

Article

Salmon Consumption Behavior Prediction Based on Bayesian Optimization and Explainable Artificial Intelligence

Zhan Wu ^{1,†}, Sina Cha ^{2,†}, Chunxiao Wang ^{1,*}, Tinghong Qu ^{1,*} and Zongfeng Zou ³ 

¹ School of Economics and Management, Shanghai Ocean University, Shanghai 201306, China; m220751363@st.shou.edu.cn

² School of Business, The University of Hong Kong, Hong Kong 999077, China; chasina@connect.hku.hk

³ School of Management, Shanghai University, Shanghai 200444, China; zfzou@mail.shu.edu.cn

* Correspondence: cxwang2@shou.edu.cn (C.W.); m220751358@st.shou.edu.cn (T.Q.)

† These authors contributed equally to this work.

Abstract: Predicting seafood consumption behavior is essential for fishing companies to adjust their production plans and marketing strategies. To achieve accurate predictions, this paper introduces a model for forecasting seafood consumption behavior based on an interpretable machine learning algorithm. Additionally, the Shapley Additive exPlanation (SHAP) model and the Accumulated Local Effects (ALE) plot were integrated to provide a detailed analysis of the factors influencing Shanghai residents' intentions to purchase salmon. In this study, we constructed nine regression prediction models, including ANN, Decision Tree, GBDT, Random Forest, AdaBoost, XGBoost, LightGBM, CatBoost, and NGBoost, to predict the consumers' intentions to purchase salmon and to compare their predictive performance. In addition, Bayesian optimization algorithm is used to optimize the hyperparameters of the optimal regression prediction model to improve the model prediction accuracy. Finally, the SHAP model was used to analyze the key factors and interactions affecting the consumers' willingness to purchase salmon, and the Accumulated Local Effects plot was used to show the specific prediction patterns of different influences on salmon consumption. The results of the study show that salmon farming safety and ease of cooking have significant nonlinear effects on salmon consumption; the BO-CatBoost nonlinear regression prediction model demonstrates superior performance compared to the benchmark model, with the test set exhibiting RMSE, MSE, MAE, R^2 and TIC values of 0.155, 0.024, 0.097, 0.902, and 0.313, respectively. This study can provide technical support for suppliers in the salmon value chain and help their decision-making to adjust their corporate production plan and marketing activities

Keywords: machine learning; SHAP model; consumption behavior prediction; influencing factors; salmon



Academic Editors: Ursula Andrea Gonzales-Barron and Sakamon Devahastin

Received: 27 November 2024

Revised: 26 January 2025

Accepted: 27 January 2025

Published: 28 January 2025

Citation: Wu, Z.; Cha, S.; Wang, C.; Qu, T.; Zou, Z. Salmon Consumption Behavior Prediction Based on Bayesian Optimization and Explainable Artificial Intelligence. *Foods* **2025**, *14*, 429. <https://doi.org/10.3390/foods14030429>

Copyright: © 2025 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/>).

1. Introduction

China is the largest consumer of aquatic products in the world, and the demand for high-quality aquatic products has increased drastically with the rising disposable income of Chinese urban residents and their increasing concern for dietary health issues [1,2]. Among the many aquatic products, salmon is increasingly popular among consumers due to its taste and rich nutritional value, especially its high-protein, low-fat properties and richness in Omega-3 fatty acids. With consumers' continuous pursuit of health and quality, the consumption trend of salmon in China will be further expanded, and the market potential is huge [3,4]. The study of factors affecting the consumers' purchasing behavior and its

prediction are not only conducive to improving the consumption level of salmon consumers in China, but also to optimizing the dietary structure of consumers. At the same time, it is beneficial for seafood suppliers to adjust their production and marketing plans according to consumers' preferences in order to increase their profits and market shares, which are of great significance in promoting the high-quality development of China's fishing industry.

In recent years, scholars at home and abroad, based on the theory of planned behavior and the theory of consumer behavior, have mainly used the structural equation model, the logistic regression (Logistic) model, and the multivariate probability ratio regression (Probit) model [5–8] to study the consumers' behavior when purchasing aquatic products and the factors influencing it from different perspectives. Meng et al. used the Tobit model to study the key factors affecting aquatic product consumption among Shanghai residents, and the study showed that income, aquatic product safety perception, and quality preference were important determinants of aquatic product purchase among Shanghai residents [9]. Fish consumption preference is negatively correlated with family size and positively correlated with age, education level, and income level [10–12]. In terms of purchasing channels, consumers have different types of behavioral preferences when purchasing at farmers' markets, supermarkets, etc., Skallerud et al. explored the consumers' store choice behavior for purchasing Norwegian salmon in a cross-cultural context, and found that the consumers' culture, education, family size, employment status, product preferences, and attitudes toward the product were highly correlated with their choices of different retail outlets when purchasing Norwegian salmon [13]. Furthermore, in terms of salmon safety, the proportion of salmon purchased with certification and traceability is increasing as the consumers' awareness continues to grow. Haghiri et al. showed that the likelihood of paying a high price for a certified product increases when food quality attributes and sustainable agricultural practices, are combined with the concept of food safety [14]. Britwum et al. used the Multinomial logit model to assess the U.S. consumers' willingness to pay for farmed salmon and found that certified farmed seafood has a higher price premium, as the people's confidence in the safety of farmed seafood increases [15]. Zhang et al. used the ordered probit model and the two-limit Tobit model to study the seafood consumption patterns of the residents of six provinces and cities in China, and found that health motivation is an important predictor of seafood consumption [16].

However, consumption behavior constitutes a multifaceted, coupled, nonlinear system, rendering traditional statistical methods inadequate for capturing its intricate nonlinear relationships and accurately forecasting consumer actions [17–19]. Conversely, machine learning models exhibit robust nonlinear fitting and adaptive capabilities, efficiently mining complex features within multidimensional and redundant datasets, thereby finding application in the domain of consumer behavior analysis [20,21]. In recent years, there has been a surge in research leveraging machine learning to predict consumer preferences and future purchasing patterns. Kumar and colleagues developed a predictive model for consumer repurchase intentions, utilizing the Artificial Bee Colony (ABC) algorithm, alongside five machine learning algorithms. Their findings revealed that the AdaBoost model surpassed neural networks and supported the vector machines' performance [22]. Trivedi and team formulated an interpretable improved stacked SVM (XISS) model to forecast consumer refill frequencies for liquefied petroleum gas (LPG), achieving a prediction accuracy of up to 81% [23]. Andoh-Odoom et al. employed classification and regression trees to anticipate factors influencing consumer preferences, purchasing behaviors, and willingness to purchase tilapia fillets, validating the model's efficacy [24].

Compared with traditional machine learning models, Extreme Gradient Boosting (XGBoost) is an integrated gradient boosting learning algorithm, which can reveal the underlying mechanism between input features and target results, and can effectively

deal with abnormal, non-linear, high-dimensional data, with advantages such as a high prediction accuracy and less overfitting [25]. Hu W. et al. showed that the Random Forest algorithm is more conducive to dealing with complex nonlinear systems that objectively reflects the indicator's contribution [26]. Lei, Y. et al. applied the XGBoost model to quantify the importance of each variable for the monthly fire risk level [27]. In addition, as interpretable machine learning becomes more popular, machine learning methods can extract unique insights from large datasets with a large number of feature variables [28,29]. Wang M. et al. used Shenzhen City as a research object, constructed an XGBoost model, combined it with Shapley additive explanatory maps and partial dependency maps, and showed that the mean building volume was a key parameter, with an average SHAP value of 0.0107 m and a contribution rate of 9.70% [30].

In conclusion, the visual and quantitative capabilities of machine learning that can be explained can be used to further study the driving mechanism that affects consumers' purchase intention. However, the current domestic and foreign literature on the factors affecting the consumption of aquatic products mainly focuses on the willingness to pay and on purchase behavior of aquatic products. Most studies focus on the broad categories of aquatic products. Few people study the purchase factors of specific seafood, and few studies use interpretable machine learning to predict and analyze the driving factors of seafood consumption behavior. Therefore, using the research on the salmon consumption behavior of Shanghai residents, a prediction model of seafood consumption behavior based on nine algorithms including neural network, XGBoost, LightGBM, CatBoost, and NGBoost, was established. SHAP was used to assess the effect sizes and interaction effects of different factors, and the Accumulated Local Effects Plot was used to analyze the specific prediction patterns of the characteristic variables, such as aquaculture safety and the individual's household income, on salmon consumption. The possible marginal contributions of this paper lie in comprehensively applying multiple machine learning methods to study the salmon consumption problem, thus enriching the related research; more accurately revealing the complex relationships among variables when exploring the influencing factors of salmon consumption by employing cutting edge Ensemble learning; and utilizing the interpretable methods in machine learning, exploring the significance of different influencing factors on salmon consumption, and analyze the specific prediction models subject to salmon quality attributes and other important influencing factors, which has important reference value and technical support for promoting the high-quality development of salmon enterprises and for the suppliers in the salmon value chain to adjust their enterprise production plans and marketing activities.

2. Materials and Methods

2.1. Bayesian Optimization Algorithm

The Bayesian optimization (BO) algorithm, as a versatile and powerful tool, has demonstrated its outstanding application value in many fields, such as optimizing robot mechanics, environmental monitoring, and material design [31–33]. Furthermore, in terms of hyperparametric optimization for machine learning algorithms, the work of Feurer and Hutter in 2019 further confirmed the effectiveness of Bayesian optimization algorithms [34].

Bayesian optimization is a global optimization algorithm proposed by Pelikan et al. [35] to find the optimal solution to minimize the objective function in a high-dimensional non-convex search space, where the posterior distribution of the objective function (Gaussian process) is updated by continuously adding sample points without knowing the internal structure and mathematical properties of the optimized objective

function until the posterior distribution fits the real distribution. The objective function of the Bayesian optimization algorithm is as follows:

$$X^{(*)} = \arg \min_{X \in S} f(X) \quad (1)$$

where $X^{(*)}$ denotes the optimal set of hyperparameters; S represents the candidate set of x ; and $f(X)$ is the objective function.

2.2. Boosting Regression Algorithm

The Extreme Gradient Boosting (XGBoost) algorithm is an ensemble learning algorithm that has been widely used in the field of machine learning in recent years, and was developed by Tianqi Chen et al. [36]. It is based on the idea of Gradient Boosting Decision Tree (GBDT), and constructs a stronger supervised model by combining multiple weak learners to reduce the learning error and updating the sample weights during each iteration.

The core strength of the XGBoost algorithm lies in the second-order Taylor expansion of the loss function and the introduction of a regularization term, which work together to prevent the overfitting of the model and thus improve its generalization ability [37]. The mathematical principle of the XGBoost algorithm can be summarized as follows:

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in F \quad (2)$$

The XGBoost constructs the model by iteratively adding CARTs (Classification and Regression Trees), where k is the number of CARTs and f_k is a function in the function space F in Equation (2). At each iteration, XGBoost adds a new tree model to reduce the residuals. The optimization objective function required for the XGBoost regression model is as follows.

$$obj^{(t)} = L(y_i, \hat{y}_i^t) + \Omega(f_t) \quad (3)$$

where, $L(y_i, \hat{y}_i^t)$ is the training loss function, which is used to measure the predictive ability of the model. $\Omega(f_t)$ is the regularization term, related to the complexity of the tree, used to control the complexity of the model to prevent overfitting. The formula is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4)$$

where T is the number of leaf nodes, w_j is the score of the j th leaf node, and γ and λ are regularization parameters controlling the number of leaf nodes and the size of leaf node weights, respectively. In this paper, we use the mean square error (MSE) as the loss function and the objective becomes as follows

$$\begin{aligned} obj^{(t)} &= \sum_{i=1}^n \left(y_i - \left(\hat{y}_i^{(t-1)} + f_t(x_i) \right) \right)^2 + \sum_{i=1}^t \omega(f_i) \\ &= \sum_{i=1}^n \left[2 \left(\hat{y}_i^{(t-1)} - y_i \right) f_t(x_i) + f_t(x_i)^2 \right] + \omega(f_t) + constant \end{aligned} \quad (5)$$

To facilitate the solution, we perform a second-order Taylor expansion of the loss function of the XGBoost model to obtain an approximation of the objective function in quadratic form.

$$obj^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) + constant \quad (6)$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (7)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (8)$$

where g_i and h_i are the first-order derivative (gradient) and second-order derivative (Hessian matrix) of the loss function with respect to the predicted value, respectively. In XGBoost, g_i and h_i denote the gradient and curvature of the current prediction value, respectively.

XGBoost uses greedy algorithms to build the tree models step by step. In each iteration step a new tree model is learned to reduce the objective function. The model updates weights through leaf split and leaf weight. At each dividing point, the gain of the dividing point is calculated. Gain indicates the reduction in error after segmentation. For the split point, the gain can be expressed as:

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (9)$$

where G_L and G_R are the sets of samples from the left and right subtrees after tree classification, $G_L^2/(H_L + \lambda)$ is the information score of the left subtree, $G_R^2/(H_R + \lambda)$ is the information score of the right subtree, and $(G_L + G_R)^2/(H_L + H_R + \lambda)$ is the information score of the current unsegmented tree.

The LightGBM algorithm, developed by Microsoft, employs a histogram-based decision tree approach to enhance computational efficiency [38]. This method involves discretizing continuous feature values into a finite number of bins, thereby simplifying the process of identifying optimal splits. By focusing on the leaf-wise growth strategy, LightGBM iteratively selects the leaf node with the highest gain for splitting, which not only accelerates tree growth but also improves model training efficiency. This strategy also contributes to superior data fitting, leading to enhanced model accuracy.

CatBoost, developed by Yandex, is a gradient boosting framework that employs oblivious decision trees as its fundamental learner [39]. The CatBoost algorithm combines classification and boosting techniques to efficiently process features according to category. It solves problems related to gradient and prediction bias. CatBoost can automatically process categorical features without converting them to numerical features.

It processes the ordering information of the categorical features through a technique called “ordered boosting” and uses this information in the tree splitting process. In addition, CatBoost features an adaptive learning rate, which can adaptively adjust the learning rate according to different features. The adaptive learning rate is calculated as follows:

$$\eta_t = \frac{1}{t + \lambda} \quad (10)$$

$$\alpha_t = \frac{\sum_{i=1}^t \eta_i}{t} \quad (11)$$

where t is the number of iterations, η_t is the learning rate of the t th round of iterations, α_t is the average learning rate of the previous t rounds of iterations, and λ is a regularization parameter. The adaptive learning rate helps the algorithm to better control the contribution of weak learners in each round of iterations, thus improving the accuracy of the whole model.

2.3. SHapley Additive ExPlanations Model and the Accumulated Local Effects Plot

The SHAP (SHapley Additive exPlanations) model is a method for interpreting the predictions of machine learning models, proposed by Lundberg and Lee in 2017 [40]. The SHAP values are based on the Shapley values from game theory, and are used to equitably distribute each participant's contribution to the collective outcome. The SHAP values have

several key properties, including efficiency, symmetry, virtuality, and additivity [41]. These properties ensure that the sum of the characteristic contributions is equal to the difference between the predicted value and the mean, and that the aggregation of predictions from a single model is equal to the predictions of all the models combined. The formula for calculating the SHAP value is shown below:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{ij}) \quad (12)$$

where y_{base} is the average value of the target variable over all samples; $f(x_{ij})$ is the SHAP value of x_{ij} . The calculation of SHAP values involves multiplying the marginal contribution of each feature by the corresponding weight and then summing. This method not only reflects the contribution of features in each sample, but also shows the positivity and negativity of the effect. In this study, the SHAP algorithm, which is based on game theory, was used to interpret and analyze the model. The algorithm possesses consistency and local accuracy and can effectively interpret the results of machine learning prediction models. This approach can provide deeper understanding and insight into consumer behavior prediction research, which can help optimize the model prediction results and improve user trust and satisfaction.

The LIME method may be influenced by randomness and sampling techniques when interpreting model predictions, potentially resulting in unstable interpretation outcomes. While the CAM method offers the model's attentional weights for various spatial locations, its interpretive scope is relatively confined to a specific deep learning model. Consequently, the SHAP method holds a relative edge over LIME and CAM in terms of interpretability, stability, and scope [42].

In this paper, ALE plots are used to describe the specific prediction patterns of key influencing factors on salmon. ALE plots eliminate the influence of variable correlation by calculating local effects. First, the range of values of the characteristic variables is divided into intervals, ensuring that each interval has the same number of data points. Second, the local effects within each interval are calculated, which results in the value of the cumulative local effects of the feature variables on the model predictions [43,44].

2.4. Model Construction and Experimental Process

In this study, the original data were randomly divided into 319 sets of training set samples and 80 sets of test set samples using python's scikit-learn, and were used as input features for nine regression prediction models, including XGBoost, LightGBM, CatBoost, and NGBoost, based on the consumers' personal characteristics and aquatic product attributes to predict salmon consumption behavior. Finally, the SHAP model and the ALE plot were used to analyze the regression prediction results of the optimal model and to explore the key factors influencing consumers to purchase salmon. The experimental environment of the study was Python 3.8, TensorFlow 1.15.5, and Pytorch 1.11.0. To ensure the consistency of the data scale, the data were normalized and uniformly mapped into the interval of [0,1]. The normalization formula is as follows:

$$X = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (13)$$

where X and X_i are the data after and before normalization, respectively.

In addition, a series of important third-party libraries are introduced. Among them, the Sklearn machine learning library provides rich machine learning algorithms and tools for data processing and model evaluation; the SHAP library provides two modules for global analysis and local analysis. To more comprehensively and accurately evaluate the effectiveness of the proposed model, several widely recognized statistical metrics were

utilized, including mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), R-squared (R^2) and Terrell's Inequality Coefficient (TIC). These indicators are commonly employed to assess the predictive performance of models, with smaller values signifying lower prediction errors and, consequently, higher prediction accuracy. The mathematical expressions for these metrics are presented as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

$$\text{TIC} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 + \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{y}_i^2}}} \quad (18)$$

where y_i and $\hat{y}_i$ are the actual and predicted values at time i , respectively; n is the number of predicted data.

3. Data Sources and Analysis

3.1. Data Sources

This study takes Shanghai residents as the research object, combines the related literature [45–47], and designs the questionnaire. The research questionnaire includes three main parts. The first part includes individual consumer characteristics, such as gender, age, education level, household income, and family demographics. The second part includes the consumers' perception of salmon, including four aspects of the commodity: nutritional quality, safety quality, sensory quality, and added value. The third part deals with the consumers' willingness to consume salmon. The formal questionnaire survey was initiated after the questionnaire was thoroughly revised and free of errors. The research data were obtained through a combination of online and offline methods: online through the Questionnaire Star web-based program; offline through the random distribution of questionnaires to passers-by in residential areas and major supermarkets, filling out the questionnaires on the spot, and then recycling them to carry out the research. A total of 450 questionnaires were distributed, of which 399 were valid, accounting for 88.67%. In addition, regarding the sample size, we followed the rule of thumb for machine learning models that the sample size should be at least 10 to 100 times the number of attributes/features considered in the study [19,48]. A total of 20 feature variables were considered in this study to predict salmon purchase intention. Therefore, 399 samples met the a priori condition.

3.2. Basic Consumer Characteristics

The basic information of the respondents is shown in Table 1. In terms of gender, males and females accounted for 40.9% and 59.1% of the total samples, respectively, indicating that females are the majority of the buyers of salmon. Next, the age group of 21 to 44 years accounted for 77.65% of the total sample. The average education level of the respondents was relatively high, with the majority holding a bachelor's degree or higher. Monthly household income was mainly between \$5000 and \$10,000, and the monthly household

income was mainly between \$10,001 and \$20,000, accounting for 7% and 9.5% of the entire sample, respectively.

Table 1. Statistical characteristics of respondents.

Demographic Variables	Items	Frequency	Percentage %
Gender	Male	163	40.9
	Female	236	59.1
Age	21–25	88	21.5
	25–34	146	36.6
	35–44	78	19.55
	45–54	67	16.8
	Over 55	20	5
Education	Lower secondary school and below	45	11.3
	Higher secondary school	79	19.8
	Bachelor	176	44.1
	Master	89	22.3
	Phd	10	2.5
Monthly household income	Less than 5000 RMB	133	33.33
	5000–10,000 RMB	161	40.35
	10,000–15,000 RMB	28	7
	15,000–20,000 RMB	38	9.5
	Above 20,000 RMB	39	9.8

4. Results

4.1. Variable Description and Assignment

This paper selects whether consumers buy salmon as an explained variable and assigned values of 0 and 1. The naming and assignment of other independent variables are shown in Table 2. In order to reproduce the experiment in this paper, the number of random seeds in the dataset partition is 291. Nine regression prediction models, such as ANN, XGBoost, LightGBM, CatBoost, and NGBoost, are set as the default parameters of the Skelearn machine learning library. The number of random seeds is set to 100. The prediction performance accuracy of the nine models is compared in the test set, and the optimal prediction model is selected.

Table 2. Salmon quality attribute variable definitions and assignments.

Features	Variable	Variable Assignment
Quality and safety attributes	Frequent Salmon Eater	1 = Strongly Disagree, 2 = Disagree, 3 = Comparatively, Disagree 4 = Average, 5 = Agree, 6 = Comparatively, Agree 7 = Strongly Agree
	Farming water is clear	1 = Strongly Disagree, 2 = Disagree, 3 = Average, 4 = Agree, 5 = Strongly Agree
	No artificial feed added to the feed	1 = Strongly Disagree, 2 = Disagree, 3 = Average, 4 = Agree, 5 = Strongly Agree
	No antibiotics	1 = Strongly Disagree, 2 = Disagree, 3 = Average, 4 = Agree, 5 = Strongly Agree
	High nutritional value	1 = Yes, 0 = No
Marketing Attributes	High quality	1 = Yes, 0 = No
	Reasonable price	1 = Yes, 0 = No
	Brand	1 = Yes, 0 = No
	Purchase Convenience	1 = Yes, 0 = No

4.2. Regression Prediction Model Performance Analysis

In addition to the random seed setting, we also paid attention to the consistency of other key parameters, including the selection of evaluation metrics, the size of the training batch, and the setting of the training period. The uniform configuration of these parameters provides a standardized environment for our model training and evaluation, making comparisons between different models fairer and more accurate. Under these strict experimental control conditions, we obtained the prediction performance results of the benchmark model, and these results are shown in Table 3.

Table 3. Evaluation of regression prediction model performance.

Model Name	RMSE	MSE	MAE	R ²	TIC
ANN	0.202	0.041	0.124	0.833	0.409
DT	0.224	0.05	0.05	0.795	0.452
RF	0.181	0.033	0.094	0.866	0.366
AdaBoost	0.255	0.065	0.209	0.734	0.515
GBDT	0.173	0.03	0.109	0.877	0.35
XGBoost	0.195	0.038	0.093	0.845	0.394
LightGBM	0.226	0.051	0.141	0.791	0.457
NGBoost	0.174	0.03	0.11	0.876	0.352
CatBoost	0.164	0.027	0.093	0.889	0.332

RMSE, MSE, MAE, R² and TIC are key metrics to evaluate the performance of a model. The lower the RMSE value, the lower the prediction error; the closer the R² value is to 1, the better the model's explanatory ability. This study compares nine machine learning models, including ANN, XGBoost, LightGBM and CatBoost. The results show that the R² value of AdaBoost model is 0.734, which is the lowest among the four models, and the TIC value is 0.515 with the largest error. The RMSE, MSE, MAE, R² and TIC values are 0.877 for the CatBoost regression prediction model. The RMSE, MSE, MAE, R² and TIC of the CatBoost regression prediction model were 0.164, 0.027, 0.093, 0.889 and 0.332, respectively, which indicated that its prediction error was relatively small, and the model prediction accuracy was better than that of the other benchmark models. Given the superior performance of the CatBoost regression prediction algorithm in the prediction of salmon consumption behavior, we chose the CatBoost model and combined it with the SHAP (SHapley Additive exPlanations) model to conduct an in-depth analysis of the key factors affecting salmon purchase behavior.

Figure 1 provides a comprehensive assessment of the nine regression prediction models' performance using Taylor plots. Taylor plots summarize multiple aspects of model performance in a single plot, including correlation coefficients, normalized standard deviations, and distances from the observations [49,50]. The gray dashed line is the RMSE contour. The red dashed line is the reference line with a standard deviation of 1, which is used to quickly determine prediction accuracy. According to Figure 1, the CatBoost model provides the best prediction performance.

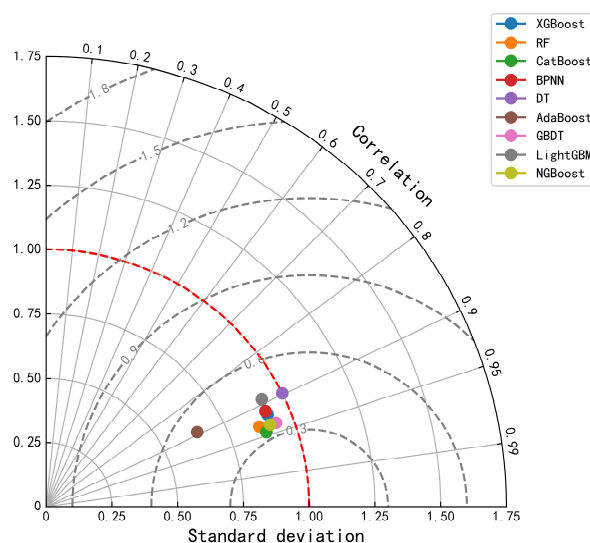


Figure 1. Taylor diagrams for regression prediction models. The gray dashed line is the RMSE contour. The red dashed line is the reference line with a standard deviation of 1, which is used to quickly determine prediction accuracy.

To avoid the overfitting problem of the CatBoost model, the model parameters must be optimized. In CatBoost hyperparameter tuning, the Bayesian optimization method automatically finds the best hyperparameter configuration to minimize the model validation error [51,52]. In this paper, we identify the most important parameters and value ranges that may affect the validity of the CatBoost model. These parameters and value ranges will be replaced in the Bayesian optimization algorithm. The combinations of the hyperparameters used for the CatBoost model prediction are as follows: the optimal range of depth is [3, 10]; the optimal range of depth is [1, 500]; and the optimal range of learning rate is [0.01, 0.3]. The specific parameters are shown in Table 4.

Table 4. Bayesian parametric optimization.

Model Name	Parameter Name	Parameter Setting
BO-CatBoost	depth	3
	iterations	267
	Learning_rate	0.136

As shown in Table 5, the RMSE, MSE, MAE, R^2 , and TIC of the BO-CatBoost prediction set are 0.155, 0.024, 0.097, 0.902, and 0.313, respectively, which are better than that of the benchmark CatBoost prediction model. The R^2 and TIC of the BO-CatBoost model are 1.46 percent better than that of the benchmark CatBoost prediction model. The R^2 and TIC of the BO-CatBoost model are 1.46% higher and 5.72% lower than those of the benchmark CatBoost model, respectively.

Table 5. BO-CatBoost Predicted Performance.

Model Name	RMSE	MSE	MAE	R^2	TIC
CatBoost	0.164	0.027	0.093	0.889	0.332
BO-CatBoost	0.155	0.024	0.097	0.902	0.313

In this paper, the tenfold cross-validation method is used to test whether the model has the overfitting phenomenon, as shown in Figure 2. The tenfold cross-validation method means that the data are divided into ten parts, nine of which are used as training data, and the remaining part is used as test data, and the cycle is repeated ten times to ensure that each set of data is tested once [53]. The final evaluation result is the average of the ten evaluations. The results of the tenfold cross-validation show that the average value of R^2 for the test set of the CatBoost model is 0.84, indicating that the CatBoost model has good regression prediction performance and robustness.

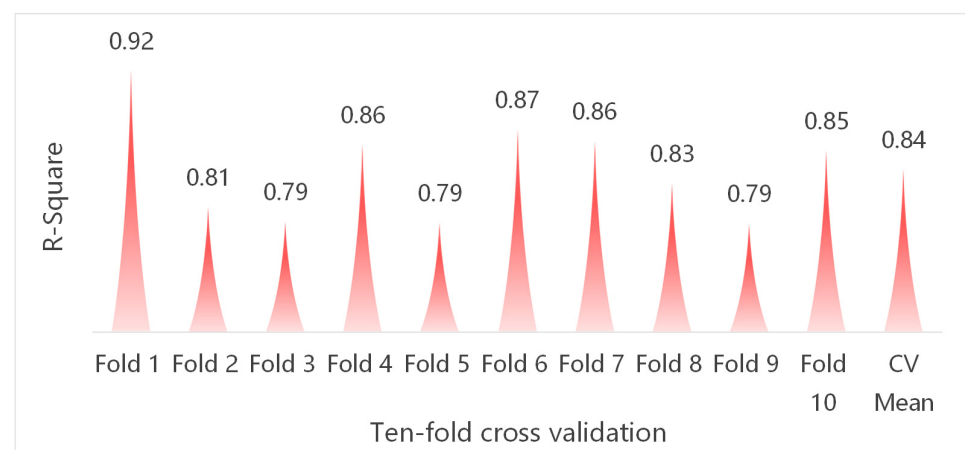


Figure 2. Cross-validation results.

4.3. SHAP Modeling and ALE Analysis

Figure 3 presents the global feature analysis of the CatBoost model using SHAP values. Key predictors of salmon consumption include consumer habits, price, and the use of antibiotics in farming practices. To gain a more intuitive understanding of how these features influence the model's output, extract valuable insights, and assist companies in implementing targeted marketing strategies, this study employs a SHAP value mapping diagram to illustrate the nonlinear relationships among variables [48]. Unlike partial dependency plots, the vertical axis of the SHAP value mapping diagram represents the SHAP value itself, rather than the labeled output value. This approach allows for a more effective analysis of the nonlinear impacts of significant predictors on salmon consumption.

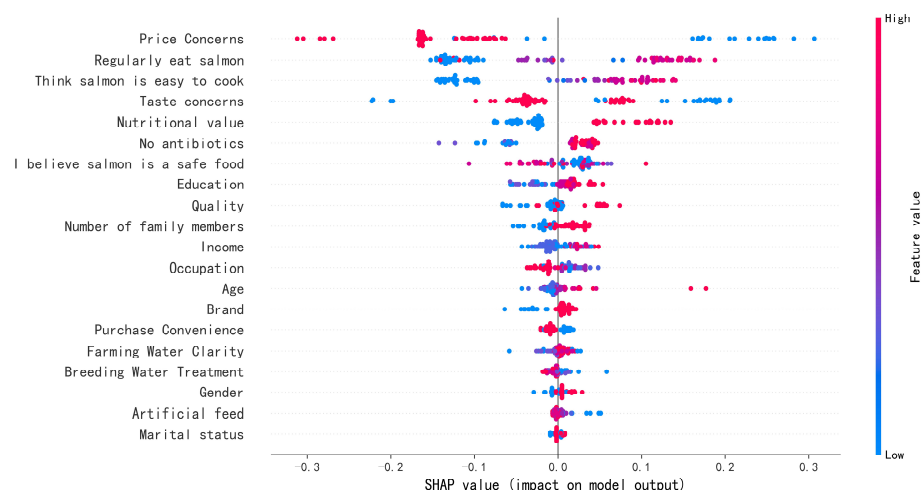


Figure 3. SHAP summary plot.

From the respondents' viewpoint on the quality and safety factors in salmon farming, salmon without medication in the farming process is more popular among consumers; as their living standard improved, the consumers began to pay more attention to the salmon's quality and safety when purchasing the product. In regard to taste, salmon with good taste presents a negative effect on purchase intention. In regard to quality, consumers are more inclined to choose salmon with good quality. As the living standard of the population has improved, the demand for aquatic products has shifted from quantity to quality, and consumers will choose the best quality products within their price range. From the perspective of respondents' marketing concerns for salmon, a reasonable price has a significant impact on the consumers' willingness to buy. The consumers' willingness to purchase is negatively impacted by reasonable pricing. Since salmon is a more expensive aquatic product than other aquatic products, there is a negative correlation between price and consumers' willingness to purchase; a lower price for salmon could cause consumers to doubt the product's quality.

Higher literacy levels correlate with increased consumption of salmon, as individuals with greater literacy are more likely to purchase medium to high-end aquatic products. This demographic tends to possess a better understanding of salmon and generally has a greater economic capacity to afford it. From an income perspective, there is a direct relationship between income levels and salmon consumption; as income rises, so does the demand for salmon. This fish is highly sought after in the domestic market, yet the costs associated with salmon farming and logistics are substantial, coupled with a short shelf life. Consequently, these factors contribute to the elevated price of salmon. Interpretive modeling is established through the design of the model, while the SHAP model elucidates the relationships among the variables. The insights derived from the SHAP model and

ALE diagrams do not merely replicate previous findings, even though they align with the conclusions drawn from studies utilizing interpretive modeling.

The SHAP method interprets the predictions of an instance by calculating the extent to which each feature in the instance contributes to the prediction. This instance-specific calculation allows SHAP to analyze the predictions of any instance, i.e., to interpret the model locally. Figure 4 shows an explanatory power diagram for a single instance that allows personalized marketing to be assigned to consumers. SHAP labels the degree of contribution of each feature, with blue indicating a negative contribution to salmon consumption and red indicating a positive contribution to salmon consumption [54]. In this instance, taste and ease of cooking provide a positive gain for consumers purchasing salmon; affordable price and safe farming provide negative gains.

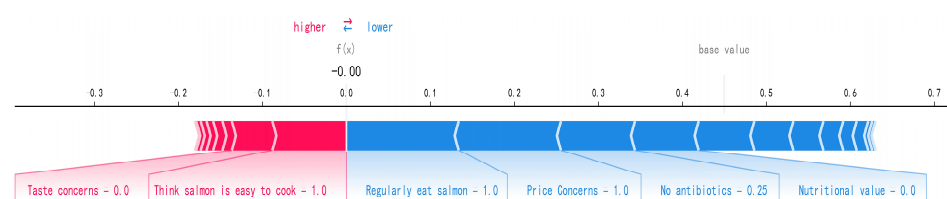


Figure 4. SHAP force plot.

The one-sample explanation helps to customize the marketing of salmon consumption, and in order to explore the impact of these features on the model output more intuitively, this paper uses SHAP value mapping plots to demonstrate the nonlinear relationship between the variables. Unlike the partial dependency plot, the vertical coordinate of the SHAP value mapping plot is the SHAP value rather than the output label value [55,56]. It can reflect the magnitude of the marginal effect between each variable and salmon consumption. As can be seen from Figures 5 and 6, when the consumer attitude is at level 1 to 3, the SHAP value is negative, and the negative effect increases. When the consumer attitude is greater than level 4, it shows a significant positive gain, indicating that salmon without medication in the aquaculture process is more popular among consumers.

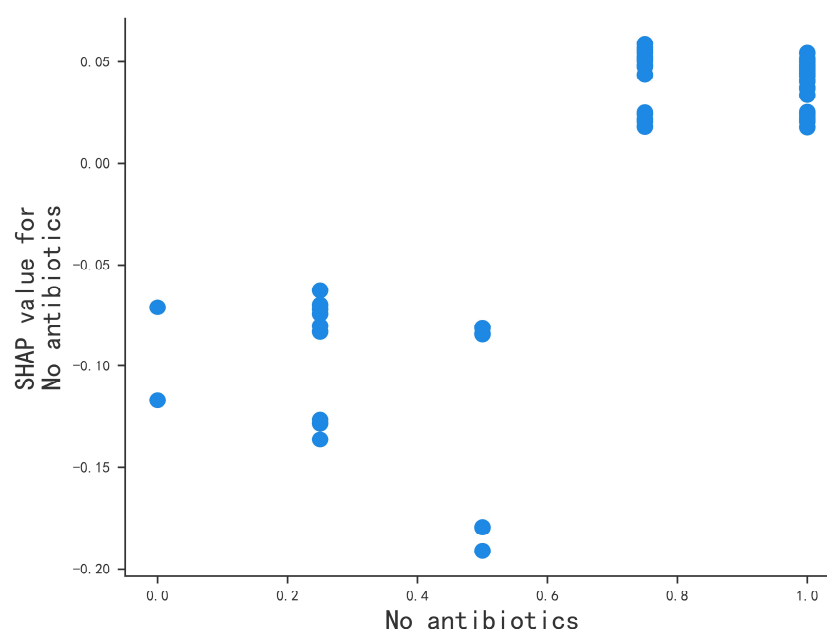


Figure 5. Analysis of SHAP feature dependency plot.

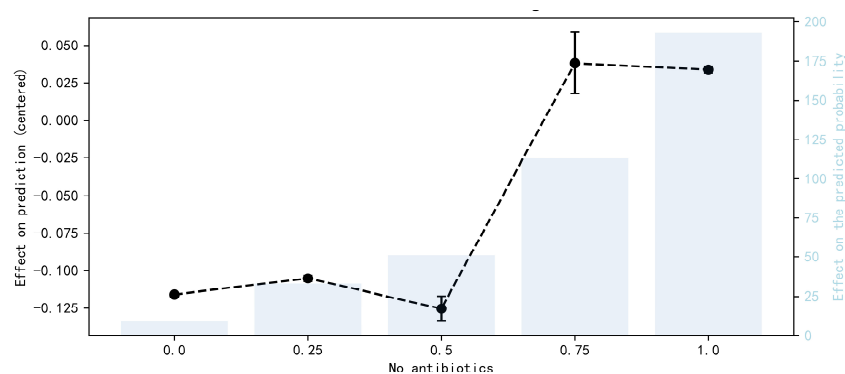


Figure 6. Analysis of accumulated local effects plot.

This paper analyzes the effects of the consumer income and the farming safety on the willingness to pay for salmon based on the SHAP model, as shown in Figure 7. When consumer income is greater than 10,000 and the external quality attributes of drug-free farmed salmon reach level 3 and above, there is a positive gain in consumer willingness to pay for salmon.

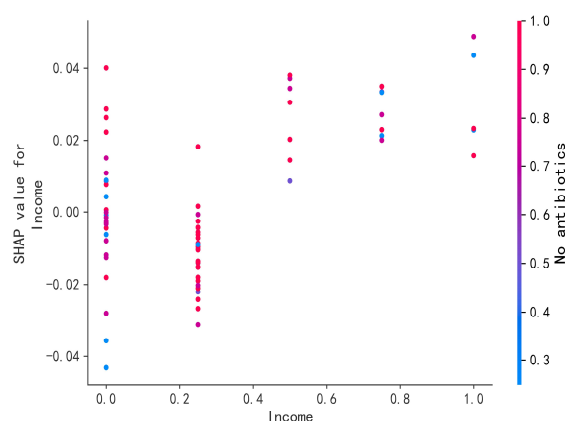


Figure 7. SHAP decision diagram.

5. Conclusions

The fishing industry is one of the most important industries in China, and its development is of great significance in promoting sustainable economic and social development in China and guaranteeing its food security. In this study, an interpretable seafood consumption behavior prediction model based on machine learning is proposed to predict salmon consumption behavior. The results show that the CatBoost model with Bayesian optimization has a high accuracy in seafood consumption behavior prediction and that the RMSE, MSE, MAE, R^2 and TIC of the test set are 0.155, 0.024, 0.097, 0.902, and 0.313, respectively. The intuitive analysis of the results of the SHAP interpretable instrument shows that gender and education level have less influence on salmon consumption; reasonable price and whether antibiotics are used in the farming process are the key factors influencing the consumption of salmon; the interaction analysis plot of the SHAP model shows that when the consumers' income is more than 10,000, and there were no antibiotics used in the farming process of salmon, the consumers' willingness to pay for salmon increases positively.

Given that seafood consumption behavior is influenced by a combination of many factors, it is a great challenge to achieve an accurate prediction of it. In order to improve the accuracy of prediction, future research should focus on the deep integration of machine learning methods with causal inference frameworks. Specifically, on the one hand, the

intrinsic correlation between variables should be explored via predictive modeling; on the other hand, the causal mechanism should be constructed and verified via explanatory modeling to achieve the full potential of the complementary advantages of the two in terms of objectives and functions, so as to further deepen and broaden the understanding of the factors influencing salmon consumption.

In addition, multimodal techniques and unsupervised machine learning methods should be combined to explore seafood consumption patterns. Incorporating more types and sources of seafood consumption sample data and focusing on the application of spatial-temporal deep learning and interpretable frameworks in seafood consumption behavior prediction, as well as providing more accurate and in-depth insights for the prediction of seafood consumption behavior should be the aim of future research.

Author Contributions: Methodology, Z.W.; Data curation, C.W.; Software, Z.W.; Writing—original draft, Z.W., S.C. and C.W.; Writing—review and editing, Z.Z., S.C., T.Q. and C.W.; Supervision, C.W. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Agricultural Research System (CARS-47-G29).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting this study's findings are available from the corresponding author upon reasonable request.

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

References

1. Budhathoki, M.; Lincen, L.; Xu, H.; Zhang, W.; Li, S.; Newton, R.; Campbell, D.; Little, D. Understanding Farmed Salmon Imports and E-Commerce Consumer Satisfaction in China: A Text Mining Approach. *J. Agric. Food Res.* **2024**, *18*, 101342. [\[CrossRef\]](#)
2. Budhathoki, M.; Campbell, D.; Belton, B.; Newton, R.; Li, S.; Zhang, W.; Little, D. Factors Influencing Consumption Behaviour towards Aquatic Food among Asian Consumers: A Systematic Scoping Review. *Foods* **2022**, *11*, 4043. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Dong, M.; Zhang, Z.; Wang, H.-P.; Huang, X.; Wang, X.; Qin, L. Discrimination and Evaluation of Commercial Salmons by Low-Molecule-Weight Compounds: Oligopeptides and Phosphatides. *Food Chem.* **2024**, *455*, 139777. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Du, Y.; Lou, X.; Oishi, T.; Liu, Y. The Influence of Quality Characteristics of Aquatic Products on Its Price Determination in China—A Case of Salmon Products in Supermarkets of Shanghai. *Aquac. Fish.* **2020**, *5*, 317–322. [\[CrossRef\]](#)
5. Williams, R.; Jorgensen, A. Comparing Logit & Probit Coefficients between Nested Models. *Soc. Sci. Res.* **2023**, *109*, 102802. [\[CrossRef\]](#)
6. Chico, J.R.; García, M.J.; Sánchez, A.R.P.; Sánchez, J.A.L. Socioeconomic Acceptance of Aquaculture Measured through Structural Equations Modeling. Are There Discrepancies between Men and Women in Spain? *Aquac. Econ. Manag.* **2022**, *26*, 439–461. [\[CrossRef\]](#)
7. Thompson, O.A.; Arifalo, S.F.; Atejiaye, A.A. Determinants of Climate Change Risk Management Strategies Among the Aquaculture Fish Farmers in Nigeria Using Multinomial Logit Model. *Fish. Aquac. J.* **2021**, *12*, 274. [\[CrossRef\]](#)
8. Lee, M.-K.; Nam, J. The Determinants of Live Fish Consumption Frequency in South Korea. *Food Res. Int.* **2019**, *120*, 382–388. [\[CrossRef\]](#)
9. Meng, T.; Wang, C.; Florkowski, W.J.; Yang, Z. Determinants of Urban Consumer Expenditure on Aquatic Products in Shanghai, China. *Aquac. Econ. Manag.* **2023**, *27*, 1–24. [\[CrossRef\]](#)
10. Saidi, A.; Cavallo, C.; Del Giudice, T.; Vecchio, R.; Cicia, G. Consumer Preferences for Finfish: A Systematic Literature Review. *Food Qual. Prefer.* **2023**, *105*, 104786. [\[CrossRef\]](#)
11. Wang, K.; Ping, Y. Analysis on Econometric Model of the Influencing factors of Aquatic Consumer Demand. *Chin. Agric. Sci. Bull.* **2012**, *28*, 102–106. [\[CrossRef\]](#)
12. Xu, H.; Wu, T.; Budhathoki, M.; Fang, D.S.; Zhang, W.; Wang, X. Consumption Patterns and Willingness to Pay for Sustainable Aquatic Food in China. *Foods* **2024**, *13*, 2435. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Skallerud, K.; Myrland, Ø.; Olsen, S.O. Buying Norwegian Salmon: A Cross-Cultural Study of Store Choice Behavior. *J. Food Prod. Mark.* **2012**, *18*, 257–267. [\[CrossRef\]](#)

14. Haghiri, M. Consumer Choice between Food Safety and Food Quality: The Case of Farm-Raised Atlantic Salmon. *Foods* **2016**, *5*, 22. [\[CrossRef\]](#)
15. Britwum, K.; Bernard, J.C. Willingness to Pay for Farmed Salmon: Balancing Sustainability and Safety Perceptions. *Aquaculture* **2024**, *596*, 741787. [\[CrossRef\]](#)
16. Zhang, H.; Sun, C.; Wang, Z.; Che, B. Seafood Consumption Patterns and Affecting Factors in Urban China: A Field Survey from Six Cities. *Aquac. Rep.* **2021**, *19*, 100608. [\[CrossRef\]](#)
17. Li, B.; Liao, M.; Yuan, J.; Zhang, J. Green Consumption Behavior Prediction Based on Fan-Shaped Search Mechanism Fruit Fly Algorithm Optimized Neural Network. *J. Retail. Consum. Serv.* **2023**, *75*, 103471. [\[CrossRef\]](#)
18. Liu, Y.; Li, B.; Yang, S.; Li, Z. Handling Missing Values and Imbalanced Classes in Machine Learning to Predict Consumer Preference: Demonstrations and Comparisons to Prominent Methods. *Expert Syst. Appl.* **2024**, *237*, 121694. [\[CrossRef\]](#)
19. Choudhury, N.; Mukherjee, R.; Yadav, R.; Liu, Y.; Wang, W. Can Machine Learning Approaches Predict Green Purchase Intention? A Study from Indian Consumer Perspective. *J. Clean. Prod.* **2024**, *456*, 142218. [\[CrossRef\]](#)
20. Shen, Y.; Hamm, J.A.; Gao, F.; Ryser, E.T.; Zhang, W. Assessing Consumer Buy and Pay Preferences for Labeled Food Products with Statistical and Machine Learning Methods. *J. Food Prot.* **2021**, *84*, 1560–1566. [\[CrossRef\]](#)
21. Carrillo, E.; González, M.; Parrilla, R.; Tarrega, A. Classification Trees as Machine Learning Tool to Explore Consumers' Purchasing Decision Pathway. A Case-Study on Parent's Perception of Baby Food Jars. *Food Qual. Prefer.* **2023**, *109*, 104916. [\[CrossRef\]](#)
22. Kumar, A.; Kabra, G.; Mussada, E.K.; Dash, M.K.; Rana, P.S. Combined Artificial Bee Colony Algorithm and Machine Learning Techniques for Prediction of Online Consumer Repurchase Intention. *Neural Comput. Appl.* **2019**, *31*, 877–890. [\[CrossRef\]](#)
23. Wen, Z.; Lin, W.; Liu, H. Consumer Segmentation and Decision: Explainable Machine Learning Insights. *J. Comput. Inf. Syst.* **2024**, 1–17. [\[CrossRef\]](#)
24. Trivedi, S.K.; Roy, A.D.; Kumar, P.; Jena, D.; Sinha, A. Prediction of Consumers Refill Frequency of LPG: A Study Using Explainable Machine Learning. *Heliyon* **2024**, *10*, e23466. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Andoh-Odoom, A.H.; Bugyei, K.A.; Atter, A.; Kyei-Baffour, V.O.; Kunadu, A.P.-H.; Saalia, F.K.; Chama, M.A.; Lee, Y.; Koivula, H.M.; Amoa-Awua, W.K. Tilapia Consumption Patterns and Consumer Preferences: Predictors and Perspectives of Consumers in Ghana. *Heliyon* **2024**, *10*, e30247. [\[CrossRef\]](#)
26. Hu, W.; Zhang, S.; Fu, Y.; Jia, G.; Yang, R.; Shen, S.; Li, Y.; Li, G. Objective Diagnosis of Machine Learning Method Applicability to Land Comprehensive Carrying Capacity Evaluation: A Case Study Based on Integrated RF and DPSIR Models. *Ecol. Indic.* **2023**, *151*, 110338. [\[CrossRef\]](#)
27. Qiu, Y.; Zhou, J.; Khandelwal, M.; Yang, H.; Yang, P.; Li, C. Performance Evaluation of Hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost Models to Predict Blast-Induced Ground Vibration. *Eng. Comput.* **2022**, *38*, 4145–4162. [\[CrossRef\]](#)
28. Lei, Y.; Shen, Z.; Tian, F.; Yang, X.; Wang, F.; Pan, R.; Wang, H.; Jiao, S.; Kou, W. Fire Risk Level Prediction of Timber Heritage Buildings Based on Entropy and XGBoost. *J. Cult. Herit.* **2023**, *63*, 11–22. [\[CrossRef\]](#)
29. Wu, Z. Evaluation of Provincial Economic Resilience in China Based on the TOPSIS-XGBoost-SHAP Model. *J. Math.* **2023**, *2023*, 1–12. [\[CrossRef\]](#)
30. Ma, X.; Hao, Y.; Li, X.; Liu, J.; Qi, J. Evaluating Global Intelligence Innovation: An Index Based on Machine Learning Methods. *Technol. Forecast. Soc. Change* **2023**, *194*, 122736. [\[CrossRef\]](#)
31. Mahboubi, N.; Xie, J.; Huang, B. Point-by-point transfer learning for Bayesian optimization: An accelerated search strategy. *Comput. Chem. Eng.* **2025**, *194*, 108952. [\[CrossRef\]](#)
32. Ma, Y.; Yu, Z.; Yu, B. CAD tool design space exploration via Bayesian optimization. In Proceedings of the 2019 ACM/IEEE 1st Workshop on Machine Learning for CAD (MLCAD), Canmore, AB, Canada, 3–4 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6. [\[CrossRef\]](#)
33. Moosavi, S.M.; Novotny, B.Á.; Ongari, D.; Elias, M.; Mehrdad, A.; Özge, K.; Charithea, C.; Ortega-Guerrero, A.; Amir, H.F.; Lev, S.; et al. A data-science approach to predict the heat capacity of nanoporous materials. *Nat. Mater.* **2022**, *21*, 1419–1425. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Marchant, R.; Ramos, F. Bayesian optimisation for intelligent environmental monitoring. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 2242–2249. [\[CrossRef\]](#)
35. Pelikan, M.; Sastry, K.; Goldberg, D.E. Scalability of the Bayesian optimization algorithm. *Int. J. Approx. Reason.* **2002**, *31*, 221–258. [\[CrossRef\]](#)
36. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
37. Chen, M.; Xin, J.; Tang, Q.; Hu, T.; Zhou, Y.; Zhou, J. Explainable Machine Learning Model for Load-Deformation Correlation in Long-Span Suspension Bridges Using XGBoost-SHAP. *Dev. Built Environ.* **2024**, *20*, 100569. [\[CrossRef\]](#)

38. Guolin, K.; Qi, M.; Thomas, F.; Taifeng, W.; Wei, C.; Weidong, M.; Qiwei, Y.; Tie-Yan, L. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
39. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6639–6649. [[CrossRef](#)]
40. Lundberg, S. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874.
41. Li, Q.; Ji, Y.; Zhu, M.; Zhu, X.; Sun, L. Unsupervised Feature Selection Using Chronological Fitting with Shapley Additive Explanation (SHAP) for Industrial Time-Series Anomaly Detection. *Appl. Soft Comput.* **2024**, *155*, 111426. [[CrossRef](#)]
42. Wang, Y.; Wang, P.; Tansey, K.; Liu, J.; Delaney, B.; Quan, W. An interpretable approach combining Shapley additive explanations and LightGBM based on data augmentation for improving wheat yield estimates. *Comput. Electron. Agric.* **2025**, *229*, 109758. [[CrossRef](#)]
43. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [[CrossRef](#)]
44. Qi, X.; Wang, X. Factors Influencing the Citizenization of Rural-Urban Migrants: A Study Based on Machine Learn in. *Chin. Rural. Econ.* **2024**, *5*, 128–150.
45. Aminizadeh, M.; Mohammadi, H.; Karbasi, A.; Rafiee, H. Predicting Consumers' Intention towards Seafood Products: An Extended Theory of Planned Behavior. *Food Qual. Prefer.* **2024**, *113*, 105061. [[CrossRef](#)]
46. Hoque, M.Z.; Akhter, N.; Chowdhury, M.S.R. Consumers' Preferences for the Traceability Information of Seafood Safety. *Foods* **2022**, *11*, 1675. [[CrossRef](#)] [[PubMed](#)]
47. Petrontino, A.; Madau, F.; Frem, M.; Fucilli, V.; Bianchi, R.; Campobasso, A.A.; Pulina, P.; Bozzo, F. Seafood Choice and Consumption Behavior: Assessing the Willingness to Pay for an Edible Sea Urchin. *Foods* **2023**, *12*, 418. [[CrossRef](#)]
48. Alwosheel, A.; van Cranenburgh, S.; Chorus, C.G. Is Your Dataset Big Enough? Sample Size Requirements When Using Artificial Neural Networks for Discrete Choice Analysis. *J. Choice Model.* **2018**, *28*, 167–182. [[CrossRef](#)]
49. Taylor, K.E. Summarizing Multiple Aspects of Model Performance in a Single Diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [[CrossRef](#)]
50. Lai, D.; Demartino, C.; Xiao, Y. Probabilistic Machine Learning Models for Predicting the Maximum Displacements of Concrete-Filled Steel Tubular Columns Subjected to Lateral Impact Loading. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108704. [[CrossRef](#)]
51. Martin, P.E.D.G.; Shigeyoshi, T. Hierarchical Bayesian Optimization Algorithm: Toward a New Generation of Evolutionary Algorithms. *SICE Annu. Conf. Program Abstr.* **2003**, *2003*, 124. [[CrossRef](#)]
52. Shi, R.; Xu, X.; Li, J.; Li, Y. Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Appl. Soft Comput.* **2021**, *109*, 107538. [[CrossRef](#)]
53. Wong, T.T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [[CrossRef](#)]
54. Adak, A.; Pradhan, B.; Shukla, N.; Alamri, A. Unboxing Deep Learning Model of Food Delivery Service Reviews Using Explainable Artificial Intelligence (XAI) Technique. *Foods* **2022**, *11*, 2019. [[CrossRef](#)] [[PubMed](#)]
55. Baptista, M.L.; Goebel, K.; Henriques, E.M. Relation between Prognostics Predictor Evaluation Metrics and Local Interpretability SHAP Values. *Artif. Intell.* **2022**, *306*, 103667. [[CrossRef](#)]
56. Arenas, M.; Barceló, P.; Bertossi, L.; Monet, M. On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results. *J. Mach. Learn. Res.* **2023**, *24*, 1–58.

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.