A BP Neural Network Used Sailboat Price Prediction Model Based on Depreciation Formula Optimization

Yang Lu †, Kaiyu Wang †, Yushuai Wu *,†

College of Civil Engineering and Water Conservancy, Qinghai University, Xining, China, 810016

* Corresponding Author Email: wysqhu@163.com

† These authors contributed equally.

Abstract. In the used sailboat market, accurate estimation of sailboat price is of great significance for both sellers and buyers. Based on the study of the weights of the factors affecting the price of used sailboats using the random forest algorithm, this paper aims to propose a BP neural network method based on the optimization of the depreciation formula for the price estimation of used sailboats. It solves the problem that the traditional multi-input, non-linear model cannot effectively predict the price of used sailboats for future years. The model accuracy can reach 0.758 for monohull sailboats and 0.904 for catamaran sailboats.

Keywords: Used Sailboat, Depreciation Formula, BP Neural Network.

1. Introduction

Shipping is an important part of the world's economic development, and ships are a fundamental element of the shipping market. [1] With the development of the market economy, the global shipping industry is gradually recovering after about 10 years of market downturn characterized by overcapacity and sharply reduced trade volume. [2] At the same time, the shipping market is on fire, which also leads to the growth of second-hand ship trading volume. [3] Used sailboat trading market in different countries under the influence of the system and policy, showing significant characteristics of the times. In coastal countries, as a more practical means of sea transportation, many people began to buy sailboats. The expensive price of new sailboats and the long service life of sailboats are characteristics that make people prefer to buy second-hand sailboats. Due to the special characteristics of marine assets, the price of a used sailboat is not only affected by factors such as manufacturer, sailboat model, displacement, canvas area, etc., but also by factors such as market rates for different types of vessels, the scope of vessel operation, and time.

In the past, the research direction on used sailboats mainly focuses on two aspects, the first one is to find the covariance relationship between the price of used boats and the factors affecting the price of used boats based on the time series model. [4] Meanwhile the research direction also includes econometric analysis. Luo Fucai established two BP neural network models for evaluating the price of second-hand ships but there are two problems: first, there is no subdividing of ships because different variables have different impacts on different ship types; second, there is no weight analysis of the factors affecting the price of second-hand ships, which makes the training effect of the BP neural network lack of a basis.Du et al. introduced a market price evaluation method applicable to different ship types. The market substitution method is used to evaluate the price of used sailboats. [5]

Through the analysis of previous studies, this paper establishes a BP neural network optimized using the depreciation formula to predict the price of used sailboats on the basis of the weights of the factors influencing the price of the studied used boat. The overfitting problem that BP neural network is prone to and the limitation of econometric model to predict the price of used sailboat without reference price are solved.

2. Used sailboat pricing model architecture

2.1. Architecture of the Random Forest Model

Random forests are a relatively new machine learning model that can achieve high levels of classification accuracy even with certain missing data, and random forests are not overfitted due to the addition of decision trees. [6] The basic principle of the Random Forest algorithm is to use Bootstrap sub-self-sampling to obtain different sets of samples for model construction, thus increasing the degree of variability between models and improving the ability to extrapolate predictions. [7] The principle of the Random Forest model is to use the Bagging algorithm in a CART decision tree to perform putative sampling in the sample set, and to extract multiple training sample sets of the same size as the original sample set to complete the construction of the integrated model. [8] Random forest is also a relatively new class of machine learning models, even in the case of a certain lack of data, the classification accuracy can be at a high level, and will not be caused by the increase in the decision tree random forest overfitting. The flowchart of the algorithm is shown in Figure 1.

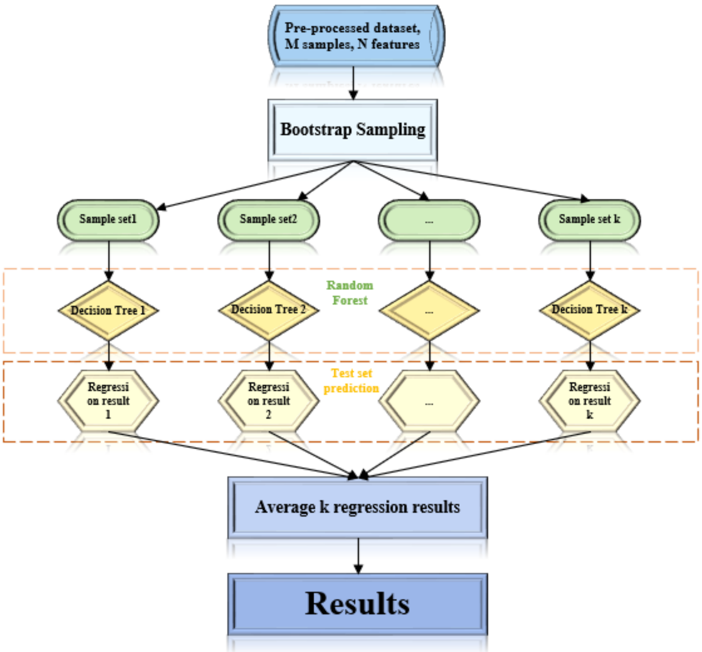

Figure 1. Flow chart of random forest regression model algorithm

The random forest model that realizes the weight analysis of the factors influencing the price of a used sailboat is a combination of multiple binary decision trees packaged together, and training a random forest is to train multiple binary decision trees. Specifically can be divided into four steps:

(1) Determine the weighted sum $G(x_i, v_{ij})$ of the impurity of each sub-node using the exhaustive method.

(2) Calculate the function $G(x_i, v_{ij})$ that measures the impurity of a node.

(3) Training of a node in a decision tree.

(4) Finding the cut-off variable and cut-off point that minimizes $G(x_i, v_{ij})$.

We can represent the above process as follows using Eq:

$$G(x_i, v_{ij}) = \frac{n_{\text{left}}}{N_s} H(X_{\text{left}}) + \frac{n_{\text{right}}}{N_s} H(X_{\text{right}}) \tag{1}$$

$$H(X_m) = \frac{1}{N_m} \sum_{\mathrm{i} \in \mathrm{N_m}} (y - \bar{y}_m)^2 \tag{2}$$

$$(x^*,v^*)=argmin_{x,v}G(x_i,v_{ij}) \tag{3}$$

$$G(x,v)=\frac{1}{N_s}\left(\sum_{y_i\in X_{\text{left}}}\left(y_i-\bar{y}_{\text{left}}\right)^2+\sum_{y_j\in X_{\text{right}}}\left(y_j-\bar{y}_{\text{right}}\right)^2\right) \tag{4}$$

2.2. Structure of the depreciation formula

The average life method, also known as the straight-line method, is a method of depreciating equipment assets evenly over their useful lives. [9] The annual depreciation rate calculated by the average life method is considered to be a constant value, and the equipment can be depreciated over its useful life, thereby addressing the effect of the time factor on the value of the equipment.

The average life method requires a predetermined expected net salvage rate of β, the original cost of the used sailboat is n, and the expected net salvage value is h:

$$h=\beta\times n \tag{5}$$

Let the annual depreciation rate be α and the estimated useful life be Y. You can obtain the annual depreciation rate ε for a used sailboat and the monthly depreciation rate η for a used sailboat:

$$\varepsilon=[\ (1-\alpha)\ /Y]\times 100\% \tag{6}$$

$$\eta=\varepsilon/12 \tag{7}$$

The monthly depreciation amount m can ultimately be calculated from the monthly depreciation rate η:

$$m=n\times\eta \tag{8}$$

The estimated useful life Y of the sailboat is 30 years, the estimated net salvage rate β is 5%, and the year of delivery of the sailboat is X. The pre-depreciation price of the used sailboat is J.

$$J=\frac{n}{(2023-x+1)}\times Y\div(1-\beta) \tag{9}$$

2.3. Architecture of BP neural network

BP (Back Propagation) Neural Network is a multilayer feed forward network consisting of input, output and hidden layers with varying number of neurons in each layer. [10] The BP neural network architecture is as follows:

(1) Implicit layer selection: h is the number of implicit layer nodes, m is the number of input layer nodes, and n is the number of output layer nodes.

$$\mathrm{h}=\sqrt{m+n}+a \tag{10}$$

(2) Signal forward propagation: input $[x_1,x_2]$, weights $[w_1,w_2]$, bias values w_0, f is the activation function.

$$z=w_0x_0+w_1x_{1+}w_2x_2=\sum_{i=0}^{2}w_ix_i \tag{11}$$

In vector form it can be expressed as:

$$z = \sum_{i=0}^{2} w_i \, x_i = w^T x \tag{12}$$

$$y = f(w^T x) \tag{13}$$

(3) The result of the output layer in error back propagation is $y = f(w^T x)$. For the calculation of the error between the predicted value and the true value, the following formula (cost function) is practical:

$$E = \frac{1}{2}(t - y)^2 \tag{14}$$

3. Results

3.1. Data processing

We have collected other relevant indicators of used sailboats and regional economic tourism conditions. Details are shown in Table 1 below.

Table1. Data and data sources

Database Names	Database Websites
Sailing data	http://www.sailboatdata.com/
GDP per capita	https://www.worldbank.org/ https://iccwbo.org/ https://www.weforum.org/
Gross GDP	https://www.worldbank.org/ https://iccwbo.org/ https://www.weforum.org/
Average cargo throughput	https://www.porttechnology.org/

By reviewing the literature and combining the special characteristics of sailboats, we mainly discussed the effects of Make, Variant, Length, Listing Price, Release year, LWL (ft), Beam (ft), Draft (ft), Displacement (lbs), Sail Area (sq ft), Average cargo throughput (tons), and longitude, latitude, GDP per capita, and tourist trips on the selling price of a used sailboat, Average cargo throughput (tons), as well as longitude, latitude, GDP per capita, tourism, and other factors that influence the sale price of used sailboats in the area where they are sold. We collect data on sailboats from sailing associations and regional economic and tourism data from the World Bank and the U.S. National Bureau of Statistics. The data collected on sailboats are roughly divided into monohulls and catamarans, which are corrected, censored, and labeled separately.

By observing the data and reviewing the information, we found that most of the sailboats were named according to their captains, but there was a small deviation between the naming of individual second-hand sailboats and the length of the sailboats. Therefore, the data in column Length is the true and accurate length of the sailboat. And the vacant values are eliminated. At the same time, the related text data, to be labeled.

3.2. Random Forest Modeling Results Presentation

The optimal number of leaves of the random forest is 5, and the optimal number tree is 100. the accuracy of monohull sailboat R is 0.89, and the accuracy of catamaran sailboat R is 0.91. We have obtained the values of each index for a typical ship model including 6 ship type indexes and 3 regional economic indexes as shown in Fig. 2.

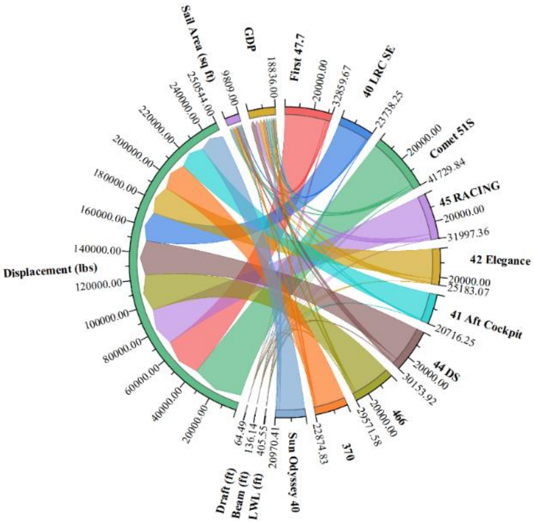

Figure 2. Chord diagram of some ship characteristic parameters

3.3. Presentation of forecast results

When the BP neural network optimized with the depreciation formula is used to predict the price of a used sailboat, the regression results are shown in Figures 3 and 4 below. The regression coefficients on the training set, test set and validation set can be up to 0.801, 0.727 and 0.752 for monohull used sailboats, and up to 0.904, 0.942 and 0.849 for catamaran used sailboats.The performance of the model is better, and the performance of the model can basically meet the requirements.

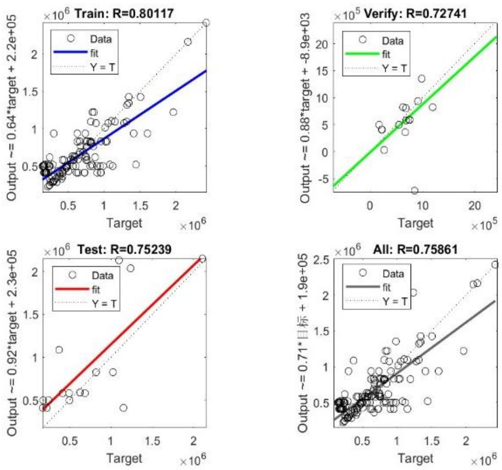

Figure 3. Regression plot for monohull sailboat prediction

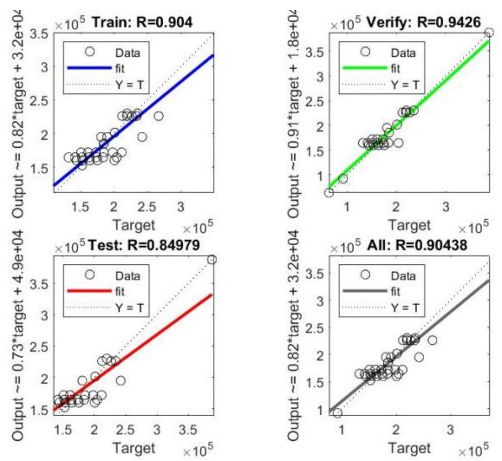

Figure 4. Predictive regression plot for catamaran sailboats

4. Conclusion

The expanding market for trading used sailboats has made determining the value and proper pricing of used sailboats a major problem for sellers of used sailboats. Past models often fail to make good predictions for such multi-input, nonlinear problems, and some of the resulting models are only suitable for predicting the trading price of used sailboats in the current year. The model established in this paper for used sailboat price prediction based on BP neural network optimized by depreciation formula not only solves the problem of predicting the price of used sailboats in future years, but also solves this multi-input, nonlinear problem at the same time. The experimental results show that the model in this paper has an accuracy of more than 75% for the prediction of different models of sailboats, and can estimate the price of used boats more accurately.

However, the accuracy of the model needs to be further improved because the amount of data is not very large. Meanwhile, because the value of sailboats in this paper is based on the selling price in 2023 as a reference, random factors such as the economic crisis are not taken into account when predicting the price of used boats in the future.

References

[1] Ma W.Y.. Research on the value assessment and investment decision of used dry bulk carriers [D]. Shanghai Jiao Tong University, 2022.

[2] Du K, Jiao F. Influencing factors in the evaluation of second-hand ship price market method [J]. Water Transport Management, 2020, 42 (11): 1-4+7.

[3] Hu Haoqiong. Second-hand ship market on fire in 2021 [J]. China Shipping Weekly, 2022 (21): 32-33.

[4] Jeroen Frederik Josef Pruyna, Eddy van de Voordeb, Hilde Meersmanb. Second hand vessel value estimation in maritime economics: a review of the past 20 Second hand vessel value estimation in maritime economics: a review of the past 20 years and the proposal of an elementary method [J]. Marine Economics & Logistics, 2011: 213-236.

[5] Xiao Q, Zhang Y. Research on second-hand ship price evaluation based on BP neural network [J]. Ship engineering, 2013 (2): 100-103.

[6] Hu, At-Wen, Zhang, T.Y., Zhang, X. & Zhou, C.Y.. (2022). Research on used car price prediction algorithm based on machine learning. Information Technology and Informatization (10), 52-55.

[7] Dong Hongyao, Wang Yidan, Li Lihong. A review of random forest optimization algorithms [J]. Information and Computer (Theoretical Edition), 2021, 33 (17): 34-37.

[8] Deng Hongwei, Luo Liang. Optimization of PPV prediction model for random forest based on SMA algorithm [J/OL]. Gold Science and Technology: 1-16 [2023-08-16].

[9] Song Changzhi. Economic analysis of offshore decommissioning equipment and facilities [J]. China Offshore Platform, 2020, 35 (02): 97-100.

[10] LI Fang, MIN Yongjun. Application of BP neural network optimized by genetic algorithm in used car price evaluation [J]. Software, 2021, 42 (08): 91-94.