Comparison and Analysis of Machine Learning Models to Predict Hotel Booking Cancellation

Yiying Chen[1] Chuhan Ding[2] Hanjie Ye[3] Yuchen Zhou[4,*]

[1] *Management School, Beijing Jiaotong University, Beijing, China*
[2] *Thurgood Marshall College, University of California San Diego, U. S. A.*
[3] *School of Social Science, University of California Irvine, U. S. A.*
[4] *School of Economics, Huazhong University of Science and Technology, Wuhan, China*
[*] *Corresponding author. Email: u201916417@hust.edu.cn*

ABSTRACT
Hotel booking cancellation prediction is crucial in conducting revenue and resource management for hotels. This paper provides three possible substitutes for the neural network including logistic regression, *k*-Nearest Neighbor (*k*-NN), and CatBoost, whereas CatBoost, is the most suitable model for hotels to do the prediction. The advantages of them are effectiveness, high accuracy, and lower cost. The dataset used in this paper was adapted from Kaggle, a set of the booking data from two types of hotels (resort hotel and city hotel) in Portugal, and the corresponding customers' information. We select some key variables as the predictor to train and test the prediction models based on three machine learning algorithms. After preprocessing the raw data, i.e., standardizing, dealing with missing data, recoding some variables, and scaling, we conduct the prediction and compare each model through three metrics (confusion matrix, accuracy score, and F_1-score). The result indicates that CatBoost has the best performance in predicting hotel booking cancellation because it has the greatest number of correct prediction samples and the highest accuracy score. We focus on the efficiency and economy of doing cancellation prediction in the hospitality industry to form a basis for future revenue and resource management for hotels.

Keywords: *CatBoost, hospitality industry, machine learning models, prediction.*

1. INTRODUCTION

In the hospitality industry, hotels prepare their rooms and resources according to room reservations. Advanced booking from customers forms a contrast between customers and hotel which ensures a stable price for customers to purchase and enjoy the service [1], but at the same time, it can increase the cost of revenue and resource management for the hotels, and increase the risk of the opportunity costs of the customers [2]. To do revenue management, in other words, to bring customers a proper product with a proper price at the right time [3], hotels need to respond quickly to booking cancellation, thus, booking cancellation is an important aspect to consider when making decisions [4]. Hotel management including managing service level and revenue is affected by the net demand, the difference between exact booking and booking cancelation of a particular hotel [5]. To help the hospitality industry better manage revenue, forecasting is considered one of the most crucial tools [6]. It is vital to guarantee an accurate forecast to conduct a proper recommendation for revenue management [1].

Hotel booking cancellation prediction is a binary classification problem as it only has two outcomes, will or will not cancel. Literature for using machine learning models to do this prediction are not much till now. One of the most popular methods to do this prediction is the neural network. However, different from using traditional machine learning models, conducting neural network to solve the prediction problem has some limitations that need to be considered. For example, the most widely known shortcoming of the neural network, 'black box', limits the ability for observers to find out the causal relationship between input and output [7]. In this specific hotel booking prediction, it is hard for hotels to explain their different revenue management strategies to each individual simply based on the outcome of a model, and it may probably lead to customer dissatisfaction. Furthermore, compared to traditional machine learning models, the cost of using neural network for predictions is much higher [8]. The cost includes both the time and financial investment spent on developing. Neural network may also take weeks to finish the training process as it usually requires a large dataset.

To enhance the prediction performance and find an effective model in this type of prediction, we conduct three machine learning models including logistic regression, *k*-NN, and CatBoost. These three algorithms require less financial cost, time, and fewer samples to do the prediction [7], [9], [10]. More specifically, logistic regression has the advantages of high efficiency since its calculating process is only related to the number of features when doing classification [7]. Therefore, it is suitable for the hospitality industry. Since *k*-NN is simple, it is easy for hotel managers or decision-makers to understand and conduct [9]. Compared to other algorithms, it is not sensitive to outliers that its outcome will not be largely affected, specifically, it normally performs better than Support Vector Machine (SVM) [9]. CatBoost is an advanced and effective machine learning algorithm of gradient boosting with binary decision trees. It can automatically deal with categorical features and can avoid overfitting during the process of machine learning and improve accuracy [10].

In our paper, we use three machine learning models including logistic regression, *k*-NN, and CatBoost to do prediction with higher accuracy, less time, and lower cost. The raw booking data of 2 types of hotels from 2015 to 2017 are adapted from Kaggle. The dataset includes customers' information. We select some key features from the customers' information as our predictor to predict whether a customer will or will not cancel the reservation. By comparing the confusion matrix, accuracy score, and F_1-score of each model, we come up with the conclusion that CatBoost is the most suitable model for this particular case.

Different from the previous study, we focus on the efficiency and economy of doing cancellation prediction in the hospitality industry to form a basis for future revenue and resource management for hotels. We compare the performance of each model not only from a theoretical aspect but also from the practical one. Furthermore, we select an advanced boosting model, CatBoost, as one of our prediction algorithms because of its particular advantages that it can avoid overfitting problems effectively and can tackle categorical features without regular human intervene.

The following part of this paper will be divided into four parts. In Data and Variables, we briefly describe the raw data, the process of data preprocessing, followed by the descriptive analysis of the processed dataset. In Methodology, we introduce the rationales of the three models separately, and three metrics that would be used to further evaluate the performance of our models. In Results and Discussion, we demonstrate how each model performs and do the comparison based on the metrics. Finally, we conclude that the most suitable model for the classification problem in hospitality industry is CatBoost.

2. DATA AND VARIABLES

2.1. Data Description

The dataset in this program is collected from hotel booking demand datasets provided by Antonio et al. [11]. It comprehends the bookings of two hotels (one is a resort hotel, the other is a city hotel) in Portugal. All reservations' arrival time is between 1st of July in 2015 to 31st of August in 2017. There are 119391 samples and 32 features in the dataset. 14 features are categorical features and the others are numerical features. Table 1 illustrates 13 key variables.

Table 1. Key variables

Name	Description
hotel	Hotel type: resort hotel or city hotel
is_canceled	A booking was canceled (1) or not (0)
lead_time	Number of days that elapsed between the date of the booking and the arrival
is_repeated_guest	Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_ cancellations	Number of previous cancelled bookings by the customer before the current booking
previous_bookings_ not_canceled	Number of previous bookings not cancelled by the customer before the current booking
booking_changes	Number of changes made to the booking from the booking to check-in or cancellation by customers
days_in_waiting_list	Number of days the booking was on the waiting list before it was confirmed to the customer
customer_type	Type of booking
required_car_ parking_spaces	Number of car parking spaces required by the customer
total_of_special_ requests	Number of special requests made by the customer
reservation_status	Reservation last status
reservation_status_ date	Date at which the last status was set

2.2. Data Preprocessing

To deal with missing data, inconsistent data, outliers, etc., we preprocessed the raw data to eliminate these problems. We convert the raw data to a standardized efficient dataset so that we can run models based on it. The sklearn.preprocessing package in Python helps us to preprocess data and get a standardized dataset.

Some of the rows in our data set missed the values on features. These missing data will cause significant effects on machine learning and the data analysis outcomes. Thus, we choose to delete the entire rows to handle the missing data. Moreover, we use the drop function in Python to drop the rows holding '0' as its value.

There are still some features that are not given as continuous values, which means that we need to convert these categorical features into integer codes. Thus, we do the data encoding step. The estimator called OrdinalEncoder allows us to transform every categorical feature into the feature of integers with the format in 0 or 1.

The final step is feature scaling. By using MinMaxScaler, we set the minimum value to 0, and the maximum value to 1 as the range of scaling features. As a result, we contain the robustness of features with a very small standard deviation and hold 0 entries in sparse data.

2.3. Descriptive Statistics

From Figure 1, we can see that the shorter the lead time is, the more people book rooms. A possible reason for this is that most people are likely to book rooms when they are certain to check-in. If the lead time is very long, there would be many uncertainties. For instance, the uncertainty of the weather, or any urgent events, may block the customer from checking in on time and affect his or her original schedule. Therefore, people are more likely to book rooms when it is up to the arrival date. In this way, they can largely avoid schedule change costs.

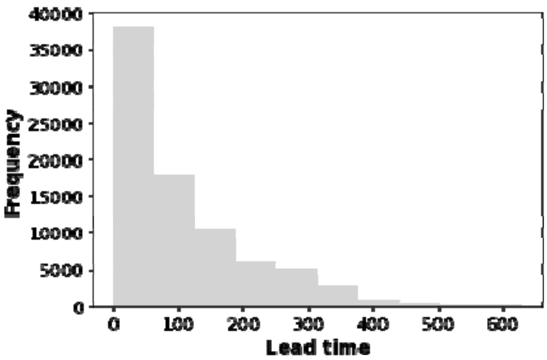

Figure 1 What time do customers books a room.

Figure 2 further confirms the previous assumption of the reason why there is a trend in Figure 1. When the lead time is long, the cancelling rate is 50 percent, a very high percentage. When the lead time is less than 7 days, the cancelling rate reaches 10 percent.

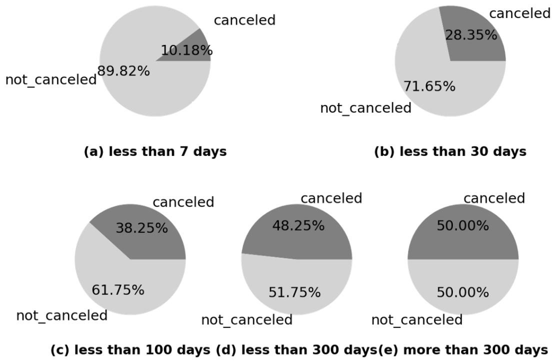

Figure 2 Cancelling rate according to 'Lead_time'.

The reason is that a longer lead time will bring more uncertainties. In this situation, people are more inclined to book hotels when the arrival date is coming even though the price might be higher than booking long before.

Table 2. Correlation coefficient between features and the label (is_canceled)

Feature Name	Correlation Coefficient
lead_time	0.314161
previous_cancellations	0.132844
required_car_parking_spaces	−0.191471
total_of_special_requests	−0.269111

To investigate feature importance, we calculate the correlation coefficient between features and the label. According to Table 2, the result shows that 'lead_time' and 'previous_cancellation' are most positively related with cancellation, which means the possibility of cancellation increases with 'lead time' and the customers who used to cancel their bookings are more likely to cancel their next order. And 'total_of_special_requests' and 'required_car_parking_spaces' are most negatively related to cancellation. Thus, it is reasonable to assume that the reservation with a more special request is less likely to be canceled.

3. METHODOLOGY

3.1. Machine Learning Models

Logistic regression was firstly introduced by Cox in 1958 [12]. It is a widely used statistical model to solve problems with sparse dependent variables. Each value of the dependent variables represents a category, for example, in our case, the customer will ('1') or will not ('0') cancel his or her booking. The essence of logistic regression is the same as linear regression, the former includes a function mapping from the features to labels which can be written as the following equation,

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}} \tag{1}$$

where $h_\theta(x)$ is the label, and $g(z)$ is called logistic function shown below

$$g(z) = \frac{1}{1+e^{-z}} \tag{2}$$

that used to standardize the label in order to make it placed between 0 and 1.

The logistic regression model can be explained by the following equation,

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i \tag{3}$$

where p is the probability of the outcome, i.e., the label, β_0 is the intercept, β_i is the corresponding coefficient, and x_i is the feature.

k-NN was proposed by Cover and Hart in 1968 [13]. It assumes that the category of one sample is the same with k nearest congeneric samples in feature space. *k*-NN classifies the samples by several nearest samples. The only necessary parameter in the model is 'k', which decides how many congeneric samples are going to be invested in order to categorize a new sample. *k*-NN is more suitable for large sample data.

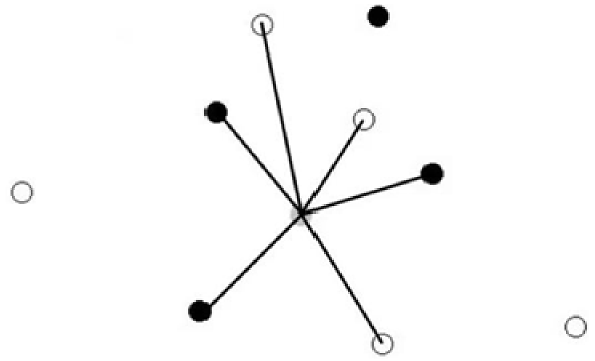

Figure 3 Visible schematic diagram of *k*-NN.

Notes: k =3. The grey point is unclassified. Black points are in class A and white points are in class B. The length of the segments shows the distance between two points.

In Figure 3, the grey point belongs to class A according to *k*-NN, because the total distance between the grey point and the three nearest black points is shorter than that between the grey point and the three nearest white points.

The model can also be represented by the following equation,

$$y = f(x) = \arg\min_c \sum_{x_i \in N_k(x)} \mathrm{II}(y_i = c) \tag{4}$$

where $N_k(x)$ is the set of k points that are closest to the point x.

CatBoost was firstly developed by Prokhorenkova et al. [14] and engineers for prediction and other tasks. It works well with disposing of diverse data types, especially for multiple categorical data. CatBoost is an algorithm of gradient boosting with binary decision trees. It solves the problem of gradient bias and prediction shift, therefore avoiding overfitting during the machine learning process and improving the accuracy. In our project, we use CatBoost classifier to build the model and test its accuracy score.

CatBoost was designed to better dispose of categorical features in GBDT (Gradient Boosting Decision Tree). In each iteration of GDBT, the loss function uses the same data set to get the gradient of the current model and then trains the base learner. However, this leads to gradient biased estimation, the problem of over-fitting the model, and the prediction shift.

Thus, to avoid prediction shift, CatBoost uses the algorithm of ordered boosting shown in Figure 4.

input : $\{(\mathbf{x}_k, y_k)\}_{k=1}^n, I$;
$\sigma \leftarrow$ random permutation of $[1, n]$;
$M_i \leftarrow 0$ for $i = 1..n$;
for $t \leftarrow 1$ **to** I **do**
  for $i \leftarrow 1$ **to** n **do**
    $r_i \leftarrow y_i - M_{\sigma(i)-1}(\mathbf{x}_i)$;
  for $i \leftarrow 1$ **to** n **do**
    $\Delta M \leftarrow$
      $LearnModel((\mathbf{x}_j, r_j) :$
      $\sigma(j) \leq i)$;
    $M_i \leftarrow M_i + \Delta M$;
return M_n

Figure 4 Ordered boosting.

Notes: This figure is adapted from Prokhorenkova et al. [14].

To get an unbiased gradient, CatBoost trains an individual model M_i for each sample x_i . M_i is obtained by training the training set that does not include sample x_i . We use M_i to get gradient prediction of the sample and use the gradient to train the base learner and obtain the final model.

However, with the traditional ordered boosting algorithm, we need to train n different models, which will increase the complexity and memory requirements by n times. Thus, in CatBoost, we modify the algorithm based on a gradient boosting algorithm with decision trees as base predictors.

CatBoost has two boosting modes in the process of building trees. One is a modified ordered boosting algorithm, and the other is the standard GBDT algorithm after the transformation of categorical features by built-in ordered TS (Target Statistic). The running process is listed as follows.

$$grad \leftarrow CalcGrandient(L, M, y), \tag{5}$$

$$r \leftarrow random(1, s), \tag{6}$$

if Mode = Plain, then

$$G \leftarrow (grand_r(i)) \ \ for \ i = 1, \ldots, n, \tag{7}$$

if Mode = Ordered, then

$$G \leftarrow (grand_{r,\sigma_{r(i)-1}}(i)) \ \ for \ i = 1, \ldots, n. \tag{8}$$

3.2. Model Evaluation Metrics

We apply three metrics to evaluate the performance of each model.

The first metric is the confusion matrix. The confusion matrix is a two factorial square matrix, and the meaning of each number is shown in the following Table 3 [15].

Table 3. Definition of elements in the confusion matrix

	Actual Positive Class	Actual Negative Class
Predicted Positive Class	Number of true positive (TP)	Number of false negative (FN)
Predicted Positive Class	Number of false positive (FP)	Number of true negative (TN)

The elements of leading diagonal (TP, TN) are much greater than the other two (FN, FP) in a well-performed model and it means that the model correctly predicts much more samples.

The second metric is the accuracy score ($Accuracy$). The accuracy score measures the fraction of the number of correct predictions of total predicted samples. The formula is shown below [15].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{9}$$

The value of $Accuracy$ is between 0 and 1. The greater the accuracy score is, the better the model performs.

The third metric is F_1-score. F_1-score is the harmonic mean of precision and recall. Precision measures how many samples of a predicted class are correctly predicted. Recall measures how many samples of an actual class are correctly predicted. F_1 of a high-quality model is close to 1. The formulas are shown below [15].

$$Precision = \frac{TP}{TP + FP} \text{ or } \frac{TN}{TN + FN}, \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \text{ or } \frac{TN}{TN + FP}, \tag{11}$$

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}. \tag{12}$$

4. METHODOLOGY

4.1. Logistic Regression Classifier

To analyze the performance of the logistic regression model in predicting hotel booking cancellation, its confusion matrix is presented in Figure 5.

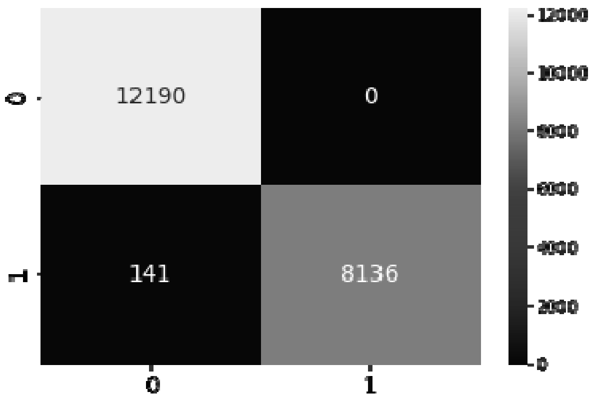

Figure 5 Confusion matrix of logistic regression.

From Figure 5, there are 12190+8136 correct predictions, specifically, 12190 samples are correctly predicted that they will cancel the booking whereas 8136 samples are correctly predicted that they will not cancel the booking. There are 141 cancelled samples that have been incorrectly predicted as not cancelled, and none of the not cancelled samples are mistakenly predicted.

Table 4. Classification report of logistic regression

	Precision	Recall	F_1-score	Support
Cancelled samples	0.99	1.00	0.9	12190
Uncancelled samples	1.00	0.98	0.99	8277
Accuracy			0.99	20467
Macro avg	0.99	0.99	0.99	20467
Weighted avg	0.99	0.99	0.99	20467

According to the information we have obtained till this stage, we can come up with a classification report of logistic regression as shown in Table 4. The accuracy score of 0.99, and the F_1-score of 0.99 indicate the good performance of this model.

The ROC curve in Figure. 6 is a line as there are only has 2 outcomes in our case ('1' for cancelled and '0' for not cancelled). Its x-axis is the false positive rate (FP rate), y-axis is the true positive rate (TP rate). The greater the area under the curve, the better performance the model is. Figure 6 shows that the area under the curve (the grey area) is 0.98. Thus, we can say that it is a well enough prediction.

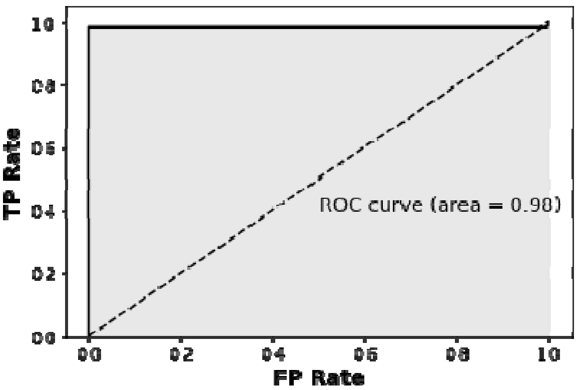

Figure 6 ROC curve of logistic regression.

4.2. k-Nearest Neighbor (k-NN)

To decide the value of k, we estimate 10 numbers (from 1 to 10) in parameter comparison. The results are shown in Figure 7.

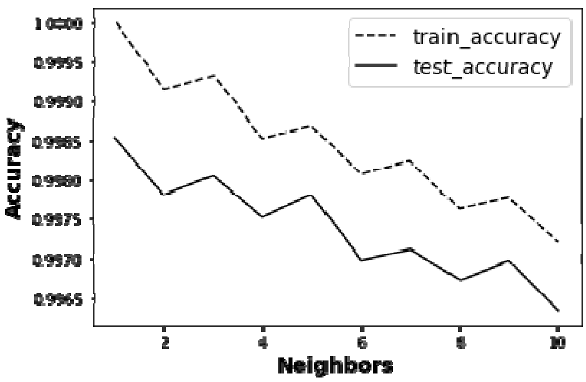

Figure 7 Accuracy score of different values of *k*.

As Figure 7 depicted, the accuracy score of both training set and test set generally decreases with the increase of the value of '*k*'. Therefore, a smaller one would be better. However, '1' may lead to over-fitting and the prediction may not be accurate when the model is applied to new data. Then it is obvious in Figure 7 that the model performed better when *k* equaled to '3' than to '2'. Consequently, we choose '3' for '*k*'.

The confusion matrix of the *k*-NN algorithm is shown in Figure 8. It illustrates that 12189 canceled samples and 8238 not canceled samples in the test set are correctly predicted. Only 39 canceled samples are mistakenly predicted to be not canceled and 1 not canceled sample is predicted to be canceled.

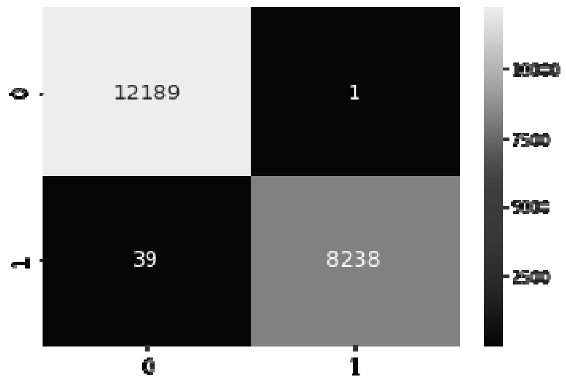

Figure 8 Confusion matrix of *k*-NN.

Based on the confusion matrix, the accuracy score and F_1-score are calculated. As presented in Table 5, the accuracy score of *k*-NN is 0.998 and the F_1-score is approximately 1.

Table 5. Classification report of *k*-NN

	Precision	Recall	F_1-score	Support
Cancelled samples	1.00	1.00	1.00	12190
Uncancelled samples	1.00	1.00	1.00	8277
Accuracy			0.998	20467
Macro avg	1.00	1.00	1.00	20467
Weighted avg	1.00	1.00	1.00	20467

It can conclude that *k*-NN can well predict the cancellation of hotel booking of this dataset.

4.3. CatBoost Classifier

We use the confusion matrix to test the accuracy of using CatBoost classifier to do the prediction.

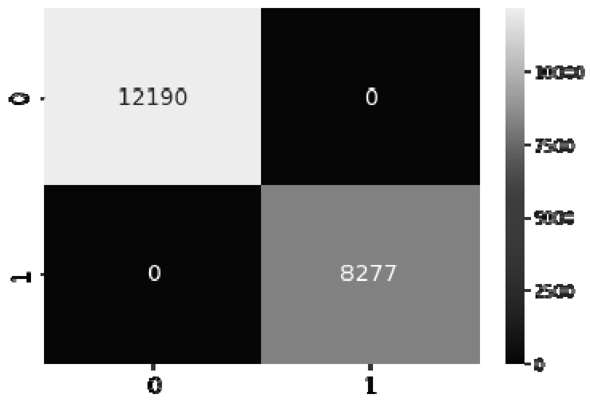

Figure 9 Confusion matrix of CatBoost.

Figure 9 above is the confusion matrix of CatBoost classifier, indicating that 12190 canceled samples and 8277 not canceled samples are correctly predicted through CatBoost. And 0 canceled samples are predicted incorrectly to be not canceled.

Table 6. Classification report of CatBoost classifier

	Precision	Recall	F_1-score	Support
Cancelled samples	1.00	1.00	1.00	12190
Uncancelled samples	1.00	1.00	1.00	8277
Accuracy			1.00	20467
Macro avg	1.00	1.00	1.00	20467
Weighted avg	1.00	1.00	1.00	20467

According to Table 6, we can see that the accuracy score of CatBoost classifier is 1.0, which means that it is the best model among the models we used to predict the cancellation of hotel booking.

4.4. Model Comparison

Based on each model's confusion matrix presented in the last section, the logistic regression model predicts 20326 (12190+8136) samples correctly, the figures of *k*-NN and CatBoost are 20427 (12189+8238), and 20467 (12190+8277), respectively. Thus, we can say CatBoost has the highest accuracy in this case.

Table 7 presents the accuracy score of the models we have conducted. The higher the accuracy score, the better the model performs. Therefore, it can be corroborated from Table 7 that CatBoost has the best performance among these three.

Table 7. Accuracy score of 3 models

Rank	Model	Accuracy score
1	CatBoost	1.000
2	*k*-NN	0.9980
3	Logistic regression	0.9931

5. CONCLUSION

By using the hotel booking demand datasets collected from hotel booking demand datasets provided by Antonio et al. [11], we have improved the existing cancellation strategy with the model which can be used to predict hotel booking cancellation. In this paper, we first analyze the datasets to find out feature importance. Then we prepare it for the models by filtrating, encoding and scaling them. Next, we split the whole datasets into a training set and a test set, the former is 80% of the original datasets and the latter is 20%. After training models with the training set, we compare all three models by their accuracy scores of predicting the cancellation of the test set. In descriptive statistics, we come up with feature importance. CatBoost is the best model among all models we discuss in this case as it has the most correctly predicted samples and highest accuracy score among the three algorithms.

Since booking status is an important factor in hotels revenue management, the prediction model can be applied to hotels RM system to assist that process. According to our paper, we suggest the hotels collect some more information about their guests including but not limited to 'previous cancellations' and 'booking changes', to increase the accuracy of cancellation prediction. After distinguishing the reservations which will be canceled or not, a priority rank is necessary to make sure that reliable reservations are before the others, in order to manage hotel resources more efficiently. The hotels can establish cancellation policies, such as setting cancellation deadlines.

There are some limitations of our project. First, most of the features in the datasets are describing the customers. Although the current model is acceptable since the behaviors of customers can always reflect some information about the hotels they booked. However, adding more features about the hotels into the model can make the prediction more credible. Moreover, the dataset contains only the information of reservations from July 1, 2015, to August 31, 2017, which is not the most recent one. We would like to find the latest data to apply to the model in future work.

AUTHORS' CONTRIBUTIONS

Yiying Chen did the data preprocessing and conducted logistic regression.

Chuhan Ding did the data preprocessing and conducted CatBoost.

Hanjie Ye did the descriptive analysis and compared each model.

Yuchen Zhou was responsible for descriptive statistics and conducted *k*-NN.

All authors wrote the paper together and approved the final version.

REFERENCES

[1] K. Talluri and G. V. Ryzin, "Revenue management under a general discrete choice model of consumer behavior," *Management Science*, 50(1), pp. 15-33, Jan. 2004.

[2] *Handbook of Research on Holistic Optimization Techniques in the Hospitality*, Tourism, and Travel Industry, António, N., Almeida, A. D., and Nunes, L. M. M., 2017, pp.141–167.

[3] S. E. Kimes and J. Wirtz, "Has revenue management become acceptable? Findings from an international study on the perceived fairness of rate fences," *Journal of service research*, 6(2), pp. 125-135, Nov. 2003.

[4] N. Antonio, A. D. Almeida, and L. Nunes, "Big data in hotel revenue management: Exploring cancellation drivers to gain insights into booking cancellation behavior," *Cornell Hospitality Quarterly*, 60(4), pp. 298-319, May 2019.

[5] C. C. Chen and K. L. Xie, "Differentiation of cancellation policies in the US hotel industry," *International Journal of Hospitality Management*, 34, pp. 66-72, Mar. 2013.

[6] W. C. Chiang, J. C. Chen, and X. Xu, "An overview of research on revenue management: current issues and future research," *International Journal of Revenue Management*, 1(1), pp. 97-128, Oct. 2006.

[7] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, 49(11), pp. 1225-1231, Nov. 1996.

[8] D. Ciobanu and M. Vasilescu, "Advantages and Disadvantages of Using Neural Networks for Predictions," *Ovidius University Annals*, *Series Economic Sciences*, vol. 13, pp. 444-449, 2013.

[9] S. B. Imandoust and M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background," *International Journal of Engineering Research and Applications*, 3(5), pp. 605-610, Sep. 2013.

[10] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363*, Oct. 2018.

[11] N. Antonio, A. D. Almeida, and L. Nunes, "Hotel booking demand datasets," *Data in brief*, 22, pp. 41-49, Feb. 2019.

[12] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), pp. 215-232, July 1958.

[13] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, 13(1), pp. 21-27, Jan. 1967.

[14] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *arXiv preprint arXiv:1706.09516*, June 2017.

[15] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, 5(2), pp. 1-11, Mar. 2015.