Prediction of fishers' income using a flexible model in Karanggongso fishers community, Trenggalek regency, Indonesia

A D Sulistyono, E Susilo, P Purwanti, N H Wardani
Department of Socio-Economy Fisheries and Marine, Faculty of Fisheries and Marine Sciences, Universitas Brawijaya
Email: agusdwistat@ub.ac.id

Abstract. The contribution of fisheries to the national GDP had increased from 2.32% in 2014 to 2.60% in 2018. However, in 2020, the threat of the Covid-19 pandemic emerged, which hit all sectors of the economy, including the fisheries sector. Many communities, especially coastal fishing communities, are complaining about economic hardship. Income has fallen dramatically because people's purchasing power has fallen significantly. Based on these problems, this research was conducted to build a fishers' income prediction model. This research took a case study on fishers in Karanggongso District Trenggalek by surveying 50 fishing households. There were 12 predictors variables, namely Boat Type (X1), Price Boat (X2), Age of the Boat (X3), Boat Power (X4), Machine Price (X5), Engine Life (X6), Fishing Equipment Price (X7), Fishing Gear Life (X8), Cool Box (X9), Trip/week (X10), Average Hours/trip (X11), and Total Expenditures/week (X12). The response variable is Income Per Week (Y). Data analysis was done by using multiple linear regression analysis and flexible modelling with a machine learning approach. Based on the results of the analysis, a multiple linear regression model had an accuracy level of $R^2 = 70.5\%$ and $MSE = 1.086 \times 10^{18}$ with the boat price was the most dominant influence on fishers' income. While flexible modelling has an accuracy level of $R^2 = 85.2\%$ and $MSE = 3.308 \times 10^{14}$. From this research, it was proven that the flexible model had a higher level of accuracy than the linear regression model. Also, the flexible model obtained the non-linear effect on the number of cool boxes and the fishing gear life.

1. Introduction
The contribution of Quarterly Fisheries GDP based on current prices 2014-2018 to National GDP shows an increasing contribution from an average of 2.32% in 2014 to an average of 2.60% in 2018. It shows an increase in value that reflects an average increase in the income of fisheries sector actors [1]. The value of Fishery's GDP increased from IDR 58.97 trillion in Q1 2018 to IDR 62.31 trillion in Q1 2019. Meanwhile, Fishers' Terms of Trade, which was at less than 106% in 2014, rose to 113.08% in May 2019 [2].

The achievements in 2019 will not be smooth without obstacles. The biggest threat continues from year to year, namely the rampant illegal fishing [3]. It is recorded that starting from 2014-2018, the Marine and Fisheries Ministry had sunk 463 foreign vessels proven to have carried out illegal fishing.
In 2020, another threat that occurs internationally will emerge. The Covid-19 pandemic hit all economic sectors around the world, including the fisheries sector. Many people, especially coastal fishing communities, complain about financial difficulties because their income has dropped dramatically. Not because of problems in fishing activities to go to sea, but people's purchasing power has decreased significantly due to the pandemic [4]. The government is trying to take steps to provide direct assistance to the community by buying fishery products directly from fishers to increase people's purchasing power. Besides, the government has also increased the construction of cold storage in several areas [5].

Based on these problems, it is important to develop a model that can be used to predict the fishers' income so they can use resources optimally. Research is needed in the field of modelling that can predict fishers' income accurately, knowing not only the significance of the predictors, but also the predicting income accurately. As stated in Universitas Brawijaya's Master Plan for Good Governance, Universitas Brawijaya has launched a research program with strategic issues of Economic Resilience and Sustainable Local Business. This research was conducted to build a Fishers' Income Prediction Model with a Machine Learning Approach, where this research took a case study on fishing rods in Karanggongso, Trenggalek Regency.

2. Material and Methods

2.1. Coastal community household production economics

According to [6], time as a farmer household economic resource can be allocated to activities which can be classified as follows:

a. Income-generating activities
b. Activities that do not generate income
c. Relaxing (leisure) and
d. Time devoted to acquiring skills

Based on [7], revenue is all receipts of all members of economical household, both in the form of goods and services. These are the admissions:

a. taking savings or savings
b. sale or procurement of goods
c. receivable receivables
d. irregular shipments or gifts from family or other parties, inheritance or grants, and others.

The results of the research showed that the income contribution of fisherwomen ranged from 2.25% to 45.45% with an average of 15.09%. This low contribution is due to the low-income category of the income earned from fisherwomen's activities. Most fisherwomen have income or income from activities as service sellers, such as "rampek" workers, "pindang" labourers, shop workers, and scavengers and sell with small capital. Household expenditure is one indicator that can provide an overview of the welfare state of the population. The higher the income, the portion of expenditure will shift from expenditure on food to non-food expenditure. Consumption expenditure is grouped into expenditure on foodstuffs and expenditure on non-food ingredients, namely:

1. Consumption of food ingredients: grains, tubers, fish, meat, eggs, milk, vegetables, nuts, fruits, oils and fats, beverage ingredients, spices, food and beverages, tobacco, and betel.
2. Consumption for non-food items consists of housing, fuel, lighting

2.2. Prediction model of fishers' income

The use of statistical data analysis methods and mathematical models that are often used to predict fishers' income is a multiple linear regression analysis. Here are some studies on the use of predictive models:
Table 1. Prediction models in several studies.

| No | Title                                                                 | Method                                      |
|----|----------------------------------------------------------------------|---------------------------------------------|
| 1  | Analysis of Factors Affecting the Income of Fishers in Bendar Village, Juvana District, Pati Regency [8] | Multiple Linear Regression Analysis         |
| 2  | Analysis of Socio-Economic Factors Affecting Net Income of Rawai Fishers in Teluk Ambon District [9] | Multiple Linear Regression Analysis         |
| 3  | Analysis of Factors Affecting the Income of Capture Fishers in Galesong Village, Galesong City, Takalar Regency [10] | Douglass Cobb Regression Analysis           |
| 4  | Income Analysis of 5-10 GT Motor Boat Fishers in Southwest Aceh District [11] | Descriptive Analysis                        |
| 5  | Analysis of Capture Fisheries Business Income: Experience From Fishers in Garut Regency, West Java [6] | Multiple Linear Regression Analysis         |
| 6  | Analysis and forecasting fisheries time series: purse seine in Indian Ocean as a case study [12] | Singular Spectrum Analysis                 |
| 7  | Forecasting Aquaculture Production in Malang Regency Using the Exponential Smoothing Method [7] | ARIMA with Exponential Smoothing            |
| 8  | Fishing and Non-Fishing Income Decisions: The Role of Human Capital and Family Structure [13] | Probit Regression                           |
| 9  | Forecasting the Amount of Fish Production Using the Backpropagation Neural Network (Case Study: UPTD Banjarmasin Fishery Port) [14] | Backpropagation Neural Network              |
| 10 | Deep learning models for the prediction of small-scale fisheries catches: Finfish fishery in the region of "bahiá Magadalena-Almejas" [15] | Non-linear Autoregressive Neural Network    |

Table 1 shows several types of modelling methods used for prediction, namely the linear and non-linear modelling approaches. The linear approach used includes multiple linear regression models and ARIMA. Meanwhile, non-linear approaches include Cobb-Douglass regression, Singular Spectrum Analysis, and Probit regression. This method has the advantage of the complexity of the model on the exogenous variables used. However, this method also has a weakness, namely a relatively large error rate. Besides, several studies also use a deep learning approach, one of which is by using an Artificial Neural Network. The advantage of this deep learning approach results in relatively lower errors because it goes through a fairly complex iteration process. However, this model also has a weakness, namely the limited exogenous variables used due to the complexity of the iteration, which adds to the difficulty, especially in the coding and iteration processes which take longer.

2.3. Machine learning for modelling

Arthur Samuel introduced the term machine learning in 1959 through his journal entitled "Some Studies in Machine Learning Using the Game of Checkers" (IBM Journal of Research and Development). Machine learning is a computer field that gives computers the ability to learn without being explicitly programmed using a model defined by certain parameters [16]. Machine learning uses statistical theories to form mathematical models. Models can be predictive or descriptive. In general, machine learning algorithms can be grouped into four categories, namely:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Deep Learning

2.4. Types and sources of data
The data used in this study were primary. Primary data used in this study were obtained from a survey of fishing rods which was carried out in July 2019. Primary data was in the form of production assets, variable costs and production costs, the outpouring of fishing time, household expenses, and fishery product income measured using the observation sheet.

2.5. Place and time of research
This research was conducted in Trenggalek Regency, from July to September 2019. The location of the research was precisely in the target areas that had the impact of coastal development, especially Tasikmadu village where there is an Indonesian fishing port in that location.

2.6. Population and sample research
Based on the preliminary study, it was found that the total population of households who work as fishers in Tasikmadu Village, Watulimo District who live on the coast of Karanggongso is 333 households. By using the Slovin formula, the results of calculating the number of samples are as follows:

\[
n = \frac{N}{1 + Nd^2} = \frac{333}{1 + 333 \times 0.15^2} = 39.21 \approx 40
\]

The number of samples that were calculated, with the level of sampling error of 15% (0.15), a total sample of 40 households was obtained. 10 samples were added to the study so that the total sample taken in this study was 50 households to anticipate failure to produce valid data. As for the determination of households to be sampled, a sampling process was carried out using simple random sampling.

2.7. Method of analysis
This study used 12 predictor variables, namely Boat Type (X1), Price Boat (X2), Age of the Boat (X3), Boat Power (X4), Machine Price (X5), Engine Life (X6), Fishing Equipment Price (X7), Fishing Gear Life (X8), Cool Box (X9), Trip in 1 week (X10), Average Hours/trip (X11), and Total Expenditures/week (X12). The response variable is Income Per Week (Y).

The method of analysis is carried out in the following steps:
1. Multiple Linear Regression Analysis
2. Flexible Model Modelling
3. Model Validation Test
4. Selection of the Best Model

The software used in this study included MS Excel, SPSS 25.0, and Eureqa 1.24.

2.8. Best model selection criteria
The selection of the best model was made by comparing the largest $R^2$ and the lowest MSE. The following equation can calculate the value of the coefficient of determination ($R^2$).

\[
R^2 = \frac{\sum_{t=1}^{n}(\hat{z}_t - \bar{Z})^2}{\sum_{t=1}^{n}(z_t - \bar{Z})^2}
\]

The greater the $R^2$ value, the tighter the relationship between $Z$ and $\hat{z}_t$ indicates a high level of prediction accuracy. MSE had a function to obtain the magnitude of the difference that appears between the actual value and the predicted value. The MSE value was obtained from the following formula:
The best model chosen is the model that had the lowest of the MSE value, the lowest of the difference or error, or in other words, a high degree of accuracy.

3. Discussion

3.1. Descriptive statistics

In general, it could be shown that there was high variability, especially in boat prices (X2), engine prices (X5), fishing gear prices (X7), and total expenses per week (X12). It could be seen from the relatively far Minimum - Maximum value and the high standard deviation value. Likewise, the income per week was also shown high variability.

Table 2. Descriptive statistics of variables.

| Variable                     | Minimum | Maximum | Mean   | Std. Deviation |
|------------------------------|---------|---------|--------|----------------|
| Boat Type (X1)               | 1       | 5       | 1.92   | 0.78           |
| Price Boat (X2)              | 2,000,000 | 700,000,000 | 75,020,000 | 152,847,795.4 |
| Age of the Boat (X3)         | 3       | 30      | 12.7   | 5.99           |
| Boat Power (X4)              | 0       | 100     | 21.8   | 14.15          |
| Machine Price (X5)           | 0       | 125,000,000 | 13,628,000 | 18,932,188.76 |
| Engine Life (X6)             | 0       | 20      | 7.14   | 4.48           |
| Fishing Equipment Price (X7) | 70,000  | 350,000,000 | 12,621,400 | 50,043,568.37 |
| Fishing Gear Life (X8)       | 1       | 240     | 48.68  | 50.61          |
| Cool Box (X9)                | 0       | 25      | 2.96   | 4.14           |
| Trip in 1 week (X10)         | 3       | 7       | 6.44   | 1.03           |
| Average Hours/trip (X11)     | 6       | 19      | 12.64  | 3.52           |
| Total Expenditures/Week (X12)| 150,000 | 399,210,000 | 14,396,860 | 56,026,604.94 |
| Income Per Week (Y)          | 2,070,000 | 9,878,925,000 | 597,514,080 | 1,667,438,144 |

3.2. Results of multiple linear regression analysis

To determine the effect of independent variables on income per week (Y) fishers in Karanggongso, Trenggalek Regency. The following are the results of testing multiple linear regression analysis:

From Table 3, we get the following regression model:

\[
Y = -1538351152 - 64840382.47X_1 + 13.233X_2 + 165219.729X_3 - 12875136.45X_4 -14.237X_5 + 19770689.88X_6 - 10.167X_7 - 2813095.801X_8 - 132338548.1X_9 + 215719582.6X_{10} + 75871961.3X_{11} - 6.405X_{12}
\]

Based on the results of the regression analysis in Table 3, it was shown that the boat price variable (X3) and fishing gear price (X7) have a p-value of less than 0.05 (p <0.05). From this test, it was shown that these two variables had a significant effect on Fishers's Income per Week (Y). The coefficient of determination (R^2) was 0.705 or 70.5%, and the Mean Square Error (MSE) was 1.086 × 10^18. The diversity of data on the effect of the independent variable on Fishers's Income per week (Y) which was explained by the regression model is 70.5%. The remaining 29.5% was explained by errors and other factors not included in the linear regression model.
Table 3. Results of multiple linear regression analysis.

| Variable                          | Coefficient β | Standardized β | t-statistics | p-value |
|-----------------------------------|---------------|----------------|--------------|---------|
| Constant                          | -1538351152   | 0.000          | -1.175       | 0.248   |
| Boat Type (X1)                    | -64840382.468 | -0.030         | -0.213       | 0.832   |
| Price Boat (X2)                   | 13.233        | 1.213          | 5.633        | 0.000   |
| Age of the Boat (X3)              | 165219.729    | 0.001          | 0.006        | 0.995   |
| Boat Power (X4)                   | -12875136.450 | -0.109         | -0.942       | 0.352   |
| Machine Price (X5)                | -14.237       | -0.162         | -0.709       | 0.483   |
| Engine Life (X6)                  | 19770689.884  | 0.053          | 0.516        | 0.609   |
| Fishing Equipment Price (X7)      | -10.167       | -0.305         | -2.528       | 0.016   |
| Fishing Gear Life (X8)            | -2813095.801  | -0.085         | -0.807       | 0.425   |
| Cool Box (X9)                     | -132338548.125| -0.328         | -1.887       | 0.067   |
| Trip in 1 week (X10)              | 215719582.622 | 0.134          | 1.414        | 0.166   |
| Average Hours/trip (X11)          | 75871961.305  | 0.160          | 1.489        | 0.145   |
| Total Expenditures Per Week       | -6.405        | -0.215         | -0.747       | 0.460   |

3.3. Flexible model analysis results

The process of estimating the flexible model parameters was carried out using the help of Eureqa 1.24 software. The generated model begins by making the mean model. Furthermore, the model generation process was carried out by considering the complexity of the model being built. In this study, the generating model process was carried out as many as 583,589 models were built. After generating a model with a stability level of 91.6%, the maturity of 85.6%, and converging of 96.3%, the 27 best possible models based on the R2 value are as follows:

Table 4. The possible best model of flexible model.

| No | Size | Model                                                                 |
|----|------|-----------------------------------------------------------------------|
| 1  | 1    | Y = X2                                                                |
| 2  | 3    | Y = X2 + X7                                                           |
| 3  | 4    | Y = 51749753.982/X1                                                   |
| 4  | 5    | Y = 228159078.383 - 103568906.823*X1                                 |
| 5  | 7    | Y = 236230638.927 - X7 - 107603040.039*X1                             |
| 6  | 8    | Y = 24.920*X2/(X1*X11)                                               |
| 7  | 10   | Y = 3.560*X2*X10/(X1*X11)                                            |
| 8  | 12   | Y = X7 + 3.446*X2*X10/(X1*X11)                                       |
| 9  | 14   | Y = 90.109 + X7 + 3.446*X2*X10/(X1*X11)                              |
| 10 | 16   | Y = 5183239.87315498*X9 + 3.11492883372325*X2/exp(X1*X9)              |
| 11 | 18   | Y = 3598643.459*X9 + 0.295*X2 + 112792944.608/exp(2.4709*X9)          |
| 12 | 20   | Y = 3520410.648*X9 + 0.310*X2 + 112205779.049/exp(0.156*X4*X9)        |
| 13 | 26   | Y = 4547054.335*X9 + 0.281*X2 + 113560573.655/exp(2.378*X9) - 0.002*X2*X8*X9 |
| 14 | 28   | Y = 4486352.435*X9 + 0.263*X2 + 114090763.082/exp(2.486*X9) - 0.0002*X2*X8*X9^2 |
| 15 | 30   | Y = 4244419.802*X9 + 0.260*X2 + 114188038.968/exp(2.500*X9) - 3.120e- |
| 16 | 32   | Y = 4245420.883*X9 + 0.259*X2 + 114228548.724/exp(2.345*X9^2) - 3.131e- |
|    |      | 6*X2*X8^2*X9^2                                                      |
The goodness of fit model test that is formed is as follows:

| No | Size | Model                                                                 |
|----|------|-----------------------------------------------------------------------|
| 17 | 34   | \( Y = 4327271.542 \times X9 + 0.257 \times X2 + 114402818.886/\exp(X9 + 1.080e-7 \times X2 \times X9) - 3.240e-6 \times X2 \times X8 \times X9 \times 2 \) |
| 18 | 38   | \( Y = 3615093.626 \times X9 + 0.294 \times X2 + 112856250.749/\exp(X9 \times 2 + 1.121e-7 \times X2 \times X9) - X8 \times X9 \times 6 \) |
| 19 | 39   | \( Y = 35895647.618 + 1.952e-9 \times X2 \times X12 \times X9 \times 2 + 2.861e-11 \times X4 \times X8 \times X2 \times 2 - 196421.938 \times X4 \times X1 \times 2 - 6.122e-5 \times X2 \times X4 \times X8 \times X11 \) |
| 20 | 41   | \( Y = 3215958.669 \times X9 + 0.365 \times X2 + 109981550.357/\exp(X9 \times (3 - X9) + X9 \times (1.170e-7 \times X2) \times X9) \) |
| 21 | 43   | \( Y = 3136263.90 \times X9 + 0.539 \times X2 + 107095038.259/\exp(X9 \times (3 - X9) + X9 \times (1.130e-7 \times X2) \times X9) - 4052739.746 \) |
| 22 | 47   | \( Y = 3440789.075 \times X9 + 0.337 \times X2 + 111136458.994/\exp(X9 \times (3 - X9) + X9 \times (1.177e-7 \times X2) \times X9) - 13177.014 \times X8 \times X9 \) |
| 23 | 49   | \( Y = X7 + 2978684.226 \times X9 + 0.380 \times X2 + 105432852.752/\exp(X9 \times (3 - X9) + X9 \times (1.146e-7 \times X2) \times X9) - 27708.279 \times X8 \times X9 \) |
| 24 | 53   | \( Y = X7 + 2621966.437 \times X9 + 0.450 \times X2 + 102637044.329/\exp(X9 \times (3 - X9) + X9 \times (1.133e-7 \times X2) \times X9) - 1.553e-5 \times X2 \times X9 \times X8 \times 2 \) |
| 25 | 59   | \( Y = X7 + 2871621.405 \times X9 + 0.363 \times X2 + (1.143e-7 \times X2) \times X9 + 106111904.289/\exp(X9 \times (3 - X9) + X9 \times (1.143e-7 \times X2) \times X9) - 25752.410 \times X8 \times X9 \) |
| 26 | 63   | \( Y = X7 + 2845554.756 \times X9 + 0.343 \times X2 + (1.149e-7 \times X2) \times X9 + 10861683.533/\exp(X9 \times (3 - X9) + X9 \times (1.149e-7 \times X2) \times X9) - 1.463e-5 \times X2 \times X9 \times X8 \times 2 \) |
| 27 | 71   | \( Y = X7 + 2901499.795 \times X9 + 0.331 \times X2 + 1.478e-5 \times X2 \times X8 \times X9 + (1.15267e-7 \times X2) \times X9 + 107347804.509/\exp(X9 \times (3 - X9) + X9 \times (1.152e-7 \times X2) \times X9) - 1.4781e-5 \times X2 \times X9 \times X8 \times 2 \) |

The goodness of fit the possible model:

| No | \( R^2 \) | Goodness of Fit | Correlation Coefficient | Mean Squared Error | Mean Absolute Error |
|----|-----------|----------------|-------------------------|--------------------|--------------------|
| 1  | 0.071     | 0.445          | 168600000               | 2.075E+15          | 26215148           |
| 2  | 0.129     | 0.497          | 164600000               | 1.946E+15          | 25231296           |
| 3  | 0.229     | 0.790          | 156850000               | 1.722E+15          | 24902926           |
| 4  | 0.618     | 0.790          | 103590000               | 8.541E+14          | 16079391           |
| 5  | 0.621     | 0.794          | 102628000               | 8.470E+14          | 15854141           |
| 6  | 0.799     | 0.905          | 83999936                | 4.497E+14          | 12142810           |
| 7  | 0.802     | 0.910          | 83999936                | 4.414E+14          | 11292639           |
| 8  | 0.804     | 0.910          | 84000090                | 4.374E+14          | 11142299           |
| 9  | 0.804     | 0.910          | 84000000                | 4.374E+14          | 11142289           |
| 10 | 0.818     | 0.920          | 84002847                | 4.074E+14          | 11198559           |
| 11 | 0.830     | 0.927          | 84000621                | 3.798E+14          | 9974782            |
| 12 | 0.832     | 0.928          | 83977856                | 3.749E+14          | 9581229            |
| 13 | 0.839     | 0.932          | 83809944                | 3.602E+14          | 9246128            |
| 14 | 0.838     | 0.932          | 83997506                | 3.613E+14          | 9474858            |
| 15 | 0.838     | 0.932          | 84003370                | 3.611E+14          | 9304512            |
| 16 | 0.839     | 0.932          | 83995055                | 3.602E+14          | 9171174            |
| 17 | 0.840     | 0.932          | 83934571                | 3.584E+14          | 9149003            |
| 18 | 0.836     | 0.932          | 83998795                | 3.663E+14          | 9067686            |
| 19 | 0.841     | 0.932          | 84330487                | 3.560E+14          | 900744             |
| No | $R^2$ Goodness of Fit | Correlation Coefficient | Maximum Error | Mean Squared Error | Mean Absolute Error |
|----|----------------------|--------------------------|---------------|-------------------|--------------------|
| 20 | 0.848                | 0.937                    | 84022765      | 3.386E+14         | 8686554            |
| 21 | 0.849                | 0.936                    | 84008063      | 3.3632E+14        | 8734292            |
| 22 | 0.850                | 0.938                    | 84000991      | 3.3444E+14        | 8384827            |
| 23 | 0.848                | 0.937                    | 83960022      | 3.3954E+14        | 7849078            |
| 24 | 0.850                | 0.938                    | 83978909      | 3.348E+14         | 7918515            |
| 25 | 0.850                | 0.937                    | 83980395      | 3.3416E+14        | 7318104            |
| 26 | 0.852                | 0.938                    | 84016544      | 3.3078E+14        | 7218226            |
| 27 | 0.852                | 0.938                    | 83999390      | 3.3078E+14        | 7199612            |

Based on the goodness of fit coefficients in the table above, it was shown that the 27th model had the highest $R^2$ value and the lowest Mean Squared Error. From this coefficient, it is proven that the 27th model has better accuracy than the other models. Therefore, the model used to predict the income of fishers in the Karanggonso area of Trenggalek Regency is as follows:

$$Y = X_7 + 2901499.795X_9 + 0.331X_2 + 1.478e - 5X_2X_8X_9 + (1.1526e - 7X_2)^X_8$$

$$+ \frac{107347804.509}{\exp(X_9^{(X_8X_8)} + X_9^{(1.1526e - 7X_2)^X_8})} - 1.478e - 5X_2X_8X_8^2$$

The following is a plot graph between the actual income with the predicted value of the selected model:

**Figure 1.** Scatter plot of actual and predicted fisher's income.

Based on Figure 1, it was shown that the red line, namely the predictive value of fishers's income ($Y$), was close to the actual value of fishers's income. This proves that the model built had a high degree of accuracy.
Likewise, when viewed from the scatter plot (Figure 2), the relationship between the actual value of fishers' income \( (Y) \) and the predicted results shows a positive relationship. The closer to the diagonal line, it indicates that there was a closer relationship, and the predicted value was closer to the true value.

### 3.4. Comparison of linear regression model with flexible model

Based on the results of modelling using linear regression analysis, a linear model was obtained with \( R^2 \) of 70.5\% and MSE of \( 1.086 \times 10^{18} \). In contrast, the flexible model with the deep learning approach obtained 1 model with \( R^2 \) of 85.2\% and MSE of \( 3.308 \times 10^{14} \). The following is a comparison of the prediction results between the linear regression model and the flexible model:

**Figure 2.** Relationship between actual with predicted fisher's income.

**Figure 3.** Comparison of the prediction results between the linear regression model and the flexible model.
Figure 3 shows the scatter plot between actual income and predicted income for each variable. Predicted income from the regression model has a higher distribution, it even has a negative value. It is different from the predicted income from the flexible model where there is no negative predictive value found. As presented in descriptive statistics, it is shown that there is no negative income per week for fishers. From this test, it is shown that the flexible model has a better performance in predicting fishers's income.

4. Conclusions
The multiple linear regression model has an accuracy rate of $R^2 = 70.5\%$ and $\text{MSE} = 1.086 \times 10^{-18}$ with the boat price variable having the most dominant influence on fishers' income. Meanwhile, based on the modelling results with the flexible modelling, a prediction model was obtained with an accuracy rate of $R^2 = 85.2\%$ and $\text{MSE} = 3.308 \times 10^{-14}$. Therefore, it is proven that the flexible model with the machine learning approach has a higher level of accuracy than the linear regression model. Besides, the flexible model also has a non-linear effect on the number of cool boxes and the age of the fishing gear.

References
[1] Rahayu E M 2019 Menteri Susi: Kinerja KKP Triwulan II 2019 Meningkat (Minister Susi: Marine and Fisheries Ministry’s Performance in Quarter II 2019 Increases) SWA 1 [in Indonesian]
[2] Pregiwati L A 2019 KKP Catat Kinerja Positif dl Semester Pertama 2019 (Marine and Fisheries Ministry Records Positive Performance in the First Semester of 2019) Kementeri. Kelaut. dan Perikan. Republik Indonesia. 1 [in Indonesian]
[3] Fajar J 2019 Apa Ancaman Sektor Kelautan Indonesia di Tahun 2019? (What are the threats to Indonesia’s marine sector in 2019?) MONGABAY Situs Ber. Lingkung. 1 [in Indonesian]
[4] Djunaedi 2020 Pandemi Covid-19, Pendapatan Nelayan Donggala Menurun (Covid-19 Pandemic, Donggala Fishermen’s Income Decreased) Metroswalawesi.id [in Indonesian]
[5] Mahrofi Z and Fardaniah R 2020 Pendapatan Nelayan Anjlok, Pemerintah Diminta Beli Tangkapan Nelayan (Fishermen’s income drops, the government is asked to buy fishermen’s catch) Antararavews.com [in Indonesian]
[6] Sari I T P and Rauf M I A 2020 Analisis Pendapatan Usaha Perikanan Tangkap : Pengalaman Dari Nelayan Kabupaten Garut Jawa Barat (Analysis of Capture Fisheries Business Income: The Experience of Fishermen in Garut Regency, West Java) Ekono Insentif 14 12–27 [in Indonesian]
[7] Supriatun F E and Rohman A N 2020 Peramalan Produksi Perikanan Budidaya di Kabupaten Malang Dengan Metode Exponential Smoothing (Forecasting Aquaculture Production in Malang Regency Using Exponential Smoothing Method) Media Akuatika J. Ilm. Jur. Budid. Perair. 5 51–8 [in Indonesian]
[8] Nugroho V R 2017 Analisis Faktor - Faktor yang Mempengaruhi Hasil Pendapatan Nelayan di Desa Bendar Kecamatan Juwana Kabupaten Pati (Analysis of Factors Affecting the Income of Fishermen in Bendar Village, Juwana District, Pati Regency) (Universitas Islam Indonesia Yogyakarta) [in Indonesian]
[9] Talakua W 2013 Analisis Faktor Sosial Ekonomi yang Mempengaruhi Pendapatan Bersih Nelayan Rawai di Kecamatan Teluk Ambon (Analysis of Socio-Economic Factors Affecting Net Income of Rawai Fishermen in Teluk Ambon District) Pena J. Ilmu Pengetah. dan Teknol. 25 22–7 [in Indonesian]
[10] Syahma A 2016 Analisis Faktor-Faktor Yang Mempengaruhi Pendapatan Nelayan Tangkap di Desa Galesong Kota Kecamatan Galesong Kabupaten Takalar (Analysis of Factors Affecting the Income of Caught Fishermen in Galesong Village, Galesong District, Takalar Regency) (Universitas Negeri Makassar) [in Indonesian]
[11] Zuriat 2016 Analisis Pendapatan Nelayan Kapal Motor 5-10 GT di Kabupaten Aceh Barat Daya
(Income Analysis of 5-10 GT Motor Boat Fishermen in Aceh Barat Daya District) *J. Perikan. Trop.* **3** 85–94 [in Indonesian]

[12] Coro G, Large S, Maglioizzi C and Pagano P 2016 Analysing and forecasting fisheries time series: purse seine in Indian Ocean as a case study *ICES J. Mar. Sci.* **73** 2552–71

[13] Eskander S M S U and Barbier E B 2018 Fishing and Non-Fishing Income Decisions: The Role of Human Capital and Family Structure *Land Econ.* **94** 114–36

[14] Razak M A 2017 *Peramalan Jumlah Produksi Ikan dengan Menggunakan Backpropagation Neural Network (Studi Kasus: UPTD Pelabuhan Perikanan Banjarmasin)* (Forecasting the Amount of Fish Production Using the Backpropagation Neural Network (Case Study: UPTD Banjarmasin Fishery)) (Institut Teknologi Sepuluh November Surabaya) [in Indonesian]

[15] Cavieses Núñez R A, Ojeda Ruiz De La Penã M Á, Flores Irigollen A, Rodríguez Rodríguez M and Jardim E 2018 Deep learning models for the prediction of small-scale fisheries catches: Finfish fishery in the region of “bahía Magadalena-Almejas” *ICES J. Mar. Sci.* **75** 2088–96

[16] Primartha R 2018 *Belajar Machine Learning Teori dan Praktik (Study Machine Learning Theory and Practice)* (Bandung: Informatika Bandung) [In Indonesian]