FishMAZE: Fish Monitoring and AI-based Zone Evaluation

Aldrin Lambon¹, Erica Faye Sagun¹, Myra Saet¹, Zyrah Maranon², and Sharmen Berlin¹

¹University of the Philippines (UP), Philippines
²Ateneo de Manila University (ADMU), Philippines

Abstract
FishMAZE, Fish Monitoring, and AI-based Zone Evaluation is an analytics and reporting solution to help solve the following: prediction of fish likelihood on a coordinate, creating a user-friendly visualization for fishermen, and recommending sustainable fishing routes. Environmental data, historical catch notes, and coordinate data were used to train a regression model for fish likelihood prediction. By combining both the prediction model and the interactive dashboard, a fishing plan based on maximum fish catch was visualized. The final model scored an overall RMSE of 8.6830 against the actual coordinates of the test data provided.

Keywords: artificial intelligence; machine learning; fishing; automatic reporting

Introduction
According to the United Nations Department of Economic and Social Affairs, the global population is projected to reach around 9.7 billion by 2050 [1]. This will create a lot of demand for food and resource management. One source of food for humanity is supplied by fisheries. The global demand for fish has roughly doubled since the turn of the 21st Century and will likely double again by 2050, assuming constant real prices for fish [2]. This study will focus on Norway, which is considered Europe’s largest fishing and aquaculture nation. This study will help provide a solution for improving the core operation of searching and catching fish in a sustainability context [3].

Smederevac-Lali et al. [4] utilized environmental parameters and catch statistics for fish catch prediction. The related study was implemented for Serbia. For this research, methodologies from the above-mentioned study were applied to the available data for Norway. This study will provide a basis for an end-to-end data product for Fish Monitoring and AI-based Zone Evaluation (FishMAZE). FishMAZE will help provide real-time fish data insights and an automated fish zone evaluation of catch likelihood based on historical patterns.

Data Details
The scope of this study will cover only ten fish species: Herring, Mackerel, Haddock, Pollock (Saithe), Acadian Redfish, European Pollock, Cod, Angler (Monkfish), Common Ling, and Ballan Wrasse for the analysis. Moreover, Catch Notes Data (2015 - 2022) for vessels larger than 15 meters was used. This contains information about the manually logged fish catch during landing. There are approximately 130 data fields and around one million catch notes each year. Salinity Data (2015 - 2022) is also provided from the SMAP Salinity V4 dataset [5]. Sea Surface Temperature (SST) data from 1981 to 2022, containing daily estimates of SST globally, is also available. Salinity, in combination with temperature, affects the growth rate of microalgae. Furthermore, Moon Phase Data consisting of dates and exact times of full moon from the year 1900 to 2050 was used.

Methodology
The model created in this study predicts the optimal coordinates a vessel should prioritize in order to maximize the likelihood of catching a specific type of fish based on historical data. Due to the limitation of the available data, the scope of the study is from 2015-2021 only. A database was utilized to store and process big data from catch notes and environmental data. Python [6] scripts were created for data transformation of different data sets. Fish zones, fishing areas, and vessel information from catch notes were identified as relevant features for the model. These features were one-hot encoded to enable the model to interpret the data correctly as categorical data. The number of days before a full moon, salinity, sea surface temperature, and wind speed were also used as model inputs to infer the behavior of different...
species. Combining these environmental variables with the catch notes produced a total of 2,067,724 data points from 2015 to 2022. Product weights for each vessel (determined by the vessel id) for each date and location were normalized by dividing the summed weights by the number of lines per vessel. This can be interpreted as the product weight per vessel per line. It was further normalized by averaging the product weight for each month and species were given environmental data. As seen in Figure 1 (left), product weights are heavily right skewed which was transformed using the Gaussian distribution shown in Figure 1 (right). These are also applied to all quantitative features, namely sea surface temperature, wind speed, salinity, air pressure, and days before a full moon.

The transformed data were split into training and testing data with a test size equal to 30 percent of the data. Training data contained 10,390 data points, and the testing data contained 4,453 data points. Note that the number of data points significantly decreased due to averaging. The year was also dropped from the analysis, which also contributed to decreased data size.

Several models and hyper-parameters were examined using the Optuna package [7], and a Random Forest Regression model provided the best results. In order to produce the final prediction in the selected dates and top coordinates, all locations for a specific evaluation date were predicted, and only the top coordinates that give the highest catch value were selected. The model was evaluated using the Root Mean Square Error (RMSE) of the predicted coordinates measured against the actual coordinates values.

Results
A summary of the RMSE metrics for evaluation dates is presented in Table 1. The overall RMSE for unseen and evaluation data was 8.6830. A visualization sample prediction is marked as blue against the ground, and the truth is marked red in Figure 2. Based on the results presented in Table 1, our model got its best prediction for 11/10/2022 with an RMSE of 6.1687, and the model performed worst on 11/15/2022 with an RMSE of 11.7475.

<table>
<thead>
<tr>
<th>Evaluation Date</th>
<th>RMSE</th>
<th>RMSE Lower CI</th>
<th>RMSE Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/10/2022</td>
<td>8.6274</td>
<td>7.3987</td>
<td>9.7018</td>
</tr>
<tr>
<td>11/10/2022</td>
<td>6.1687</td>
<td>4.9735</td>
<td>7.1673</td>
</tr>
<tr>
<td>12/10/2022</td>
<td>7.7304</td>
<td>5.7628</td>
<td>9.2902</td>
</tr>
<tr>
<td>13/10/2022</td>
<td>7.3753</td>
<td>5.5526</td>
<td>8.8295</td>
</tr>
<tr>
<td>14/10/2022</td>
<td>8.0852</td>
<td>7.0854</td>
<td>8.9744</td>
</tr>
<tr>
<td>15/10/2022</td>
<td>11.7475</td>
<td>7.4257</td>
<td>14.8616</td>
</tr>
<tr>
<td>16/10/2022</td>
<td>9.8868</td>
<td>7.7700</td>
<td>11.6244</td>
</tr>
<tr>
<td>Overall</td>
<td>8.6830</td>
<td>6.6465</td>
<td>10.3254</td>
</tr>
</tbody>
</table>

Table 1: RMSE Evaluation

Conclusion
The catch pattern of the model prediction appears to cluster near the actual fish zone and fishing areas. Additional environmental and COVID-related parameters may improve the model’s accuracy. The predictions were utilized to build a weekly fishing plan dashboard in Figure 3 that can be used by fishermen and vessel operators. Future iterations of the model could explore advanced machine learning models and introduce real-time analytics to the dashboard.

Conflict of interest
There is no conflict of interest.
References


