HiLL PUBLISHING

Advances in Computer and Communications, 2023, 4(6), 358-362
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Research on Hotel Reservation Scheme Based on Random Forest Model Prediction

Junchao Zhang, Dong Li, Huiling Lan, Lei Tang, Mengmeng Guo*

Guilin University of Technology at Nanning, Chongzuo, Guangxi, China.

How to cite this paper: Junchao Zhang, Dong Li, Huiling Lan, Lei Tang, Mengmeng Guo. (2023) Research on Hotel Reservation Scheme Based on Random Forest Model Prediction. *Advances in Computer and Communication*, 4(6), 358-362.
DOI: 10.26855/acc.2023.12.003

Received: November 25, 2023
Accepted: December 22, 2023
Published: January 17, 2024

***Corresponding author:** Mengmeng Guo, Guilin University of Technology at Nanning, Chongzuo, Guangxi, China.

Abstract

The X Hotel in a certain city is mainly used for conferences and tourists. Guest rooms are reserved by telephone or the Internet. This kind of reservation has great uncertainty, and customers are likely to cancel the reservation due to various reasons. The May Day holiday is approaching. In order to strive for greater profits, Hotel X must win customers on the one hand, and reduce the losses suffered by customers who cancel reservations on the other hand. To this end, Hotel X has adopted a number of measures. First of all, the guest room is required to prepay the first day's rent as a deposit. If the customer cancels the reservation before noon of the previous day, the deposit will be refunded in full, otherwise, the deposit will be forfeited. Secondly, Hotel X adopts variable prices and adjusts prices according to market demand. Generally speaking, the prices in peak tourist seasons are relatively high, and the prices in off-seasons are slightly lower. Then use the random forest model to predict the booking demand in a specific period in the future, which will help the hotel to adjust the room rate reasonably. Establishing a reservation cancellation prediction model to predict the probability of a customer canceling a reservation under certain conditions will help the hotel adjust the reservation limit and cancellation policy.

Keywords

Random Forest, Hotel Reservation, Mathematical Modeling

1. Introduction

With the development of the social economy, people's quality of life is constantly improving, the number of tourists is increasing, and the demand for hotel rooms is also increasing. The hospitality industry is a highly competitive and challenging industry, and the accuracy and efficiency of hotel booking solutions are critical to providing quality customer service and maximizing revenue. With the rapid development of big data and machine learning technology, the prediction method based on a random forest model has become the focus of attention in the field of hotel reservation research. The random forest model has many advantages, such as good generalization ability, strong interpretability, and ability to handle high-dimensional data. So, it is widely used in the prediction and decision-making of hotel reservation programs [1].

Statistics show that the normal demand for domestic hotels is 3.295 billion room nights, and after the epidemic stabilizes, the demand has reached 3.45 billion room times. The demand for guest rooms during holidays is even more in short supply. Hotel room reservations are mainly made by phone or major APPs, which brings convenience but also brings a lot of uncertainty to the hotel. Some guests may cancel the reservation, which directly leads to the loss of hotel profits. In order to maximize profits, hotels usually require customers to prepay the first day's rent as a deposit to reduce losses. Major hotels usually adjust their prices according to market demand. During the peak tourist season, the room rate is high, while in the off-season, the room rate is low [2]. In the peak tourist season, the number

of rooms that exceed the actual number of rooms will be booked to maximize profits. However, this method may It will lead to disputes caused by exceeding the number of rooms and the need to bear the corresponding reputation risk. In order to better accurately judge the market demand, grasp the market trend in time, establish a reasonable room price and plan, reduce disputes caused by exceeding the number of rooms, and maximize the interests of the hotel, it is established by establishing a mathematical model study in depth.

This paper aims to explore future research on predicting hotel reservation scenarios based on random forest models. First, we introduce the importance and challenges of hotel reservation scenarios, and the limitations of traditional methods when faced with complex problems. Then, we will introduce the principle and characteristics of the random forest model in detail, and explain its advantages in the prediction of hotel reservation plans. Next, we discuss several directions for future research, including dataset expansion, feature engineering improvement, model optimization, interpretability and explainability, scheduled policy optimization, and real-time prediction and dynamic adjustment. Through an in-depth exploration of these aspects, we are expected to further improve the accuracy and practicability of hotel reservation plan prediction based on the random forest model [3].

The findings of this study have important implications for hotel business decision-makers and managers. Accurately predicting hotel reservation plans can help hotels optimize resource utilization, make preparations in advance, and improve customer satisfaction and business benefits. Through the prediction method based on the random forest model, the hotel can better respond to market changes and fluctuations in customer demand, and formulate more targeted reservation strategies. Therefore, the results of this study will provide strong support for the development and competition of the hotel industry.

2. Model establishment and solution

It is clearly known that the ideal situation is that all the people who booked online will check in normally + the number of people who did not check in directly through online booking is exactly equal to the number of rooms, but this situation is basically impossible, and the probability is very, very high Low. The most basic reason for the uncertainty is: that Hotel X cannot guarantee how many people will check in directly without booking every day. People who book in advance may not actually check in, which may cause vacant rooms or insufficient rooms. Compensation or upgrades are required for the rooms, which are very bad for a hotel. Therefore, it is necessary to calculate the probability of reservation cancellation and conduct prediction research by building a model [4].

2.1 The establishment and prediction of random forest cancellation reservation prediction model

Using web crawler technology, the historical data of presidential suites, luxury suites, and standard rooms canceled by customers are collected, that is, the number of days between Y's booking and arrival x1, the year of the arrival date, the month of the arrival date x2 [5], the number of weeks of the arrival date, and the arrival date Date, number of weekend stays (Saturday or Sunday) that the guest stayed or booked to stay at the hotel, number of nights of the week (Monday to Friday) that the guest stayed or booked to stay at the hotel, indicating whether the booking name is from a repeat guest (1) or not (0), the number of previous reservations canceled by the customer before the current booking, the number of previous reservations that the customer did not cancel before the current booking, the reservation room type code [6], the room type assigned to the reservation is the training input, and the training output is the reservation canceled (1) is not canceled (0). Use random forests to build a predictive model to predict the probability of a customer canceling a reservation under certain conditions. This will help the hotel adjust booking limits and cancellation policies.

2.2 Solving the Random Forest Unbooking Prediction Model

In the process of random forest parameter tuning, select estimators, that is, the largest number of decision trees in the random forest; max depth is used to limit the maximum depth of the decision tree; "max samples" defines the samples that will be drawn from the training set of each base estimator Number [7], max features This parameter controls the maximum number of features selected.

When using the default parameters, the effects of the three types of suites on the test set are as follows:

Presidential Suite: MSE=13.409795, RMSE=4.838367, MAE=3.385053, R2=0.49925

Deluxe Suite: MSE=10.409795, RMSE=2.834367, MAE=5.385053, R2=0.679725

Standard room: MSE=18.409795, RMSE=6.876667, MAE=5.885053, R2=0.57825

The optimal parameters obtained after Bayesian optimization are:

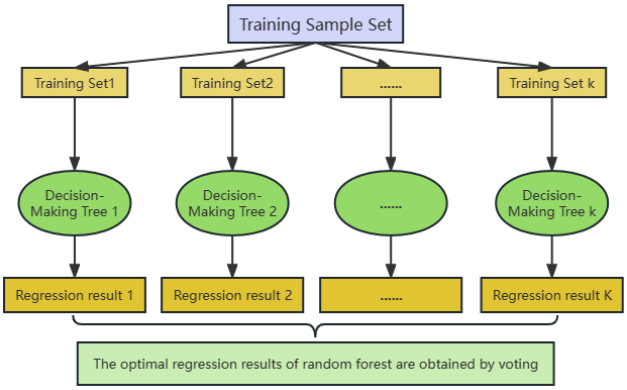

Figure 1. Schematic diagram of random forest algorithm.

Table 1. Random forest optimal parameter

$max\_samples$	$max\_features$	$n\_estimators$	max_depth
presidential suite			
1	0.7	242	70
Luxury Suite			
4	0.9	242	70
Standard Suite			
4	0.9	242	70

After optimization, the effects of the three types of suites have been significantly improved, as shown below:
Presidential Suite: MSE=1.190021, RMSE=1.033321, MAE=0.633452, R2=0.819925
Deluxe Suite: MSE=2.161021, RMSE=1.09221, MAE=0.533444, R2=0.856374
Standard room: MSE=3.140221, RMSE=1.093381, MAE=0.733552, R2=0.883554

3. Conclusion

Use the trained Random Forest of presidential suites, luxury suites, and standard rooms, and use the Python built-in probe ( ) function to obtain the cancellation probability of presidential suites, luxury suites, and standard rooms, and use the obtained cancellation probabilities a1, a2, a3, That is, the presidential suite requires at least (1-a1)*20 sets [8], the luxury suite requires at least (1-a2)*100 sets and the standard room requires at least (1-a3)*500 sets. For example, on March 1, 2023, the president the number of canceled reservations for suites was 13.7%, the probability of canceled reservations for deluxe suites was 8.33%, and the probability of canceled reservations for standard rooms was 5.14%. That is, at least 17 presidential suites are required, 91 deluxe suites are required, and 473 standard rooms are required. At least prepare the above data. For example, on May 1, 2023 [9], the number of canceled reservations for the presidential suite is -11.7%, the probability of canceled reservations for luxury suites is -13.53%, and the probability of canceled reservations for standard rooms is -20.14%. That is, the presidential suite requires at least 24 There are at least 114 luxury suites and at least 601 standard rooms. The hotel must at least prepare the above data, otherwise it will have to upgrade the room level or pay compensation to settle the dispute. For this reason, Hotel X has to bear the reputation risk. When the probability is negative, it means that the hotel will face compensation or upgrade suites.

In the peak tourist season when the demand is high, for Hotel X, the proposed strategy provides the number of reservations for presidential suites from 20 to 31, deluxe suites from 50 to 82, and standard rooms from 500 to 705. Exceeding this range can lead to compensation or suite upgrades, which will have an impact on the hotel's reputation.

4. Discussion

Future research on predicting hotel reservation plans based on the random forest model can be carried out from the following aspects:

Data set expansion: In future research, more abundant data can be considered, including but not limited to hotel historical booking data, customer evaluation and feedback, marketing activities, etc. By increasing the size and diversity of the dataset, the accuracy and generalization of the model can be improved [10].

Feature Engineering Improvements: Random forest models are critical to the selection and processing of features. Future research can explore more effective feature engineering methods, such as feature selection, feature combination, and feature dimensionality reduction, to improve the performance and efficiency of the model.

Model optimization: The performance of random forest models can be optimized through parameter tuning and improvement of ensemble strategies. Future research can explore more efficient parameter tuning methods, such as grid search, random search, and Bayesian optimization, etc., to find the best combination of parameters. In addition, you can try to introduce other integration methods, such as gradient boosting trees (Gradient Boosting Trees) or stacking models (Stacking), to further improve the prediction performance [11].

Interpretation and interpretability: The random forest model has advantages in terms of interpretation and interpretability, and can explain the prediction results through methods such as feature importance evaluation and visualization. Future research can further delve into the interpretability of the random forest model and explore how to improve the interpretability of the model to meet the needs of users for understanding and trusting the prediction results [12].

Predetermined policy optimization: In addition to using random forest models for prediction, future research could explore how to translate predictions into actual predetermined policies. The best reservation strategy can be formulated through optimization algorithms and decision rules, taking into account factors such as hotel resource utilization, revenue maximization, and customer satisfaction.

Real-time prediction and dynamic adjustment: The random forest model is usually used for offline batch prediction, and future research can try to apply it to real-time prediction and dynamic adjustment scenarios. It can be explored how to integrate the model with the real-time data stream to realize the update and adjustment of the real-time booking plan and improve the real-time accuracy of the booking decision.

In short, future research can explore in depth in terms of data set expansion, feature engineering improvement, model optimization, interpretability and interpretability, reservation strategy optimization, real-time prediction, and dynamic adjustment, etc., to further improve the hotel reservation scheme based on the random forest model Predictive performance and availability.

References

[1] Lan, H. & Pan, Y. (2019, June). A crowdsourcing quality prediction model based on random forests. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, (pp. 315-319). IEEE.

[2] Subramanian, R. R., Reddy, M. P., Kousik, K., Rupesh, S., Rohith, S., & Kumar, B. S. (2022, December). ClassHotel: Application of data analytic techniques for online hotel recommendation. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, (pp. 1021-1026). IEEE.

[3] Gómez-Ramírez, J., Ávila-Villanueva, M., & Fernández-Blázquez, M. Á. (2020). Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. *Scientific Reports*, *10*(1), 1-15.

[4] Thomas, E., Ferrer, A. G., Lardeux, B., Boudia, M., Haas-Frangii, C., & Agost, R. A. (2019). Cascaded machine learning model for efficient hotel recommendations from air travel bookings. In *Proceedings of Proceedings of the 12th ACM Conference on Recommender Systems, ACM RecSys Workshop on Recommenders in Tourism (RecTour 2019 vol 2435). CEUR Workshop Proceedings Copenhagen* (pp. 9-16).

[5] Dai, P., Chang, W., Xin, Z., Cheng, H., Ouyang, W., & Luo, A. (2021). A retrospective study on the influencing factors and prediction of hospitalization expenses for chronic renal failure in China based on random forest and LASSO regression. *Frontiers in Public Health*, *9*, 678276.

[6] Lee, M., Kwon, W., & Back, K. J. (2021). Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. *International Journal of Contemporary Hospitality Management*, *33*(6), 2117-2136.

[7] Tekin, A. T., & Cebi, F. (2020). Click and sales prediction for digital advertisements: Real world application for otas. In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making: Proceedings of the INFUS 2019 Conference, Istanbul, Turkey, July 23-25, 2019* (pp. 205-212). Springer International Publishing.

[8] El Hamdaoui, H., Boujraf, S., El Houda Chaoui, N., Alami, B., & Maaroufi, M. (2021). Improving heart disease prediction using random forest and adaboost algorithms. *International Journal of Online & Biomedical Engineering*, *17*(11).

[9] Wang, Y., Huang, C., Zhao, M., Hou, J., Zhang, Y., & Gu, J. (2020). Mapping the population density in mainland China using NPP/VIIRS and points-of-interest data based on a random forests model. *Remote Sensing*, *12*(21), 3645.

[10] Abdulkareem, K. H., Mohammed, M. A., Salim, A., Arif, M., Geman, O., Gupta, D., & Khanna, A. (2021). Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment. *IEEE Internet of Things Journal*, *8*(21), 15919-15928.

[11] Cakmak, T., Tekin, A., Senel, C., Coban, T., Uran, Z. E., & Sakar, C. O. (2019). Accurate prediction of advertisement clicks based on impression and click-through rate using extreme gradient boosting. In *ICPRAM* (pp. 621-629).

[12] Avdeef, A., & Kansy, M. (2022). Predicting solubility of newly approved drugs (2016–2020) with a simple ABSOLV and GSE (Flexible-Acceptor) consensus model outperforming random forest regression. *Journal of Solution Chemistry*, *51*(9), 1020-1055.