

A study on Station-based Carsharing Service's Price Prediction Methods

Beibut Amirgaliyev, Dana Rakym and Shynar Bolatova

Astana IT University, CI, Mangilik el str, Astana, 010000, Kazakhstan

Abstract

One of the types of carsharing service consumption is station-based carsharing which has no parking lots and management companies. The paper studies station-based carsharing service's price and methods of price prediction for such kind of services. The relevance of the paper is to define the optimal method of predicting car rent price according to the data provided by the carsharing company "Turo" in US. The purpose of the paper is to overview of different methods of price forecasting which can contribute to the increase in profitability of carsharing services by predicting appropriate prices in accordance with key factors affecting the demand for this kind of service.

The research defines characteristics that affect pricing in carsharing services and explores the application of three price prediction methods for the service of car renting & car sharing in order to define an appropriate way of price formulation. To accomplish the research the following steps have been performed: analysis of the price formulation factors in market of car renting, exploring and analysis of data or indicators for better pricing, finding and optimizing the way of indicating price under demand uncertainty. There the methods of theoretical research were used: literature review, observation, market research, statistical analysis.

Keywords ¹

Carsharing, carsharing price, car rent price, car rent budget, price forecasting.

1. Introduction

Carsharing is a type of short-term car rental with a minute or hourly payment, usually used for short trips around the city or the surrounding area. Car sharing harmoniously combines the advantages of a personal car and the absence of the need to spend money on its maintenance, it is more convenient than public transport and cheaper than a taxi. There are several types of carsharing service. One of them is "free-floating", when cars are freely placed around the city without the use of stations. To use car sharing services, the user needs to select the nearest free car on the city map through the mobile application and book it remotely. In this paper station-based carsharing was considered.

Traditionally, station-based carsharing is similar to a regular car rental, but there are no parking lots and management companies. And the service does not own the car. The service allows customers to rent their own car or rent someone else's in a short time: the lessor registers on the site and fills out a questionnaire, indicating the details of his car, and tenants can choose a suitable car for themselves. In the case of mutual agreement, the owner of the car and the lessee draw up a rental agreement between themselves. Advances in technology provide an improvement for carsharing: apps or web-sites are often used to book carsharing services on the go, providing fast checkout, a personalized experience for users, and continuous monitoring and collection, analysis of usage data for businesses (Laoutaris et al, 2014). [1]. This method of carsharing has obvious advantages - the prices are very variable, depending on the brand of the car, its condition and year of manufacture. In addition, the choice of cars is much wider than that of classic carsharing, which are often mono-branded. The prices of this service are a key factor influencing the demand for the service and directly on the company's profit. Implementing trip pricing

Information Technology and Implementation (IT&I-2022), November 30 - December 02, 2022, Kyiv, Ukraine

EMAIL: amirgaliyev@gmail.com (A.1), rakymdana@gmail.com (A.2), bolatova.shynar@gmail.com (A.3)

ORCID: 0000-0003-0355-5856 (A.1),



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

greatly increases profitability. Therefore, it is important to research and identify ways to determine the prices of the car sharing service. To do this, the price of each service must be selected and determined in an optimal way. Research related to pricing service in one-way carsharing systems suggests different solutions. Particularly, they belong to the following categories: strategic, tactical and operational decisions. Strategic decisions are the location, number and capacity of carsharing stations, which should be made over a long period of time. Tactical decisions include fleet size and trip price, which typically are more flexible to modify than strategic decisions. The relocations of vehicles and personnel are an operational decision that can change from day to day according to actual demand. The number of station ride requests varies by hour, day, and week due to inherent demand uncertainty. It is therefore challenging to jointly optimize the long-term strategic/tactical decisions and the real-time operation decisions considering uncertainty in demand (Huang, An, Correia, Rich & Ma, 2021) [2].

The secret to forecasting user demand is having a thorough grasp of the travel preferences of carsharing customers. Scholars have carried out a number of research for the examination of user travel characteristics. Kang et al. used transaction data from a carsharing operator in Seoul, South Korea, to perform multiple linear regression modeling on factors that influence users' travel behavior. They used the quantity of carsharing transactions as a dependent variable and three groups of independent variables: environment, demographics, and transportation variables [3].

Sioui et al. [4], Luca et al. [5], and Efthymiou et al. [6] used questionnaires to collect statistical data on users' travel patterns and discovered that various variables, including age, gender, and income, had an impact on users' travel choices. Tian Lijun and colleagues developed a dynamic model of individual route choice in order to investigate the impact of personal preference on the evolution of route flow [7]. The introduction of the user's preferences allowed for continuous model rule updating. The outcome demonstrated that psychological elements, both subjective and objective, will influence travelers' choice of route. (Wang, C., Bi, J., Sai, Q., & Yuan, Z., 2021) [8].

2. Related work

In one-way station-based carsharing services there is a problem of instability in the demand for car rental services depending on time and place. Inappropriate prices for services can lead to losses and instability in the company's income. The main way to remedy this situation is to individually assign prices charged to customers, depending on the indicators affecting the demand for a particular service, such as the number of available vehicles, time, weather conditions and other factors. This method of income adjustment can bring higher profits through the selection of the highest optimal prices for services. In the paper of Jorge, Molnar and Correia (2015) [9] there was proposed a model that considers demand as a function of price and searches for the prices that maximize the profit of the daily operation of a one-way carsharing company.

Since the model is non-linear and the objective function is non-concave, they used ILS as a metaheuristic to solve the problem. For setting the prices, stations were grouped into zones and time was divided into time intervals. Therefore, trip prices varied between each OD pair of zones according to the time interval in which the trip begins. In this article (Jorge et al., 2015) [9] there was developed a mixed-integer nonlinear programming method, that determines which prices to select and charge for a given period of time to maximize profits given all-day trips and price elasticity of demand. Authors of the article (Jorge et al., 2015) [9] demonstrated that system balancing has a very important role in reducing the costs and increasing profitability of carsharing systems. Their perfectly balanced solution has a higher profit than the imbalanced one.

The paper of Huang et al. (2021) [2] determines a novel strategy that combines long-term prices, real-time relocations and access trips for the demand-supply imbalance problem under demand uncertainty. The carsharing vehicle fleet size and the price of trips are optimized in order to maximize the total profits of the operator by anticipating the vehicle relocations and access trips via walking or biking in the operational planning under uncertain demand. A two-stage stochastic programming is formulated on the basis of a service rate. A gradient search algorithm, a genetic algorithm and an iterated local search algorithm are proposed as a means to solve the program. To reduce computation time, parallel computing is used for solving different demand scenarios. A case study is conducted to demonstrate the applicability of the algorithms and to generate insights with respect to the management of one-way carsharing systems. The case study is based on a real traffic network and randomly

generated demand formations from Poisson distributions. Huang, Kai & An, Kun & Homem de Almeida Correia, Gonçalo & Rich, J. & Ma, Wanjing. (2021) [2]. According to the results of optimization in the article “An innovative approach to solve the carsharing demand-supply imbalance problem under demand uncertainty” which shows that the price of carsharing plays an important role in solving the problem of long-term imbalance. A higher price reduces demand at high demand stations during peak hours while keeping the system profitable. Taking into account the operational decisions for a typical day, moving vehicles in real time to virtual movement zones can reduce the problem of imbalance. For the entire car sharing system, using a pricing strategy, performing real-time vehicle movements and allowing access roads, more than 84% of all carsharing requests can be served.

Giorgione, Ciari & Viti (2019) [10] in their paper described the steps of analysis by examining the properties of various demand-side pricing policies and what trends observe. To meet the goal of the paper (Giorgione et al., 2019) [10], there were introduced two new attributes: income and Value of Time (VOT). The introduction of this variable influences the reaction of users to pricing schemes and helps to better model the behavior of different groups of users. And concluded that carsharing users with an average VOT tends to take resources from users with a lower VOT that will migrate to other means of transportation while pricing only slightly affects high VOT users.

3. Methods

This study uses quantitative techniques to perform exploratory data analysis. The method of exploratory data analysis includes the data containing the information analyzed to answer the research question, data visualization and supervised algorithms such as Random Forest, XGBoost and Decision Tree for price forecasting of the carsharing service.

3.1 XGBoost Algorithm

This machine learning algorithm XGBoost is based on a decision tree and uses a gradient boosting framework. The gradient boosting method creates new models that perform the task of predicting the errors and residuals of all previous models, which are then summed in turn and then the final prediction is made.

3.2 Decision Tree Algorithm

The tree algorithm, as the name says, has a tree structure and functions in a tree-like manner, like a stem grows, starting from the collection of information from the root node and ending with the decision made by the leaves.

3.3 Random-Forest Algorithm

In this study, in order to get higher accuracy during prediction, prediction with random forest algorithm is applied. The random forest algorithm, which consists of a large number of individual trees in the form of decisions, is well suited for research where you need to make a choice from a large amount of data. It is also convenient in cases where the sequence of performing a finite number of actions necessary to find a solution is unknown. Each individual tree in the random forest produces a class prediction, and the class with the most votes become our model's prediction. The grid search approach is often too costly as many combinations are tested. In these cases, it is easier to use randomized search, which evaluates only a user-specified number of random combinations for each hyperparameter at each iteration. Thus, this algorithm makes it possible to test more hyperparameters, which makes it possible to see a more complete and accurate picture.

4. Data acquisition

In this article, algorithms for analyzing and predicting prices for car sharing services were carried out using car rental price data from Turo, the best existing peer-to-peer car rental and car sharing service

based in San Francisco, USA. The data was taken from the kaggle platform from the profile of Christopher Lambert and contains information about 8475 vehicles in the car rental industry in a JSON file. There was also a lack of publicly available data on carsharing, which made it impossible to make a comparative analysis. The first step was to collect data suitable for analysis. They were originally not prepared for training. Manipulations were performed to transform the raw dataset into a more usable one. In addition, steps were taken to reduce duplicate data, and a certain amount of data was grouped together to show consistency and increase versatility.

The dataset has 19 main columns and its subcolumns: "distance", "reviewCount", "businessClass", "renterTripsTaken", "rating", "images", "deliveryLabel", "distanceLabel", "owner", "rentableFromSearchedAirport", "rate", "distanceWithUnit", "freeDeliveryPromotion", "location", "responseTime", "vehicle", "newListing", "responseRate", "instantBookDisplayed".

Here are descriptions of the main useful datapoints in the dataset:

- Response Rate: The Response Rate is the number of times on average a host replied to a rental request (in percent). In case no data was present, it was imputed with a 0 percent.
- Make, Model, Year: Represents the Make of the Car, Model of the Car, and the year in which the car was released.
- Renter Trips Taken: This represents the number of times the car has been rented from that particular renter.
- Scalar: The distance driven during the rental.
- Instant Book Displayed: Whether the car could be booked without any request processing time to the renter.
- Average Daily Price: The price for each particular car rental. This is our response variable.

5. Results

After conducting an exploratory data analysis, we tried to determine whether there is a statistically significant relationship between the variables in the dataset and the average daily price. As a result of studying the relationship of attributes in the dataset, we determined the correlation of data with the average price (avgDailyPrice) of the service (Figure 1).

When it comes to prediction using machine learning, there are no one-size-fits-all solutions. As data scientists, we consider a variety of heavy-duty algorithms to identify the best ones. But this is not enough. It is necessary to tune the hyperparameter selection algorithm. Therefore, after preparing the data for testing and choosing a learning model, we use three supervised learning algorithms of the prediction algorithm were applied: Random Forest, XGBoost and Decision Tree algorithms.

For regression issues, three standard evaluation metrics were used:

Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{N} \cdot (Y_i - \hat{Y}_i)^2$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}}$$

or

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

Because MAE is the average error, it is the most straightforward of these metrics to comprehend. Since MSE "punishes" larger errors, which tends to be useful in the real world, it is more widely used than MAE. Due to its ability to be interpreted in "y" units, RMSE is even more widely used than MSE. These are all loss functions because minimizing them is what we want to do. As a result, there were evaluated following indicators of evaluation metrics (Table 1).

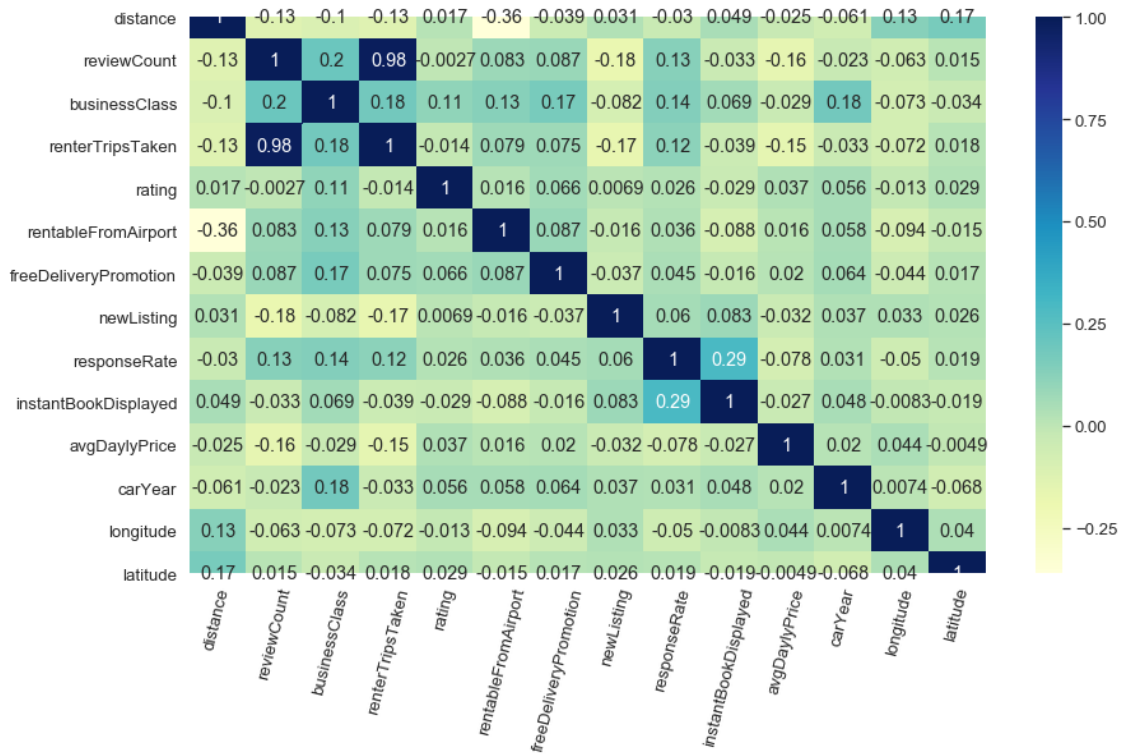


Figure 1: Correlation Matrix with key values

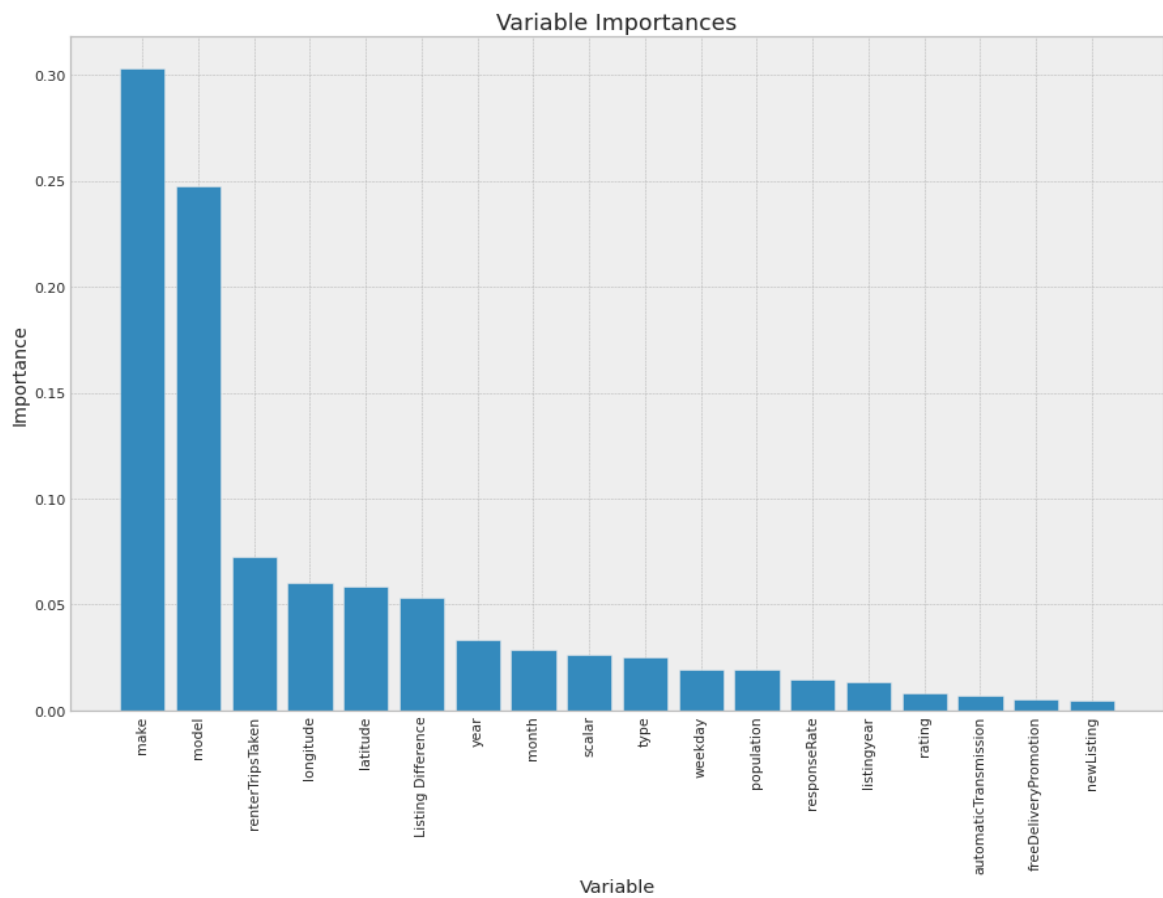


Figure 2: Prioritized key parameters of correlation

Table 1
Results of evaluation metrics for regression problems

	Model	RMSE_mean	RMSE_std	MAE_mean	MAE_std	r2_mean	r2_std
0	XGBoost	66.086674	381.914817	34.457647	0.683037	0.653628	0.015805
1	Random Forest	66.958574	441.365148	33.020088	0.742532	0.646644	0.029640
2	Decision Tree	93.374384	1049.653586	40.614379	1.473933	0.305412	0.099051

According to the mean error indicators it seems that decision tree model gives the least accurate result, since its error is the lowest. Moreover, at this point we are only evaluating our models on the train set, so the risk of overfitting for decision trees is quite high. To understand better the performances of the models there can be used an alternative strategy: cross-validation.

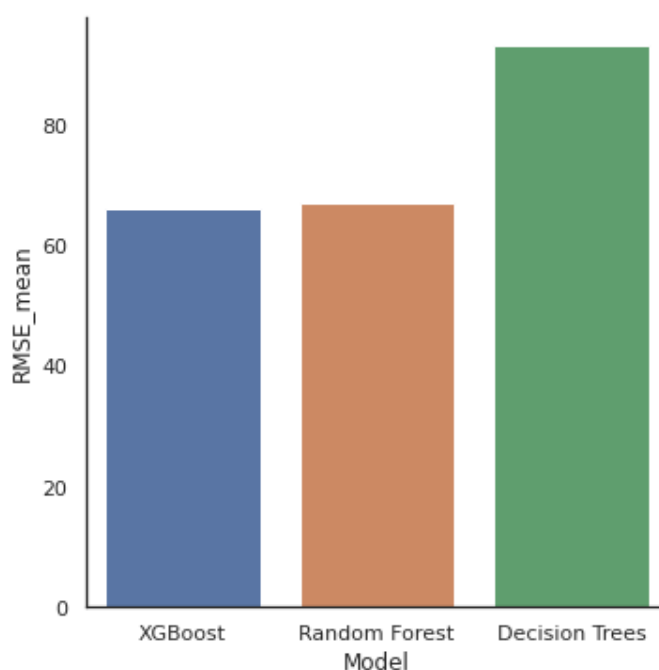


Figure 3: Comparison of RMSE mean by regression methods

6. Conclusion

During the research made on exploring the effectiveness of cost prediction algorithms of the carsharing system was found out that among three considered algorithms: XGBoost, random forest and decision tree algorithms – XGBoost and random forest showed significantly less error level and appeared to be more effective than decision trees. In the future we would like to make more automation: write a couple of functions to speed up the pre-processing part; build a pipeline to automate the preprocessing transformations and use more models: we would like to try SVM, NN and MLP.

7. Acknowledgement

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. BR10965311 "Development of the intelligent information and telecommunication systems for municipal infrastructure: transport, environment, energy and data analytics in the concept of Smart City").

8. References

- [1] Cici, B., Markopoulou, A., Frias-Martinez, E., & Laoutaris, N. (2014). Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 201–211).
- [2] Huang, K., An, K., de Almeida Correia, G. H., Rich, J., & Ma, W. (2021). An innovative approach to solve the carsharing demand-supply imbalance problem under demand uncertainty. *Transportation Research Part C: Emerging Technologies*, 132, 103369.
- [3] Kang, J., Hwang, K., & Park, S. (2016). Finding factors that influence carsharing usage: Case study in seoul. *Sustainability*, 8(8), 709.
- [4] Sioui, L., Morency, C., & Trépanier, M. (2013). How carsharing affects the travel behavior of households: a case study of Montréal, Canada. *International journal of sustainable transportation*, 7(1), 52-69.
- [5] De Luca, S., & Di Pace, R. (2015). Modelling users' behaviour in inter-urban carsharing program: A stated preference approach. *Transportation research part A: policy and practice*, 71, 59-76.
- [6] Efthymiou, D., & Antoniou, C. (2016). Modeling the propensity to join carsharing using hybrid choice models and mixed survey data. *Transport Policy*, 51, 143-149.
- [7] Tian, L., Jiang, X., Liu, T., & Zhao, Y. (2016). Study on daily travel behavior considering path preference based on Dogit model. *Transp. Syst. Eng. Inf*, 16, 228-235.
- [8] Wang, C., Bi, J., Sai, Q., & Yuan, Z. (2021). Analysis and prediction of carsharing demand based on data mining methods. *Algorithms*, 14(6), 179.
- [9] Jorge, D., Molnar, G., & de Almeida Correia, G. H. (2015). Trip pricing of one-way station-based carsharing networks with zone and time of day price variations. *Transportation Research Part B: Methodological*, 81, 461-482.
- [10] Giorgione, G., Ciari, F., & Viti, F. (2019). Availability-based dynamic pricing on a round-trip carsharing service: an explorative analysis using agent-based simulation. *Procedia Computer Science*, 151, 248-255.