



# Article The Study of Fishing Vessel Behavior Identification Based on AIS Data: A Case Study of the East China Sea

Bowen Xing <sup>1,†</sup>, Liang Zhang <sup>1,†</sup>, Zhenchong Liu <sup>2</sup>, Hengjiang Sheng <sup>3</sup>, Fujia Bi <sup>4</sup> and Jingxiang Xu <sup>1,\*</sup>

- 1 College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; bwxing@shou.edu.cn (B.X.); lz15254123876@163.com (L.Z.)
- 2 Shanghai Zhongchuan NERC-SDT Co., Ltd., Shanghai 201114, China; lzc\_just@163.com
- 3 Daishan County Transportation Bureau, Zhoushan 316299, China; shenghj0a@163.com
- 4 Beijing Mingzhou Technology Co., Ltd., Beijing 316299, China; mingzhou2020@163.com
- Correspondence: jxxu@shou.edu.cn

+ These authors contributed equally to this work.

Abstract: The goal of this paper is to strengthen the supervision of fishing behavior in the East China Sea and effectively ensure the sustainable development of fishery resources. Based on AIS data, this paper analyzes three types of fishing boats (purse seine operation, gill net operation and trawl operation) and uses the cubic spline interpolation algorithm to optimize the ship trajectory and construct high-dimensional features. It proposes a new coding method for fishing boat trajectory sequences. This method uses the Geohash algorithm to divide the East China Sea into grids and generate corresponding numbers. Then, the ship trajectory is mapped to the grid, the fishing boat trajectory points are associated with the divided grid, and the ship trajectory ID is extracted from the corresponding grid. The extracted complete trajectory sequence passes through the CBOW (continuous bag of words) model, and the correlation of trajectory points is fully learned. Finally, the fishing boat trajectory is converted from coordinate sequence to trajectory vector, and the processed trajectory sequence is trained by the LightGBM algorithm. In order to obtain the optimal classification effect, the optimal superparameter combination is selected. We put forward a LightGBM algorithm based on the Bayesian optimization algorithm, and obtained the classification results of three kinds of fishing boats. The final result was evaluated using the F1\_score. Experimental results show that the F1\_score trained with the proposed trajectory vectorization method is the highest, with a training accuracy of 0.925. Compared to XgBoost and CatBoost, the F1\_score increased by 1.8% and 1.2%, respectively. The results show that this algorithm demonstrates strong applicability and effectiveness in fishery area evaluations and is significant for strengthening fishery resource management.

Keywords: AIS; interpolation algorithm; ensemble learning; Bayesian optimization algorithm

# 1. Introduction

With the increasing number of fishing boats and the continuous improvement of fishing technology, the fishing intensity in coastal waters is increasing, and the problems of marine ecological environment destruction and fishery resources shortage are prominent [1]. With this background, understanding the distribution and fishing intensity of different types of fishing boats in coastal waters is helpful in indirectly revealing changes in coastal fishery resources, identifying illegal fishing activities, and providing necessary decision support for marine spatial planning and marine ecological protection. China has introduced a series of fishery supervision measures and made significant efforts to restore coastal fishery resources and regulate fishing activities.

There are various ways to restore coastal fishery resources [2], such as setting a fishing moratorium [3], regularly throwing fry into the sea, and limiting the minimum size of trawler nets.



Citation: Xing, B.; Zhang, L.; Liu, Z.; Sheng, H.; Bi, F.; Xu, J. The Study of Fishing Vessel Behavior Identification Based on AIS Data: A Case Study of the East China Sea. *J. Mar. Sci. Eng.* 2023, *11*, 1093. https://doi.org/ 10.3390/jmse11051093

Academic Editor: Carlos Guedes Soares

Received: 7 April 2023 Revised: 19 May 2023 Accepted: 19 May 2023 Published: 22 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). In order to regulate fishing activities, the following must occur: monitoring the behavior of fishing boats through satellite remote sensing technology [4], vessel monitoring system (VMS) [5] and automatic identification system (AIS) [6].

Some satellite remote sensing systems have the potential to detect ships and offer value in fishery monitoring. Meteorological satellite data has been utilized to detect marine vessel activities. Currently, research technology can accurately detect target ships at sea through satellite remote sensing. However, identifying the types of ships from numerous targets remains challenging. Satellite synthetic aperture radar (SAR) systems [7] can also detect ships at sea and possess the advantage of all-weather operation. Nonetheless, SAR images have limitations dependent on the SAR type, radiance, spectral resolution (typically very low), and the material and geometry of the detected object. Generally, optical images have higher spatial resolution compared to most SAR images [8]. High-resolution optical sensors can detect and characterize ships in low-cloud daytime conditions [9]. However, their primary disadvantages are their inability to be utilized at night or in adverse weather conditions. Furthermore, satellite remote sensing images can only capture the instantaneous state of fishing boats, and the revisit period is long, making it challenging to continuously monitor the navigation and fishing activities of fishing boats over an extended period. This limitation fails to meet the real-time requirements for fishing boat supervision.

Shipping big data such, as VMS data and AIS data, are widely used to monitor and identify fishing boats in real time. However, VMS data has some limitations, including low time resolution and limited access. In recent years, the popularization and application of AIS has brought new opportunities for the study of marine fishing activities. The advantages of timeliness, high accuracy and convenient collection can effectively make up for the deficiency of VMS data and better serve the characteristic mining of regional fishery activities.

In December 2000, the Navigation Sub-committee of the International Maritime Organization officially issued a proposal that ships must be equipped with AIS equipment, which clearly required that ships built after 2002 and ships operating since 2008 must be equipped with an AIS system [10]. According to the Technical Rules for Statutory Survey of Domestic Seagoing Ships (2020), ships should be equipped with an automatic identification system (AIS) according to the following requirements:

1. All passenger ships and cargo ships of 500 gross tonnage and above should be equipped with an A-class AIS equipment.

2. Domestic cargo ships with a gross tonnage of less than 500 tons shall be equipped with A-class or B-class AIS equipment as required.

At present, many coastal areas have established automatic identification system (AIS) base stations and coastal radar stations, generating a significant amount of ship information daily. This information includes both static details such as ship type, tonnage, and port of departure, as well as dynamic information like position, direction, and speed. Under normal circumstances, the real-time status of ships at sea can be easily determined using the aforementioned information. This provides an excellent data source for monitoring and managing marine fishing activities and assessing ecological pressure. It plays a crucial role in assisting managers and researchers in swiftly predicting and analyzing the distribution of fishery resources and locating central fishing grounds.

However, illegal, unreported, and unregulated fishing poses a threat to food safety and marine biodiversity [11]. It is estimated that these illegal fishing activities capture between 11 million and 26 million tons of fish annually, accounting for approximately 15% of global fish consumption [12,13]. To combat such illegal activities, the maritime warning system must be capable of real-time locating and identifying fishing boats within a specific area. This enables a better understanding of various fishing grounds and provides a foundation for implementing fishing regulations and cracking down on illegal, unreported, and unregulated fishing activities.

However, there is no standardized and mandatory reporting system for fishing boat activities or catches. Some jurisdictions require recording the date range, location, and

catch weight. Nevertheless, these records are typically not submitted until the ship returns to its landing site, and there is no central system for sharing this data. Therefore, correctly classifying and identifying unknown fishing vessels based on AIS data has become an important task.

Researchers both domestically and abroad have mostly used Bayesian Hidden Markov Models (HMMs) [14] and machine learning algorithms [15] for analyzing fishing behavior and fishery area assessment. Erico N. et al. used global fishing vessel trajectory data from 2011 to 2015 and vessel speed as the input variable to generate an HMM with an 83% recognition rate [16]. However, this model's algorithm efficiency is low and cannot achieve real-time analysis of AIS data. For research, Zheng Qiaoling et al. used 15 gillnet fishing vessels, 39 trawling fishing vessels and 24 Chinese-style large mesh fishing vessels in offshore China. A Backpropagation neural network model was used with the corresponding fishing vessel speed and heading data in the Vessel Monitoring System (VMS) as input to achieve accuracy rates of 96.6% and 91% based on the speed and heading models, respectively, but this model recognizes a small amount of data, and the recognition rate for gillnet fishing vessels is only 70% [17]. Tang Xianfeng et al. [18] extracted voyage information of fishing vessels from BeiDou Satellite Data. They batched the trajectory graph of each voyage using longitude and latitude data and trained a deep convolutional neural network model (CNN) with transfer learning and fine-tuning methods. Their tenlayer CNN model achieved a precision rate of 94.3% for identifying trawl and towed fishing boat operation types. However, this model's hidden layer has numerous layers, resulting in low efficiency and poor real-time analysis.

Yuan Feng et al. [19] used a neural network model with VMS data from shrimp fishing boats within the range of 25°–35° N, 120°–130° E as parameters to identify towing fishing vessels. They achieved a model training accuracy of 79%. Nevertheless, this model only identifies fishing vessels of a single type and cannot provide a macro perspective for multiple fishing vessels.

Yang Shenglong et al. [20] conducted a study on trawling fishing vessels in Xiangshan Port. They mined 1508 fishing vessel position, heading, and speed data with BeiDou terminals to determine the fishing vessel's speed threshold in the fishing state and heading variation. They used window filtering correction to obtain the fishing strength in each fishing area grid. However, this method does not account for the impact of latitude and longitude on fishing area grids. Near an island reef, when fishing vessels pass by, the speed and heading change may resemble their fishing state, but the fishing effort is zero. Therefore, accurately determining whether the fishing vessel operates within the fishing area is necessary.

Tang Fenghua et al. [21] determined the fishing status of trawl fishing vessels based on the threshold difference between navigation speed and heading. They used cumulative fishing time as the fishing intensity in a particular area for a while. Tracing the location of the fishing vessel, fishing ground, and fishing port allows obtaining the source and cumulative fishing time of aquatic products. The traceability of aquatic products can be achieved with information such as fishing grounds and fishing areas. However, this method also counts the time when the fishing vessel is not engaged in fishing operations, and the defined fishing intensity range for the fishing ground is too large, leading to decreased data accuracy and an overestimation of the catch.

Yang et al. [22] used VMS data from 8 single otter trawl vessels in the Zhoushan fishing ground from September to December 2012. They simulated trawl trajectories at different intervals using the cubic Hermite spline interpolation method to determine vessel activity (fishing and non-fishing).

Chen Renli et al. [23] used AIS data from offshore fishing vessels and the Gaussian mixture model (GMM) to identify fishing behavior and determine the speed threshold during fishing activities. They proposed a combined mapping method using kernel density estimation (KDE) and hotspot analysis (HSA) for fishing ground mapping. This combined method has proven effective and efficient for fishing ground mapping. However, the

drawn fishing ground area is rather general, and it is uncertain what kind of fishing boat is operating in that fishing ground.

In view of the aforementioned issues, this paper utilizes a data-driven approach based on machine learning to address the challenges. It optimizes fishing boat trajectories using the cubic spline interpolation algorithm and constructs high-dimensional features. Recognizing the limitations of traditional latitude and longitude features in fully characterizing fishing boat trajectories, an encoding method for fishing boat trajectory sequences is proposed to capture dynamic information and the relationships between trajectory points. Additionally, the Bayesian optimization (Boa) method is employed to optimize the parameters of the LightGBM (Light Gradient Boosting Machine) fishing boat classification model, enhancing model accuracy, reducing complexity, and improving parameter optimization efficiency. The specific steps are as follows:

1. To address missing values in ship trajectories, this paper applies the cubic spline interpolation algorithm to restore missing key data points, ensuring data integrity and accuracy of fishing vessel classification results.

2. Traditional statistical features are insufficient in fully characterizing fishing boat characteristics. Therefore, this paper constructs high-dimensional features based on the navigation patterns of the three types of fishing boats. These features are utilized in model training to enhance classification accuracy.

3. Recognizing the limited representation ability of conventional latitude and longitude statistics for fishing boat trajectories, a trajectory sequence coding method is devised. This method divides the fishing boat working area into grids using the Geohash algorithm [24] and represents the fishing boat trajectory using grid numbers. This approach reduces data complexity. The CBOW (continuous bag of words) algorithm [25] is employed to vectorize the trajectory sequence encoded by characters, and the vectorized data serves as input for the LightGBM algorithm, further enhancing algorithm execution efficiency.

4. To achieve optimal classification results for fishing boats, a LightGBM algorithm based on Bayesian optimization is proposed. This method selects appropriate values from numerous hyperparameters to optimize the model training process.

The organization of this paper is as follows:

Section 1 introduces the significance and research status of the classification of fishing boats.

Section 2 introduces the overall architecture of the system.

Section 3 is the behavior analysis and trajectory optimization of fishing Vessels, according to the law of ship motion. This paper uses the cubic spline interpolation algorithm to fill in the missing data and increase the integrity of the data;

Section 4 introduces the optimization of algorithm, it is mainly carried out from three aspects: firstly, extracting the high-dimensional features of fishing boats and improving the feature set; secondly, in order to efficiently process the massive AIS data, this paper proposes a composite coding method, which grids the fishing grounds in the East China Sea and vectorizes the fishing boats' trajectory as the training data set; finally, in order to obtain the best classification effect, we propose a LightGBM algorithm based on Bayesian optimization, which Maximizes the classification performance of the model.

Section 5 introduces the experimental data and classification results.

Section 6 is the relevant conclusion of the study and the prospects for future study.

# 2. Overall Architecture

In Figure 1, the flowchart shows the fishing vessel operation classification process, divided into two parts: data processing and algorithm optimization.

Data Processing: Firstly, extract AIS data from the database. Due to the influence of weather and other factors on communication between AIS devices and shore stations, there are abnormal points in the data. Therefore, data filtering is necessary. The filtered data are then optimized using the cubic spline interpolation algorithm three times to complete the ship's sailing trajectory by filling in missing points.

Algorithm Optimization: In this paper, we establish conventional statistical characteristics such as the mean, variance, median, as well as high-dimensional information quantity features of speed, longitude, latitude, and heading. We construct features from different perspectives to enrich the feature set. Additionally, we propose an encoding method suitable for ship trajectories, which vectorizes fishing boat trajectories. This method takes into account the correlation between each trajectory point of the ship, thus enhancing classification accuracy.



Figure 1. Flowchart of fishing vessel operation classification based on ensemble learning.

# 3. Behavior Analysis and Trajectory Optimization of Fishing Vessels

Definition of Fishing Activity by Gear Type

As shown in Figure 2, gillnet fishing vessels [26,27] use gillnets as fishing gear to fish in the ocean. Its navigation trajectory characteristics are as follows:

Relatively stable: Gillnet fishing vessels [28] usually fish within a certain area, with a cruising speed of 0–2 knots during operation. The navigation trajectory is relatively stable, unlike trawl or seine fishing vessels, which frequently change direction and speed.

Straight-line navigation: Gillnet fishing vessels usually navigate in a straight line to maintain the tension and stability of the gillnet and facilitate fishing.

Brief pause: Gillnet fishing vessels need to briefly pause during fishing to wait for the fish to enter the gillnet. Therefore, there may be brief pause points in the navigation trajectory [29].

Multi-point navigation: Within a fishing ground, gillnet fishing vessels typically fish in different locations so that the navigation trajectory may show characteristics of multi-point navigation [30].

As shown in Figure 3, trawling involves dragging one or more nets behind a fishing vessel on the seafloor (bottom trawl) or in the water (pelagic or midwater trawl). When trawling, the fishing vessel usually slows down and maintains a constant speed to keep the tension of the trawl net as even as possible. The duration of trawling primarily depends on the density of prey, ranging from a few minutes to several hours. Here, trawling activity is defined from the moment the net is cast until it is retrieved. The characteristics of trawl fishing vessels are usually slow and stable, with a cruising speed of 2.5–5.5 knots during operations. The distribution of AIS speed data determines these speed thresholds and corresponds to similar values obtained in the literature [31,32].

Characteristics of trawling vessel trajectories [33]:

Regular cruising of vessels: In the East China Sea, trawl fishing vessels usually cruise regularly to search and catch fishery resources. The trajectories of these regular cruises often exhibit certain patterns and periodicity.

Slow speed: Trawling vessels typically use a slow trawling method during operations, resulting in a relatively slow vessel movement speed.

As shown in Figure 4, a purse seine is a long net hung vertically on floats by fishing vessels or independent small boats on the water's surface [34,35]. To avoid fish escaping, purse seining needs to be quickly set at an average high speed of about 10 knots. Once the net completely encircles the fish, the bottom of the net is pulled up, and the net is towed away, with a cruising speed of 2–8 knots during operations. Fish drift with the net and are picked up and transferred to the vessel. The duration of this process depends on the quantity caught, ranging from one hour to several hours.

Characteristics of purse seine fishing vessel trajectories:

High vessel aggregation: Purse seine fishing vessels in the sea often gather around specific fishing grounds with abundant fishery resources, which are typically favored due to favorable marine environmental conditions, water depth, and seafloor terrain, resulting in a high concentration of vessels in those areas.

Traditional cruising of vessels: Purse seine fishing vessels typically follow regular cruising patterns in the sea as they search for and catch fishery resources. These regular cruises often exhibit specific patterns and periodicity.

Limited operational range of vessels: Compared to trawl fishing vessels, purse seine fishing vessels have a relatively restricted operational range, usually confined to the purse seine area.

Longer vessel parking time: Purse seine fishing vessels often need to remain in specific fishing grounds with abundant fishery resources for a certain period, waiting for fish to enter the purse seine. Therefore, during the parking period, the vessel's trajectory tends to be relatively stable.

Flexible operational time periods of vessels: The operational time of purse seine fishing vessels is typically not fixed, as they need to adapt their operations based on the movement of fish schools, allowing for flexibility in their operational periods.



Figure 2. The trajectory of gillnet fishing vessels in the East China Sea.



Figure 3. The trajectory of trawl fishing vessels in the East China Sea.



Figure 4. The trajectory of purse seine fishing vessels in the East China Sea.

Due to the influence of factors such as weather on the signal transmission of AIS devices on ships, the trajectories generated by data with abnormal values removed often have many missing points. However, complete trajectories are an essential factor in determining the type of ship operation. Therefore, we must use more advanced interpolation methods to obtain smoother and more accurate fishing boat trajectories. Furthermore, the information obtained through these methods can be used for more detailed research to understand the spatial distribution of fishing efforts and their impact on benthic organisms [36]. Therefore, this paper uses the cubic spline interpolation algorithm to optimize ship trajectories.

In some developed countries, there have been studies on ship trajectory interpolation [37]. These interpolation methods can be roughly divided into two categories: linear

8 of 25

and nonlinear interpolation methods [38], such as spline interpolation. Linear interpolation can connect continuous data points and handle discontinuous data points. It is the simplest and fastest interpolation method. When the time interval is short, a straight line can be considered as the trajectory of a fishing boat. However, when the time interval is long, the difference between the actual trajectory and the replaced straight-line trajectory will be large, affecting the further analysis of fishing effort. Linear interpolation cannot simultaneously consider speed and heading, which can result in a significant deviation between the simulated and true paths. Spline interpolation is commonly used to mitigate the problem of linear interpolation, which considers the complexity of data structures and is an efficient algorithm. It also considers heading and speed and maximizes the use of data. There are many types of spline interpolation methods, each with its own advantages and disadvantages. Many spline functions require four data points to interpolate the trajectory between two continuous points [39]. Compared with linear interpolation, which only contains two points, spline interpolation often performs better in terms of interpolation accuracy.

Description of the cubic spline interpolation algorithm:

Step 1: Given n + 1 data points on the interval [a,b], divide [a,b] into n sub-intervals: that is,  $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$ .

Step 2: Calculate the step size,  $h_i = x_i + 1 - x_i$ .

Step 3: Solve equations to obtain m<sub>i</sub>.

Step 4: Calculate the parameters of each sub-interval:

 $a_i, b_i, a_i = y_i, b_i = (y_i + 1 - y_i)/2 - (h_i m_i)/2 - h_i (m_i + 1 - m_i)/6, c_i = m_i/2, d_i = (m_i + 1 - m_i)/6h_i.$ 

Step 5: Generate the spline function.  $S_i(x_i) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ . The optimization results are shown in Section 5.1.

#### 4. Algorithm Optimization

#### 4.1. Build High-Dimensional Features

Currently, existing methods for recognizing fishing vessel behavior typically analyze single-speed data, leading to low accuracy and ineffective identification of fishing behavior. This is particularly evident during turns, where reduced speed can cause misjudgments of the operational trajectory and lower the accuracy of determining fishing areas. Moreover, relying solely on low-information features such as heading angle, longitude and latitude, and distance for identifying fishing vessel behavior can result in misjudgments under different navigation states.

To address these limitations, feature engineering is conducted by analyzing the original AIS data, as the number of available features is limited. Feature engineering involves discovering potential and valuable variables through the analysis of raw data. Common feature engineering methods include discrete data coding, custom methods, function transformation methods, and statistical value construction methods [40]. The statistical value construction method involves obtaining new features by calculating statistical quantities of the existing features. The function transformation method involves obtaining new features by squaring, square-rooting, exponentiating, logarithmic, differential transformations, etc. Discrete data coding involves one-hot encoding or binary encoding of discrete data.

In this article, we establish standard statistical features, such as the mean, variance, and median of speed, latitude and longitude, and heading, as well as high-dimensional informative features such as percentiles, skewness, and kurtosis, to comprehensively describe fishing vessel behavior from various angles by enriching the feature set [41].

The Figure 5 shows that high-dimensional statistical features (such as  $x\_bin$ ,  $x\_y$ , std) are extracted for longitude, latitude, and speed. The longitude and latitude ranges where fishing vessels are divided into multiple intervals, and the trajectory data features for each interval are combined to form compound features. These compound features balance the differences between attribute values. After introducing them to the training model, the model can better learn the fishing vessel behavior, resulting in a significant improvement in

accuracy. Finally, the fishing vessel trajectory is vectorized. Most traditional classification algorithms use one-hot encoding for coding, but this method does not fully consider the connection between trajectory points, leading to inaccurate classification results. Therefore, this article proposes a coding method suitable for ship trajectories by referring to the encoding method, which vectorizes fishing vessel trajectories. This method considers the correlation between each trajectory point of the vessel, which is conducive to improving classification accuracy.



**Figure 5.** The proportion of feature importance of the algorithms. x represents longitude, y represents latitude, v represents speed, k represents the ratio of latitude to longitude, b represents the difference between latitude and longitude multiplied by the average value of k. Other values include *k\_mean*: the mean value of k, min: minimum value, max: maximum value, mean: mean value, 1/4: 1/4 percentile, 3/4: 3/4 percentile, std: standard deviation, cov: covariance, kurt: kurtosis, skew: skewness.

#### 4.2. Regional Gridding of Fishing Ground and Vectorization of Fishing Boat Trajectory

Facing the massive amount of fishing boat trajectory data, the conventional statistical information based on latitude and longitude has limited ability to accurately characterize the fishing boat trajectory. To address this limitation and capture the dynamic information of the trajectory along with the relationships between the points the fishing boat traverses, we have developed a trajectory sequence coding method based on the specific sea areas through which the fishing boat navigates.

Firstly, we unpack the map of the sea area that the fishing boat passes through into rectangular grids by using the Geohash algorithm [42,43], and the size of the divided rectangular grids is determined by the character coding length of Geohash. Based on the size of the fishing boat, we choose the character length of 7 to determine the grid area size, and the area size determined by the Geohash algorithm is 153 m  $\times$  153 m. The coded fishing ground area is composed of a series of numbered grids, and the trajectory points where fishing boats are composed of these numbers.

The basic principle of Geohash is as follows: for example, the latitude and longitude are [31.19993, 121.0007], the latitude interval of the earth is [-90, 90]. Divide this interval

Finally, the binary number is converted into characters, and all points of the fishing boat trajectory are coded by this method, so that the fishing boat trajectory points are associated with the divided rectangular grid points, and the complete fishing boat trajectory is coded by characters.

For the sake of understanding, we use a digital grid to represent the trajectory sequence of fishing boats.

Then, we use the CBOW algorithm to vectorize the trajectory sequence and add it to the subsequent training [44], and show the relationship between the trajectory points in the form of a heat map.

Table 1. Base32 code table.

Decimal	0	1	2	3	4	5	6	7	8	9	10	11	12
Base32	0	1	2	3	4	5	6	7	8	9	b	с	d
Decimal	13	14	15	16	17	18	19	20	21	22	23	24	25
Base32	e	f	g	h	j	k	m	n	р	q	r	s	t
Decimal	26	27	28	29	30	31							
Base32	u	v	W	x	У	Z							

As shown in Figure 6, the specific steps are as follows:

First, map each trajectory to a grid, where the ID of the grid corresponding to each trajectory point represents a word in the vocabulary, and each trajectory corresponds to a document. Secondly, extract vessel trajectory IDs on the corresponding grid. The extracted grid ID sequence is trained by the CBOW model [45] to fully learn the relevance of trajectory points and output as a single trajectory point. The input layer of the CBOW model is the context  $\{x_1, x_2, ..., x_c\}$  encoded by one-hot, and the calculation process is as follows: The first step is to calculate the hidden layer h:

$$h = \frac{1}{c} w \left( \sum_{i=1}^{c} x_i \right) \tag{1}$$

The second step is to calculate the input of each node in the output layer:

U

$$_{i} = V_{wi}^{T} * h \tag{2}$$

 $V'_{wj}^{T}$  is the *j*th column of the output matrix.  $y_{c,j}$  is represented as:

$$y_{c,j} = p(w_{y,j}|w_1, \dots, w_c) = \frac{\exp(u_j)}{\sum\limits_{j'=1}^{v} \exp(u'_j)}$$
(3)



Figure 6. The proportion of feature importance of the algorithms.

As shown in Figure 6, the top right corner of the graph is the correlation heat map of each grid point, represented by the correlation vectors of each grid point. The darker the color, the higher the correlation, and the lighter the color, the lower the correlation. Thus, the fishing boat trajectory is converted from a coordinate sequence to a trajectory vector.

#### 4.3. LightGBM Algorithm

Gradient Boosting Decision Tree [46] (Gradient Boosting Decision Tree, GBDT [47,48]) is an integrated algorithm based on the decision tree combined with the idea of gradient boosting, which is easy to visualize and has strong generalization ability. LightGBM was proposed by Microsoft in 2017. The objective of this proposal is to address the issues associated with training GBDT [49], such as high computational complexity, time consumption, and space utilization. These challenges arise from the requirement of traversing all data and features for each node split, hindering parallelization and overall efficiency.

The fishing boat data set is defined as  $D = \{(X_i, Y_i)\}_{i=1}^n$ , where n is the sample size,  $Y_i$  is the ship type label, and  $X_i$  is the classification label. The specific steps are as follows:

1. Initialize the classification state  $F_{k,0}(X_i)$  of  $X_i$  on the k<sup>th</sup> class, and the number of classification types is k = 1, 2, 3.

$$F_{k,0}(X_i) = 0 \tag{4}$$

- 2. Iterative calculation, the number of iterations is m = 1, 2, ..., M.
- 3. Calculate the probability that X corresponds to each class.

$$p_{k,m-1}(X_i) = \frac{e^{p_{k,m-1}(X_i)}}{\sum\limits_{k=1}^{K} e^{p_{k,m-1}(X_i)}}$$
(5)

4. Calculate the negative gradient value Y of  $X_i$ , the true probability of X is P.

$$\tilde{y}_{i,k} = y_{i,k} - p_{k,m-1}(X_i)$$
 (6)

5. Calculate the leaf node value  $\gamma_{j,k,m}$  after node splitting, the sample set on the leaf node is  $R_{j,k,m}$ , and the number of leaf nodes is j = 1, 2, ..., J.

$$\gamma_{j,k,m} = \frac{K-1}{K} \frac{\sum_{N_i \in R_{j,k,m}} \tilde{\gamma}_{i,k}}{\sum_{X_i \in R_{j,k,m}} \left| \tilde{\gamma}_{i,k} (1 - \left| \tilde{\gamma}_{i,k} \right|) \right|}$$
(7)

6. Update the model with a learning rate of  $\eta$ .

$$F_{k,m}(X_i) = F_{k,m-1}(X_i) + \eta \sum_{j=1}^{J} \gamma_{j,k,m} I$$
(8)

7. Finally, obtain the model.

$$F_k(X) = \eta \sum_{m=1}^M \sum_{j=1}^J \gamma_{j,k,m} I$$
(9)

The improvement of LightGBM [50] on GBDT is as follows: use the growth strategy based on the Leaf-wise algorithm to build trees, reduce unnecessary calculations, and improve the accuracy of fishing boat classification; use the Histogram algorithm to traverse samples and store eigenvalues to reduce time and space. In addition, LightGBM uses gradient-based unilateral sampling and mutually exclusive feature bundling technology to reduce redundancy and improve efficiency in the two dimensions of sample size and features. In addition, if the learning rate is too large, the algorithm will fail to converge, and if it is too small, the iteration speed of the model will be too slow. Therefore, it is necessary to set an appropriate learning rate; if LightGBM uses the Leaf-wise strategy to grow the tree, the tree will grow too deep, resulting in overfitting. Therefore, it is necessary to control the number of leaf nodes, the depth of the tree and the minimum number of data on the leaves. In order to obtain better model training results, we propose a LightGBM algorithm based on Bayesian optimization. From many hyperparameters, select the optimal parameter combination in order to optimize the effect of model training.

#### 4.4. Bayesian Optimization Algorithm

Bayesian optimization [51] is suitable for solving black-box problems where the derivative of the objective function is unknown and the cost of evaluating the objective function is high. The specific process of Bayesian optimization is as follows:

First, random sampling is performed in the parameter space, and then a preliminary objective function distribution is established using a probabilistic proxy model, and the next point to be evaluated is determined by maximizing the collection function, and the newly obtained point evaluates the objective function. After adding the value to the existing set of evaluation points, the probabilistic proxy model is updated, and the cycle is repeated until the end of the iteration. Therefore, probabilistic surrogate models and acquisition functions are two important components of Bayesian optimization. When applying the LightGBM model to classify fishing vessel types, it is necessary to determine the optimal parameter combination. In order to realize the fast and efficient parameter optimization process and avoid falling into the local optimal solution, this paper introduces the Bayesian global optimization algorithm to optimize the model parameters.

#### 4.5. Probabilistic Proxy Model

Probabilistic proxy models [52] are divided into parametric models and non-parametric models. Compared with parametric models with a fixed number of parameters, non-

parametric models are more flexible and highly scalable, among which, Gaussian Processes (GP) are widely used. The Gaussian process is a random combination of parameters to be adjusted in the LightGBM model, which can be expressed as:

$$f(x) \sim GP(\mu(x), k(x, x')) \tag{10}$$

In the formula: f(x) is the objective function,  $\mu(x)$  is the mean function,  $\mu(x) = E(f(x)), k(x, x')$  is the covariance function.

Assuming that the known parameter combination is  $D_t = \{(x_i, f_i)\}_{i=1}^t$ , where  $f_t = f(x_t)$ , the next sampling point is  $x_{t+1}$ , and assuming that the mean value of the prior distribution is 0, the joint distribution of f and  $f_{t+1}$  can be expressed as:

$$\begin{pmatrix} f \\ f_{t+1} \end{pmatrix} \sim N \begin{pmatrix} 0 & \begin{bmatrix} K & K_{t+1} \\ K_{t+1}^T & K_{t+1,t+1} \end{bmatrix} \end{pmatrix}$$
(11)

In the formula, *K* is a matrix composed of k(x, x),  $k_{t+1}$  is a matrix composed of  $k(x, x_{t+1})$ , and  $K_{t+1,t+1}$  is a matrix composed of  $k(x_{t+1}, x_{t+1})$ . Then, the posterior distribution  $P(f_{t+1}|D_t, x_{t+1})$  of  $f_{t+1}$  can be expressed as:

$$P(f_{t+1}|D_t, x_{t+1}) = N(\mu(x_{t+1}), \sigma_t^2(x_{t+1}))$$
(12)

where the mean of the posterior distribution of  $f_{t+1}$  is  $\mu_t(x_{t+1}) = K_{t+1}^T K^{-1} f$  and the variance is  $\sigma_t^2(x_{t+1}) = K_{t+1,t+1} - K_{t+1}TK^{-1}K_{t+1}$ .

#### 4.6. Acquisition Function

After establishing the posterior distribution of the Gaussian process, the next sampling point  $X_{t+1}$  is determined by the acquisition function [53]. There are three common acquisition functions, PI, EI (expected improvement), and UCB (upper confidence bound). Among them, the PI (Probability of Improvement) function is easy to use. The sampling points based on the PI strategy are selected in this paper as follows:

$$x_{t+1} = \arg\max\Phi\left(\frac{\mu_t(x) = f_t(x^+) - \varepsilon}{\sigma_t(x)}\right)$$
(13)

In the formula,  $\Phi()$  is the standard normal cumulative distribution function,  $f_t(x^+)$  is the maximum value of the current objective function, and  $\varepsilon$  is the trade-off coefficient for balanced exploration and development.

As shown in Figure 7, the process of the fishing boat classification model based on Bayesian optimization is as follows:

1. Normalize and standardize the sample data, and divide the data set into a training set and a test set according to the ratio of 4:1.

2. After initializing the model, the Gaussian process regression is used to calculate the maximum value of the AC function. If the target value is met, it is output; if not, return to the Gaussian process to continue calculation.

3. Use the Bayesian optimization algorithm to optimize the learning rate, tree depth, and number of classifiers of the model to obtain the optimal hyperparameters, and set the parameters of the LightGBM algorithm.

4. Determine whether the hyperparameters have reached the target maximum accuracy. If so, set the parameters of the model as the optimal hyperparameters. Otherwise, return to steps (2), (3) and repeat.

5. Test the classification effect of the model through the test set, determine the classification accuracy of the model, and evaluate the model. As shown in Table 2, we conducted 30 iterative experiments on the model according to the above process, and the target is the test score of each combination. We found that the result of the 14th iteration is the best; therefore, the selected parameter combination is as follows: bagging\_fraction:0.8297, lambda\_l1:0.5224, learning\_rate:0.4271, max\_depth:20, num\_leaves:26.

Table 2. Hyperparameter Training Results.

Iteration	Target	Bagging_Fraction	Lambda_l1	Learning_Rate	Max_Depth	Num_Leaves
1	0.9508	0.2748	1.739	0.411	47.5	39.38
2	0.9569	0.4646	1.747	0.3693	30.93	45.1
3	0.9361	0.3118	1.287	0.7563	29.6	90.55
4	0.9594	0.8648	1.57	0.7333	13.37	16.99
5	0.9545	0.7073	0.1431	0.7922	14.75	26
6	0.9569	0.9284	1.442	0.6866	14.14	17.46
7	0.9165	0.2	2.619	1	5.779	14.34
8	0.9569	0.8178	2.142	0.7896	15.19	21.55
9	0.9214	0.261	3	1	17.84	14.83
10	0.9349	0.2251	0.24	0.5378	12.25	21.02
11	0.9569	1	1.946	0.909	16.27	24.2
12	0.9557	0.8462	1.527	0.6978	17.27	21.04
13	0.9643	1	0.1117	0.3489	18.75	23.91
14	0.9655	0.8297	0.5224	0.4271	19.85	25.48
15	0.9532	0.8031	1.342	0.9805	22.41	23.98
16	0.9557	0.3588	1.665	0.1169	18.76	28.09
17	0.9581	0.4522	0.3603	0.1994	23.62	30.4
18	0.9618	1	0.1	1	20.35	27.71
19	0.963	0.907	0.8223	0.2752	22.02	34.44
20	0.9643	1	0.1	1	26.47	33.91
21	0.9594	0.5902	1.167	0.2274	25.67	38.11
22	0.9545	1	3	1	25.82	33.22
23	0.9581	0.6216	0.12	0.1	29.74	34.11
24	0.9508	1	0.1	1	30.91	36.57
25	0.9214	0.2	0.1	1	20.71	36.57
26	0.9569	0.6714	1.374	0.4674	22.87	30.67
27	0.9643	1	0.1	0.1	28.17	30.16
28	0.9569	0.387	0.1	0.1131	28.97	29.45
29	0.9594	1	0.1	1	31.98	30.85
30	0.9618	0.9854	2.839	0.9269	28.43	27.03



Figure 7. Flow chart of Bayesian optimization fishing vessel classification model.

#### 5. Experiment and Result

#### 5.1. Experimental Environment and Data

In Figure 8, the image is provided by Global Fishing Watch (access date is 1 October 2022) (https://globalfishingwatch.org/). This paper uses the trajectory data of fishing boats in Zhoushan Fishery of the East China Sea in October 2021, with a spatial range of 20°–35° N and 120°–130° E. The sea area is relatively shallow, and rich in nutrients and bait. Zhoushan Fishery is the largest fishery in China, with abundant fishery resources and a long history of marine fishing. Coastal fishing ports and docks are densely distributed, with tens of thousands of motorized fishing boats. During the fishing season, the intensity of fishing activities is high, and the fishing activities are concentrated. Frequent encounters with commercial ships occur, and the maritime traffic conditions are complex, making it an ideal place to conduct AIS data application research. The dataset used in this paper is provided by the Transportation Bureau of Daishan County, Zhejiang Province. There are 2109 fishing boats in the dataset, with over 200 million records in total, of which 2034 fishing boats are of a determined type, and 75 fishing boats have no operation mode specified.The data format of the fishing vessel is shown in Table 3. The experiment environment is shown in Table 4. The optimization results are shown in Figures 9 and 10.

MMSI	Vessel Length	Vessel Width	Lat (°)	Lon (°)	SOG (kn)	COG (°)
410405500	24		20.154(5.N	100 00770 5		1(( )
412425783	24 m	6 m	30.17465 N	122.03770 E	5	166.9
422426552	40 m	7 m	30.13483 N	122.12875 E	4.6	278.0
412425118	37 m	7 m	30.14063 N	122.13303 E	6.1	96.0
412425828	28 m	6 m	30.13437 N	122.16628 E	3.9	158.0
412425526	41 m	7 m	30.26035 N	122.16628 E	8.6	169.0
412426203	25 m	5 m	30.15163 N	122.177795 E	9.5	52.0

Table 3. Partial Fishing Vessel Data Sheet.

 Table 4. Experimental Environment.

CPU Memory NVIDIA GeForce GTX 3070. 8 G	Intel(R)Core(TM) i7-11800H. 2.3 GHz 32 G Graphics Card
Operating System	Windows11
Programming Language	Python



**Figure 8.** Original trajectory heat map. The figure shows the visualization of the partial data used in this paper; red represents the heat map of the ship's trajectory, and the more vivid the color, the busier the route.



**Figure 9.** Optimized through three rounds of cubic spline interpolation. The green line represents the original trajectory with many gaps. The yellow area indicates the optimized region where the trajectory will be improved. The red line on the graph represents the optimized trajectory.



**Figure 10.** Optimized through three rounds of cubic spline interpolation. The green line represents the original trajectory with many gaps. The yellow area indicates the optimized region where the trajectory will be improved. The red line on the graph represents the optimized trajectory.

#### 5.2. Experimental Result

According to the "Management of Summer Fishing Rest Period in Putuo District in 2021" issued by the People's Government of Putuo District, Zhoushan City:

The fishing rest period for single-boat trawl nets (pole trawl shrimp nets), cage traps, gill nets, and light fishing (lay) nets is from 12:00 on 1 May to 12:00 on 1 August. The fishing rest period for small-scale anchored net fishing vessels is from 12:00 on 1 May to 12:00 on 16 August. The fishing rest period for single-anchor gill nets (sail-type gill nets), trawls, and other unlisted marine fishing operations is from 12:00 on 1 May to 12:00 on 16 September.

With the decline of resources in the East China Sea region and the increasing awareness of resource protection, problems such as poor selectivity of trawl and gill nets, serious damage to young fish of economic species, and excessive fishing intensity beyond the carrying capacity of nearshore waters have become prominent. Therefore, according to the "Regulations on the Management of Fishing Licenses" issued by the Ministry of Agriculture and Rural Affairs, fishing vessel management will be stricter starting in 2019, and large and medium-sized fishing vessels are not allowed to operate within the banned fishing zone line inside the wheel bottom trawl, and small-scale fishing boats are not allowed to operate outside the banned fishing zone line. At the same time, in order to control the intensity of marine fishing, the approval and manufacture of trawls, single-anchor gill nets, and single-boat large deep-water sac nets for fishing operations have been strictly prohibited since 2019.

The classification result is shown in the Figure 11, with 66 purse seiners, 5 trawlers, and 4 gill netters. Figures 12–14 are the operating trajectories of gillnets, trawls, and seiners, respectively. This figure is drawn using the open source framework kepler.gl. From the figure, we can roughly determine the respective operating areas of the three fishing vessels. It provides some assistance for fisheries resource management. It should be noted that the data collected in this article are for some fishing vessels in Daishan County in October 2021. Therefore, the number of fishing vessels of different types may vary greatly, which does not represent the actual number of fishing vessels in the local area. At the same time, due to changes in fishery management policies and fishery resources, the number of fishing vessels may also undergo significant changes.



Figure 11. The result of Classification.



Figure 12. Classified area for gillnet fishing vessels.



Figure 13. Classified area for trawler fishing vessels.



Figure 14. Classified area for Purse seine fishing vessels.

#### 5.3. Model Evaluation

The confusion matrix belongs to the category of model evaluation and essentially represents the judgment of the model results; it counts the number of each situation. As shown in Table 5, the confusion matrix mainly includes the four fundamental indicators (primary indicators) described below: True Negative (TN) and True Positive (TP), False Negative (FN), and False Positive (FP). Among them, FP is called the first type of error in statistics (Type I Error), and FN is called the second type of error in statistics (Type II Error). The four fundamental indicators are explained in the table below.

Table 5. Description of basic indicators.

Index	Symbol	Description
1	TP	both are positive
2	TN	both are negative
3	FP	The model judged positive, but the actual situation was negative
4	FN	The model judged negative, but the actual situation was positive

Precision refers to the proportion of true positive (*TP*) cases correctly identified out of all positive cases identified. The formula for Precision is defined as follows:

$$P = \frac{TP}{TP + FP} \tag{14}$$

Recall refers to the proportion of true positive (*TP*) cases correctly identified out of all positive cases. The formula for Recall is defined as follows:

$$R = \frac{TP}{TP + FN} \tag{15}$$

*F1\_score* is the harmonic mean of precision and recall, ranging from 0 to 1. The formula is defined as follows:

$$F_1 = 2 * \frac{P * R}{P + R} \tag{16}$$

When sample class imbalance occurs, precision and recall may not be applicable. For example, in medical science, if we consider individuals with cancer as positive cases and

the rest as negative cases, the proportion of positive cases is very small. If our model predicts all samples as positive cases, the recall will be 1, but precision will be low. In such cases, these two evaluation metrics are not used, and *F1\_score*, the harmonic mean of precision and recall, is more reliable. Therefore, *F1\_score* is widely used for evaluating the performance of classification models.

As shown in Figure 15, the training iteration is 100. The black curve represents the  $F1\_score$  curve of the LightGBM algorithm, the red curve represents the  $F1\_score$  curve of the XgBoost algorithm, and the blue curve represents the  $F1\_score$  curve of the CatBoost algorithm. The black curve in the figure is the optimized  $F1\_score$  curve of the algorithm used in this paper, and the red and blue curves are the  $F1\_score$  curves that have not been optimized using the algorithm proposed in this paper. It can be seen from the figure that the XgBoost algorithm has the smallest  $F1\_score$ , with a value of around 0.907; the  $F1\_score$  of the CatBoost algorithm is around 0.913, and the  $F1\_score$  of the LightGBM algorithm is the highest, with a value of about 0.925.

As shown in the Figure 16, this is the  $mlog_loss$  curve of the LightGBM algorithm. The training iteration is 100, and the  $mlog_loss$  curve for multi - class gradually converges with the number of iterations, with a final  $mlog_loss$  of 0.22. This shows that the algorithm model we proposed is convergent, not divergent, and the algorithm can meet the requirements for classification of fishing boats.

Figure 17 shows the feature importance ratio of various features in three algorithms. From the graph, we can see that the feature importance ratios of the *lightGBM* algorithm are relatively balanced compared to the other algorithms, with most ratios being below 0.05. However, in *XgBoost* and *CatBoost* algorithms, some features have a disproportionately large weight. For example, the weights of features like '*dis*' and '*h\_kurt*' exceed 0.1. When the feature learning rate and other hyperparameters are the same, a high weight of a single feature indicates that the algorithm strongly depends on that feature during training. This often leads to some misclassifications and reduces the accuracy of the algorithm.



Figure 15. F1\_score of three classification algorithms.



Figure 16. Multiclass loss curve.



**Figure 17.** The proportion of feature importance of the three algorithms. x represents longitude, y represents latitude, v represents speed, k represents the ratio of latitude to longitude, b represents the difference between latitude and longitude multiplied by the average value of k. Other values include  $k\_mean$ : the mean value of k, min: minimum value, max: maximum value, mean: mean value, 1/4: 1/4 percentile, 3/4: 3/4 percentile, std: standard deviation, cov: covariance, kurt: kurtosis, skew: skewness.

# 6. Conclusions

In this paper, we propose a classification method for fishing boat operation types, aiming to supervise fishing operation behavior in the East China Sea and promote sustainable development of fishery resources. The method involves optimization and construction of high-dimensional features. Additionally, a new encoding method for fishing boat trajectory sequence is introduced. This method utilizes Geohash to divide the East China Sea into grids and assign corresponding numbers to each grid. The ship trajectory is then mapped to these grids, enabling the association and extraction of fishing boat trajectory points. The complete trajectory sequence is obtained and passed through the CBOW model to capture the correlation among trajectory points. Finally, the fishing boat trajectory is converted from a coordinate sequence to a trajectory vector. The processed trajectory sequence is trained using the LightGBM algorithm. To achieve optimal classification performance and select the best combination of hyperparameters, we propose a LightGBM algorithm based on Bayesian optimization. The classification results for the three operating fishing boats are obtained. Experimental results demonstrate that the proposed method achieves the highest F1\_score during training, with a training accuracy of 0.925. Compared to XgBoost and CatBoost, the F1\_score has increased by 1.8% and 1.2%, respectively. The method proposed in this paper effectively strengthens the supervision of fishing operations by fishing vessels and contributes to the sustainable development of fishery resources. Therefore, Zhoushan fishery can benefit from the methods proposed in this paper to enhance the supervision of fishing operations and contribute to the sustainable development of fishery resources. Future research will focus on further studying fishing behavior management of fishing vessels.

**Author Contributions:** Conceptualization, L.Z. and B.X.; Methodology, L.Z. and B.X.; Formal analysis, L.Z. and B.X. and J.X.; Data curation, L.Z. and Z.L. and F.B.; Software, L.Z.; Writing—original draft preparation, L.Z. and B.X. and J.X.; Writing—review and editing, L.Z.; Supervision, B.X. and H.S. and Z.L.; Project administration, B.X. and H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Shanghai Science and Technology Committee (STCSM) Local Universities Capacity-building Project (No. 22010502200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available on request.

**Acknowledgments:** The authors would like to express their gratitude for the support of the Fishery Engineering and Equipment Innovation Team of Shanghai High-level Local University and Daishan County Transportation Bureau.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- Pham, T.D.T.; Huang, H.W.; Chuang, C.T. Finding a balance between economic performance and capacity efficiency for sustainable fisheries: Case of the Da Nang gillnet fishery, Vietnam. *Mar. Policy* 2014, 44, 287–294. [CrossRef]
- 2. Bell, J.D.; Leber, K.M.; Blankenship, H.L. A new era for restocking, stock enhancement and sea ranching of coastal fisheries resources. *Rev. Fish. Sci.* 2008, *16*, 1–9. [CrossRef]
- Ding, L.; Lu, M.; Xue, Y. Driving factors on implementation of seasonal marine fishing moratorium system in China using evolutionary game. *Mar. Policy* 2021, 133, 104707. [CrossRef]
- Pettorelli, N.; Laurance, W.F.; O'Brien, T.G. Satellite remote sensing for applied ecologists: Opportunities and challenges. *J. Appl. Ecol.* 2014, *51*, 839–848. [CrossRef]
- 5. Gerritsen, H.; Lordan, C. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J. Mar. Sci.* **2011**, *68*, 245–252. [CrossRef]
- Fournier, M.; Casey Hilliard, R.; Rezaee, S. Past, present, and future of the satellite-based automatic identification system: Areas of applications (2004–2016). WMU J. Marit. Aff. 2018, 17, 311–345. [CrossRef]

- 7. Chen, F.; Lasaponara, R.; Masini, N. An overview of satellite synthetic aperture radar remote sensing in archaeology: From site detection to monitoring. *J. Cult. Herit.* 2017, 23, 5–11. [CrossRef]
- Hou, X.; Ao, W.; Song, Q. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* 2020, 63, 1–19. [CrossRef]
- Gallego, A.J.; Pertusa, A.; Gil, P. Automatic ship classification from optical aerial images with convolutional neural networks. *Remote Sens.* 2018, 10, 511. [CrossRef]
- Harati-Mokhtari, A.; Wall, A.; Brooks, P. Automatic Identification System (AIS): Data reliability and human error implications. J. Navig. 2007, 60, 373–389. [CrossRef]
- 11. Christensen, J. Illegal, unreported and unregulated fishing in historical perspective. Perspect. Oceans Past 2016,30, 133–153.
- 12. Long, T.; Widjaja, S.; Wirajuda, H. Approaches to combatting illegal, unreported and unregulated fishing. *Nat. Food* **2020**, *1*, 389–391. [CrossRef]
- 13. European Commission. Fighting Illegal Fishing: Commission Warns Taiwan and Comoros with Yellow Cards and Welcomes Reforms in Ghana and Papua New Guinea; European Commission: Brussels, Belgium, 2015.
- 14. Deng, Q.; Söffker D. A Review of the current HMM-based Approaches of Driving Behaviors Recognition and Prediction. *IEEE Trans. Intell. Veh.* **2021**, *7*, 21–31. [CrossRef]
- 15. Mahesh, B. Machine learning algorithms—A review. Int. J. Sci. Res. (IJSR) 2020, 9, 381–386.
- 16. Zheng, Q.L.; Fan, W.; Zhang, S.M. Identification of fishing type from VMS data based on artificial neural network. *South China Fish. Sci.* **2016**, *12*, 81–87.
- 17. Tang, X.; Zhang, S.; Fan,W. Fishing type identification of gill net and trawl net based on deep learning. *Mar. Fish.* **2020**, *42*, 233–244.
- Feng, Y.; Zhao, X.; Han, M. The study of identification of fishing vessel behavior based on VMS data. In Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering, Tokyo, Japan, 9–12 November 2019; pp. 63–68.
- 19. Zhang, S.M.; Yang, S.L.; Fan, W. Algorithm of fishing effort extraction in trawling based on Beidou vessel monitoring system data. *J. Fish. China* **2014**, *38*, 1190–1199.
- 20. Shengmao, Z.; Fenghua, T.; Heng, Z. Research ontrawling tracing based on BeiDou vessel monitoring system data. *South China Fish. Sci.* **2014**, *10*, 15–23.
- 21. Wang, Y.; Wang, Y.; Zheng, J. Analyses of trawling track and fishing activity based on the data of Vessel Monitoring System (VMS): A case study of the single otter trawl vessels in the Zhoushan fishing ground. *J. Ocean Univ. China* **2015**, *14*, 89–96. [CrossRef]
- 22. Chen, R.; Wu, X.; Liu, B. Mapping coastal fishing grounds and assessing the effectiveness of fishery regulation measures with AIS data: A case study of the sea area around the Bohai Strait, China. *Ocean Coast. Manag.* **2022**, 223, 106136. [CrossRef]
- Ke, G.; Meng, Q.; Finley, T. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 2017, 30, 3149–3157.
- 24. Huang, K.; Li, G.; Wang, J. Rapid retrieval strategy for massive remote sensing metadata based on GeoHash coding. *Remote Sens. Lett.* **2018**, *9*, 1070–1078. [CrossRef]
- Liu, B. Text sentiment analysis based on CBOW model and deep learning in big data environment. J. Ambient. Intell. Humaniz. Comput. 2020, 11, 451–458. [CrossRef]
- 26. Mpomwenda, V.; Tómasson, T.; Pétursson, J.G. Adaptation Strategies to a Changing Resource Base: Case of the Gillnet Nile Perch Fishery on Lake Victoria in Uganda. *Sustainability* **2022**, *14*, 2376. [CrossRef]
- 27. Liu, Q.; Chen, Y.; Wang, J. An example of fishery yield predictions from VMS-based navigational characteristics applied to double trawlers in China. *Fish. Res.* **2023**, *261*, 106614. [CrossRef]
- Li, J.; Qiu, Y.; Cai, Y. Trend in fishing activity in the open South China Sea estimated from remote sensing of the lights used at night by fishing vessels. *ICES J. Mar. Sci.* 2022, 79, 230–241. [CrossRef]
- 29. Bærum, K.M.; Anker-Nilssen, T.; Christensen-Dalsgaard, S. Spatial and temporal variations in seabird bycatch: Incidental bycatch in the Norwegian coastal gillnet-fishery. *PLoS ONE* **2019**, *14*, e0212786. [CrossRef] [PubMed]
- Glemarec, G.; Kindt-Larsen, L.; Lundgaard, L.S. Assessing seabird bycatch in gillnet fisheries using electronic monitoring. *Biol. Conserv.* 2020, 243, 108461. [CrossRef]
- 31. Zhang, H.; Yang, S.L.; Fan, W. Spatial analysis of the fishing behaviour of tuna purse seiners in the western and central Pacific based on vessel trajectory data. *J. Mar. Sci. Eng.* **2021**, *9*, 322. [CrossRef]
- 32. O'Neill, F.G.; Mutch, K. Selectivity in trawl fishing gears. Scott. Mar. Freshw. Sci. 2017, 8, 1890–1891.
- 33. Wardle, C.S. Fish behaviour and fishing gear. Behav. Teleost Fishes 1986, 1,463–495.
- He, P.; Chopin, F.; Suuronen, P. Classification and illustrated definition of fishing gears. In FAO Fisheries and Aquaculture Technical Paper; FAO: Rome, Italy, 2021; pp. 1–11, 13–17, 19–29, 31–37, 39–41, 43–59, 61–69, 71–91, 93–94.
- Chang, S.K.; Yuan, T.L. Deriving high-resolution spatiotemporal fishing effort of large-scale longline fishery from vessel monitoring system (VMS) data and validated by observer data. *Can. J. Fish. Aquat. Sci.* 2014, 71, 1363–1370. [CrossRef]
- 36. Deng, R.; Dichmont, C.; Milton, D. Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Can. J. Fish. Aquat. Sci.* 2005, *62*, 611–622. [CrossRef]
- Guo, S.; Mou, J.; Chen, L. Improved kinematic interpolation for AIS trajectory reconstruction. *Ocean Eng.* 2021, 234, 109256. [CrossRef]

- Habermann, C.; Kindermann, F. Multidimensional spline interpolation: Theory and applications. *Comput. Econ.* 2007, 30, 153–169. [CrossRef]
- Merghadi, A.; Yunus, A.P.; Dou, J. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* 2020, 207, 103225. [CrossRef]
- 40. Marler, R.T.; Arora, J.S. Function-transformation methods for multi-objective optimization. *Eng. Optim.* 2005, 37, 551–570. [CrossRef]
- 41. Zhang, L.; Xing, B.; Chen, X. Ensemble Learning based Fishing Behavior Analysis for Vessels around Zhoushan Islands Erea//Journal of Physics: Conference Series. *IOP Publ.* **2022**, 2213, 012012.
- 42. Zhou, C.; Lu, H.; Xiang, Y. GeohashTile: Vector geographic data display method based on geohash. *ISPRS Int. J. Geo-Inf.* 2020, *9*, 418. [CrossRef]
- Liu, J.; Li, H.; Gao, Y. A geohash-based index for spatial data management in distributed memory. In Proceedings of the 2014 22nd International Conference on Geoinformatics, Kaohsiung, Taiwan, 25–27 June 2014; pp. 1–4.
- 44. Chowdhary, K.; Chowdhary, K.R. Natural language processing. Fundam. Artif. Intell. 2020, 29, 603–649.
- 45. Kenter, T.; Borisov, A.; De Rijke, M. Siamese cbow: Optimizing word embeddings for sentence representations. *arXiv* 2016, arXiv:1606.04640.
- Chen, T.; He, T.; Benesty, M. Xgboost: Extreme Gradient Boosting; R Package Version 0.4-2; Journal=R Package Version 0.4-2. 2015; Volume 1, pp. 1–4. Available online: https://cran.microsoft.com/snapshot/2015-10-20/web/packages/xgboost/xgboost.pdf (accessed on 18 May 2023).
- Ke, G.; Xu, Z.; Zhang, J. DeepGBM: A deep learning framework distilled by GBDT for online prediction tasks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; Volume 8, pp. 384–394.
- Prokhorenkova, L.; Gusev, G.; Vorobev, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* 2018, 31, 6638–6648.
- 49. Liang, W.; Luo, S.; Zhao, G. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics* **2020**, *8*, 765. [CrossRef]
- 50. Al Daoud, E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *Int. J. Comput. Inf. Eng.* **2019**, 13, 6–10.
- 51. Frazier, P.I. A tutorial on Bayesian optimization. arXiv 2018, arXiv:1807.02811.
- 52. Shahriari, B.; Swersky, K.; Wang, Z. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* 2015, 104, 148–175. [CrossRef]
- 53. Frazier, P.I. Bayesian optimization//Recent advances in optimization and modeling of contemporary problems. *Informs* **2018**, *18*, 255–278.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.