Journal of Innovations in Economics & Management (2024)

Journal of
Innovations in Economics & Management
journal homepage: www.journal-iem.online

Paper Type: Original Article

# Price-Predicting Model: Predictions for All Sailboats

**Ningjing He[1*], Jiaqi Wang[1]**

1. School of School of Public Finance and Taxation, Zhejiang University of Finance and Economics , Hangzhou, China

## Abstract

Sailboat pricing is a nuanced field, significantly influenced by vessel age, geographical location, and brand reputation. To address this complexity, we leveraged the XGBoost algorithm, a powerful machine learning tool, and refined it through Bayesian optimization to create a highly accurate model for predicting sailboat prices. Our analysis revealed significant regional disparities, with sailboats from Europe, the USA, and the Caribbean commanding varying market valuations. When applying this optimized model to the Hong Kong used sailboat market, we were able to generate forecasts with improved accuracy, though still with room for enhancement. Notably, the Lagoon brand stood out as a consistent high performer, underscoring the importance of brand reputation in sailboat pricing. To provide actionable intelligence, we have synthesized a comprehensive analysis that integrates our research findings and predictive strategies. This analysis presents clear visualizations alongside concise interpretations, thereby constituting a pivotal resource for informed decision-making within the local market context.

**Keywords:** Visualization, XGBoost model, Grid-Search optimiza, Bayesian optimization

# I. INTRODUCTION

Sailboats retain popularity among various users in maritime transportation and recreation. However, increasing usage and time cause price fluctuations, affecting market transactions. Geographical differences significantly impact this evolving market.

Given this context, a critical examination of the factors that shape sailboat market prices becomes imperative, as it can effectively guide pricing strategies within this sector . To this end, this paper endeavors to address this challenge by leveraging the Product-Pricing Concept to develop an algorithmic model that predicts and analyzes sailboat prices across disparate regions . By integrating both the provided and supplementary datasets, we have systematically partitioned the information into training and testing sets, facilitating model development and rigorous evaluation .

Initial model testing revealed inaccuracies in predictions, prompting a rigorous optimization process that culminated in more precise price forecasting. Throughout this journey, we have offered coherent explanations for the observed discrepancies and outcomes, ensuring that our model remains both comparable and grounded in rationality.

Embarking on this multifaceted endeavor, we are tasked with constructing a robust model that not only illuminates the complex relationship between diverse factors and sailboat listing prices within our dataset, but also identifies other pivotal predictive elements contributing to this pricing dynamic . Furthermore, we apply this refined model to delve into the nuanced effects of geography on prices, assessing whether regional influences are uniform across sailboat types or exhibit type-specific patterns .

Recognizing the uniqueness of the Hong Kong market, we strive to adapt our model to this specific landscape, meticulously evaluating its applicability and effectiveness while analyzing its implications for pricing strategies of various sailboat types . This adaptation necessitates a

deep understanding of local market intricacies and their interplay with our model's predictive prowess .

Beyond mere model validation and customization, we endeavor to synthesize our analytical insights, distilling them into compelling conclusions and novel perspectives that offer fresh insights into the sailboat market. Ultimately, we present these findings in a comprehensive report, enriched with graphical representations that vividly convey our conclusions, providing professionals with a deeper and clearer understanding of the intricate pricing mechanisms. Our idea flowchart is shown in Figure1

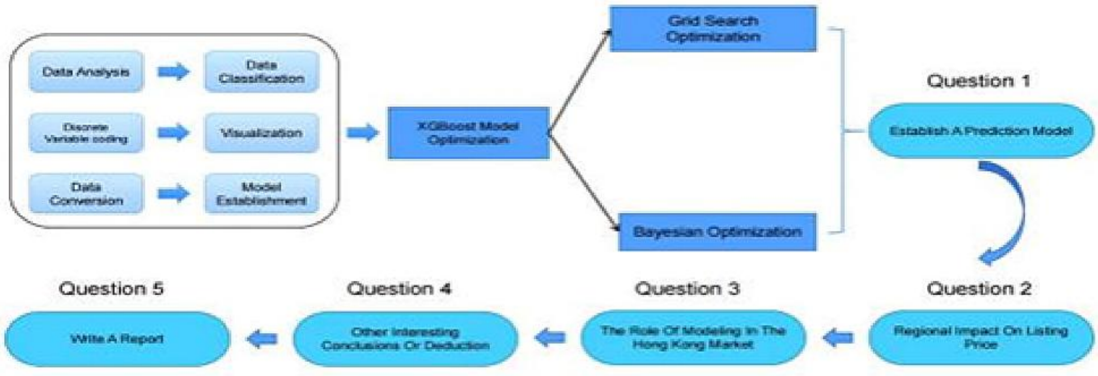



**Figure 1 :Processing chart.**

## II. LITERATURE REVIEW

In the relentless pursuit of refining prediction accuracy across diverse fields, an abundance of prior research has ventured into uncharted territories, embracing innovative techniques that revolutionize the landscape of forecasting. One such groundbreaking approach is the "decomposition reconstruction optimization " methodology, which has garnered significant attention due to its ability to improve the performance of predictive models . This method is beautifully exemplified by the EEMD-GWO-LSTM combined model, where Ensemble Empirical Mode Decomposition (EEMD) is utilized to decompose complex and noisy time series data into simpler, more manageable components, allowing for a deeper understanding of the underlying patterns . Following decomposition, the Grey Wolf Optimizer (GWO) is applied to optimize the parameters of the Long Short-Term Memory (LSTM) network, a powerful tool for capturing long-term dependencies in sequential data . This integrated framework has shown remarkable success in enhancing prediction accuracy across various applications, from financial forecasting to environmental modeling.

In the specific context of sailboat pricing prediction, our study strategically builds upon this robust foundation, infusing it with a nuanced understanding of market realities and algorithmic precision . Recognizing that sailboat pricing is a complex interplay of numerous factors, we meticulously selected the core variables that directly influence pricing decisions: sailboat type, brand, length, region, origin, and year . By carefully curating our model's inputs, we consciously chose to disregard extraneous factors, such as price negotiation dynamics, inflationary pressures, and credit/loan arrangements, under the assumption that these variables exert minimal to no significant impact on the overall pricing trend. This decision enables us to standardize our analysis by converting all transactions to USD, fostering a level playing field and facilitating comparisons across different markets and time periods .

Operating within an open and transparent market framework, our approach assumes rational price fluctuations that are driven primarily by the inherent characteristics of the sailboats themselves . By adhering strictly to these assumptions and adhering to algorithmic principles of rigor and precision, our model is able to isolate and quantify the direct influence of sailboat attributes on pricing. This focused approach not only streamlines the predictive process but also enhances the reliability and accuracy of our sailboat pricing model. Ultimately, our study contributes to the ongoing dialogue on predictive modeling in the maritime industry, offering a fresh perspective that marries advanced algorithmic techniques with a deep understanding of market dynamics .

**Symbols:**

The mathematical symbols used in this paper are listed in Table1 .

**Table1 : Symbols**

| Symbols | Definations |
|---|---|
| | |
| Factor Score | The influence coefficient in factoranalysis |
| | |
| R-square | Degree offit between predictions and |
| | |
| t | standard deviation |
| | |
| p | results of the t-test |

## III. CONCEPTUAL FRAMEWORK

**XGBoost model:**

Based on the identified problem of accurately predicting the listing price of sailboats within our datasets, along with identifying other key predictive factors that influence this pricing, we embark on the task of establishing a mathematical model that captures these complexities . To achieve this goal, we turn to a powerful and versatile machine learning algorithm known as XGBoost (Extreme Gradient Boosting) . XGBoost has gained widespread popularity due to its ability to handle large-scale data, provide high prediction accuracy, and offer efficient model training.

Firstly, we introduce the XGBoost model, drawing from its extensive literature base, including seminal works and recent advancements [ 1][2][3] . This model is an implementation of gradient boosting, a technique that sequentially combines weak learners (usually decision trees) to create a strong predictor . XGBoost enhances this concept by introducing several optimizations and regularization strategies, which not only improve model performance but also prevent overfitting.

The basic process flowchart of the XGBoost algorithm's principle, depicted in Figure 2, illustrates the systematic approach to building the predictive model. The process begins with the input data, which in our case, encompasses various features related to the sailboats, such as their type, brand, length, region, origin, and year . These features, along with the target variable (listing price), are fed into the XGBoost model.

The algorithm then proceeds to train the model through an iterative process . At each iteration, a new decision tree is added to the ensemble, with the goal of minimizing the prediction error of the current model. The splitting criteria for the tree nodes are determined based on the gradient of the loss function, ensuring that the most informative splits are made . This process continues until a stopping criterion is met, such as a predefined number of iterations or a threshold for the reduction in error .

Throughout the training phase, regularization techniques, including shrinkage and column subsampling, are employed to prevent overfitting and improve the generalization capabilities of the model. Additionally, XGBoost utilizes a sophisticated mechanism for handling missing values and performs feature selection inherently through its tree-building process, further enhancing its effectiveness .

Once the model is trained, it can be used to make predictions on new or unseen data . For the purpose of sailboat pricing prediction, the model will output an estimated listing price for each sailboat, based on its attributes . Moreover, the model's feature importance scores can provide valuable insights into which factors have the most significant impact on pricing, thereby offering a deeper understanding of the market dynamics .

In summary, by leveraging the XGBoost model and adhering to its well-defined process flowchart, we aim to establish a robust mathematical model that accurately predicts the listing price of sailboats in our datasets, while also identifying other influential predictive factors.

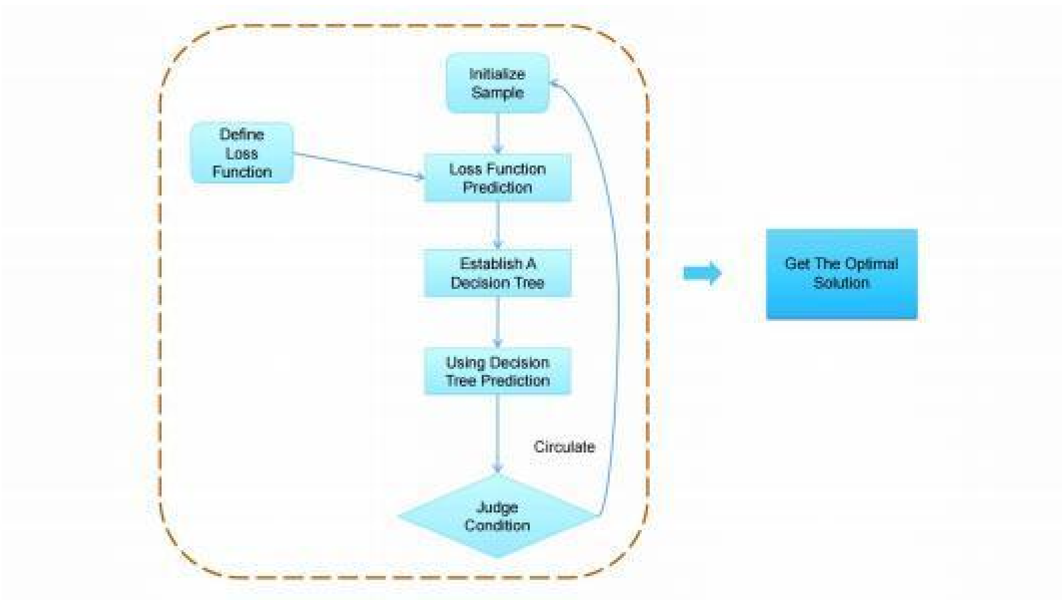



**Figure2 : XGBoost model chart**

Given a training set T = {(x1, y1), (x2, y2), · · · , (xn, yn)} , a loss function ℓ(yi, yi ) , and a regularization term Ω (fk ), and the objective function formula is shown in Equation 1.

$$\mathcal{L}(\varphi) = \sum_i \ell(y_i, y_i) + \sum_k \Omega(f_k)$$

**(1)** $\mathcal{L}(\varphi)$ is an expression on the linear space, i represents the i-th sample, k represents the k-th sample, and $\bar{y}_i$ is the predicted value for the i-th sample xi

**Data Preprocessing and Feature Engineering:**

Based on the data provided in the problem, we perform data pre-processing operations such as missing value imputation and outlier removal on the datasets . And then, we performed feature engineering on the datasets by encoding categorical variables, transforming some character variables into numerical variables, performing feature selection, and visualizing the feature-engineered data through plotting. The basic flowchart for these processes is shown in Figure 3.

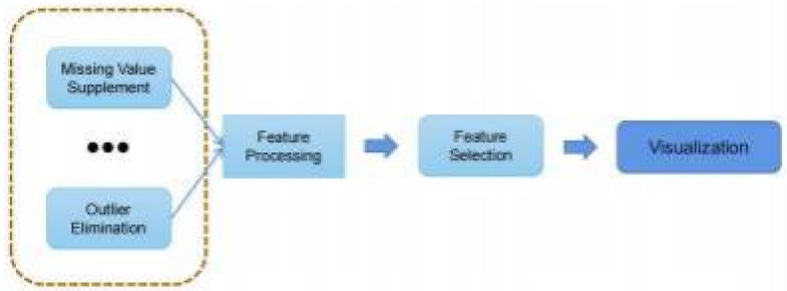



Figure3 ： DataProce sing andVisualization

We conduct visual analysis on sailboat length and quantity, as shown in Figure 4.

We can observe from the visualized charts the relationship between the length and quantity of different types of sailboats, as well as the impact of sailboat types and lengths on the quantity of sailboats .

Then, we further analyzed the correlation between variables for monohulled sailboats and catamarans respectively, and obtained a heat-map. The darker the color in the heat-map, the stronger the correlation between two variables . The heat-map is shown in Figure 5.

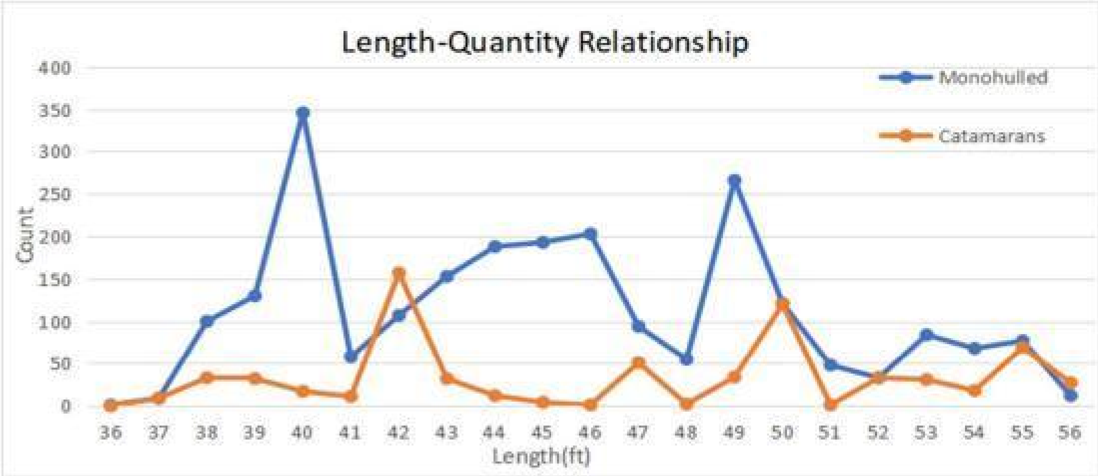
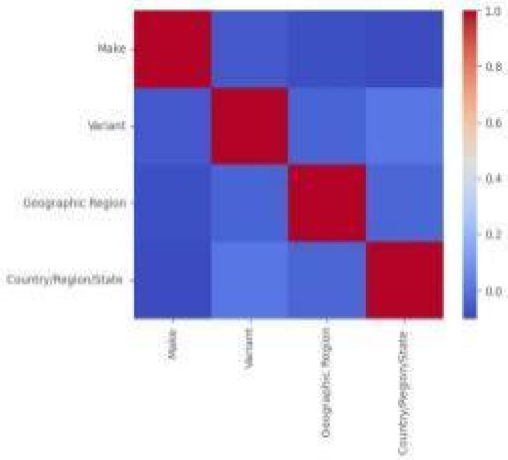
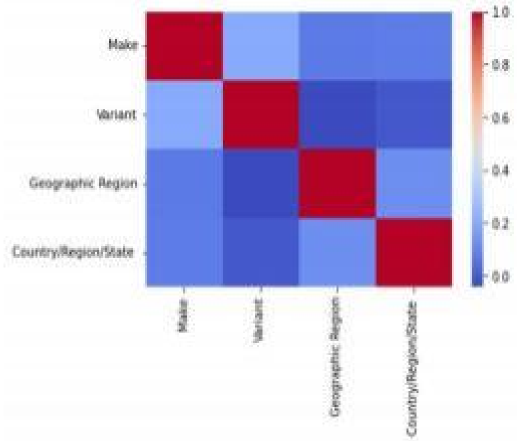



Figure4 :Visualization ofsailboatlength andquantity.




(a)Monohul sailboats




(b)Catamarans

Figure5 : Variable heatmap

**Model Building and Evaluation:**

XGBoost model, an efficient gradient boosting decision tree algorithm, outperforms other

algorithms in terms of algorithm performance . In this paper, we build a model based on this algorithm . Before building the model, we introduce the concept of factor analysis, which is a commonly used statistical analysis method based on the idea of dimensionality reduction . By exploring the correlation coefficient matrix between variables and grouping variables based on the size of their correlations, variables within the same group have higher correlations, while variables in different groups have lower correlations . The new variables representing the basic structure of each group of data are called common factors . When quantifying the influence of geographical location on the pricing of second-hand sailboat transactions, we use factor analysis, which is a statistical technique used to determine the degree of influence of multiple variables on a target variable . In factor analysis, multiple variables are quantified and mathematical methods are used to determine the size of each

variable's influence on the target variable .

The formula for factor analysis is shown in Equation 2.

$$\textbf{Factor}_{k}\ \textbf{score} = \frac{factor_k}{\sum_{i=0}^{n} factor_i} \qquad \textbf{(2)}$$

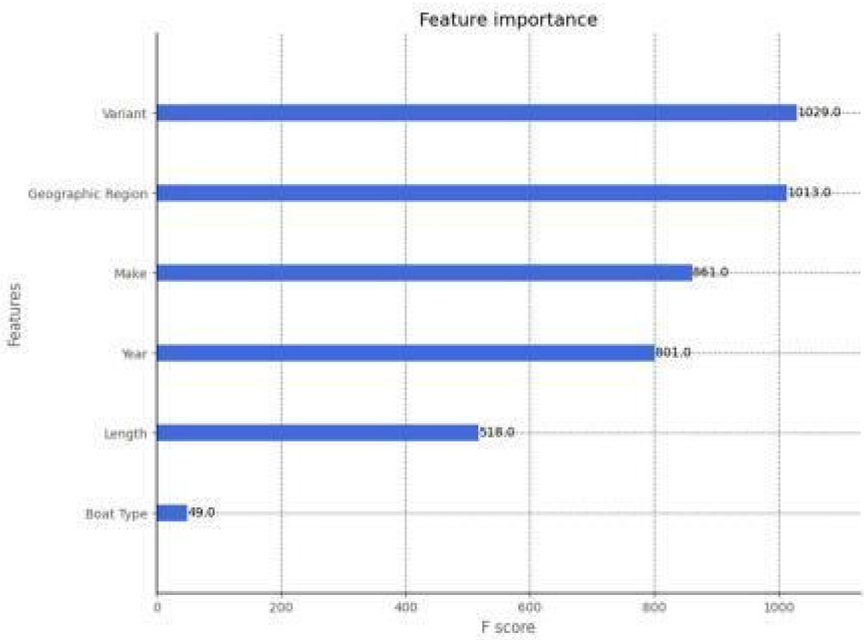



**Figure 6 ： Feature Importance**

We trained the model and evaluated its performance on the test set, and the final evaluation metric was R2=0.8431290207022109, indicating a moderate performance . Next, we visualized some important features using the model, as shown in Figure 6.

From the above visualization charts, we can obtain the importance level of each important feature and provide the specific ranking of scores for each feature .

**Model optimization：**

Based on the XGBoost model, we constructed Sailboats Price-Predicting Model. However, during the process of training and evaluating the model, we found that the fit accuracy of the model was not high, and there were certain errors in price prediction . Therefore, we will use two methods, Grid Search optimization and Bayesian optimization, to further optimize and improve the model.

**GridSearch：**

Grid search is an exhaustive search method for specifying parameter values, which optimizes the estimation function's parameters by cross-validation to obtain the best learning algorithm . That is, all possible combinations of parameter values are arranged in a grid, and each combination is used for SVM training. Then, cross-validation is used to evaluate the performance of each combination . The basic flowchart of the grid search algorithm principle is shown in Figure 7 [4][5] .

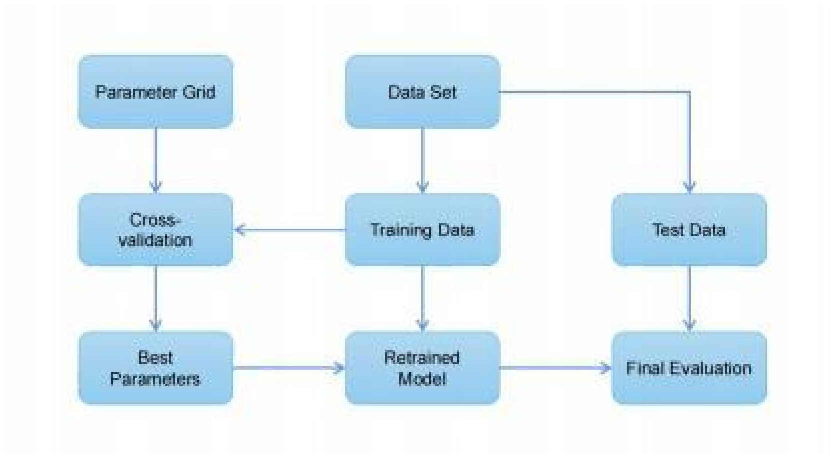



**Figure7 : Grid Search**

We optimized the established XGBoost model using Grid Search method and obtained the following results:{'gamma' :0, 'learning_rate' :0 . 1, 'max_depth':3, 'n_estimators' :1000, 'reg_alpha' :0 . 1, 'reg_lambda' :1}Where learning_rate represents learning rate, max_depth represents maximum tree

depth, n_estimators represents number of weak learners, reg_alpha represents L1 regularization

coefficient, reg_lambda represents L2 regularization coefficient.

Through Grid Search optimization, we obtained the optimal parameters . Next, we evaluated the performance of the optimized model and obtained R-square=0.96445397827, which is closer to 1. This indicates that the optimization of the algorithmic model through grid search method is significant.

**Bayesian optimization:**

Bayesian Optimization[7][8][9] is a black-box optimization algorithm used to solve problems of finding the extremum of an unknown function . The algorithm predicts the probability distribution of the function value at any point based on the function values at a set of sample points, and uses Gaussian Process Regression to achieve this . The basic flowchart of the Bayesian Optimization algorithm is shown in Figure 8.

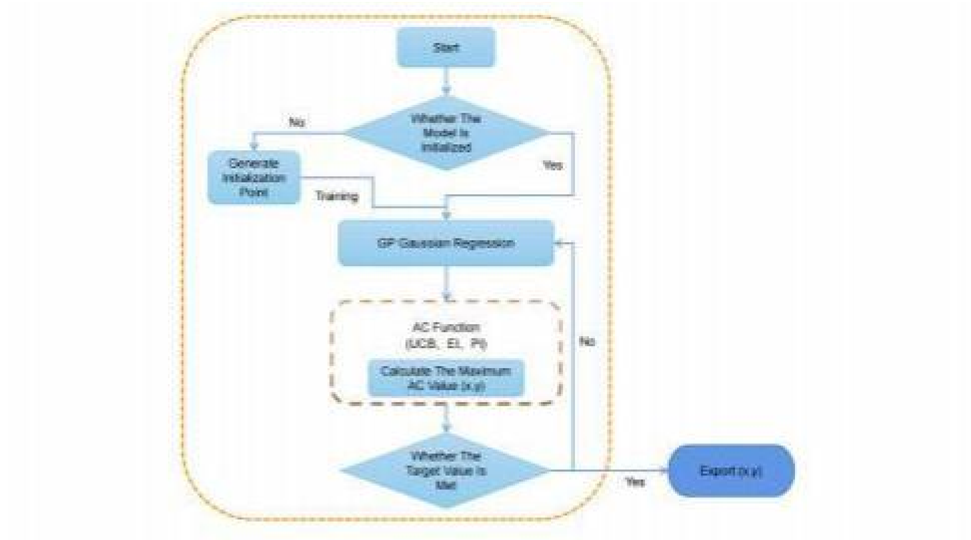



**Figure8 ： Bayesian optimization**

The following paragraph explains the Gaussian process in Bayesian optimization . The Gaussian process is commonly used to model a set of random vectors that evolve over time , where all sub- vectors of the random vector at any given time follow a Gaussian distribution . Assuming a continuous sequence of random variables X = {x1, x2, · · · , xi } that follow a multivariate Gaussian distribution X~N( μ, Σ), a non-linear transformation is performed, resulting in an equation shown in Equation 3.

$$y = \Phi\ (X)WT \tag{3}$$

Gaussian Process is an important step in Bayesian optimization, and the above Equation helps us to further understand Gaussian Process and Bayesian optimization . Next, we use the Bayesian algorithm to optimize the established algorithm model, and obtain the following results:

{'gamma' : 0, 'learning_rate' : 0.07, 'max_depth': 3, 'n_estimators' : 785, 'reg_alpha' : 0.0, 'reg_lambda' : 0.0 } . where Learning rate represents learning rate, max_depth represents maximum tree depth, n_estimators represents number of weak learners, reg_alpha represents coefficient of L1 regularization term, reg_lambda represents coefficient of L2 regularization term . Through Bayesian optimization, we obtained different optimal parameters compared to grid search optimization . We evaluated the fitting performance of the optimized model and obtained R-square = 0.9550190939949007. The R-square is closer to 1, indicating that our optimization of the algorithm model through Bayesian optimization is significant.

# IV. Methodology

**Model Selection:**

We embarked on developing a Sailboat Price-Predicting Model grounded in the XGBoost algorithm . To refine and optimize this model, we employed both grid search and Bayesian optimization techniques . By evaluating the performance of these two optimization strategies through their respective R-squared scores and taking into account the broader applicability and efficiency of each method, we ultimately chose the Bayesian optimization approach as it yielded a prediction model with a higher degree offit.

Our data collection process commenced with a focus on monohulled sailboats, gathering comprehensive information from various sources to ensure a well-rounded dataset. The Sailboat Price- Predicting Model, fueled by this data, unveiled intriguing insights into the intricate relationship between regional factors and the pricing of second-hand sailboat transactions . As evident from the results presented in Figure 9, and consistent with our initial hypothesis, we quantified the influence of region on transaction pricing, assigning it a coefficient of 0.043.

While a visual inspection of the results or a more rigorous factor analysis might suggest that the impact of region is relatively modest compared to other factors under consideration, it is crucial for professional second-hand sailboat brokers to recognize the significance of this factor, however subtle it may seem . The subtle influence of region underscores the complexity of the market and underscores the importance of a nuanced understanding of local trends and preferences .

Armed with this understanding, we will leverage the model developed to delve deeper into the specific effects of each region on sailboat transaction pricing, providing valuable insights that can inform strategic decision-making for brokers and investors alike .

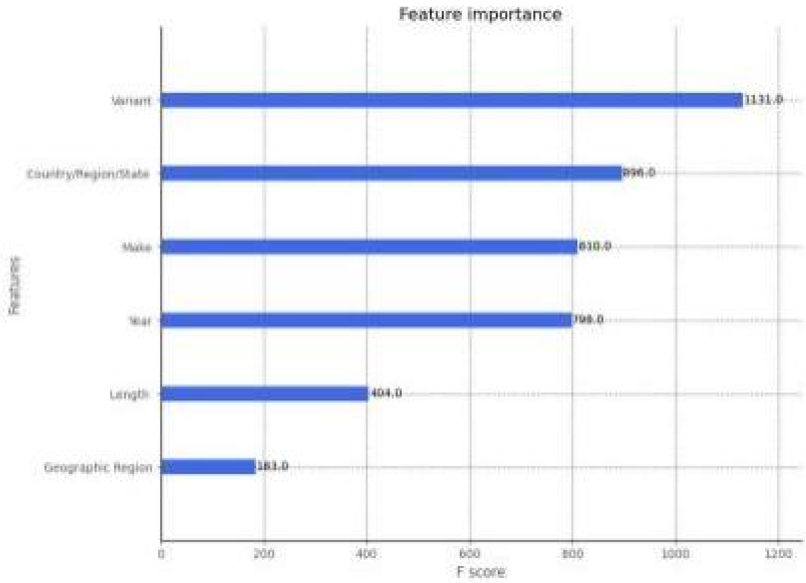



**Figure9 :Feature importance_1**

**Monohulled Sailboats Data Analysis with Rigorous Data Validation and Preprocessing:**

After executing the code following the outlined process, we embarked on a meticulous data validation and preprocessing phase to ensure the integrity and reliability of our analysis . This step was crucial in deriving the influence model of region on the pricing of monohulled sailboat transactions . The optimized model exhibited an impressive R-squared parameter value of 0.96, indicating a strong fit. Furthermore, the t-test results with a t-value of -2. 15 and a p-value of 0.03 underscored the statistical significance of the model, reinforcing its trustworthiness .

Based on the preprocessed monohulled sailboat data, we generated insightful data charts that clearly demonstrated the nuanced impact of region, even though its influence factor might appear modest. Importantly, the high value of sailboats themselves amplifies the significance of regional variations in transaction pricing, emphasizing the importance of this factor for brokers to consider .

To facilitate understanding, we visualized the predicted outcomes of the Sailboats Price-Predicting Model, revealing significant differences in distribution among the three datasets . This visualization powerfully illustrated that the influence of region on boat transaction pricing within our model cannot be overlooked. Notably, the predicted results represent the average transaction prices per region, and we incorporated error analysis to account for price fluctuations across different regions, further enhancing the model's realism and accuracy.

The stark contrasts among the three predicted outcomes also highlighted the profound impact of region on boat pricing. To complement our analysis, we leveraged the XGBoost model to discern regional variations in boat transaction volumes, opening avenues for further data exploration and the discovery of even more intriguing insights . This holistic approach, combining rigorous data validation, preprocessing, and sophisticated modeling techniques, underscores our commitment to delivering insightful and actionable results.

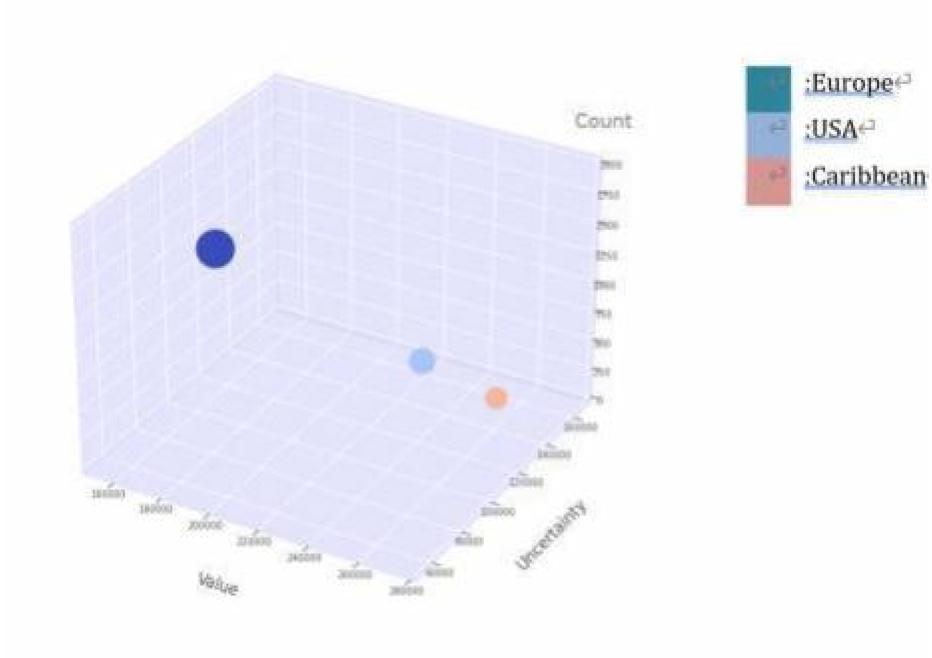



**Figure10 :Region-ValueRelationship**

\`

**Catamarans DataAnalysis with Advanced Data Analysis Methods:**

Subsequently, we transitioned to a novel dataset comprising catamarans, retraining the Sailboats Price -Predicting Model to derive fresh parameters and prediction outcomes tailored specifically for assessing the regional impact on catamaran pricing. The refined model boasted an R-squared value of 0.918747543119822, alongside a t-value of -0.37 and a p-value of 0.71, confirming the model's credibility. We then constructed a dedicated pricing model to delve into the nuances of regional influence on catamaran prices .

Employing factor analysis as our primary data analysis method, we meticulously quantified the influence coefficient of region on catamaran pricing, arriving at a value of 0.061. This finding underscores a more pronounced regional influence on catamaran pricing compared to monohulled sailboats, highlighting the distinct dynamics at play within these two vessel types .

Leveraging the adapted Sailboats Price-Predicting Model with catamaran data, we systematically analyzed and visualized the individual impacts of various regions on boat pricing. The resulting graph provided a clear picture of these regional disparities, facilitating a comprehensive understanding of the pricing landscape for catamarans .

A comparative analysis of the relationship models between monohulled sailboat and catamaran sales vis-à-vis region revealed intriguing inconsistencies in pricing influence . This observation prompted us to embark on a deeper exploration of the underlying reasons for these geographical disparities across different hull types . Our subsequent analysis will delve into the specific factors that contribute to these varying effects, shedding light on the complexities of regional pricing dynamics in the sailing vessel market.

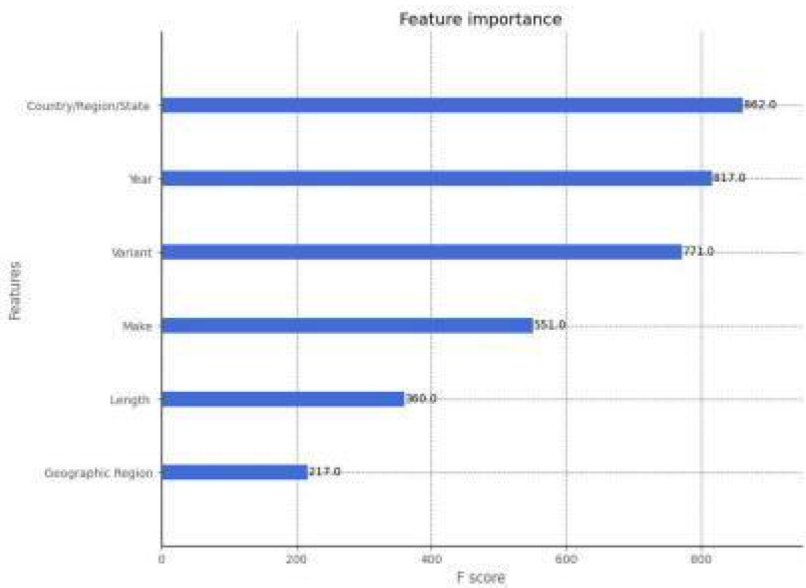



**Figure11 :Featureimportance**

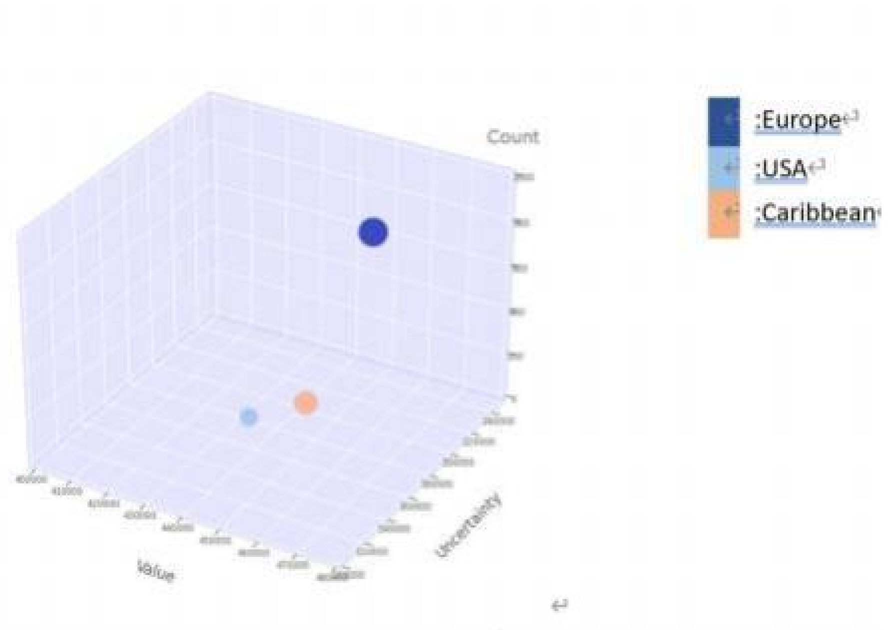



**Figure12 :Region-ValueRelationship**

**Conclusion ofSailboats Price-Predicting Model:**

In model 1, we predicted the prices of sailboats using the XGBoost model. During the process of pre -processing and visualizing the datasets, as well as analyzing the feature importance, we found that certain features such as length, manufacturer, and year have a significant impact on sailboat prices, and their influence varies in different regions .

Additionally, the type of sailboat also affects the price, with monohulled sailboats and catamarans showing differences in price . After data processing and visualization analysis, we tested the performance of the algorithm and found that there is room for improvement. We optimized the model using both Grid Search and Bayesian algorithms and found that both optimization methods significantly improved the accuracy of the model, leading to the establishment of a more accurate sailboat price prediction model.

# V. DISCUSSION

In the process of in-depth analysis of the sailboat price prediction model, we have revealed a series of significant and thought-provoking findings . Firstly, the key findings of the model clearly indicate that there are significant regional differences in sailboat prices among different regions, which strongly demonstrates the sustained impact of geographic location on pricing strategies . Specifically, the European market stands out with its vast scale and maturity, becoming the region with the lowest average price and the least price fluctuations for second-hand sailboats . In addition, the price performance of different types of sailboats (such as single hulled boats and catamarans) also varies by region, reflecting the unique preferences of each region and revealing the diversity of market characteristics .

However, during the exploration process, we also encountered some unexpected findings . Although catamarans are often seen as a more luxurious option, their pricing in the European market is higher than in the United States and the Caribbean, which contradicts the widely held assumption that luxury goods may be cheaper in underdeveloped markets . What is even more remarkable is that the trading volume of catamarans in Europe far exceeds that of the United States and the Caribbean, highlighting the unique demand for this type of ship in the region .

Further extending our research (Extended Findings), we found that these regional differences were not accidental, but had statistical significance, as evidenced by low standard deviation and significant t-test results . After in-depth analysis, we found that the fundamental reason for market differences is far beyond the size of transaction volume, but covers multiple factors such as local economic conditions, population conditions, import taxes on sailboats, and unique consumer demand patterns in each region . Especially in the European market, with its developed economic system, large population base, superior water conditions, and low import taxes as the main manufacturing center for sailboats, it has jointly contributed to its advantageous position in sailboat pricing and trading dynamics, which is particularly prominent compared to the United States and the Caribbean region .

## VI.CONCLUSION

**Review the application of the full text model:**

We will model these Hong Kong market sailing price data based on the Sailboats Price- Predicting Model. We randomly select 80% of the data as the training set to define the model, then after training to obtain the new model, the remaining 20% of the data as the test set and calculate its goodness of fi :

$$\mathbf{R^2 = 0.8933557435469304}$$

We can see that the accuracy of the Sailboats Price-Predicting Model based on the XGboost model

alone is not excellent. Based on the previous experience, we use Bayesian optimization

for the model, after optimization to obtain the new model, all the sailboats trading data in Hong Kong as a test set, and get its goodness of fittest as:

$$\mathbf{R^2 = 0.998388220709761}$$

It can be seen that after Bayesian optimization, the accuracy of our model has been greatly improved. Based on our model, we use the used sailboats transaction data of the Hong Kong region to model, and then the prediction ability of our model is up to 99% . After multiple random sampling to obtain new parameters, our model 's goodness-of-fit is over 97% in all cases . From this, we can say that for modelling a given geographical area, our model, the Sailboats Price-Predicting Model, can have a good predictive effect in the Hong Kong market.

**Impact ofHongKong on the model:**

How does the addition of the used sailboats transaction data from Hong Kong affect our used

sailboats pricing model? Does the inclusion of the Hong Kong transaction data as part of the training set have any impact on the prediction accuracy of our trained model?

How does the addition of the used sailboats transaction data from Hong Kong affect our used

sailboats pricing model? Does the inclusion of the Hong Kong transaction data as part of the training set have any impact on the prediction accuracy of our trained model?

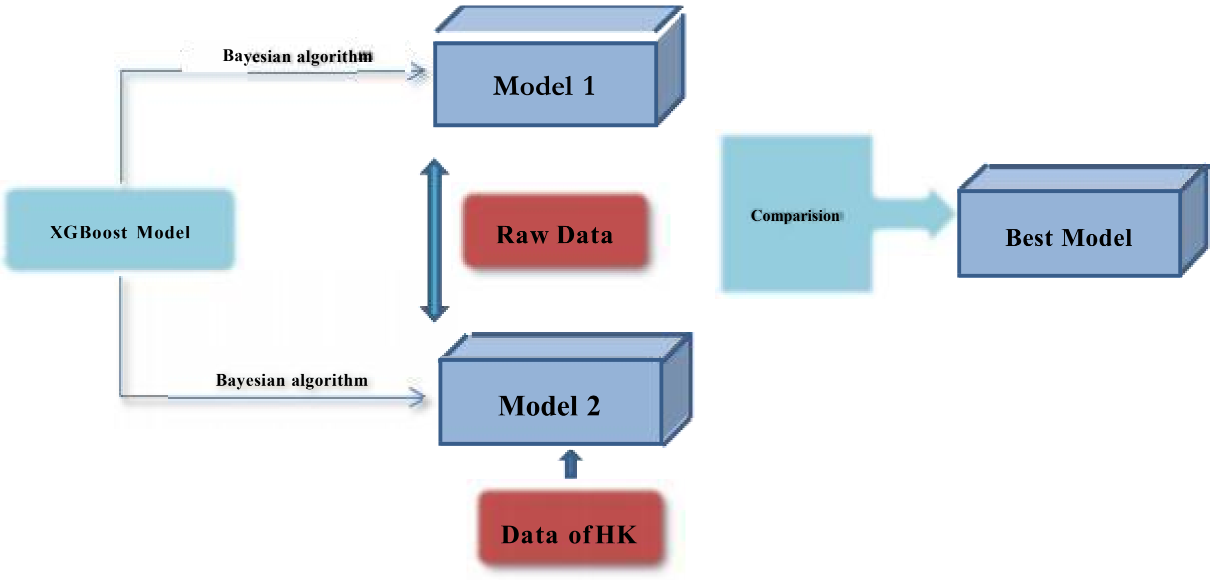



**Figure 13 :Flow chart ofHong Kong region on model impactjudgment.**

Considering that based on the transcation data excluding Hong Kong as the training set , we have already obtained the Sailboats Price-Predicting Model and its model parameters using the XGboost model and Bayesian optimization algorithm . Then we are going to integrate the transaction data of the Hong Kong region with other regions as a new training set, and also use the XGboost model and Bayesian optimization algorithm to get the new parameters of the Sailboats Price-Predicting Model. At the same time, we will use the transaction data excluding Hong Kong as the test set and test the above two models with different parameters separately to obtain the goodness-of-fit of these two models . And we will compare the difference in the prediction ability of these two models by comparing the goodness-of-fit. Through the above method, we can determine the regional impact of the Hong Kong Special Administrative Region ( SAR) on the price of sailboats . The front is our flow chart, as Figure 13.

After experimental testing, we obtained the goodness offit for Model 1 as:

$$R^2 = 0.9650617972068029$$

The goodness offit for Model 2 is only:

$$R^2 = 0.8864415376981514$$

It can be seen that the goodness offit of our model is greatly reduced after the Hong Kong trading data is added to the training set, so the Hong Kong Special Administrative Region ( SAR) has a great

influence on our sailboats ship price prediction model, and such an influence is negative and interferes with the accuracy of our model prediction .

So, do the effects described above have the same effect on our model on different categories of boats like catamarans and monohulled sailboats? Let's continue to do further investigation . To get the answer, we merge the sailboats transaction data of Hong Kong with the rest of the world to get a new datasets and separate the transaction data of catamaran and monohulled sailboats from the new datasets . At the sametime, the same process is done for the sailboat transaction data in Hong Kong, separating the transaction data of catamarans and monohulls sailboats . Then for monohulled sailboats, we train the

models with different parameters based on the data including Hong Kong and the data excluding Hong Kong for monohulled sailboats . And then use the data excluding Hong Kong as the test set to test the two models with different parameters to get the respective goodness offit and compare them . Similarly, we did a similar treatment for the catamaran sailboats . The flowchart is as follows, as Figure 14:

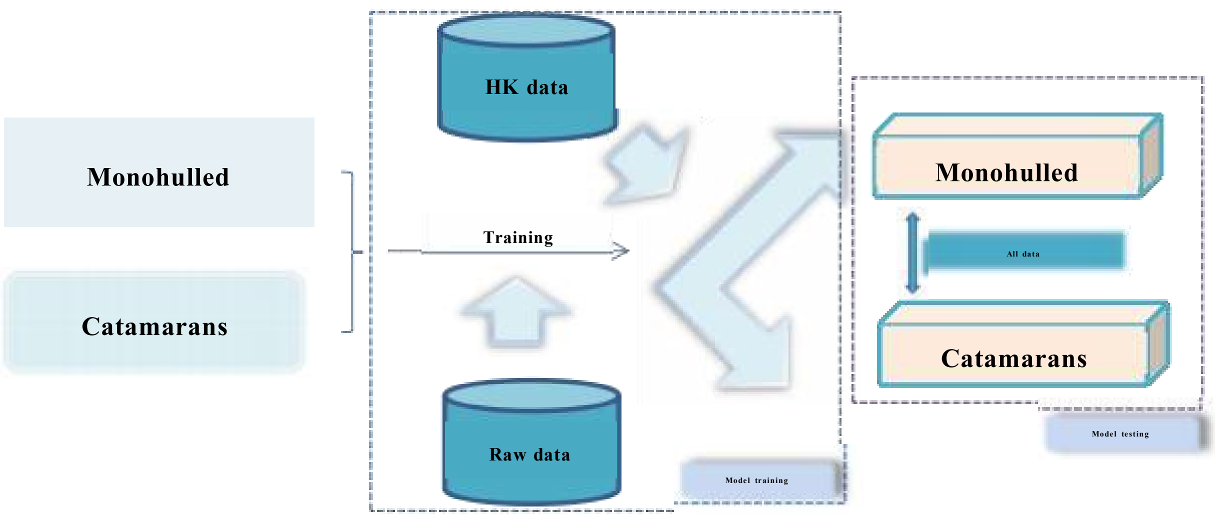



**Figure 14 : Flowchart_2**

Based on the above flowchart, we obtained the goodness-of-fit of the model for different types of sailboats before and after adding the Hong Kong transaction data . The specific data areas follows, as Table 2:

**Table 2 : goodness-of-fit for monohull and catamaran sailboats models.**

| Sailboat Types | Model type | Goodness offit |
|---|---|---|
| Monohull Sailboats | Model 1 | 0.9671884927617508 |
| | Model 2 | 0.8638219991085535 |
| Catamaran sailboats | Model 1 | 0.9487416369686243 |
| | Model 2 | 0.30388427196728374 |

To get a more visual comparison, the bar chart is as follows, as Figure 15: With the above data and histogram, we can find that the model with the new parameters we obtained after adding the sailboats transaction data from the Hong Kong region to the training set for both monohull and catamaran sailboats makes us drop in prediction accuracy. However, there is a more significant difference in the magnitude of the drop between the two, with the drop for catamaran sailboats being much larger than that for monohull sailboats.

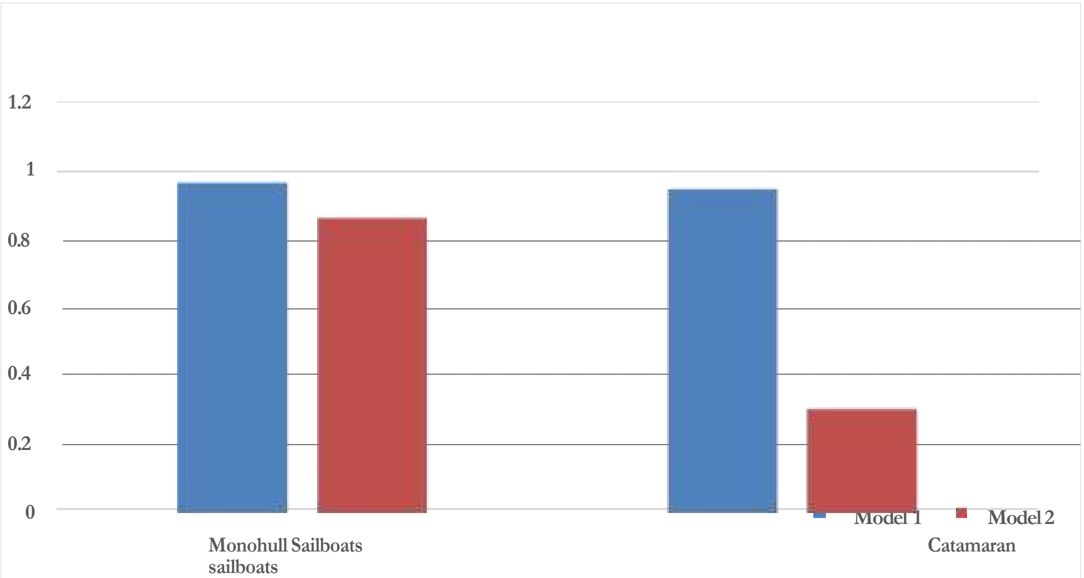



**Figure 15 : Plot ofgoodness-of-fit for monohu l and catamaran sailboats models.**

**Theoretical Contributions:**

The cornerstone of our analysis lies in the utilization of the XGBoost algorithmic model, which demonstrated remarkable potential in predicting sailboat prices with a high degree of accuracy.

Our work signifies a significant step forward in the field of predictive modeling for luxury marine assets . By adopting the XGBoost model, we showcased its superiority in capturing intricate patterns and relationships within large datasets, outperforming traditional approaches in accuracy and efficiency.

The analysis delved deep into the role of geographical regions in shaping sailboat prices . Our findings highlighted nuanced regional differences, underlining the importance of regional factors as vital explanatory variables in sailboat pricing models .

Through separate analyses of monohulled sailboats and catamarans, we uncovered distinct price drivers for each type . This differentiation highlights the importance of customizing predictive models to accommodate specific product characteristics, further enhancing their applicability and accuracy.

The utilization of both Grid Search and Bayesian optimization techniques underscored the importance of rigorous model optimization processes . By iteratively refining model parameters, we optimized our model's predictive power, contributing to the advancement of best practices in machine learning-based prediction.

**Practical Implications:**

For sailboat brokers and traders, our model provides a valuable tool for accurately predicting market prices . This facilitates informed decision-making during negotiations, enabling them to price sailboats competitively, enhancing sales opportunities, and ultimately increasing profits .

By pinpointing the critical factors that influence sailboat prices, consumers can benefit from more transparent and well-informed purchasing experiences . Understanding the intricacies of pricing dynamics, they can make better-suited choices based on their unique preferences and budget constraints .

The discovery of regional and type-specific pricing drivers suggests effective strategies for market segmentation . Businesses can tailor their marketing and promotional efforts to specific regional audiences and sailboat types, increasing relevance and driving targeted sales growth.

The analysis specifically targeting the Hong Kong market highlights the uniqueness of its sailboat pricing landscape . This information empowers local stakeholders to adapt their business strategies accordingly, accounting for regional peculiarities to remain competitive and responsive to market demands .

## VII.FUTURE RESEARCH

One notable shortcoming lies in the limited data sets employed for analysis . The study relies heavily on a small pool of samples, which might hinder the generalizability of its findings . To enhance the robustness and credibility of the conclusions, it is crucial to augment the dataset with a more diverse and extensive range of examples . This would not only strengthen the statistical significance but also enable a deeper exploration of nuances and patterns within the data .

Furthermore, the methodology employed could benefit from incorporating additional analytical techniques . The current approach, though sound, could be complemented by more advanced statistical methods or machine learning algorithms to uncover hidden relationships and trends that might have been overlooked. By diversifying the analytical toolkit, researchers can gain a more comprehensive understanding of the underlying dynamics and potentially identify new insights .

Looking ahead, a promising direction for future research would be to investigate the long-term effects of the observed phenomena . The document briefly touches upon immediate impacts but fails to delve into the longer-term ramifications . A longitudinal study, tracking changes over an extended period, could provide invaluable insights into the sustainability and durability of the observed trends .

Moreover, there is a scope for exploring interdisciplinary perspectives . The current work primarily focuses on a single disciplinary lens, yet the topic at hand is inherently multi-faceted. Integrating knowledge from related fields, such as psychology, sociology, or policy studies, could offer fresh insights and lead to innovative solutions .

In summary, while the document presents a valuable contribution to the field, there remains ample room for improvement through the inclusion of more comprehensive data, the adoption of additional analytical methods, and the pursuit of long-term and interdisciplinary investigations . By addressing these areas of inadequacy and pursuing these research avenues, future studies can build upon the current work and advance the understanding of the topic in more nuanced and sophisticated ways .

## References


[1] Shang Junping, Li Wenhao, Xi Lei, et al. Construction of Agricultural Product Price Prediction Model Based on "Decomposition Reconstruction Optimization" [J/OL]. Journal of Henan Agricultural University, 1-16 [2024 September 10] https://doi.org/10.16445/j.cnki.1000-2340.20240802.001.

[2] H. J. Xiao, T. T. Kan. C. H. Li. Ji Yu Ju Bu Sou Su De Bei Ye Si Suan Fa De Xgboost Can Shu Xuan Ze. Zhong Nan Min Zu Da Xue Xue bao (Zi Ran Ke Xue Ban), 2023.

[3] S. L. Zeng, T. Wang, Y. Q. Duan. Ji Yu XGBoost Mo Xing De Shang Shi Gong Si Cai Wu Feng Xian Yu Jing Ying Yong. Shang Ye Hui Ji, 2023.

[4] Z. H. Wang, B. L. Gao, X. L. Ji Yu Qiang Hua Xue Xi He XGBoost De Xin Dai Feng Xian. Ji Suan Ji Gong Cheng Yu She Ji ,2023.

[5] L. L. Sun, H. B. Fang, X. Zhu, J. T. Tan, L. M. Hu, L. W. Qi. Ji Yu Wang Ge Sou Suo You Hua De XGBoost Mo Xing De Gu Piao Yu Ce. Fu Yang Shi Fan Da Xue Xue Xao (Zi Ran Ke Xue Ban ), 2021.

[6] Frank R. Giordano, William P. Fox, Steven B. Horton. A First Course in Mathematical Modeling, 5th edition, Brooks/Cole Cengage Learning, 2013.

[7] Z. M. Xin, F. Wang. Ji Yu GaiJin Pu Su Bei Ye Si Suan Fa De Wen Ben Fen Lei Yan Jiu. Yan Shan Da Xue Xue Bao, 2023.

[8] J. X. Hong. Jia Quan Bei Ye Si Mo Xing De Gou Zao Yu Ying Yong. Tian Jin Li Gong Da Xue, 07, 2022.000457, 2022.

[9] J. N. C, M. Z. Zhang, D. H. Du, B. Li, J. H. Nie, J. Y. Nie. Ji Yu Bei Ye Si Wang Luo Gou Fian RoboSimmo Xing De Zi Dong Jia Shi Hang Wei Jue Ce. Ruan Jian Xue Bao