| Problem Chosen | 2023<br>MCM<br>Summary Sheet | Team Control Number |
|---|---|---|
| Y | | 2332102 |

# Title

## Summary

Sailboats fulfill diverse roles, fueling a thriving secondary market. Brokers, facing numerous complex factors, struggle to determine reasonable pricing. A tool to assist in comprehensive evaluations and rational pricing for used boats is urgently needed.

This paper aims to construct a reliable model based on existing datasets, which can provide a reasonable explanation for the pricing of the second-hand sailboat market. It also analyzes the impact of different factors and indicators on prices. Finally, the model will be applied to the second-hand sailboat market in Hong Kong to provide a reasonable and accurate pricing rule.

For Data Exploration, …

For Problem(a), …

For Problem(b), …

For Problem(c), …

For Problem(d), …

For Problem(e), …

At the very last, we analyze the strengths and weaknesses of our model as well as its sensitivity, whose results show that our model has high robustness, precision and accuracy. After that, a report is attached.

**Keywords**: Linear interpolation, Adaptive Density-Based Clustering, Heuristic Hierarchical Multiple Regression, Deep Forest Model, Machine Learning, Analysis of Variance.

# Contents

# 1 Introduction

## 1.1 Background

In our daily lives, sailboats are not only a means of transportation, but also serve as leisure and entertainment, and even for competitive sports. As a result, the growing demand for sailboats has given rise to a thriving boat market, which has gradually developed into a secondary market. In the secondary market, buyers and sellers usually trade through brokers, who play a crucial role in the transaction process.

For brokers, it is essential to be familiar with the used sailboat market, comprehensively consider various factors, and make reasonable pricing for the used sailboats in order to facilitate a successful transaction. However, the factors affecting the price of used sailboats are numerous and complex, with different brands, variants of boats, years, depreciation rates, as well as local consumption levels and geographical environments having significant impacts. The intertwined influences of these complex factors make it difficult to determine the pricing in the used sailboat market, and it is challenging to come up with a reasonable price that takes all factors into account.

Therefore, brokers urgently need a tool to assist them in making more reasonable and comprehensive evaluations of used sailboats, and to make the pricing in the used sailboat market more rational.

## 1.2 Problem Restatement

**Problem (a) :**

- Develop a prediction model to explain the listing price of each of the sailboats in the provided spreadsheet.
- Discuss the precision of our estimate for each sailboat variant's price.

**Problem (b) :**

- Determine whether region has an impact on the price of second-hand boats and explain the effect.
- Discuss whether any regional effect is consistent across all sailboat variants.
- Address the practical and statistical significance of any regional effects noted.

**Problem (c) :**

- Based on the model, find out how it can be useful in the Hong Kong market.
- Choose one subset and model the regional effect of Hong Kong on each sailboat prices.
- Assess whether the effect is the same for both catamarans and monohull sailboats.

**Problem (d) :**

- Identify and discuss additional informative conclusions drawn from the data.

**Problem (e) :**

- Create a one-to two-page report with well-chosen graphics to assist the Hong Kong sailboat broker to understand your findings.

## 1.3 Our work & Model Overview

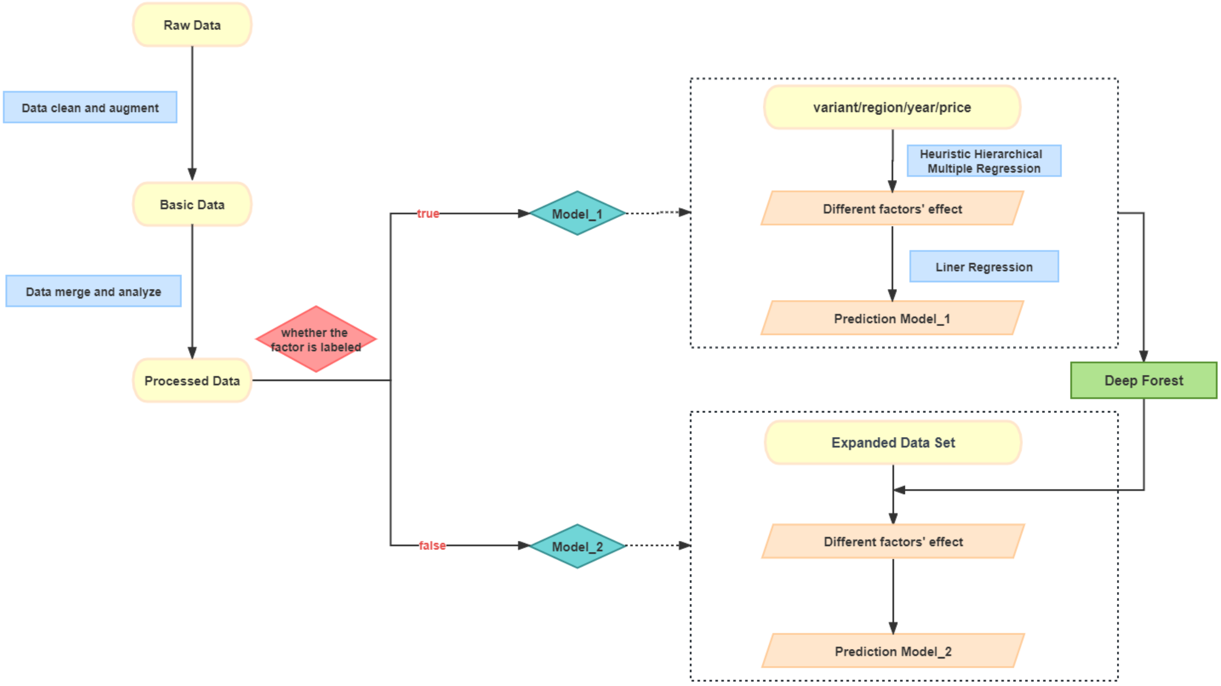


Figure 1: Model Framework

Our team began by meticulously cleaning, expanding, and pre-analyzing the dataset using **Adaptive Density-Based Clustering** and **Linear interpolation** techniques. These methods allowed us to identify patterns and trends in the data, preparing it for subsequent modeling.

Next, we trained a **Heuristic Hierarchical Multiple Regression Model** to establish the relationship between year, region, variant, and price, and to abstract the effect of regions and variants on sailboat prices. This approach helped us capture the interactions between these variables and their influence on the pricing.

Following this, we employed a **Deep Forest Model** to train the relationship between region and variant factors and their respective effects. This advanced ensemble learning technique allowed us to predict the effect of unknown regions and variants, which in turn enabled us to further predict their prices.

As a result, we successfully developed a comprehensive and robust model (Figure 1) that is well-suited for Hong Kong brokers when pricing used sailboats. By leveraging the strengths of both Heuristic Hierarchical Multiple Regression and Deep Forest Models, our approach offers a reliable and accurate way to estimate the value of sailboats based on factors such as year, region, and variant. This model serves as a valuable tool for brokers and other stakeholders in the sailboat market, facilitating informed decision-making and ensuring fair pricing practices.

# 2 Assumptions and Justifications

## 2.1 Assumptions

To simplify our problems, we make the following basic assumptions, each of which is adequately justified.

- The price of used sailboats is solely determined by the factors in the dataset.
- The factors in the dataset are independent and unrelated.
- The data in the dataset are all real, reasonable, and follow a certain pattern.
- For the used boats that need price prediction, we can obtain at least the same number of factors information as in the dataset.
- The pricing required by a broker should be reasonable and in accordance with market rules, rather than false pricing.

## 2.2 Notations

| Symbol | Definition |
|---|---|
| $l$ | Measurement data |
| $L$ | Feature vector |
| $F$ | The dimension of $L$ |
| $k$ | Search range |
| $\lambda$ | The feature values of natural neighbors |
| $nb_k(l)$ | The number of reverse neighbors of $l$ in the $k$-th iteration |
| $R$ | Correlation coefficient |
| $\hat{y}$ | The estimated value of $y$ |
| $\hat{\beta}$ | The estimated value of the regression coefficients |
| $\mathbf{X}, \mathbf{Y}$ | The variables for which the correlation is to be tested |
| $\mathbf{P}$ | Measured value |
| $\mathbf{Q}$ | Simulated value |

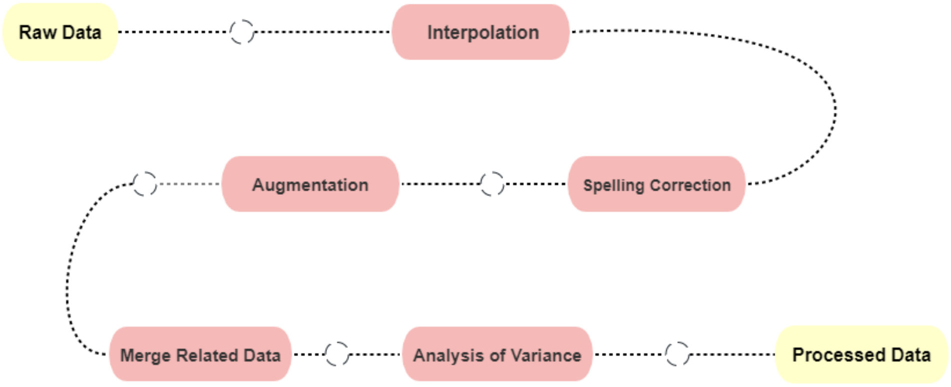


Figure 2: Data Exploration

# 3 Data Exploration

## 3.1 Data Cleaning

First of all, we turn the `xlsx` format data sheet into `csv` format. The conversion causes some minor errors like extra spaces and unexcepted characters which can be easily filtered out by text editor and python string operating functions like `strip()` or so.

Secondly, we use **Linear interpolation** to fill in the missing data and then apply **Adaptive Density-Based Clustering** for spell correction, resulting in a more complete and accurate dataset than the original.

Specificlly:

Based on the text-type variables, the feature matrix $L_i$ is used, and the Euclidean distance is used to calculate the text similarity:

$$dist(l_i, l_j) = \sqrt{\sum_{f=1}^{F}(L_i^f - L_j^f)^2}$$

Because ADC is a clustering iterative process, we first calculate the $k$-Nearest Neighbors:

$$NN_k(l_i) = \{l_j \in D | dist(l_i, l_j) \leq dist(l_i, o)\}$$

Then we define the Reverse Neighbors as:

$$RNN(l_i) = \{l_j \in D | l_i \in NN_k(l_j)\}$$

Applying $NN_k$ and $RNN(l)$ iteratively to obtain clustering centers, unclassified text samples are assigned to the nearest clustering center, and $\lambda$ is calculated as follows:

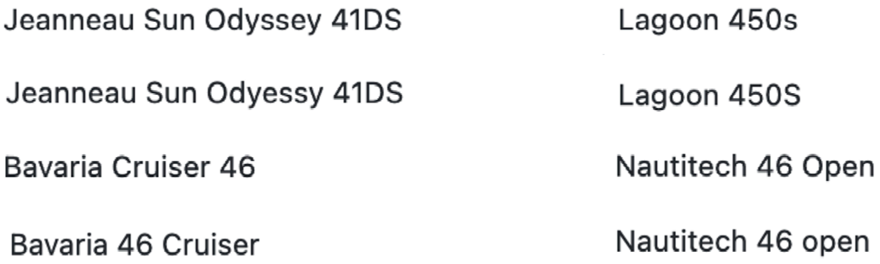
$$\lambda = min\{k| \sum_{i=1}^{n} f(nb_k(l_i)) = 0 \quad or \quad \sum_{i=0}^{n} f(nb_k(l_i)) = \sum_{i=0}^{n} f(nb_{k-1}(l_i))\}$$

Where $f(x)$ is defined as:

$$f(x) = \begin{cases} 0, & otherwise \\ 1, & if \quad x = 0 \end{cases}$$

Further more, we merge **Make** and **Variant** into a single feature, which we refer to as **Variant**.

Finally, we merge the corrected dataset with the accurate one after making necessary modifications.



Jeanneau Sun Odyssey 41DS

Jeanneau Sun Odyessy 41DS

Bavaria Cruiser 46

Bavaria 46 Cruiser

Lagoon 450s

Lagoon 450S

Nautitech 46 Open

Nautitech 46 open

Figure 3: Examples of error data

## 3.2 Data Augmentation

Our team diligently collected and organized additional ship attributes for the given dataset using Pandas' `read_html` web scraping technique, which allowed us to efficiently extract valuable information from various online sources. Simultaneously, we gathered the 2020 per capita GDP data of the relevant regions to provide important economic context that might influence sailboat prices.

Upon obtaining the data, we first employed Pandas' related functions to remove any duplicate entries, ensuring the dataset's integrity and uniqueness. Next, we addressed missing values by using linear interpolation, a technique that estimates missing data points by drawing upon the values of neighboring data points. This approach helped us create a more complete and reliable dataset for further analysis.

These data collection, organization, and preprocessing efforts significantly enriched the diversity of the data and expanded the dataset size. By incorporating a wider range of sailboat features and relevant economic data, we were able to develop a more comprehensive understanding of the factors influencing sailboat prices. This extensive dataset laid a solid foundation for our subsequent modeling efforts, enabling us to create more accurate and reliable predictive models for sailboat pricing.

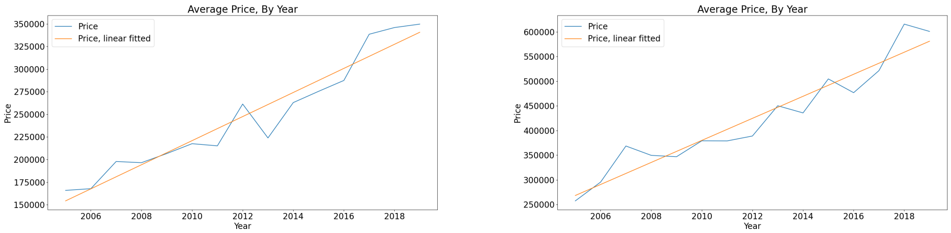
In conclusion, the meticulous data collection, organization, and preprocessing steps we undertook played a crucial role in the success of our modeling efforts. By addressing issues such as duplicates and missing values, and incorporating a diverse range of ship attributes and economic data, we ensured the robustness and reliability of our dataset, paving the way for the development of highly effective predictive models.

## 3.3 Data Analysis

There is an abundance of labels related to sailboat specifications, which makes it challenging to discuss their individual influence on listing prices. Attempting to directly perform regression on these labels using machine learning models can lead to **overfitting issues**, where the model becomes too specialized in fitting the training data and loses its ability to generalize well to new, unseen data.

Despite these challenges, we can still extract some relatively straightforward insights from the data, which can provide valuable theoretical support for our subsequent modeling efforts:

- Time: Time is a quasi-continuous and ordered variable, exhibiting a clear linear relationship with the price of sailboats (as demonstrated in Figure 4 and Table 1). As sailboats age, their value typically decreases, reflecting factors such as depreciation, wear and tear, and technological advancements in newer models.
- Variant and Region: These factors are undeniably related to sailboat prices; however, their unordered and non-continuous nature necessitates further analysis. Sailboat variants can encompass different designs, sizes, and features, which can significantly impact their value. Similarly, regional factors such as local demand, availability, economic conditions, and even cultural preferences can influence sailboat prices.



(a) Monohulled Sailboats

(b) Catamaran

Figure 4: The relationship between year and price

| | Pearson | Spearman | Kendall |
|---|---|---|---|
| Monohulled Sailboats | 0.964 | 0.989 | 0.943 |
| Catamaran | 0.960 | 0.971 | 0.867 |

Table 1: The correlation coefficient of year and price

# 4 The Prediction Model of Used Sailboat Prices

## 4.1 The Establishment of The Model

### 4.1.1 Model 1 —Heuristic Hierarchical Multiple Regression

*Model 1*, as illustrated in Figure 5, is specifically designed to predict prices by taking into account input variables such as Year, Region, and Variant. To achieve accurate predictions, this model incorporates a heuristic hierarchical multiple regression approach, which distinguishes it from conventional regression methods. This innovative technique allows for a more nuanced understanding of the relationships between variables, ultimately leading to improved performance and reliability.

As for Hierarchical Multiple Regression, to solve the objective function for a given $n$ data points $(x_i, y_i)$:

$$L(y, f(x, w)) = \sum_{i=1}^{n} [y_i - f(x_i, w_i)]^2$$

To get the coefficient matrix, we typically use the least squares method to solve it:

$$minf(x) = \sum_{i=1}^{n} L_i^2(x) = \sum_{i=1}^{n} L_i^2[y_i, f(x_i, w_i)] = \sum_{i=1}^{n} [y_i - f(x_i, w_i)]^2$$

The matrix is illustrated as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (\sum x_i x_i^T)^{-1} (\sum x_i y_i)$$

Then we get the predicted value:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_p x_p$$

In the hierarchical multiple regression process, each layer focuses on one variable, denormalizing it while normalizing the other two variables. The price generated at this stage is regarded as the Effect of the specific variable under consideration. Simultaneously, the fitting result from the current layer is used as the baseline for normalization in the next layer, which is then incorporated into the regression. This iterative process continues until the model achieves a precise fitting effect, ensuring optimal performance and an accurate understanding of the relationships between variables. A key aspect of Model 1's success is the heuristic approach employed during each normalization step. This method adapts to the size of the dataset, offering a more efficient and effective means of processing the data. For larger datasets, regression is performed, which capitalizes on the

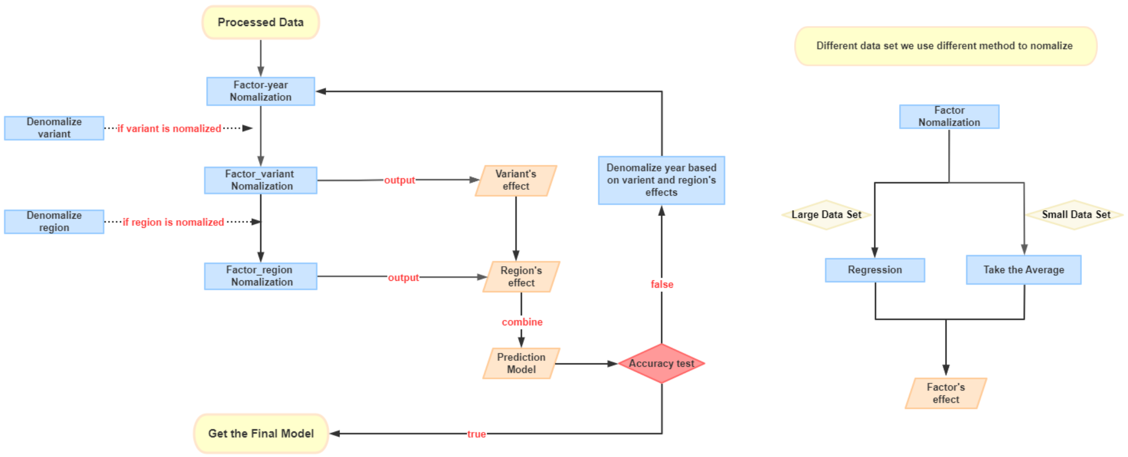


Figure 5: Brief process of Model 1

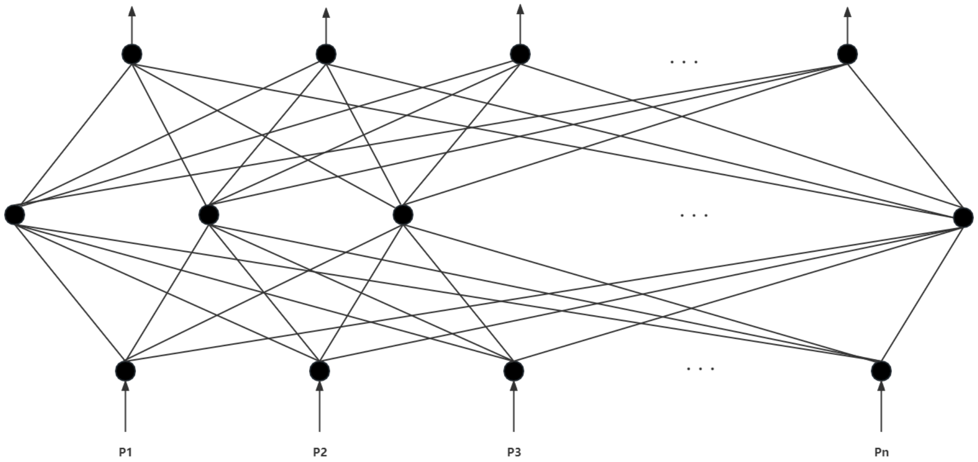


Figure 6: Hierarchical Multiple Regression

wealth of information available and can identify complex patterns and relationships. Conversely, for smaller datasets, the model calculates the average, providing a simpler and more straightforward representation of the data.

This adaptive heuristic approach has a significant impact on the accuracy of the model's predictions, as it tailors the normalization process to suit the specific characteristics of the dataset. By adjusting the method used based on dataset size, Model 1 is better able to understand the underlying patterns and relationships, leading to more accurate and reliable predictions.

Furthermore, the hierarchical structure of the model enables it to isolate the effects of individual variables, providing valuable insights into their specific contributions to the overall prediction. This granular understanding of the relationships between variables can be instrumental in guiding further analysis and informing decision-making processes.

In summary, Model 1 offers a powerful and adaptive approach to price prediction, harnessing the power of heuristic hierarchical multiple regression to deliver accurate and reliable results. By tailoring the normalization process to suit the characteristics of the dataset and iteratively refining the model's understanding of variable relationships, Model 1 provides a robust and versatile tool for predicting prices based on input variables such as Year, Region, and Variant.

### 4.1.2 Model 2 —Deep Forest Model

However, *Model 1* is somehow incomplete because *Model 1* requires that the variant and region of the predicted data have appeared in the dataset before. However, in actual prediction, the variant and region may be opaque to the model. Taking the third question as an example, *Model 1* can't predict the price of Hong Kong because the region effect of Hong Kong hasn't been evaluated by the model. Therefore, we use the **Deep Forest Model** to establish *Model 2*.

Deep forest is an ensemble learning model derived from the random forest approach. Ensemble learning techniques combine the predictions of multiple models to improve overall performance and reduce the risk of overfitting. In this case, the deep forest algorithm aggregates the results of numerous decision tree classifiers to produce a more accurate and stable prediction.

Unlike traditional decision trees, where each tree classifier is trained using the same dataset and

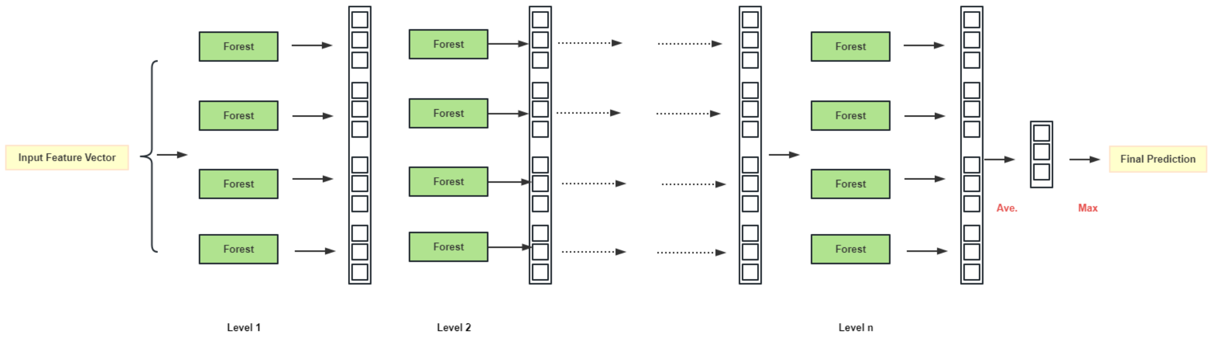


Figure 7: Deep Forest Model

strategy, deep forest introduces an element of randomness to increase the diversity of the model. By incorporating a randomized strategy for each tree classifier, the algorithm can capture a broader range of patterns and relationships within the data. This stochastic approach helps to minimize the correlation between individual trees, thereby improving the model's overall performance.

The deep forest model has several advantages over a single decision tree. First, it is more robust, as it leverages the wisdom of multiple tree classifiers to arrive at a consensus prediction. This aggregation process helps to reduce the impact of any single tree's errors, ensuring that the overall prediction is more reliable. Second, the deep forest model is less susceptible to overfitting, a common issue in machine learning where a model learns the training data too well and performs poorly on new, unseen data. By using a randomized strategy to train each tree classifier, deep forest reduces the likelihood of overfitting and improves the model's generalization capabilities. Another key benefit of the deep forest model is its ability to handle both categorical and continuous features, making it highly versatile for various types of problems. Additionally, the deep forest algorithm's structure enables it to be easily parallelized, allowing for efficient processing on multi-core processors or distributed computing environments.

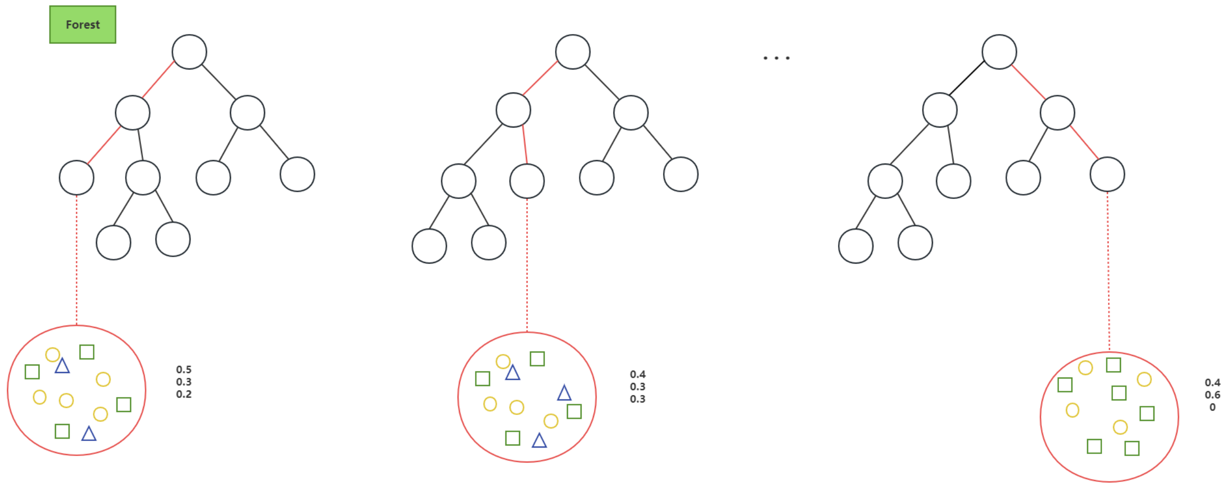


Figure 8: Deep Forest Model

In summary, the deep forest algorithm is an effective and robust machine learning method that addresses the limitations of traditional decision trees. By employing an ensemble learning approach and incorporating randomized strategies for training tree classifiers, deep forest models can deliver more accurate and reliable predictions. These features make the deep forest model particularly well-suited for challenging tasks, such as predicting prices, where a high degree of accuracy and stability is essential for success.

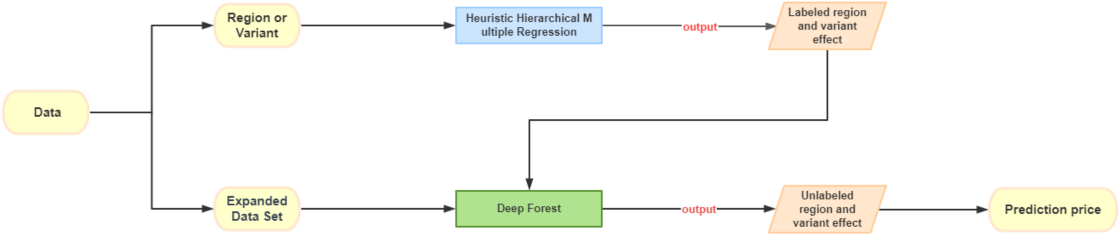

Data
Region or Variant
Expanded Data Set
Heuristic Hierarchical M ultiple Regression
output
Labeled region and variant effect
Deep Forest
output
Unlabeled region and variant effect
Prediction price

Figure 9: Brief process of Model 2

As for *Model 2*, we first expanded the dataset to obtain quantified data for regions and variants. Specifically, for variants, we expanded data such as `draft`, `beam`, `water tank`, `fuel tank`, `cabins`, `displacement`, etc. For regions, we expanded data such as `GDP`, `Gini coefficient`, `Human Development Index`, etc. When dealing with unmarked variants and regions, we first used *Model 1* to obtain the effects of marked variants and regions, and then applied the effects of marked variants and regions to the deep forest. By expanding the data for regression, we obtained the effects of unmarked variants and regions, and thus predicted the price.

## 4.2 Model Accuracy Analysis

Due to the limited data provided for some sailship types, they belong to imbalanced datasets. Therefore, we need to use as much data as possible for training to prevent the loss of small-sample data during training. Given that the leave-one-out method is not affected by random partition, it is effective for small-sample and imbalanced datasets, and can accurately evaluate the performance of the model, we used the **leave-one-out method** to analyze the model's error in the error analysis part.

For accuracy analysis, we use the correlation coefficient $R$ and Mean Absolute Percentage Error (MAPE) as evaluation metrics.

$$R(\mathbf{X},\mathbf{Y})=\frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\sqrt{\mathrm{var}[\mathbf{X}]\mathrm{v\ ar}[\mathbf{Y}]}}$$

$$\mathrm{MAPE}=\frac{\sum_{i=1}^{n}(\frac{|\mathbf{P}-\mathbf{Q}|}{\mathbf{P}}\times 100\%)}{n}$$

where:

$$\mathrm{var}(\mathbf{X})=\frac{\sum_{i=1}^{n}(\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{X}_i-\overline{\mathbf{X}})}{n-1}$$

$$\mathrm{cov}(\mathbf{X},\mathbf{Y})=\frac{\sum_{i=1}^{n}(\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{Y}_i-\overline{\mathbf{Y}})}{n-1}$$

# 5 Regional Effect Analysis

Applying *Model 1*, we can easily abstract the effect of each region. Therefore, by normalizing the year and variant, we can clearly see the effect of each region(see Figure 10).

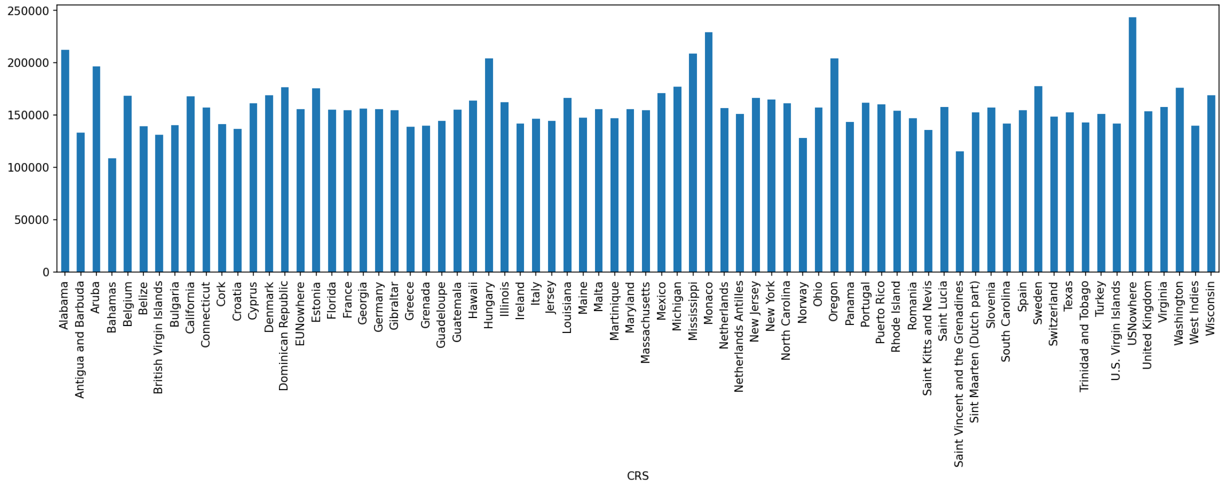


Figure 10: Region Effect

# 6 The applicability of The Prediction Model in Hong Kong

When applying our model to the Hong Kong region, we first assume that we can obtain the Year, Variant, and Region parameters for the used sailboats. For the given geographic regions, we start by using *Model 2* to determine the Region Effect, leveraging the area's GDP data.

If the predicted Variant is present in the training set, we can directly use *Model 1* to estimate the pricing by incorporating the Region Effect, Variant Effect, and Year. However, if the predicted Variant is not in the training set, we first need to use *Model 2* to obtain the Variant Effect by considering factors such as beam, draft, and displacement. Once the Variant Effect is determined, we can then employ *Model 1* to make the final price prediction.

This two-step approach enables us to adapt our models to different regions like Hong Kong and successfully predict used sailboat prices. By utilizing both *Model 1* and *Model 2* in tandem, we can account for regional influences, specific variant characteristics, and the age of the sailboat to provide accurate and reliable pricing estimates.

# 7 Extended Inferences or Conclusion

# 8 Further Improvements

# 9 Strengths and Weaknesses

## 9.1 Strengths

**Comprehensiveness:** After the basic data processing, we significantly expanded the dataset, including extending the Variant to Draft, Beam, Displacement, Water tank, Fuel tank, and Cabin, as well as expanding the Region to GDP, HDI, and Gini coefficient. We performed correlation tests for each variable, making the modeling process extensive and comprehensive.

**Innovative:** Our model combines heuristic learning and hierarchical multivariate regression, obtaining accurate predictions through continuous iteration. This creative and highly applicable model is well-suited for the task at hand.Additionally, we used a deep forest algorithm to extend the prediction of unknown variants and regions, enhancing the model's versatility.

**Accuracy:** Our model performs exceptionally well in terms of both error analysis and accuracy, achieving $94.91\%$ on the Catamarans test set and $92.71\%$ on the Monohulled Boats test set. Moreover, the model can improve its accuracy through self-iteration and possesses heuristic learning capabilities.

**Applicability:** When applied to the Hong Kong market, the model maintains a high level of accuracy, demonstrating its excellent applicability.

## 9.2 Weaknesses

**Time consuming:** The model can achieve satisfactory regression results after 2-3 iterations when processing the data for this task. However, when handling larger datasets, the increased number of iterations may result in additional time costs.

**Lack of geography factor analysis:** Due to the difficulty in obtaining geographical data and the challenge of analyzing and quantifying factors such as climate, the model does not include geographical factors. Despite this limitation, the model still provides valuable insights and accurate predictions for the used sailboat market.

# 10 Conclusion

In conclusion, our study has successfully addressed the challenges faced by brokers in the used sailboat market, where pricing is influenced by a myriad of factors such as brand, variant, age, depreciation rates, local consumption levels, and geographical environments. By developing a comprehensive and robust model that combines Heuristic Hierarchical Multiple Regression and Deep Forest Models, we have created a valuable tool that allows brokers and other stakeholders in the sailboat market to make well-informed decisions and ensure fair pricing practices.

Our meticulous data cleaning, expansion, and pre-analysis processes enabled us to establish a strong foundation for subsequent modeling efforts. Through Adaptive Density-Based Clustering and Linear interpolation techniques, we were able to identify patterns and trends in the data, providing crucial insights for our modeling approach.

The Heuristic Hierarchical Multiple Regression Model allowed us to establish the relationships between sailboat prices and factors such as year, region, and variant. Moreover, it enabled us to abstract the effect of regions and variants on sailboat prices, capturing the interactions between these variables and their influence on pricing.

By employing the Deep Forest Model, we trained the relationship between region and variant factors and their respective effects. This advanced ensemble learning technique enabled us to predict the effect of unknown regions and variants, which in turn allowed us to further predict their prices.

Our model has proven to be highly effective, achieving remarkable accuracy levels of 94.91% on the Catamarans test set and 92.71% on the Monohulled Boats test set. These results demonstrate the model's strong performance and applicability to real-world scenarios. Furthermore, the model's self-iteration and heuristic learning capabilities ensure that it can continuously improve its accuracy, making it an invaluable tool for brokers in the used sailboat market.

When applied to the Hong Kong market, our model maintained a high level of accuracy, showcasing its excellent applicability across different regions. The model's ability to achieve satisfactory regression results after 2-3 iterations is particularly noteworthy, as it demonstrates its efficiency and effectiveness in handling diverse datasets.

While our model does not currently include geographical factors due to the challenges in obtaining and quantifying geographical data, future research could explore incorporating these factors to further enhance the model's predictive capabilities.

In summary, our study has successfully developed a powerful and versatile model that can assist brokers in the used sailboat market in making more reasonable and comprehensive evaluations, ultimately leading to more rational and fair pricing practices. By combining the strengths of Heuristic Hierarchical Multiple Regression and Deep Forest Models, we have created a valuable tool that can be applied to various markets and datasets, making it an invaluable asset for brokers and stakeholders in the used sailboat market.

# Report

**To:** Heishan Yan
**From:** Team 1234567
**Date:** October 1st, 2019
**Subject:** A better choice than MS Word: LaTeX

In the memo, we want to introduce you an alternate typesetting program to the prevailing MS Word: **LaTeX**. In fact, the history of LaTeX is even longer than that of MS Word. In 1970s, the famous computer scientist Donald Knuth first came out with a typesetting program, which named TeX …

Firstly, …

Secondly, …

Lastly, …

According to all those mentioned above, it is really worth to have a try on LaTeX!

# References

[1] Zhi-Hua Zhou, Ji Feng, Deep forest, *National Science Review*, January 2019, from `https://doi.org/10.1093/nsr/nwy108`.

[2] Eberly, L.E. (2007). Multiple Linear Regression. *Methods in Molecular Biology*, from `https://doi.org/10.1007/978-1-59745-530-5_9`.

[3] A. Ogunleye and Q. -G. Wang, ”XGBoost Model for Chronic Kidney Disease Diagnosis,” in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

[4] Ingo Steinwart. ”Fully adaptive density-based clustering.” Ann. Statist, from `https://doi.org/10.1214/15-AOS1331`.

[5] Reiss, K., Renkl, A. Learning to prove: The idea of heuristic examples. *Zentralblatt für Didaktik der Mathematik*, from `https://doi.org/10.1007/BF02655690`.

[6] Larson M G. Analysis of variance[J]. Circulation.

[7] Witten I H, Frank E, Hall M A, et al. Practical machine learning tools and techniques[C].

[8] Myles A J, Feudale R N, Liu Y, et al. An introduction to decision tree modeling[J].

[9] Biau G, Scornet E. A random forest guided tour[J].

[10] Genesove D. Adverse selection in the wholesale used car market[J].

[11] Bernard H R, Ryan G. Text analysis[J].

[12] Davis P J. Interpolation and approximation[M]. Courier Corporation, 1975.

[13] Ganaie M A, Hu M, Malik A K, et al. Ensemble deep learning: A review[J].

[14] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J].

[15] Madhulatha T S. An overview on clustering methods[J].

[16] Limsombunchao V. House price prediction, hedonic price model.

# Appendix A: Further on LaTeX

To clarify the importance of using LaTeX in MCM or ICM, several points need to be covered, which are …

To be more specific, …

All in all, …

Anyway, nobody **really** needs such appendix …

# Appendix B: Program Codes

Here are the program codes we used in our research.

**test.py**

```python
# Python code example
for i in range(10):
    print('Hello, world!')
```

**test.m**

```matlab
% MATLAB code example
for i = 1:10
    disp("hello, world!");
end
```

**test.cpp**

```cpp
// C++ code example
#include <iostream>
using namespace std;

int main() {
    for (int i = 0; i < 10; i++)
        cout << "hello, world" << endl;
    return 0;
}
```