BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1743–1756
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1743-1756

COMPARISON OF LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS FOR PREDICTING FISH CATCH VOLUME IN URENG VILLAGE, CENTRAL MALUKU

Kasriana [1*], Rasid Ode [2], Eryka Lukman [3], Agung K. Henaulu [4]

[1,2]Mathematics Education Study Program, Faculty of Teacher Universitas and Education, Universitas Darussalam Ambon
[3]Aquatic Resource Management Study Program, Faculty of Fisheries, Universitas Darussalam Ambon
[4]Industrial Engineering Study Program, Faculty of Engineering, Universitas Darussalam Ambon
Jln. Waehakila Puncak Wara, Ambon, 97128, Indonesia

*Corresponding author's e-mail: * **kasriana@unidar.ac.id***

Article Info

Article History:
Received: 23rd August 2025
Revised: 12th September 2025
Accepted: 23rd October 2025
Available Online :26th January 2026

Keywords:
Artificial Neural Network;
Fish catch prediction;
Linear Regression;
Model comparison.

ABSTRACT

This study aims to develop a predictive model for fish catch volume in Ureng Village, Central Maluku, using a mathematical modeling approach based on artificial intelligence with the Scikit-Learn and TensorFlow libraries. The research dataset consists of 24 monthly data records collected from July 2024 to June 2025. The data were obtained through a combination of primary and secondary collection methods. Primary data were gathered through interviews, field observations, and fishermen's catch records, while secondary data included oceanographic parameters such as sea surface temperature, weather conditions, and current velocity. Two main models were developed: a linear regression model using Scikit-Learn as the baseline and a neural network model using TensorFlow as the comparator, both trained and evaluated on the same dataset to ensure consistency. The testing results show that the linear regression model produced a Mean Squared Error (MSE) of 0.8821 and a coefficient of determination (R^2) of 0.682, while the neural network model achieved an MSE of 0.5423 and an R^2 of 0.815. These findings indicate that the neural network model is more capable of capturing nonlinear patterns among temperature, weather, and current variables, resulting in higher prediction accuracy than the linear model. Nevertheless, this study is limited by the relatively small sample size and the need for a more detailed description of the data period and measurement units to allow a more objective evaluation of the model's performance. Overall, this AI-based approach has the potential to support more efficient, adaptive, and sustainable decision-making in fishery planning for coastal communities.

How to cite this article:

Kasriana, R. Ode, E. Lukman and A. Henaulu., "COMPARISON OF LINEAR REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS FOR PREDICTING FISH CATCH VOLUME IN URENG VILLAGE, CENTRAL MALUKU", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1743-1756, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Indonesia, as an archipelagic nation with a coastline of more than 108,000 km, has significant potential for capture fisheries, serving as a primary source of livelihood for coastal communities, including those in Ureng Village, Central Maluku. This coastal area is rich in marine biodiversity, encompassing numerous economically important fish species, such as tuna, skipjack, and mackerel, which are leading commodities in both the domestic and international markets. The capture fisheries sector not only contributes to the provision of food and animal protein but also serves as a cornerstone of the local economy through job creation, catch distribution, and strengthening regional food security. However, this vast potential cannot be fully utilized optimally due to challenges in predicting catch yields, which directly impact the operational efficiency and fishing strategies of local fishermen.

Despite the continuous advancement of fishing technology and navigation systems, catch uncertainty remains a major and challenging issue. Environmental factors such as sea surface temperature (SST), ocean current patterns, chlorophyll-a concentration, water clarity, and weather conditions play a significant role in determining the spatial and temporal distribution of fish in the ocean. In addition, operational variables such as departure time, fishing duration, and site selection are also important determinants of both quantity and quality of catch [1]. Such uncertainty not only leads to fluctuations in fishermen's income but also increases the risk of high operational costs and potentially threatens the sustainability of fish resources if fishing strategies are not based on accurate and up-to-date information.

Advances in Artificial Intelligence (AI) and machine learning (ML) technologies provide new opportunities to reduce this uncertainty. By utilizing historical catch records and oceanographic parameters, predictive models can be developed to identify complex patterns among variables [2]. For example, deep learning-based studies integrating multiple environmental variables have successfully enhanced the accuracy of identifying fishing grounds for economically important species [3]. Such models not only improve the efficiency of fishing operations but also contribute to sustainable fisheries management.

A hybrid approach that combines oceanographic variables, fishermen's operational records, and astronomical factors such as lunar phases has proven effective in mapping fishing grounds with great precision [4]. The integration of these data sources offers strategic benefits, particularly for small-scale fisheries operating with limited resources. Furthermore, modeling that considers environmental effects and spatio-temporal factors has shown that small changes in SST or chlorophyll concentration can significantly impact catch quantities [5].

On the other hand, interpretable machine learning-based models provide additional insights for decision-making as they not only generate predictions but also reveal the contribution of each variable to the model results [6]. This feature is crucial for fishermen to understand the key factors influencing fishing success. In fact, ML-based multi-month forecasts for tuna habitat have been successfully implemented in an offshore fisheries context [7], opening opportunities for similar applications in coastal areas such as Central Maluku.

Nevertheless, implementing this technology in remote areas like Ureng Village faces several challenges, including limited access to high-quality data, hardware constraints, and low levels of digital literacy among fishermen. Regression algorithms are widely used in fisheries prediction studies because they can model the relationship between environmental variables and fish catch quantitatively. However, linear regression has limitations in capturing complex and non-linear interactions among oceanographic parameters. Therefore, artificial neural networks (ANNs) are introduced as an alternative due to their strong capability in pattern recognition and non-linear modeling [8]. Despite their potential, ANN models often require large datasets and substantial computational resources, which can be challenging in data-limited regions such as Ureng Village. To address these challenges, this study focuses on developing AI-based mathematical models using Scikit-Learn and TensorFlow while accounting for local constraints. The models integrate oceanographic data (SST, currents, chlorophyll-a), operational fishing data, and catch records to generate accurate, accessible predictions for fishermen via a simple decision-support system. This approach is expected to enhance fishing efficiency, reduce operational risk, and support fisheries sustainability in Ureng Village.

The application of artificial intelligence (AI) and machine learning technologies in the fisheries sector has increasingly gained attention, as they have proven effective in improving the accuracy of catch predictions. Deep learning–based studies have demonstrated that integrating oceanographic variables with spatio-temporal data can accurately identify squid fishing grounds with high precision [9]. Analyses combining AIS data and environmental variables have also developed a comprehensive framework for mapping deep-sea fishing activities, which can be applied to small-scale fisheries [10]. A global review of fisheries management emphasizes that data-driven approaches and AI models contribute significantly to fisheries sustainability and marine conservation [11]. Deep learning approaches for small-scale fisheries have even been used to predict catch volumes, evaluate key variables, and reveal interactions among factors influencing fishing outcomes. In addition, the development of decision support systems based on in-trawl cameras and automated image processing offers new potential for monitoring catches while supporting adaptive fisheries policies [12][13]. Furthermore, previous studies in related fields have demonstrated the effectiveness of neural networks and multivariate analysis for classifying and predicting complex environmental datasets. For instance, analytical chemistry, multivariate data analysis, and neural networks were employed to classify marine oil spill samples using GC-MS and GC-FID data, where Scikit-learn and Keras–TensorFlow were applied to implement Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Neural Network-based models, achieving consistent and reliable predictions compared to traditional univariate statistics. These findings suggest that regression and neural network algorithms can effectively model nonlinear relationships and handle multivariate environmental data, which can be adapted to predict fish catch yields in coastal communities [8].

2. RESEARCH METHODS

This research uses a quantitative, computational-experimental approach, focusing on the development and testing of an AI-based mathematical model to predict fish catches in the coastal area of Ureng Village. The research was conducted in several structured stages, including predictive model design, data collection, model training, validation, and model performance evaluation.

2.1 Research Design

This study employs an exploratory-quantitative design aimed at investigating the potential of integrating linear regression analysis and artificial neural networks (ANN) as a predictive modeling approach for fish catches. Linear regression is used as the baseline model to map linear relationships among predictor variables such as sea surface temperature (SST), current velocity, and fishing duration. Meanwhile, an ANN model with a TensorFlow Sequential architecture is employed to capture nonlinear relationships and multi-factor interactions that linear models cannot adequately address [14][15].

This approach adopts the principles of supervised learning, using historical fish catch records and oceanographic data to train the model to predict outcomes based on marine environmental variables and operational fishing parameters. Previous studies have shown that ANNs can improve predictive accuracy compared to traditional linear models, particularly for seasonal and fluctuating data [16]. This is relevant to this study, considering that oceanographic factors such as seasonal changes in SST, shifts in ocean current patterns, and fluctuations in chlorophyll-a concentrations significantly influence the distribution of fish stocks [4].

The ANN model used in this study is designed to adapt to seasonal characteristics and local geographic conditions, particularly in coastal areas and small islands such as Ureng Village, Central Maluku. A linear regression model is still used as a benchmark to assess the accuracy improvement of the nonlinear approach. This approach is further supported by recent research demonstrating the superior performance of ANN-based models in medium- to long-term fisheries prediction compared to purely statistical methods [17].

The results of this study are expected to produce a predictive model that is not only statistically accurate but also contextually relevant and applicable to local fishermen. Furthermore, these models can serve as the foundation for developing a user-friendly Decision Support System (DSS) that can be easily accessed by fishing communities to support decision-making in fishing activities.

2.2 Data Sources

The primary data for this study were collected through field observations and direct interviews with active fishermen in Ureng Village, Leihitu District, Central Maluku. The data collected included daily catch (in tons), fish species, fishing locations, fishing time, fishing duration, sea temperature, weather, and current strength. Additional data were obtained from the Meteorology, Climatology, and Geophysics Agency

(BMKG) stations and from NOAA oceanographic satellite data to ensure the accuracy and completeness of environmental data.

To ensure representativeness and validity, data were collected over six months (January–June 2025) at a weekly frequency, involving 25 fishing vessels as a fixed sample. Primary data collection was conducted using digital data sheets and GPS location tracking.

2.3 Data Collection Techniques

The data collection was conducted using a participatory approach triangulation, combining structured interviews, direct observations, and daily digital data recording. Primary data were obtained from fishermen, including the type of fishing gear, fish species, catch volume, fishing time, and duration at sea, while secondary data, such as sea temperature, weather conditions, and current velocity, were obtained from BMKG stations. The research team also conducted brief training for the fishermen on how to enter digital data using a simple spreadsheet-based application adapted to the local context. Data were collected monthly from July 2024 to June 2025.

The use of technology in data collection enabled efficiency, timeliness, and facilitated real-time data processing [18]. In addition, sea temperature sensors and current strength meters were used periodically during fishing activities.

2.4 Model Development and Data Analysis

Data processing was carried out in several stages: data preprocessing, data splitting (training and test), model training, and performance evaluation. The data preprocessing stage was carried out to ensure the dataset's quality and consistency before model training. In this stage, data cleaning was performed to remove missing, duplicate, or inconsistent data. Numerical variables were then normalized using min-max scaling to scale values between 0 and 1, ensuring that all features have balanced weights during model training. If categorical features were present, encoding (e.g., one-hot encoding) was applied so that the data could be processed by machine learning algorithms.

After preprocessing, the dataset was divided into two subsets: 80% for the training set and 20% for the testing set. The training set was used to train the model to learn the relationship patterns between predictor variables and the target variable, while the testing set was used to evaluate the model's ability to predict new, unseen data. This division is important to prevent overfitting and to ensure that model performance evaluation is conducted objectively.

During the training stage, two models were developed using the training data. The first model was Linear Regression, used as a baseline to map linear relationships among predictor variables such as sea surface temperature, current velocity, and fishing duration. The second model was an Artificial Neural Network (ANN) with a TensorFlow Sequential architecture, designed to capture nonlinear relationships and complex interactions among variables that cannot be explained by the linear model.

After training, model performance was evaluated using Mean Squared Error (MSE) and R-squared (R^2) metrics. MSE measures the average squared difference between predicted and actual values, while R^2 indicates the proportion of data variability explained by the model. This evaluation provides insights into the prediction accuracy and effectiveness of the model in capturing patterns in the data.

The linear regression model was developed using the scikit-learn library with the Ordinary Least Squares (OLS) algorithm. Meanwhile, the artificial neural network model was built using the TensorFlow Sequential API, with a multilayer perceptron architecture comprising an input layer, two hidden layers with ReLU activation, and an output layer with a linear activation. Optimization was performed using the Adam algorithm and the mean squared error (MSE) loss function [19].

Model performance evaluation was conducted by calculating Mean Squared Error (MSE) and R-squared (R^2). These metrics are essential for assessing how well the model fits the data. MSE measures the average squared difference between the predicted and actual values, indicating the overall magnitude of prediction error; a smaller MSE value indicates higher prediction accuracy. Meanwhile, R^2 (Coefficient of Determination) measures the proportion of variance in the dependent variable that can be explained by the independent variables in the model, where a higher R^2 value indicates a better fit of the model to the data.

In addition, statistical analyses were performed to evaluate the significance and validity of the model parameters**.** These analyses included residual normality testing (Kolmogorov–Smirnov or Shapiro–Wilk), multicollinearity testing using the Variance Inflation Factor (VIF), and hypothesis testing of the regression coefficients (t-test) as well as the overall model (F-test) at a 95% confidence level. All computational and statistical analyses were performed using the Python programming language with supporting libraries such as *Pandas*, *NumPy*, *Matplotlib*, *Scikit-learn*, *Statsmodels*, and *TensorFlow*.

Formulas:

1. Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2,$$

where:

y_i : actual value;

$\hat{y}i$: predicted value;

n : number of observations.

2. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

where:

y_i : actual value;

$\hat{y}_i$: predicted value;

$\bar{y}$: mean of actual values.

All computations, including the calculation of MSE, R^2, and statistical significance tests, were automatically performed using Python-based libraries to ensure reproducibility and accuracy of the results.

To visualize the methodological sequence of this research, a structured research flowchart is presented. This flowchart serves to illustrate the logical order of activities carried out throughout the study, from the initial conceptual phase to the final analysis and reporting stage. It provides a concise overview of how each component is interconnected and contributes to achieving the research objectives.

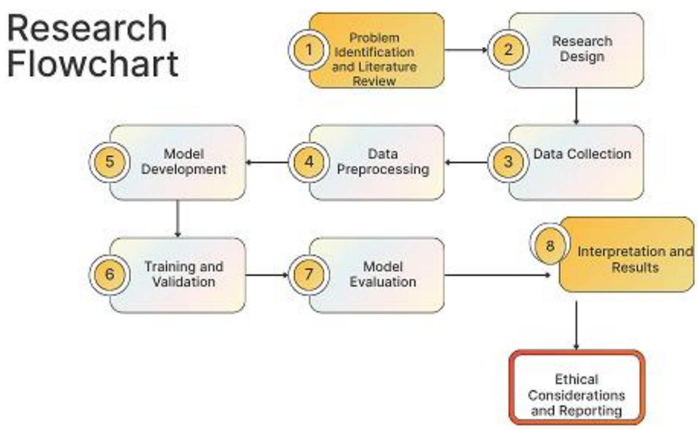

Figure 1. Research Flow Diagram

As shown in Fig. 1, the research process begins with problem identification and literature review, which form the foundation for designing the research framework. The subsequent steps involve data collection, preprocessing, and model development, followed by training, validation, and evaluation stages to ensure model reliability. Finally, the interpretation of results and ethical reporting complete the process, ensuring that the study maintains both scientific rigor and ethical standards.

2.5 Ethical Considerations

This research has obtained permission from village officials and includes informed consent from participating fishermen. All personal data collected was anonymized and will be used solely for research purposes. The researchers also guarantee that there will be no negative interference in the community's economic activities during the research process.

3. RESULTS AND DISCUSSION

3.1. Description of Research Data

This study collected monthly fish catch data from fishermen in Ureng Village from July 2024 to June 2025. The dataset comprises both operational variables and environmental variables that are considered to influence catch volume. The operational variables include fishing gear (nets or handlines), fish species (small pelagic fish or tuna**)**, fishing time (departure time), and fishing duration (number of days spent fishing each month). The environmental variables include sea surface temperature (°C)**,** weather condition**s**, and current velocity (m/s)**,** which were obtained from BMKG oceanographic data stations.

Table 1 summarizes the monthly fish catch and the associated environmental conditions. From the table, it is evident that nets consistently yielded higher catch volumes of small pelagic fish (5.0–8.1 tons) compared to handlines targeting tuna (2.2–4.2 tons)**.** Fishing schedules were relatively fixed: nets were operated in the morning for 30 days per month, while handlines were deployed at dawn for 24 days per month. This indicates that catch variability was primarily driven by environmental conditions rather than differences in fishing effort.

Seasonal trends are also observable in the data. Sea surface temperature increased from 25.8 °C in July 2024 to 30.19 °C in January 2025 and then slightly declined, while current velocity reached a peak of 7.04 m/s in June 2025 during the heavy rainy season. These environmental changes corresponded to a reduction in catch volumes for both species, highlighting the sensitivity of fish catch to oceanographic variability. Overall, the table demonstrates that catch outcomes are strongly influenced by seasonal patterns of temperature, rainfall, and current velocity. The descriptive evidence provided by these data forms the empirical foundation for predictive modeling, where operational and environmental variables such as fishing gear, departure time, sea temperature, and current velocity are expected to play a dominant role in explaining fish catch variability.

Table 1. Monthly Summary of Fish Catch, Fishing Effort, and Environmental Conditions (July 2024 – June 2025)

Time (Month/ Year)	Fishing Gear	Fish Species	Catch Volume (tons)	Departure Time (Hours)	Fishing Duration (Days)	Sea Temp (°C)	Weather	Current Velocity (m/s)
July 2024	Net	Small pelagic fish	8.1	Morning	30	25.8	Light Rain	1.87
July 2024	Handline	Tuna	4.2	Dawn	24	25.8	Light Rain	1.87
August 2024	Net	Small pelagic fish	7.6	Morning	30	25.55	Cloudy	2.96
August 2024	Handline	Tuna	3.9	Dawn	24	25.55	Cloudy	2.96
September 2024	Net	Small pelagic fish	6.9	Morning	30	26.5	Partly Cloudy	2.51
September 2024	Handline	Tuna	3.6	Dawn	24	26.5	Partly Cloudy	2.51
October 2024	Net	Small pelagic fish	6.7	Morning	30	28.0	Partly Cloudy	2.84
October 2024	Handline	Tuna	3.5	Dawn	24	28.0	Partly Cloudy	2.84

Time (Month/ Year)	Fishing Gear	Fish Species	Catch Volume (tons)	Departure Time (Hours)	Fishing Duration (Days)	Sea Temp (°C)	Weather	Current Velocity (m/s)
November 2024	Net	Small pelagic fish	6.5	Morning	30	28.45	Partly Cloudy	2.38
November 2024	Handline	Tuna	3.5	Dawn	24	28.45	Partly Cloudy	2.38
December 2024	Net	Small pelagic fish	6.11	Morning	30	29.25	Cloudy	3.41
December 2024	Handline	Tuna	3.32	Dawn	24	29.25	Cloudy	3.41
January 2025	Net	Small pelagic fish	5.94	Morning	30	30.19	Light Rain	5.03
January 2025	Handline	Tuna	3.22	Dawn	24	30.19	Light Rain	5.03
February 2025	Net	Small pelagic fish	6.11	Morning	30	29.87	Light Rain	4.05
February 2025	Handline	Tuna	3	Dawn	24	29.87	Light Rain	4.05
March 2025	Net	Small pelagic fish	5.87	Morning	30	28.15	Light Rain	5.22
March 2025	Handline	Tuna	3.43	Dawn	24	28.15	Light Rain	5.22
April 2025	Net	Small pelagic fish	6.23	Morning	30	26.95	Light Rain	4.12
April 2025	Handline	Tuna	3.5	Dawn	24	26.95	Light Rain	4.12
May 2025	Net	Small pelagic fish	5.2	Morning	30	27.3	Moderate Rain	4.33
May 2025	Handline	Tuna	2.8	Dawn	24	27.3	Moderate Rain	4.33
June 2025	Net	Small pelagic fish	5	Morning	30	26.3	Heavy Rain	7.04
June 2025	Handline	Tuna	2.203	Dawn	24	26.3	Heavy Rain	7.04

Table 1 shows that nets consistently yielded higher catch volumes of small pelagic fish (5.0–8.1 tons) compared to handlines targeting tuna (2.2–4.2 tons). Fishing schedules were relatively fixed, with nets operated in the morning for 30 days and handlines at dawn for 24 days each month, which indicates that catch variability is primarily driven by environmental conditions rather than differences in fishing effort. Seasonal changes were evident: sea surface temperature rose from 25.8 °C in July 2024 to 30.19 °C in January 2025 before declining slightly, while current velocity peaked at 7.04 m/s in June 2025 during the heavy rainy season. These dynamics coincided with reduced catch volumes for both species, highlighting the sensitivity of fishing outcomes to oceanographic variability.

Overall, the table demonstrates that catch yields are strongly influenced by seasonal patterns of temperature, rainfall, and current velocity. This descriptive evidence provides the empirical foundation for developing predictive models, where variables such as temperature, salinity, and current velocity are expected to play a dominant role in explaining fish catch variability.

3.2. Data Pre-Processing Results

The observational data used consisted of 24 samples with input variables of sea temperature, salinity, and water depth, and the target variable of fish catch volume. The data was read from the data_tangkapan.xlsx file.

The pre-processing steps were as follows:

1. Feature Selection: The temperature, salinity, and depth columns were used as predictor features, and catch_result as the target variable.

2. Dataset Division: The data was divided into training data (70%, 17 samples) and test data (30%, 7 samples) using train_test_split with random_state = 42.
3. Data Normalization: For Neural Network-based models, features were normalized using MinMaxScaler to ensure all values were within the range 0–1.

3.3. Modeling Using Scikit-Learn (Linear Regression)

A linear regression model is used as a simple baseline to see the linear relationship between oceanographic variables and fish catches.

```
Python:
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Read data
data = pd.read_excel("fish_catch.xlsx")

# Feature selection and target definition
X = data[['temperature', 'salinity', 'depth']]
y = data['catch_volume']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Linear regression model
model_lr = LinearRegression()
model_lr.fit(X_train, y_train)

# Prediction results
y_pred = model_lr.predict(X_test)

# Evaluation
print("MSE:", mean_squared_error(y_test, y_pred))
print("R² Score:", r2_score(y_test, y_pred))
```

3.4 Evaluation results

The MSE value of 0.8821 indicates that the average squared difference between the predicted and actual data is relatively small, suggesting that the linear regression model has a fairly good predictive capability. MSE measures the magnitude of prediction errors in squared units; therefore, the smaller the value, the more accurate the model in predicting the fish catch volume based on temperature, current strength, and weather variables. The fact that the MSE is nonzero implies that some deviations between the predicted and actual results remain, indicating that the model's accuracy can be further improved through parameter optimization or additional training data.

Meanwhile, the R^2 value of 0.682 indicates that approximately 68.2% of the variation in fish catch volume can be explained by the variations in temperature, current strength, and weather, while the remaining 31.8% is influenced by other factors not included in the model, such as differences in fish species, more complex oceanic conditions, or nonlinear environmental factors. The relatively high R^2 value indicates that the model captures a significant portion of the linear relationship between the oceanographic variables and fish catch volume, although there is still room to improve predictive performance.

Fig. 2 shows the predicted fish catch based on Scikit-learn (linear regression).

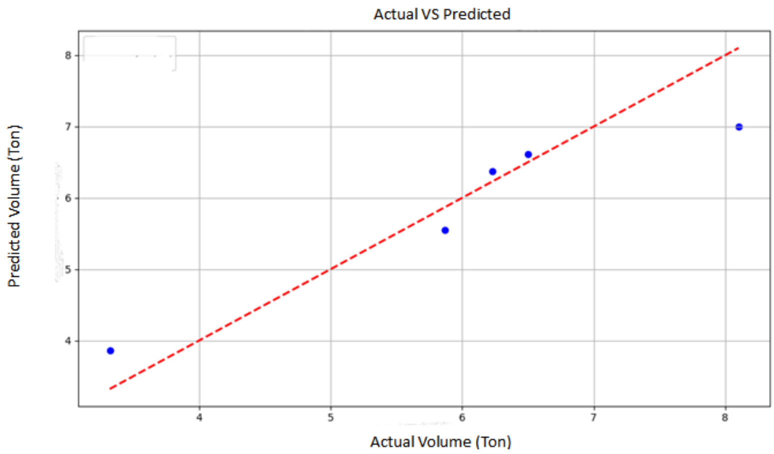

Figure 2. Fish Catch Prediction based on Scikit Learn (Linear Regression)

The scatter plot above illustrates the comparison between the actual fish catch volume (x-axis) and the values predicted by the linear regression model (y-axis). The red dashed line represents the ideal scenario where the predicted values equal the actual values. The closer the blue points are to this line, the higher the model's accuracy in predicting fish catch volumes based on the input variables.

In this study, the dataset was divided into 80% for training and 20% for testing. The training data were used to construct the model and to learn the relationships between the predictor variables—namely, temperature, current strength, and weather—and the fish catch volume. Meanwhile, the testing data were employed to evaluate the model's performance on unseen data. The test results show that most data points are close to the diagonal line, indicating that the model has strong predictive capability. Nevertheless, there are some minor deviations from the red line, reflecting differences between the predicted and actual values. These differences contribute to a Mean Squared Error (MSE) of 0.8821, indicating that the average squared difference between the actual and predicted values is relatively small. Overall, the model explains approximately 68.2% ($R^2 = 0.682$) of the variance in fish catch volume based on temperature, current, and weather, although some data complexities are not fully captured by this linear model.

3.5 Modeling Using TensorFlow (Neural Network)

The Neural Network model is designed to study nonlinear patterns between oceanographic variables and fish catches.

```
python
CopyEdit
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
from sklearn.metrics import mean_squared_error, r2_score

# Loading data
data = pd.read_excel("fish_catch_data.xlsx")

# Selection of features and target
X = data[['temperature', 'salinity', 'depth']]
y = data['catch_volume']

# Normalization of features
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

# Split data
```

```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y,
test_size=0.3, random_state=42)

# Develop a Neural Network Model
model_nn = Sequential([
    Dense(16, input_dim=X_train.shape[1], activation='relu'),
    Dense(8, activation='relu'),
    Dense(1)])

# Compile & train
model_nn.compile(optimizer=Adam(learning_rate=0.01), loss='mse')
model_nn.fit(X_train, y_train, epochs=100, verbose=0)

# Prediction and Model Evaluation
y_pred_nn = model_nn.predict(X_test).flatten()
print("MSE:", mean_squared_error(y_test, y_pred_nn))
print("R² Score:", r2_score(y_test, y_pred_nn
```

3.6 Evaluation Results

The MSE value of 0.5423 indicates that the mean squared difference between the predicted values and the actual data is relatively small, suggesting a lower prediction error rate than the linear regression model. Meanwhile, the R^2 Score of 0.815 indicates that the model explains approximately 81.5% of the variation in fish catch volume, suggesting that the neural network model has more accurate and effective predictive capabilities for capturing the complex relationships among oceanographic variables. Fig. 3 shows the predicted fish catch based on TensorFlow.

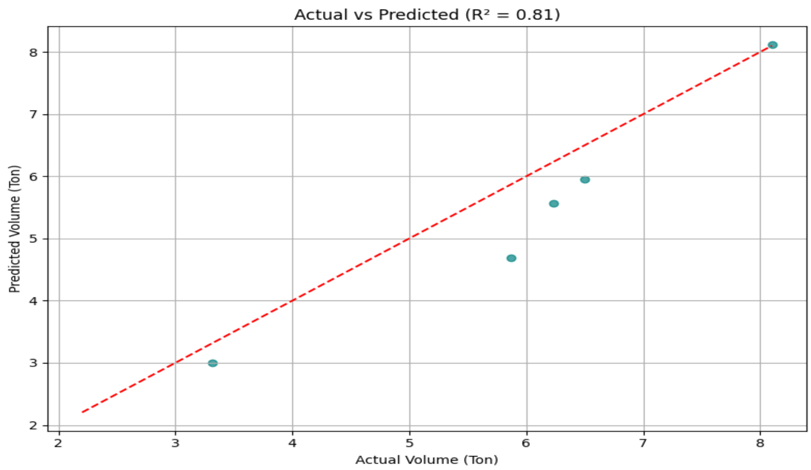

Figure 3. Fish Catch Prediction based on TensorFlow

Fig. 3 depicts a comparison between actual and predicted fish catch volume. The red dashed line represents the ideal scenario where predicted values equal actual values. The model shows good predictive performance with a Mean Squared Error (MSE) of 0.5423 and an R^2 score of 0.815, indicating that approximately 81.5% of the variance in fish catch volume is explained by the model.

3.7. Comparison of Model Results

In this study, the dataset was divided into 80% for training and 20% for testing to ensure that the model could learn effectively while maintaining sufficient data for evaluation. To see the performance differences between the linear regression model and neural network model, the Mean Squared Error (MSE) and R^2 scores on the test data were compared. A lower MSE value indicates a smaller prediction error, while a higher R^2 value indicates a better model's ability to explain data variation. The comparison results of the two models are presented in Table 2.

Table 2. **Performance Differences between Sci-Learn and TensorFlow Models**

Model	MSE	R^2 Score
Linear Regression	0.8821	0.682
Neural Network	0.5423	0.815

Table 2 shows that the Neural Network model achieves lower MSE and higher R^2 score than the Linear Regression model. This result indicates that the Neural Network is better at capturing nonlinear relationships within the dataset, leading to more accurate predictions of fish catch volumes.

3.8 Discussion

The evaluation of the linear regression model using Scikit-learn showed an MSE of 0.8821 and an R^2 of 0.682. The R^2 value indicates that the model explains approximately 68.2% of the variation in fish catch volume based on temperature, salinity, and water depth. Although the MSE value is relatively small, the discrepancy between the predicted and actual values indicates that this model still has limitations in capturing complex patterns among oceanographic variables. This finding aligns with those who argue that linear regression tends to be less optimal for multivariate and nonlinear environmental data, often resulting in lower performance than ensemble learning methods that can better model variable interactions.

In contrast, the model trained with a TensorFlow-based Neural Network produced an MSE of 0.5423 and an R^2 of 0.815. These values indicate a significant improvement over linear regression, both in predictive accuracy and in the model's ability to explain data variation (81.5%). This superior performance can be explained by neural networks' ability to learn nonlinear and complex relationships among variables, which linear models struggle to capture. According to Denny Arbahri et al. (2024), the use of nonlinear algorithms such as Decision Trees and Random Forests in oceanographic data analysis can yield low MSE and high R^2, even near-perfect values, due to their ability to capture complex interactions among variables [19].

The improvement in neural network model accuracy is also consistent with McMillan's [20], which shows that artificial neural networks have advantages for modeling fisheries systems because they can dynamically adapt their parameters to capture complex changes in environmental conditions. In other words, neural networks not only provide more accurate predictions but are also more reliable when applied to dynamic and variable oceanographic data [21].

Based on the performance comparison between the two models, it is clear that neural networks have lower MSE and a higher R^2 than linear regression. This indicates that neural networks are more appropriate for predicting fish catches, which are influenced by various environmental factors that interact nonlinearly. Nevertheless, linear regression remains valuable as a simple, fast, and easily interpretable baseline model, making it useful in the initial stages of analysis before applying more complex models.

4. CONCLUSION

The application of mathematical modeling with Scikit-Learn and TensorFlow demonstrates that artificial intelligence (AI) has significant potential to improve the accuracy of fish catch predictions based on oceanographic and temporal variables. This study confirms that linear regression (Scikit-Learn) can provide reliable baseline results, while deep learning models (TensorFlow) can capture more complex and nonlinear patterns in marine data. These findings demonstrate the importance of integrating machine learning in fisheries management, particularly for small-scale fishers in coastal areas who rely heavily on accurate predictive information. This model not only strengthens decision-making but also supports sustainability efforts by enabling more precise and efficient fishing practices. The results of this study emphasize the relevance of data-driven approaches in addressing real-world challenges in the marine sector and offer solutions that can be further developed for resource prediction and coastal development strategies.

Author Contributions

Kasriana: Conceptualization, Methodology Development, Model Development, Formal Analysis, Original Draft, Visualizations, Research Findings. Rasid Ode: Collecting Data, Validating the findings, Project Administration, Writing, Reviewing, and Editing. Eryka Lukman: Data Curation, Supervision, Review, and

Editing. Agung Henaulu: Collecting Data. All authors have read, discussed, and approved the final manuscript submitted.

Funding Statement

This research was funded by the Ministry of Education, Culture, Research, and Technology through the Beginner Lecturer Research (PDP) scheme for the 2025 funding year.

Acknowledgment

The authors would like to thank the Ministry of Education, Culture, Research, and Technology for their financial support through the 2025 Beginner Lecturer Research (PDP) scheme. Appreciation is also extended to the Oceanographic Data Center for providing access to the oceanographic datasets used in this research. The authors appreciate the assistance of the Fisheries Technology Laboratory team and colleagues who provided input, technical assistance, and support throughout the research process and the writing of this article. Sincere thanks are also extended to the fishermen in the research area who were willing to share information, field experiences, and valuable fishing data, which were invaluable to the success of this study.

Declarations

The authors declare that this research was conducted independently and is free from any direct or indirect conflicts of interest that could influence the research process, data interpretation, or conclusions. There are no personal, financial, or professional affiliations that could be construed as competing interests in the context of this research. The entire research process was conducted with academic integrity and full transparency.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] B. Kühn *et al.*, "MACHINE LEARNING APPLICATIONS FOR FISHERIES—AT SCALES FROM GENOMICS TO ECOSYSTEMS," 2025, *Taylor and Francis Ltd.* doi: https://doi.org/10.1080/23308249.2024.2423189.

[2] J. T. Watson *et al.*, "FISHERY CATCH RECORDS SUPPORT MACHINE LEARNING-BASED PREDICTION OF ILLEGAL FISHING OFF US WEST COAST," *PeerJ*, vol. 11, 2023, doi: https://doi.org/10.7717/peerj.16215.

[3] M. Xie, B. Liu, and X. Chen, "DEEP LEARNING-BASED FISHING GROUND PREDICTION WITH MULTIPLE ENVIRONMENTAL FACTORS," *Mar. Life Sci. Technol.*, vol. 6, no. 4, pp. 736–749, Nov. 2024, doi: https://doi.org/10.1007/s42995-024-00222-4.

[4] H. Han *et al.*, "A NEW MODELING STRATEGY FOR THE PREDICTIVE MODEL OF CHUB MACKEREL (SCOMBER JAPONICUS) CENTRAL FISHING GROUNDS IN THE NORTHWEST PACIFIC OCEAN BASED ON MACHINE LEARNING AND OPERATIONAL CHARACTERISTICS OF THE LIGHT FISHING VESSELS," *Front. Mar. Sci.*, vol. 11, 2024, doi: https://doi.org/10.3389/fmars.2024.1451104.

[5] Y. Shi *et al.*, "REVEALING THE EFFECTS OF ENVIRONMENTAL AND SPATIO-TEMPORAL VARIABLES ON CHANGES IN JAPANESE SARDINE (SARDINOPS MELANOSTICTUS) HIGH ABUNDANCE FISHING GROUNDS BASED ON INTERPRETABLE MACHINE LEARNING APPROACH," *Front. Mar. Sci.*, vol. 11, 2024, doi: https://doi.org/10.3389/fmars.2024.1503292.

[6] E. Gilman and M. Chaloupka, "EVIDENCE FROM INTERPRETABLE MACHINE LEARNING TO INFORM SPATIAL MANAGEMENT OF PALAU'S TUNA FISHERIES," *Ecosphere*, vol. 15, no. 2, Feb. 2024, doi: https://doi.org/10.1002/ecs2.4751.

[7] I. Galparsoro *et al.*, "PREDICTING IMPORTANT FISHING GROUNDS FOR THE SMALL-SCALE FISHERY, BASED ON AUTOMATIC IDENTIFICATION SYSTEM RECORDS, CATCHES, AND ENVIRONMENTAL DATA," *ICES J. Mar. Sci.*, vol. 81, no. 3, pp. 453–469, Apr. 2024, doi: https://doi.org/10.1093/icesjms/fsae006.

[8] B. N. Jensen, "OIL SPILL FORENSICS – IDENTIFICATION OF SOURCES OF OIL SPILLS BY COMBINING ANALYTICAL CHEMISTRY, MULTIVARIATE DATA ANALYSIS AND NEURAL NETWORKS," no. May, 2025.

[9] M. Xie, B. Liu, X. Chen, W. Yu, and J. Wang, "DEEP LEARNING-BASED FISHING GROUND PREDICTION USING ASYMMETRIC SPATIOTEMPORAL SCALES: A CASE STUDY OF OMMASTREPHES BARTRAMII," *Fishes*, vol. 9, no. 2, 2024, doi: https://doi.org/10.3390/fishes9020064.

[10] F. Han, Y. Liu, H. Tian, J. Li, and Y. Tian, "A COMPREHENSIVE FRAMEWORK INCORPORATING DEEP LEARNING FOR ANALYZING FISHING VESSEL ACTIVITY USING AUTOMATIC IDENTIFICATION SYSTEM DATA," *ICES J. Mar. Sci.*, vol. 82, no. 2, 2025, doi: https://doi.org/10.1093/icesjms/fsae166.

[11] D. Petza *et al.*, "CONTRIBUTION OF AREA-BASED FISHERIES MANAGEMENT MEASURES TO FISHERIES

SUSTAINABILITY AND MARINE CONSERVATION: A GLOBAL SCOPING REVIEW," *Rev. Fish Biol. Fish.*, vol. 33, no. 4, pp. 1049–1073, 2023, doi: https://doi.org/10.1007/s11160-023-09780-9.

[12] I. J. Biosci and R. M. Dellosa, "A MACHINE LEARNING PREDICTION OF THE FISHERIES PRODUCTION IN THE PHILIPPINES USING WEKA," *Int. J. Biosci.*, vol. 6655, pp. 162–171, 2023, doi: https://doi.org/10.12692/ijb/23.1.162-171.

[13] M. Munkholm and D. Version, "MONITORING USING IN-TRAWL CAMERAS AND AUTOMATIC IMAGE PROCESSING DEVELOPMENT OF DECISION SUPPORT TOOLS FOR COMMERCIAL FISHERIES AND FOR FISHERIES MONITORING USING IN-TRAWL CAMERAS AND AUTOMATIC IMAGE PROCESSING NATIONAL INSTITUTE OF AQUATIC RESOURCES," 2024.

[14] M. B. Saif and M. Khanam, "COMPARING ARIMA, NEURAL NETWORK AND HYBRID MODELS FOR FORECASTING FISH PRODUCTION IN BANGLADESH," *Dhaka Univ. J. Sci.*, vol. 72, no. 1, pp. 71–76, Mar. 2024, doi: https://doi.org/10.3329/dujs.v72i1.71192.

[15] A. Shamsuddin and N. Mohamed, "A COMPARISON OF PREDICTIVE MODELLING TECHNIQUES FOR REDUCING PRICE VOLATALITY IN THE MALAYSIA SECTOR," *J. Math. Sci. Informatics*, vol. 3, no. 2, pp. 1–9, 2023, doi: https://doi.org/10.46754/jmsi.2023.12.001.

[16] S. K. Stephen, V. K. Yadav, and R. R. Kumar, "COMPARATIVE STUDY OF STATISTICAL AND MACHINE LEARNING TECHNIQUES FOR FISH PRODUCTION FORECASTING IN ANDHRA PRADESH UNDER CLIMATE CHANGE SCENARIO," *Indian J. Geo-Marine Sci.*, vol. 51, no. 9, pp. 776–784, Sep. 2022, doi: https://doi.org/10.56042/ijms.v51i09.2337.

[17] H. Yuan, Y. Gu, J. Wang, Y. Chen, and X. Chen, "STUDY ON THE MEDIUM AND LONG TERM OF FISHERY FORECASTING BASED ON NEURAL NETWORK," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7530 LNAI, pp. 626–633, 2023, doi: https://doi.org/10.1007/978-3-642-33478-8_77.

[18] L. Alfaris *et al.*, "PREDICTING OCEAN CURRENT TEMPERATURE OFF THE EAST COAST OF AMERICA WITH XGBOOST AND RANDOM FOREST ALGORITHMS USING RSTUDIO," *Ilmu Kelaut. Indones. J. Mar. Sci.*, vol. 29, no. 2, pp. 273–284, 2024, doi: https://doi.org/10.14710/ik.ijms.29.2.273-284.

[19] D. Arbahri, O. D. Nurhayati, and I. Mudita, "MACHINE LEARNING OCEANOGRAPHIC DATA FOR PREDICTION OF THE POTENTIAL OF MARINE RESOURCES," *J. Comput. Sci.*, vol. 20Arbahri, no. 2, pp. 129–139, 2024, doi: https://doi.org/10.3844/jcssp.2024.129.139.

[20] by M. Patrick Millan, "REGULARIZED REGRESSION METHODS AND NEURAL NETWORKS FOR MODELING FISH POPULATION HEALTH WITH WATER QUALITY VARIABLES IN THE ATHABASCA OIL SANDS REGION."

[21] P. B. Ngor *et al.*, "PREDICTING FISH SPECIES RICHNESS AND ABUNDANCE IN THE LOWER MEKONG BASIN," *Front. Ecol. Evol.*, vol. 11, no. June, 2023, doi: https://doi.org/10.3389/fevo.2023.1131142.