the Multiple Linear Regression method. The accuracy of the prediction results will be measured using the Mean Absolute Percentage Error (MAPE). The results of this study indicate that in almost every subdistrict, especially pond aquaculture, the MAPE value is <20%, which means it has good accuracy. However, exceptions are found in the Siempat Nempu Hulu subdistrict, which has a MAPE value of 34.29%, and the Silahisabungan subdistrict, which has a MAPE value of 43.78%. Despite these values, they are still categorized as sufficient since they are <50%. The lower the MAPE value, the more accurate the prediction results. The findings of this research show that the multiple linear regression method can be considered correct. For future predictions, some results show negative values. For instance, in Silimapunggapungga subdistrict, a decline in production is predicted for 2024 with -114.779 tons and 2025 with -134.316 tons. The pessimistic prediction results are caused by the decrease in the X2 variable (area size), leading to a minor Y (production) value, potentially becoming negative if the contribution of X2 is no longer sufficient to balance the values of X1 (b1) and a. On the other hand, the Lae Parira subdistrict is predicted to experience an increase in production in 2024 by 87.024 tons and in 2025 by 84.380 tons. This system is implemented using the Python programming language. It is expected to help relevant stakeholders understand production trends and enhance the efficiency of fisheries resource management in the Dairi Regency.

Keywords: *Production Prediction, Multiple Linear Regression, Fisheries, MAPE, Python.*

1. Introduction

One of the supporters of Indonesia's economy is the fisheries sector, which has a high potential for development. Fisheries are a highly potential subsector for growth in Indonesia. One of the most protein-rich foods and always consumed by the entire community, fish is the primary commodity of the fisheries sector. Because fish protein is easily digested and meets the amino acids in humans, the protein in fish is essential for humans [1].

In practice, types of fish farming such as ponds, rice fields, and floating nets have different impacts on production yields. Factors such as land area, rainfall, altitude above sea level, and technology utilization tend to contribute variably to productivity.

Based on data from the Central Statistics Agency of Dairi Regency, three types of fish farming are still actively used in each sub-district, including ponds/freshwater ponds and rice fields/paddy fields. The production from capture fisheries is no longer reliable, so aquaculture development is being enhanced. As part of the effort to conduct research on fish farming in Dairi Regency, which consists of 15 sub-districts, this study focuses on each sub-district in the area.

The advancement of technology, such as web-based or mobile applications, has made the collection and reporting of fish farming production data easier. This data includes information about the types of fish farming, the area used, and the production data of fish farming. In such cases, the accurate collection of previous production data dramatically facilitates the management of subsequent production results.

Production is the process of yielding results. So, it can be interpreted that production is the process of generating goods or services to meet human needs. Fish production refers to the cultivation activities aimed at producing optimal results so that they can be consumed or sold. One of the supporters of Indonesia's economy is the fisheries industry, which has a high potential for development [1].

Fish production predictions play an essential role in the design of fish farming. With the prediction results based on the type of aquaculture and the area size per sub-district, the government and households managing fish farming can determine whether the production quantity each year is increasing or decreasing. The multiple linear regression method is used to predict fish production and optimize the overall production results of the fish farming sector.
Regression is a statistical method to determine the relationship between the dependent and independent variables. Regression approaches can be conducted using parametric, semiparametric, and nonparametric methods. Linear regression is used to analyze the relationship between the dependent and independent variables so that a linear equation is obtained to observe the data [2].

2. Literature Review

2.1. Previous Research

According to the research by Miftahul Huda et al. (2024), "Prediction of Used Car Prices Using Multiple Linear Regression Algorithm," the dataset used to build the prediction model with the linear regression algorithm was sourced from the Kaggle repository, which contains data on used car prices. The analysis results were visualized using RapidMiner. The prediction process was conducted with a data split of 90% for training and 10% for testing. Based on the prediction process, a model yielded a predicted price value of 1637.49. The prediction model was evaluated using RMSE and Relative Error. For the Price category, the evaluation showed an RMSE value of 1637.49 and a Relative Error value of 11.89%. The input variables include Age (24, X1), Kilometer (783764, X2), Horsepower (100, X3), Transmission (Automatic: 0, X4), Engine Capacity (1500 CC, X5), and Vehicle Weight (1300 kg, X6). Using these inputs, the predicted result was 8438.61, calculated using the formula: $Y = b1 + b2X1 + b3X2 + b4X3 + b5X4 + b6X5 + b7X6$ [3].
According to Lailiyah et al. (2023), "Prediction of Raw Material Inventory for Processed Food Production at 'Sanggar Krispi' Using Multiple Linear Regression Method," In the data collection for this research, field study and literature study methods were used. A Flowchart was employed as a system design aid, with the attributes being date, day, and weather. The date variable predicts a sales amount of 13 boxes with an x-value influence of 0.01; the day variable predicts sales of 12 boxes with an x-value influence of 0.03; and the weather variable predicts a sales amount of 9 boxes with an x-value influence of 9.16. This method's calculation process uses sales data spanning 15 days as input data. Following this procedure, the results of the multiple linear regression method were tested, yielding MSE = 4.14, RMSE = 2.03, and MAPE = 10.35%. According to this analysis, the system's accuracy was 89% since it accurately predicted 4 out of 7 data points [4].
According to Aryani (2020), "Sales Information System with Multiple Linear Regression Method in Predicting Company Revenue". The factor underlying this research is that the company where the study was conducted still uses a manual system for processing goods, managing inventory, and recording sales transactions, which has not been well documented. The research uses the SPSS version 21 application to aid calculations. In this study, with values (T-test, D-test, and R2) considering sig < 0.05, the regression equation that was derived is $Y = 11257.187 – 3.427(X1) + 12.501(X2) – 2.076(X3)$. It may be inferred from the individual T-test hypothesis testing findings that variables X1, X2, and X3 affect income. With a significance level of less than 0.05, the F-test results indicate that factors X1, X2, and X3 all impact the company's revenue. At 1.000, the Coefficient of determination (R2) yielded a very high value. The less the independent variable influences the dependent variable, the lower the Coefficient of determination (R Square) value. On the other hand, the influence will increase if the R Square value becomes closer to 1 [5].
According to Maharadja et al. (2021), "The Application of Multiple Linear Regression Method for Predicting State Losses Based on Corruption Cases." Research Results: This research uses data on corruption crimes from 2013-2020. From the calculations, the constant value obtained is 284645.5891073216, and the coefficient values are 139837.38007863 and 363493.06049751. From the data partitioning conditions, an RMSE value of 8447373.485 was obtained for the training data and 9769609.026 for the testing data. Meanwhile, the Coefficient of determination value obtained was 0.579 for the training data, indicating a strong relationship between the variables, and 0.662 for the testing data, indicating a strong relationship between the variables [6].
According to Saragih (2021), "The Application of Data Mining to Predict Palm Oil Production Using Multiple Linear Regression Method." Research Results: This research was conducted at PT. Padasa Enam Utama. This research was conducted because, at the company, the amount of palm oil production harvest is not automatically known whether the production harvest amount is increasing or decreasing. The data used are the harvest production data from 2017-2019, with 235 data points, of which 205 are used for training data and 35 for testing data. The results of this study obtained a value of 12.308% using the Mean Absolute Percentage Error (MAPE), indicating that this system is feasible for use [7].
According to Hendra Di Kesuma et al. (2022), "Implementation of Data Mining for Predicting New Students Using Multiple Linear Regression Algorithm." Research Results: This research was conducted at STMIK Bina Nusantara Jaya Lubuklinggau. This research was conducted because the object does not yet have a system to predict the number of new students in the upcoming period. The variables used are 3, namely cost, applicants, and new students. In this study, the results obtained were as follows: in 2023, there were 38 people, and then the results were tested using Ravidminer 5.0, where the predicted number of new students in 2023 was 38 [8].
According to Abdullah et al. (2021), "Prediction of Drug User Rates Using Web-Based Multiple Linear Regression Method." Research Results: This study aims to assist the legal community and the general public by creating a drug user prediction application using the multiple linear regression method. This research uses data on drug users from 2015 to 2019, totaling 619 user data points taken from the Aceh Tamiang Police Resort (Polres). In performing the calculations using this method, the equation, $Y = 4.492312 + 1.018444 X1 + 1.181143 X2$ was obtained, where X1 is methamphetamine and X2 is marijuana. If the number of methamphetamine users is 108 and cannabis users is 76, then it is predicted that there will be 195 cases of drug users with a MAPE value of 68.48% [9].

2.2. Data Mining

Data mining is collecting and processing data to extract meaningful information from it. Data collection and information extraction can be done using software with statistical calculations, mathematics, or Artificial Intelligence (AI) technology. According to Abdullah et al. (2019), data mining is a process that uses one or more computer learning techniques to analyze and extract knowledge automatically. Data mining can be divided into several groups based on the tasks performed: description, estimation, prediction, classification, clustering, and association [10].

According to the Tachi & Andri module (2021), data mining is divided into six groups [11], namely:

1. Description
   The function of description in data mining is to gain a deeper understanding of the patterns and trends from the collected data. By conducting a process, it is hoped that one can understand the behavior or characteristics of the data.
2. Estimation
   Estimation is almost the same as classification; the difference between estimation and classification lies only in the variables, as estimation is more oriented towards numbers than categories. The value of the target variable is used as the predicted value to form the modeling, which is created through complete records.
3. Prediction
   Prediction, classification, and estimation are also quite similar, but the value of the prediction will be in the future. For example, predicting rice prices in the next 3 months and the unemployment rate in 5 years.
4. Clustering
5. Clustering is the grouping, observing, or paying attention to and forming classes of objects that have similarities. Clustering differs from classification in that there is no target variable in clustering.
6. Classification
   In classification, the target variable is in the form of classes or categories to infer an unknown object's class.
7. Association
   Association has the function of finding attributes that appear at a given time. It is often referred to as market basket analysis in the business world. For example, it is important to find items in a supermarket that are purchased together and never purchased together.

2.3. Prediction

Prediction is a structured process of estimating what is likely to happen in the future based on data and information from the past and present to minimize errors or discrepancies between the predicted results and reality. Predictions do not always provide definite answers, but predictions strive to find the closest possible outcomes that will occur [12]. Forecasting techniques are one way to conduct production planning and control to face future uncertainties. Generally, forecasting methods are divided into two types: quantitative and qualitative.

2.4. Multiple Linear Regression

Multiple linear regression is a prediction method that involves two or more variables, consisting of independent (free) variables and dependent (bound) variables [6]. To predict the dependent variable (Y), all independent variables must have their values known. Here is the multiple linear regression equation model:

$$Y = a + B_1.X_1 + B_2.X_2 + \cdots + B_n.X_n \quad (1)$$

The values a, B1, B2, ..., Bn can be calculated using the least squares method, namely:

$$B_1 = \frac{(\sum X_{2^2}).(\sum X_2.Y)-(\sum X_1.X_2).(\sum X_2.Y)}{(\sum X_{1^2})(\sum X_{2^2})-(\sum X_1.X_2)^2} \quad (2)$$

$$B_2 = \frac{(\sum X_{1^2}).(\sum X_2.Y)-(\sum X_1.X_2).(\sum X_1.Y)}{(\sum X_{1^2})(\sum X_{2^2})-(\sum X_1.X_2)^2} \quad (3)$$

$$a = \frac{\sum Y-(B_1\sum X_1)-(B_2\sum X_2)}{n} \quad (4)$$

Where:

$$\sum X_{1^2} = \sum X_{1^2} - \frac{(\sum X_1)^2}{n} \quad (5)$$

$$\sum X_{2^2} = \sum X_{2^2} - \frac{(\sum X_2)^2}{n} \quad (6)$$

$$\sum Y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} \quad (7)$$

$$\sum X_1.Y = \sum X_1.Y - \frac{(\sum X_1).(\sum Y)}{n} \quad (8)$$

$$\sum X_2.Y = \sum X_2.Y - \frac{(\sum X_2).(\sum Y)}{n} \quad (9)$$

$$\sum X_1.X_2 = \sum X_1.X_2 - \frac{(\sum X_1).(\sum X_2)}{n} \quad (10)$$

Description:

$X_1, X_2 \ldots X_n$ = Independent Variable

$B_1.B_2$ = Coefficient

Y = Dependent Variable

a = Constant

After the above calculations are completed, the results will be tested using the Mean Absolute Percentage Error (MAPE) method.
Mean Absolute Percentage Error (MAPE) is an error that uses the absolute error for each period, which is then divided by the value of the actual observation [13]. Meanwhile, according to [14], MAPE is the difference in calculations between actual and forecasted data. The difference from the calculation is made absolute to become positive and then calculated into a percentage of the original data. Thus, the mean value of the percentage results is obtained. The MAPE formula is as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y-y_i|}{n}\ x\ 100\% \quad (11)$$

Description:

y = Actual data

y_i = Forecasted data for period i

n = Total number of data points

The value of the MAPE percentage is considered good if the result is less than 20%, as explained in the following range table:

Table 1. MAPE Range

MAPE Range	Description
<10%	Very Accurate Accuracy Level
10 – 20%	Good Accuracy Level
20 – 50%	Fair/Acceptable Accuracy Level
>50%	Inaccurate Accuracy Level

2.5. Python

The Python programming language is currently very popular, and Guido van Rossum was the first person to create the Python programming language at the Centrum Wiskunde & Informatica (CWI), Amsterdam, in 1991. The name Python itself comes from Monty Python; when Guido van Rossum created it, he was reading the script of Monty Python's Flying Circus on BBC. Python is a development programming language inspired by the ABC programming language that was being developed then. One of the main differences between Python and other programming languages is that Python's development involves millions of programmers, researchers, and users from various backgrounds, not just from the IT field, because Python is Open Source [15].

Python is a programming language that uses an interpreter to execute its code. Python is a computer programming language used to build websites software/applications, automate tasks, and perform data analysis. Python is a general-purpose programming language, which can be used to create various programs, not just for specific problems. This programming language has versatile and user-friendly characteristics, making Python the most widely used, especially among beginners.

3. Research Methods

The research conducted by the author focuses on the stages involved in predicting the production of plantation crops in Aceh Barat. The author uses the research stages as a guideline to ensure that the results stay aligned with the established objectives and do not deviate from the intended purpose.

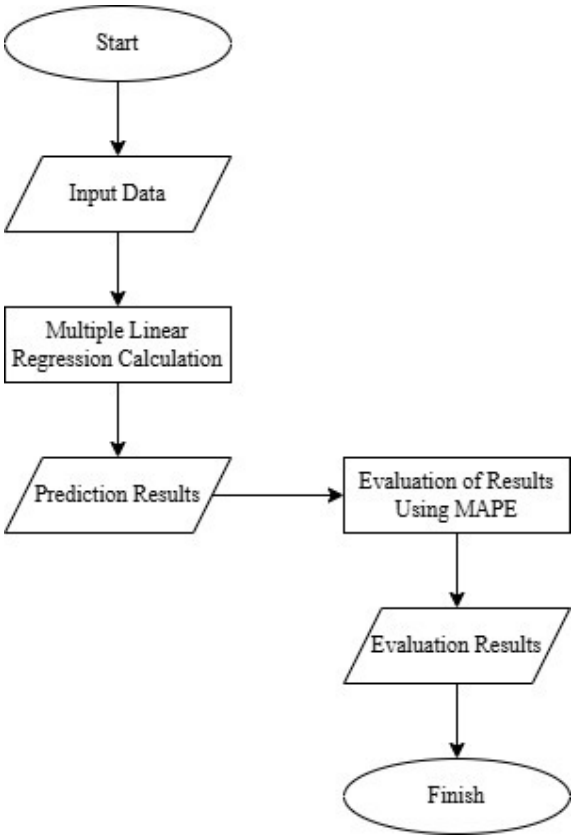

Fig 1. Flowchart System

Explanation:

1. Input production data and area size from data collected by the Central Bureau of Statistics of Dairi Regency from 2017-2023.
2. The flowchart below illustrates the calculation of multiple linear regression problem-solving using multiple linear regression.

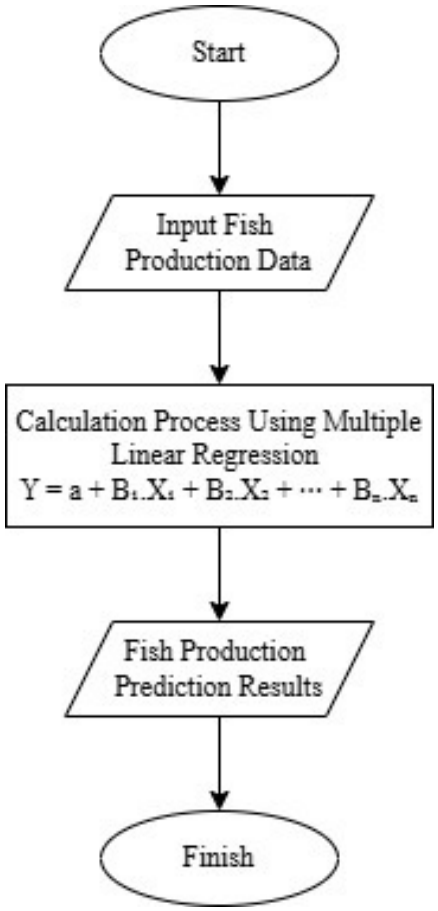

Fig 2. Solution Using the Multiple Regression Method

3. Then, from the calculation results using the linear regression method, the output of the calculations was obtained, which will later be used to predict the results.
4. After obtaining the output value, the results are evaluated using MAPE with the following formula.

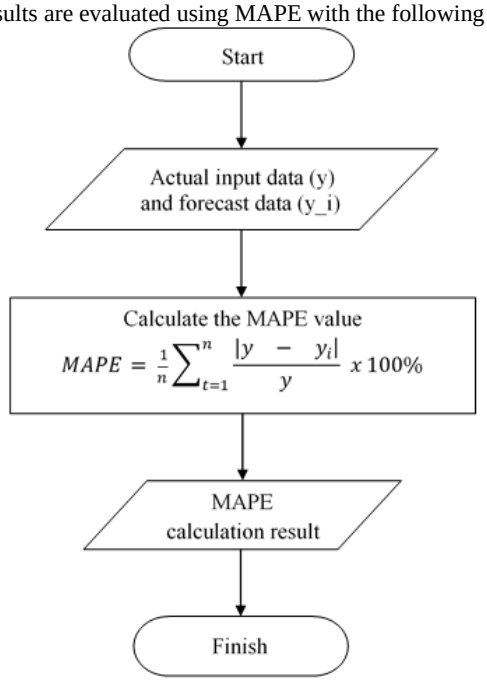

Fig 3. MAPE Calculation System

5. The prediction results were obtained from the calculations using the MAPE method.

4. Result and Discussions

4.1. Dataset

This study's data includes all the sub-districts in Dairi Regency, totaling 15 sub-districts. The data used was taken from the years 2017-2023. This data provides essential information that includes the fish production amount in each Dairi Regency district. In the regression analysis, the independent variables used are X1 as rainfall, X2 as area size, and X3 as altitude above sea level, while the dependent variable Y is the production quantity.

Table 2. Data on Pond Aquaculture Fish Production

No	District	Year	X_1	X_2	X_3	Y
1	Sidikalang	2017	25,82	35,5	1068	68,5
2		2018	28,49	31,5	1068	63,5
3		2019	27,73	31,5	1068	63,5
4		2020	23,97	30	1068	62,45
5		2021	23,97	29	1068	64,57
6		2022	23,97	18,34	1068	53,97
7		2023	23,97	18,34	1068	65,97
…	…	…	…	…	…	…
99	Tigalingga	2017	25,82	7,2	472	11,5
100		2018	28,49	5	472	8
101		2019	27,73	5	472	5,5
102		2020	23,97	5	472	5,23
103		2021	23,97	4,5	472	5,23
104		2022	23,97	4,28	472	4,37
105		2023	23,97	4,28	472	5,11

Table 3. Data on Rice Field Aquaculture Fish Production

No	District	Year	X_1	X_2	X_3	Y
1	Sidikalang	2017	25,82	20	1068	50,5
2		2018	28,49	15	1068	36
3		2019	27,73	15	1068	36,8
4		2020	23,97	30	1068	62,45
5		2021	23,97	29	1068	64,57
6		2022	23,97	18,34	1068	53,97
7		2023	23,97	18,34	1068	65,97
…	…	…	…	…	…	…
99	Tigalingga	2017	25,82	6,5	472	6,5
100		2018	28,49	6,2	472	6
101		2019	27,73	6	472	6,5
102		2020	23,97	5,6	472	6
103		2021	23,97	5,4	472	6
104		2022	23,97	5	472	5,9
105		2023	23,97	5	472	6,5

4.2. Prediction Result

In this study, the author uses the Multiple Linear Regression method to predict the amount of fish production based on the type of cultivation. The variables that predict fish production consist of rainfall, area size, and elevation above sea level, which are the X values, and actual output, which are the Y values. Multiple linear regression is one way to make predictions involving two or more variables. The variables used must be interrelated or have a cause-and-effect relationship. From the calculations using the multiple linear regression method, an equation will be produced that will serve as a reference for predicting the value of the dependent variable in the future.

Table 4. Results of Pond Cultivation Type Predictions for the Upcoming Period

No	Year	District	Prediction Data (Tons)
1	2024	Sidikalang	53,401
	2025		53,180
2	2024	Berampu	22,395
	2025		20,033
3	2024	Gunung Sitember	4,372
	2025		20,168
4	2024	Lea Parira	87,024
	2025		84,380

No	Year	District	Prediction Data (Tons)
5	2024	Parbuluan	-41,530
	2025		-28,594
…	…	…	…
11	2024	Silima Pungga-pungga	-114,779
	2025		-134,316
12	2024	Sitinjo	18,429
	2025		18,447
13	2024	Sumbul	-318,737
	2025		-298,939
14	2024	Tanah Pinem	5,083
	2025		7,471
15	2024	Tigalingga	4,452
	2025		11,815

Table 5. Results of Rice Cultivation Type Predictions for the Upcoming Period

No	Year	District	Prediction Data (Tons)
1	2024	Sidikalang	55,20810
	2025		31,95307
2	2024	Berampu	4,54651
	2025		7,32082
3	2024	Gunung Sitember	4,59820
	2025		11,20341
4	2024	Lea Parira	11,44493
	2025		14,90642
5	2024	Parbuluan	11,07327
	2025		22,99939
…	…	…	…
11	2024	Silima Pungga-pungga	7,66723
	2025		11,65335
12	2024	Sitinjo	3,41647
	2025		6,66781
13	2024	Sumbul	1,97322
	2025		3,02119
14	2024	Tanah Pinem	8,28162
	2025		3,98608
15	2024	Tigalingga	5,99009
	2025		6,43064

The table above shows the results of the predicted fish production calculations for 2024 and 2025 based on the previous equations.

4.3. Evaluation Prediction Result

The prediction evaluation results using MAPE aim to determine how well the method can produce the best predictions. The comparison results of plantation crops are shown in the following table.

Table 6. Evaluation of Pond Cultivation Type Prediction Results MAPE

No	District	Year	X1	X2	X3	Y	Prediction	MAPE
1	Sidikalang	2017	25,82	35,5	1068	68,5	66,48	0,42
2		2018	28,49	31,5	1068	63,5	63,84	0,08
3		2019	27,73	31,5	1068	63,5	64,09	0,13
4		2020	23,97	30	1068	62,45	64,68	0,51
5		2021	23,97	29	1068	64,57	64,24	0,07
6		2022	23,97	18,34	1068	53,97	59,57	1,48
7		2023	23,97	18,34	1068	65,97	59,57	1,39
			Amount					**4,08**
…	…	…	…	…	…	…	…	…
99	Tigalingga	2017	25,82	7,2	472	11,5	11,24	0,32
100		2018	28,49	5	472	8	7,00	1,79
101		2019	27,73	5	472	5,5	6,83	3,46
102		2020	23,97	5	472	5,23	6,03	2,18
103		2021	23,97	4,5	472	5,23	4,93	0,81
104		2022	23,97	4,28	472	4,37	4,45	0,27

No	District	Year	X1	X2	X3	Y	Prediction	MAPE
105		2023	23,97	4,28	472	5,11	4,45	1,84
			Amount					**10,67**

Table 7. Evaluation of Production Results of MAPE Rice Cultivation Type

No	District	Year	X1	X2	X3	Y	Prediction	MAPE
1	Sidikalang	2017	25,82	20	1068	50,5	50,48	0,01
2		2018	28,49	15	1068	36	34,52	0,59
3		2019	27,73	15	1068	36,8	38,59	0,70
4		2020	23,97	30	1068	62,45	63,67	0,28
5		2021	23,97	29	1068	64,57	63,34	0,27
6		2022	23,97	18,34	1068	53,97	59,83	1,55
7		2023	23,97	18,34	1068	65,97	59,83	1,33
			Amount					**4,08**
…	…	…	…	…	…	…	…	…
99	Tigalingga	2017	25,82	6,5	472	6,5	6,32	0,38
100		2018	28,49	6,2	472	6	6,28	0,67
101		2019	27,73	6	472	6,5	6,25	0,55
102		2020	23,97	5,6	472	6	6,19	0,45
103		2021	23,97	5,4	472	6	6,16	0,38
104		2022	23,97	5	472	5,9	6,10	0,48
105		2023	23,97	5	472	6,5	6,10	0,88
			Amount					**10,67**

Based on the overall prediction research results, almost all districts achieved MAPE values of less than 20% for both pond and rice field fish farming production, reflecting a high level of accuracy. For instance, Sidikalang District recorded a MAPE value of 4.08%, indicating high prediction accuracy. However, in Siempat Nempu Hulu District, the MAPE value reached 34.28%; in Silahisabungan District, it was 43.78% for pond farming. Despite these higher values, they are still considered reasonably accurate as they remain below 50%.

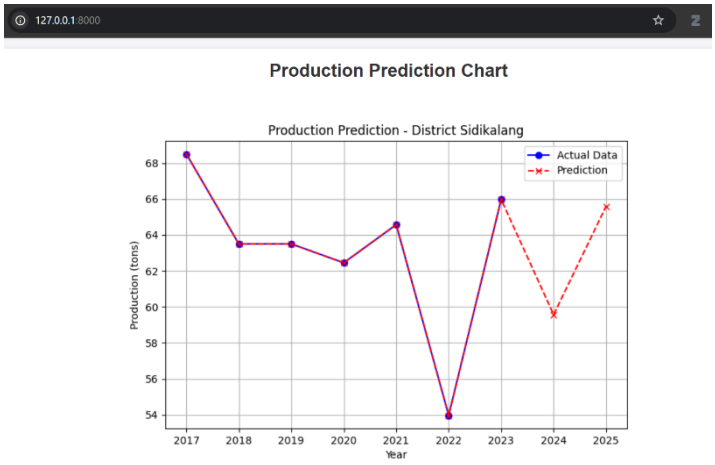

Fig 4. Sidikalang Pond Prediction Graph

5. Conclusion

Based on the research that has been conducted, the author has drawn the following conclusions:

1. The multiple linear regression method is used to predict fish production, and the Mean Absolute Percentage Error (MAPE) is used to determine the accuracy level of the prediction. In this study, the multiple linear regression method was deemed accurate in making predictions overall. For example, in the Sidikalang District, a MAPE value of less than 10% was obtained, indicating an excellent forecasting accuracy level.
2. Based on the overall prediction research results, almost all sub-districts obtained MAPE values <20% for pond and rice field fish farming production. For example, in the Sidikalang sub-district, a MAPE value of 4.08% was obtained, indicating a very accurate level of accuracy. However, in the Siempat Nempu Hulu sub-district, a MAPE value of 34.28% was obtained, and in the Silahisabungan sub-district, a MAPE value of 43.78% was obtained for pond farming. Nevertheless, these MAPE values are still considered entirely accurate because they are <50%.

References

[1] A. N. Suyono, Rifki, and Kaukab Elfan M, "Pengaruh Harga Jual, Luas Lahan dan Biaya Produksi Terhadap Pendapatan Usaha Budidaya Ikan Konsumsi Air tawar," 2022.

[2] S. F. Asnhari *et al.*, *2019 7th International Conference on Information and Communication Technology (ICoICT) : July 24-26, 2019, Kuala Lumpur, Malaysia*. 2019.

[3] D. Miftahul Huda, G. Dwilestari, and A. Rizki Rinaldi, "Jurnal Informatika dan Rekayasa Perangkat Lunak Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 150–157, 2024.

[4] S. Lailiyah, A. Yusnita, and L. Hariri, "Prediksi Persediaan Bahan Baku Untuk Produksi Makanan Olahan 'Sanggar Krispi' Menggunakan Metode Regresi Linear Berganda," *Simkom*, vol. 8, no. 2, pp. 84–94, 2023, doi: 10.51717/simkom.v8i2.141.

[5] Y. Aryani, "Sistem Informasi Penjualan Barang Dengan Metode Regresi Linear Berganda Dalam Prediksi Pendapatan Perusahaan," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 2, no. 2, pp. 39–51, 2020, doi: 10.52005/jursistekni.v2i2.47.

[6] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.

[7] L. Saragih, "Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode Regresi Linear Berganda (Studi Kasus: PT. Padasa Enam Utama)," 2021.

[8] Hendra Di Kesuma, D. Apriadi, H. Juliansa, and E. Etriyanti, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 62–66, 2022, doi: 10.52303/jb.v4i2.74.

[9] D. Abdullah, M. Maryana, and M. Ani, "Prediksi Tingkat Pengguna Narkoba Dengan Metode Regresi Linear Berganda Berbasis Web," *TECHSI - J. Tek. Inform.*, vol. 13, no. 2, p. 41, 2021, doi: 10.29103/techsi.v13i2.3738.

[10] D. Abdullah *et al.*, "Application of Data Mining Prediction of Electricity Deviation Flow Using Metode Backpropogation at PLN Binjai Area," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1363/1/012067.

[11] E. Tachi and N. & Andri, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," 2021.

[12] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.

[13] H. Elvaningsih and F. Tawakal, *Seminar Nasional Sains dan Teknologi Informasi (SENSASI) Prediksi Stok Obat Menggunakan Metode Backpropagation (Studi Kasus: Puskesmas Dumai Barat)*. 2021. [Online]. Available: http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C228

[14] M. R. Fauzi and I. D. Mulyana, "Implementasi Data Mining Menggunakan Metode Least Square untuk Memprediksi Penjualan Lampu LED pada PT. Sumber Dinamika Solusitama," vol. 1, 2021, [Online]. Available: http://sostech.greenvest.co.id

[15] S. Rahman *et al.*, "Python : Dasar Dan Pemrograman Berorientasi Objek Tahta Media Group," 2023.