The hedonic price model of online short-term rental market based on machine learning

Jin Xin
Shanghai Maritime University, Shanghai, China
jinxin@shmtu.edu.cn

Lei Xue
Shanghai Maritime University, Shanghai, China,
lx18756469220@163.com, Corresponding author

ABSTRACT

Under the background of the sharing economy, the online short-term rental market contains huge business opportunities, but the development of online short-term rental models in various regions is uneven, and there is a lack of reasonable pricing. Therefore, based on the Beijing Airbnb online short-term rental data set, this paper adopts the advanced machine learning method AutoGluon model to predict the price range, and finally analyzes the solvability of the model. First, data preprocessing is performed on the initial dataset. Second, an initial exploration of the data found that prices are correlated with housing type, location, and surrounding environment. Then, based on the existing features, an AutoGluon hedonic price model is established to predict the price range; finally, interpretable analysis of the model is performed to identify key factors. Geographical location and room type undoubtedly have the greatest impact on the online short-term rental market price, providing a reference for online short-term rental platforms and homeowners to reasonably customize house prices and improve service quality.

CCS CONCEPTS

• **Theory of computation** → Design and analysis of algorithms.

KEYWORDS

Hedonic price, AutoGluon model, online short-term rental, machine learning

ACM Reference Format:

Jin Xin and Lei Xue. 2022. The hedonic price model of online short-term rental market based on machine learning. In *7th International Conference on Cyber Security and Information Engineering (ICCSIE2022), September 23–25, 2022, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3558819.3565227

1 INTRODUCTION

The sharing economy refers to the effective reconfiguration of idle resources around them through the network platform, to maximize the utilization of resources. As the concept of sharing is deeply rooted in the hearts of the people, this model has developed rapidly in China; at the same time, with the increasing number of tourists in the tourism industry, the traditional accommodation model has been difficult to meet the individual needs of residents. Based on this, online short-term rentals came into being. Online short-term rental, as the name suggests, refers to the process in which homeowners transfer the right to use idle houses to short-term renters to obtain economic returns in the context of the sharing economy. This housing rental method has been recognized by the market once it has been promoted. The online short-term rental model was initially started by Airbnb in the United States, while the online short-term rental model in my country started late, but with its continuous development, it has been able to catch the eye. Taking Beijing as an example, this paper studies the hedonic price model of the online short-term rental market, hoping to provide a reference for online short-term rental platforms and homeowners to reasonably customize house prices and improve service quality.

The core of this paper is to solve the problem of price range prediction in online short-term rental scenarios through machine learning methods. At present, there are many studies on the pricing model of houses and the factors affecting the price, but there are not much related literatures applied to the online short-term rental scenario. Therefore, this paper summarizes the research on online short-term rental as follows:

From the perspective of both supply and demand, first, on the user side: Wu Jiang [1], based on the SOR model and combined with the theory of clue utilization, studied the effect of pictures of idle houses in online short-term rental platforms on consumers' future participation willingness. They believe that product involvement plays a significant positive moderating role between task information and consumers' own willingness. Lu, C. [2] used web crawler technology and content analysis to capture and interpret Airbnb's online reviews. They believed that tourists' motivation for choosing online short-term rentals is more inclined to the price of the house, the surrounding environment, services, etc.; Tourists hope to have a home-like consumption experience. In addition, online short-term rentals will provide opportunities to interact with local people, so that tourists can experience local customs and release emotions. Not only that, but tourists can also experience the culture of different places through online short-term rentals, which is also beyond the reach of traditional accommodation. Liu Yingjie [3] empirically analyzes the influencing factors of online short-term homestay rentals in China and the United States and consumers' own purchasing behavior in their respective contexts. She believes that there is regional heterogeneity in the influencing factors.

In terms of the platform: Ling Chao [4] discussed the main challenges faced by the development of online short-term rentals in China under the background of the sharing economy and put forward corresponding strategies accordingly. Wang Chen [5] combined the operating conditions, cash flow, and actual development status of Xiao Zhu short-term rental platform in the past three

ICCSIE2022, September 23–25, 2022, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-9741-4/22/09…$15.00
https://doi.org/10.1145/3558819.3565227

Table 1: Main features of the data source

Feature name	Feature explanation	Counts	Type
id	housing id	28452	Numerical
name	housing name	28451	Character
host_id	owner id	28452	Numerical
host_name	Owner's name	28452	Character
neighbourhood	Affiliated Community	28452	Character
latitude	latitude	28452	Numerical
longitude	longitude	28452	Numerical
room_type	housing type	28452	Character
price	price	28452	Numerical
minimum_nights	Minimum rental days	28452	Numerical
number_of_reviews	Number of comments	28452	Numerical
last_review	Last comment time	17294	Character
reviews_per_month	Percentage of monthly comments	17294	Numerical
calculated_host_listings_count	Number of Homeowners' Listings	28452	Numerical
availability_365	Number of rental days per year	28452	Numerical

years, and provided it with personalized IPO exit suggestions. Song Lin [6] based on the theoretical model of traditional evolutionary game, analyzes the multi-party game behavior between online short-term rental platforms, to analyze the operation mode of short-term rental platforms and how it affects the game mechanism. Xu Yan [7] based on the business model canvas model, took Xiao Zhu short-term rental as the case study object, compared the elements of the Tujia short-term rental business model, and found that their business models are heterogeneous, and each has its own merits. Liang Jiaming [8] provided a new computing framework based on the short-term rental platform trust problem.

To sum up, previous research on online short-term rentals includes users' purchase intention and its influencing factors, the business model of online short-term rental platforms, the game between platforms, the growth path of the platform, and the tax calculation of the platform, and the trust of the platform. Most of them are concentrated between users and platforms, but there are fewer hedonic price models for the online short-term rental market. Therefore, based on the background of the sharing economy and taking Beijing as an example, this paper uses machine learning methods to find out the influencing factors of prices, builds a characteristic price model for online short-term rentals, finds key features, and reasonably customizes online short-term rental platforms and homeowners Provide a reference for housing prices and improving the service quality.

2 DATA SOURCES AND PROCESSING

2.1 Data Sources

The data in this article comes from the Beijing Airbnb short-term rental data set of Ali Tian chi. The data set has fifteen features, namely housing id, housing name, owner id, owner name, community, latitude, longitude, housing type, price, the minimum rental days, the number of reviews, the last review time, the proportion of monthly reviews, the number of homeowners' listings, and the number of rental days per year. Table 1 shows the main characteristics of the specific data sources.

2.2 Data preprocessing

1) Null value handling. It can be seen from Table 1 that there are null values for the three characteristics of housing name, last comment time, and monthly comment volume ratio. First, analyze the house name, it is missing a row of data, delete it directly; secondly analyze the last comment time, it is timestamp data, this paper believes that this feature is not significant for the next data exploration and prediction, so delete it, and finally come to Analyzing the proportion of monthly comments, this paper analyzes the characteristics of the number of comments and finds that when the number of comments is zero, the proportion of monthly comments is null, so it is non-randomly missing data. The monthly comment volume ratio data is assigned zero.

2) Outlier handling. There are usually two sources of outliers, namely outliers in the data itself and errors caused by format conversion. First, analyze the outliers in the data itself. Through simple descriptive statistics, it is found that the minimum value of the housing price and the number of rental days per year is zero, which is considered unreasonable in this paper and is an outlier, and is deleted. In addition, the standard deviation of the housing price is too large, indicating that the data The fluctuation is large, and there is a large value that affects the normal data. By analyzing the maximum price, when the price is equal to 68983, the house name says that the house is no longer rented, so this row of data is meaningless, delete it, and then analyze the maximum price, the house name is normal.

3) Other processing. This article believes that the role of the owner id and the owner's name for our subsequent data exploration and prediction is not significant, and these two unimportant columns are deleted. In summary, 25,826 samples are remaining in this article.

3 PRELIMINARY EXPLORATION

3.1 Price and type of housing

First of all, this paper makes a drawing analysis of the core characteristic price. The drawing result is shown in Figure 1. It can be

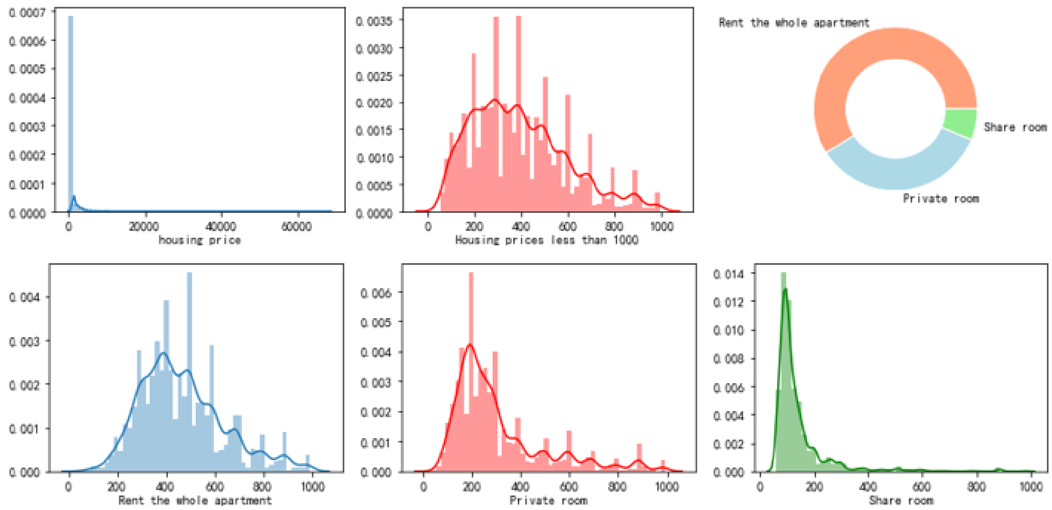

Figure 1: Distribution of prices and housing types

seen that most of the housing prices in the whole sample are within 10,000, and no detailed distribution can be seen. Considering that Airbnb's business philosophy is to share and use idle houses, and most of the idle houses provided by the owners are ordinary, so the price is not too high, so the short-term rental price range is reduced to less than 1000, of course, the housing price Outliers above 1000 also exist objectively. According to Figure 1, it is found that the housing prices within 1000 are more in the range of 200 to 700 per night, which is more in line with expectations. Secondly, this paper conducts statistical analysis on the types of housing whose housing price is less than 1000 and finds that there are 13,771 rented apartments, 8,116 private rooms, and 1,427 shared rooms. The number of rented apartments is the largest, and the number of shared rooms is the smallest. In order to understand the price distribution of housing types, the plots were analyzed separately, and the results are shown in Figure 1. First of all, the price distribution of the whole rental apartment is relatively even, there are both low-cost and high-priced houses, and there is also a certain proportion of high-priced houses. It is relatively cheap at around 100 to 300.This may be due to the high utilization rate, while the short-term rental prices of shared rooms are mostly concentrated at around 50 to 150, which is the cheapest among housing types. It may be because most of these rooms are mainly shared houses such as multi-person homestays and youth hostels, and most of the prices only reach the standard of one bed. From the above, this paper concludes that online short-term rental market prices are related to housing types.

3.2 Price and Geographical Location

In order to determine the impact of geographic location on the online short-term rental market price, this paper includes housing id, longitude, latitude, and online short-term rental price data based on PYCHARTS, and the results are shown in Figure 2. As can be seen from the picture, most of the short-term rentals are mainly concentrated in Chaoyang District, Dongcheng District, Haidian District, Xicheng District, Fengtai District, and Tongzhou District, and the rest of the areas tend to be closer to concentrated areas. In terms of price, the price of the concentrated area is mainly around 500, which is relatively evenly distributed, but there are also many short-term rentals with prices above 1001, mainly concentrated in Changping District, Shunyi District, Huairou District, Miyun District, and Yanqing District. In scattered, and relatively less developed areas such as Yanqing District, Miyun County, and Huairou District, the online short-term rental prices are relatively high, probably because of the rise of farmhouse and resort activities in recent years, short-term rental prices are relatively high. Housing in the 500 to 1000 price range is also more evenly distributed. From the above, this paper concludes that the online short-term rental market price is related to geographic location.

3.3 Price and environment

To determine the impact of the environment on the online short-term rental market price, this paper extracts the keywords from the housing names and divides them into each area, and selects the top ten keywords that appear frequently in the area. It is found that Huairou District, Changping District, Miyun District, and Yanqing District all have well-known names of scenic spots, such as Mutianyu Great Wall, Yanqi Lake, and Hongluo Temple, Ming Tombs Scenic Spot, Gubei Water Town, and Longqing Gorge, etc. These short-term rentals are mainly close to famous scenic spots. With the rise of tourism, the prices of short-term rentals in these areas will increase accordingly. Although Dongcheng and Xicheng also have famous scenic spots such as the Forbidden City and Tiananmen Square, due to the dense population of the two cities, the rental market is active. The market competition is large, so the online short-term rental prices are also higher, but the short-term rental prices are not as high as those in Huairou District; Chaoyang District, Haidian District, Fengtai District, and Tongzhou District are also active in the rental market, with dense houses, but keywords appear more frequently. Most of them are subways, hospitals, and

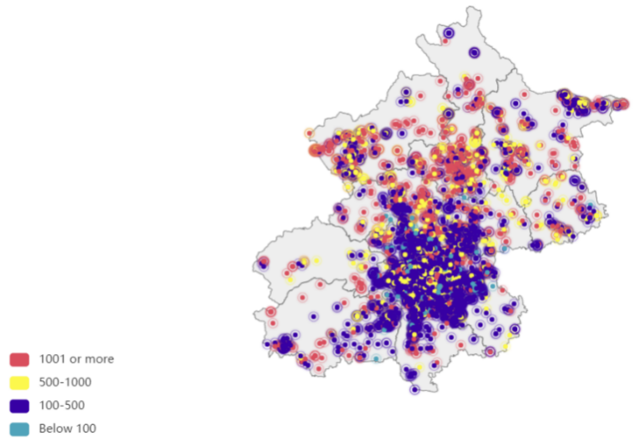

Figure 2: Price distribution of online short-term rental market in Beijing

universities, indicating that resources such as transportation, medical care, and education also play a role in short-term rental housing, and the short-term rental prices in these areas are relatively low. In summary, this paper finds that online short-term rental prices are related to housing type, location, and surrounding environment. Next, the features that can be used for prediction are preprocessed, and prices are partitioned to prevent outliers from interfering with the data. As shown in Figure 2, it is divided into four categories: less than 100, 100 to 500, 500 to 1000, and more than 1000. The remaining features are price range, community, housing type, longitude, latitude, minimum rental days, number of reviews, percentage of monthly reviews, number of owner listings, and rental days per year. There are ten features in total. These features are used as the input and output of the next model predictions.

4 HEDONIC PRICE MODEL CONSTRUCTION

4.1 Model method

With the feature data, the next step is to build a model. The model construction uses the more popular automated machine learning methods, and the AutoGluon model, Liu F. [9] also confirms this, it has good robustness. It automates data preprocessing without the need to indirectly convert the housing types and neighborhoods in this paper into dummy variables. The AutoGluon model features ensemble and multi-layer stacking. The integration technology of multi-model combination may not sound new, such as the common decision tree models such as LightGBM, CatBoost, XGBoost, and random forest, which use voting. But can the accuracy be further improved on top of multi-model combinations? The answer is yes. A new type of multi-layer stacking integration is introduced inside the AutoGluon model, which has a higher prediction effect. In this paper, ten features are input, divided into training set and test set samples, with a ratio of 7:3 respectively, and the order of the samples is shuffled. With the automatic Stacking parameter selected, the model is built.

4.2 Empirical Analysis of Hedonic Price Prediction

1) Model evaluation
After the model is built, run the model. First, model evaluation is performed. The accuracy of the model on the test set reached 0.759, and the MCC value reached 0.504. The MCC value is a multi-class evaluation indicator designed based on the confusion matrix. Its value is [-1,1]. When it is -1, the classification is completely wrong, and when it is 1, the classification is completely correct. The larger its value, the better the classification performance and the more accurate the price range classification. Secondly, the top ten internal models during the fitting period are then called out for display, as shown in Table 2. From the table, it can be seen that XGBoost is indeed a model with relatively good performance, which also verifies the research of Cao Rui [10], but XGBoost_BAG_L2 in this paper improves the prediction performance of the model, and the accuracy rate reaches 0.771.
2) Interpretability
After model evaluation, the contribution of each feature to the model prediction performance is called out, as shown in Table 3, which demonstrates the interpretability of the model. From the report of the model, among them, the characteristic data of longitude, latitude, and housing type are undoubtedly the largest contribution to the prediction performance of the model. As mentioned above, they have an impact on the price prediction, which is verified again; secondly, the homeowner The number of housing listings, the community they belong to, the number of rental days per year, the proportion of monthly reviews, and the minimum number of rental days are also significant for the prediction of the model, but the effect is not significant.
To sum up, this paper finds that the market price of online short-term rental may be related to the characteristics of housing type, geographical location, surrounding environment, the number of homeowners' listings, the community they belong to, the number of rental days per year, the proportion of monthly reviews, and the minimum number of rental days related.

5 CONCLUSION

The online short-term rental model has emerged and developed rapidly, but there are also many challenges. The existing researchers mainly focus on the commercial operation of the online short-term rental model, the growth path, and the willingness of consumers. However, there are few kinds of research on the hedonic price model for the online short-term rental market, and most of the machine learning models used in the existing research lack the interpretability of the model. Therefore, this paper is based on the Airbnb online short-term rental data set in Beijing and uses the automatic machine learning method AutoGluon model to predict the price range, and finally analyzes the solvability of the model. First, data preprocessing is performed on the initial dataset. Second, a preliminary exploration of the data was conducted to find that prices were correlated with housing type, geographic location, and surrounding environment. Then, based on the existing features, the AutoGluon hedonic price model is established to predict the price range. The AutoGluon model has built-in mainstream models, and the best model can be compared and selected. Finally, the

Table 2: Performance evaluation of each model

Model	Score	Pred_time	Stack_level
XGBoost_BAG_L2	0.771	7.052	2
WeightedEnsemble_L3	0.771	7.053	3
LightGBMXT_BAG_L2	0.769	6.878	2
LightGBM_BAG_L2	0.769	6.897	2
LightGBMLarge_BAG_L2	0.769	6.951	2
CatBoost_BAG_L2	0.769	6.817	2
WeightedEnsemble_L2	0.768	4.761	2
NeuralNetFastAI_BAG_L2	0.768	7.276	2
ExtraTreesGini_BAG_L2	0.767	7.564	2
ExtraTreesEntr_BAG_L2	0.767	7.591	2

Table 3: Interpretability of the model

Feature	Importance	Standard Deviation	P Value
longitude	0.059	0.003	0.000
dimension	0.112	0.023	0.007
housing type	0.088	0.011	0.003
Number of Homeowners' Listings	0.057	0.014	0.010
Affiliated Community	0.027	0.006	0.007
Number of rental days per year	0.023	0.010	0.030
Percentage of comments per month	0.017	0.010	0.048
Minimum rental days	0.007	0.001	0.005
number of comments	0.002	0.006	0.286

interpretable analysis of the model is carried out to identify the key factors. Geographical location and housing type are undoubtedly the biggest influences on online short-term rental market prices. In the future, more price features and samples of online business data can be combined to allow the model to learn more information; in addition, more detailed feature engineering can be performed before the model is introduced to reduce data noise; finally, automated machine learning methods are the focus of future research, AutoGluon taken in this paper is one of them, and more detailed parameter settings can also be performed. I hope the research in this paper can improve the reference for the development of the domestic online short-term rental market.

ACKNOWLEDGMENTS

This work was financially supported by Shanghai Philosophy and Social Science Planning General Project (No. 2020BGL025).

REFERENCES

[1] J. Wu and M. Jin, "The influence of online short-term rental listing pictures on consumer behavior intention," Data Analysis and Knowledge Discovery, vol. 1, no. 12, pp. 10-20, 2017.

[2] C. Lu and S. Lin, "Motivation of tourists choosing online short-term rental accommodation," Business and management journal, vol. 40, no. 12, pp. 153-167, 2018.

[3] Y. Liu, "Comparison of Influencing Factors of Online Homestay Short-term Rental Purchase Intention between China and the United States," Economic Geography, vol. 40, no. 01, pp. 234-240, 2020.

[4] C. Ling and Z. Zhang, "Research on the Development Path of "Sharing Economy" in China——Taking Online Short-term Rental as an Example," Modern Management Science, no. 10, pp. 36-38, 2014.

[5] C. Wang, "Feasibility study on the exit method of venture capital IPO in the online short-term rental industry——Taking the "Little Pig Short-term Rental" website as an example," Shandong Social Sciences, no. S1, pp. 255-256, 2016.

[6] L. Song, "Game behavior analysis of online short-term rental economy under different operating modes," Dongyue Tribune, vol. 39, no. 02, pp. 96-104, 2018.

[7] Y. Xu and F. Dai, "Research on the innovation of online short-term rental business model canvas under the sharing economy——Based on the comparative analysis of Xiaozhu short-term rental business model and Tujia short-term rental," Price:Theory & Practice, no. 06, pp. 137-140, 2019.

[8] J. Liang, J. Zhao, P. Zheng, L. Huang, M. Ye, and Z. Dong, "An online short-term rental platform trust computing framework integrating image and text analysis under feature selection " Data Analysis and Knowledge Discovery, vol. 5, no. 02, pp. 129-140, 2021.

[9] F. Liu, X. K. Jiang, and M. Zhang, "Global burden analysis and AutoGluon prediction of accidental carbon monoxide poisoning by Global Burden of Disease Study 2019," (in English), ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH, vol. 29, no. 5, pp. 6911-6928, JAN 2022.

[10] R. Cao, B. Liao, M. Li, and R. Sun, "Online short-term rental market price prediction and feature analysis model based on XGBoost," Data Analysis and Knowledge Discovery, vol. 5, no. 06, pp. 51-65, 2021.