Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 3 Number 4 December 2024

AI-Based Hotel Customer Churn Prediction Model

Jiajun Cheng[1]

[1] Hyatt Hotel Management Co., Ltd., Jinan, Shandong 250109, China
Correspondence: Jiajun Cheng, Hyatt Hotel Management Co., Ltd., Jinan, Shandong 250109, China.

doi:10.56397/JPEPS.2024.12.03

Abstract

This study delves into the application of artificial intelligence technology in predicting hotel customer churn, aiming to reduce churn rates and enhance customer satisfaction through predictive analytics. By employing a comprehensive array of statistical and machine learning models, including logistic regression, random forests, neural networks, and support vector machines, we analyzed key data features such as customer behavior, transaction history, and service interactions. The findings indicate that artificial intelligence technology can effectively predict customer churn, providing a basis for the hotel industry to implement personalized marketing and customer loyalty enhancement programs. Furthermore, this study assessed the effectiveness of intervention strategies and optimized them through A/B testing and customer feedback loops. Ultimately, we propose long-term customer relationship management strategies to continuously enhance customer value. The research emphasizes the importance of utilizing data effectively while protecting customer privacy and points to future research directions, including algorithm innovation, improving model interpretability, and addressing data privacy and security challenges.

Keywords: customer churn prediction, artificial intelligence, machine learning, data analysis, hospitality industry, Customer Relationship Management (CRM), personalized marketing, customer loyalty, data privacy, model interpretability

1. Introduction

In today's competitive hospitality industry, customer churn has become a key factor affecting business profitability and market position. Customer churn not only means the loss of direct revenue but also involves the erosion of brand reputation and market share. Therefore, effectively predicting and managing customer churn is crucial for the sustainable development of the hospitality industry.

Background and current situation of customer churn in the hospitality industry: With the advancement of globalization and digitization, the hospitality industry is facing unprecedented challenges. On the one hand, consumer demands are becoming more diversified and personalized, with increasing expectations for service quality and experience; on the other hand, emerging online travel platforms and sharing economy models pose a huge impact on traditional hotels. These factors lead to decreased customer loyalty and increased churn rates. According to industry reports, the cost of acquiring new customers is much higher than retaining existing ones, and a small reduction in customer churn rates can significantly improve hotel profit margins. Therefore, understanding the background and current situation of customer churn is crucial for formulating effective customer retention strategies.

Value and challenges of customer churn prediction: The value of customer churn prediction lies in its ability to help the hospitality industry identify potential churn customers in advance and take timely intervention measures, thereby reducing customer churn rates and enhancing customer satisfaction and loyalty. Through predictive analytics, hotels can more accurately target high-risk customer groups, implement personalized marketing and service strategies, optimize resource allocation, and improve marketing efficiency. However, customer churn prediction also faces multiple challenges, including data collection, model accuracy, and cross-departmental collaboration. How to extract valuable information from massive data, build effective prediction models, and transform prediction results into practical actions are key issues the hospitality industry needs to address.

Research objectives and questions: This study aims to explore how to use artificial intelligence technology to build a hotel customer churn prediction model and evaluate its effectiveness in practical applications. The main objectives of the research include: (1) analyzing the impact of customer churn on the hospitality industry and identifying key risk factors; (2) comparing the performance of different artificial intelligence models in customer churn prediction; (3) proposing intervention strategies based on prediction results to reduce customer churn rates; (4) providing practical guidance and decision support for customer churn management in the hospitality industry. Research questions focus on how to accurately capture the precursors of customer churn, how to build and optimize prediction models, and how to design effective customer retention strategies.

Through in-depth analysis and empirical research, this study will provide the hospitality industry with a theoretical basis and practical tools for customer churn prediction, helping businesses better cope with market changes, enhance customer relationship management levels, and strengthen competitiveness.

2. The Importance of Customer Churn

2.1 The Financial Impact of Customer Churn on the Hospitality Industry

The financial impact of customer churn on the hospitality industry is profound and severe. Firstly, the Customer Acquisition Cost (CAC) is typically much higher than the Customer Retention Cost. This means that when a customer churns, hotels not only lose the future revenue from that customer but also need to invest additional resources to attract new customers to fill the gap. Additionally, the loss of Customer Lifetime Value (CLV) is another critical factor. CLV refers to the total net profit a customer brings to a business over their entire lifecycle. Customer churn means not only an immediate reduction in revenue but also the loss of potential future profits. Therefore, by reducing the customer churn rate, hotels can significantly improve profit margins and return on investment. (PwC., 2019)

2.2 The Impact of Customer Churn on Brand Reputation

Customer churn not only affects the financial performance of hotels but can also damage brand reputation. In the digital age, dissatisfied customers may share their negative experiences through online reviews, social media, and word of mouth. The spread of this negative word of mouth can quickly influence the decisions of potential customers, reducing brand appeal and increasing the difficulty of customer acquisition. Moreover, negative impacts on social media can quickly amplify, leading to brand image damage and even affecting stock prices and market competitiveness. Therefore, preventing customer churn is not only about maintaining revenue but also about preserving brand reputation and market position.

2.3 The Strategic Significance of Predicting Customer Churn

The strategic significance of predicting customer churn lies in its ability to help hotels timely identify and respond to signals of customer dissatisfaction, thereby taking effective intervention measures. Through predictive analytics, hotels can discover potential churn customers in advance and implement personalized marketing campaigns or service improvement measures to increase customer satisfaction and loyalty. This timely intervention can not only reduce customer churn but also enhance customers' positive feelings towards the brand, increasing the likelihood of word-of-mouth recommendations. Furthermore, effective customer churn prediction can also help hotels optimize Customer Relationship Management (CRM) strategies by more accurately targeting market segments and allocating resources, improving marketing efficiency and customer experience. Ultimately, this leads to higher customer retention rates and stronger competitive advantages. (PwC., 2019)

By deeply understanding the importance of customer churn, hotels can place greater emphasis on the formulation and implementation of customer retention strategies, thus maintaining a leading position in fierce market competition. This study will further explore how to use artificial intelligence technology to predict and manage customer churn to help hotels achieve this goal.

3. Predictive Models

In the study of customer churn prediction, selecting the right predictive models is crucial for improving the accuracy of predictions and implementing effective intervention measures. This chapter will detail the application of statistical and machine learning models in customer churn prediction and discuss the criteria for model evaluation.

3.1 The Application of Statistical Models in Customer Churn Prediction

Logistic Regression Model: Logistic regression is one of the commonly used statistical methods for predicting customer churn. It can estimate the impact of one or more independent variables on a binary dependent variable, such as whether a customer churns or not. In the hospitality industry, logistic regression models can be used to predict the probability of customer churn based on features like customer behavior data, transaction history, and service interactions. The model is simple in form and easy to understand and interpret, but its limitation lies in the limited modeling capability for non-linear relationships.

Survival Analysis: Survival analysis is a statistical method used to analyze the expected duration of time or the time until an event occurs. In the context of customer churn, survival analysis can be used to predict the timing of customer churn, that is, the potential duration a customer may continue before churning. This method is particularly suitable for dealing with “right-censored” data, where events (customer churn) have not occurred within the observation period. Survival analysis models, such as the Cox Proportional Hazards Model, can identify key factors affecting the timing of customer churn and provide in-depth insights into customer churn risk.

3.2 Introduction and Selection of Machine Learning Models

Random Forest: Random forest is an ensemble learning method that improves the accuracy and robustness of predictions by constructing multiple decision trees and outputting the average results. In customer churn prediction, random forests can handle a large number of data features and identify complex non-linear relationships. Additionally, random forests provide feature importance assessment, helping hotel managers understand which factors have the greatest impact on customer churn. (IBM, 2020)

Neural Networks: Neural networks, especially deep learning models, are widely applied to various prediction tasks due to their strong non-linear fitting capabilities. In customer churn prediction, neural networks can capture complex patterns and interaction effects, especially when dealing with high-dimensional data. By adjusting the network structure and parameters, neural networks can be optimized to fit specific data distributions and business needs.

Support Vector Machine (SVM): Support Vector Machine is a supervised learning algorithm that seeks the optimal decision boundary in feature space to maximize the margin between different classes. The advantage of SVM in customer churn prediction lies in its excellent generalization ability, especially when the data dimension is much larger than the number of samples. SVM can also handle non-linear problems through kernel tricks, making it suitable for complex customer churn prediction scenarios.

3.3 Model Comparison and Evaluation Criteria

Accuracy, Recall, and F1 Score: Accuracy, recall, and F1 score are key metrics for evaluating the performance of classification models. Accuracy measures the proportion of correct predictions made by the model, recall measures the model’s ability to identify positive classes (customer churn), and the F1 score is the harmonic mean of accuracy and recall, providing a comprehensive measure of performance. These metrics help compare the predictive effects of different models and select the most suitable model for business needs.

Confusion Matrix and ROC Curve: The confusion matrix is a table used to describe the relationship between the predicted results and actual results of a classification model. It includes true positives, false positives, true negatives, and false negatives. The Receiver Operating Characteristic (ROC) curve is a graphical tool used to display the changes in true positive rate (TPR) and false positive rate (FPR) of the model at different thresholds. The area under the ROC curve (AUC) provides a quantitative measure of the model's overall performance; the higher the AUC value, the stronger the model's predictive power.

By using these model comparison and evaluation criteria, researchers and practitioners can systematically evaluate and select the most suitable model for customer churn prediction, thereby improving the accuracy of predictions and the effectiveness of interventions.

4. Data Features

In the process of building a customer churn prediction model, the selection and processing of data features are crucial steps that directly affect the model's predictive capability and accuracy.

4.1 Data Collection and Feature Engineering

Types and Sources of Customer Behavior Data: Customer behavior data is one of the most critical types of data in prediction models. This data includes the frequency of customer bookings, accommodation preferences, spending patterns, etc., which can be collected from the hotel's CRM system, online booking platforms, and customer loyalty programs. Additionally, customer interactions on social media, website clickstreams, and mobile app usage are also important data sources. These data help us understand customer preferences and behavior patterns, providing rich contextual information for the prediction model.

Integration of Transaction History Data: Transaction history data records every purchase behavior of the customer, including booking dates, room types, prices, payment methods, and consumption amounts. These data are usually stored in the hotel's central reservation system or financial database. When integrating these data, it is necessary to ensure data consistency and completeness while removing or filling in missing values. Transaction history data can reveal the customer's purchasing power and consumption habits, which are important bases for predicting customer churn. (Deloitte, 2018)

Extraction of Service Interaction Data: Service interaction data involves all touchpoints between the customer and the hotel, including front desk reception, room service, dining experiences, and customer service, etc. These data can be extracted from customer satisfaction surveys, service evaluation systems, and customer feedback records. The quality of service interaction data directly affects customer satisfaction and loyalty, making it significant for predicting customer churn.

4.2 Identification of Key Data Features

Customer Behavior Patterns: Customer behavior patterns include the frequency of accommodation, average length of stay, preferred room types, and seasonal booking patterns of customers. These patterns can be identified from customer behavior data through clustering analysis or sequence pattern mining techniques. Understanding these patterns helps hotels identify high-risk churn customer groups and develop targeted retention strategies.

Transaction Frequency and Amount: Transaction frequency and amount are direct indicators of customer value. Customers with high transaction frequency and amount usually contribute more to the hotel, thus warranting more attention. By analyzing customer spending trends, the risk of customer churn can be predicted, and timely measures can be taken to prevent churn.

Service Feedback and Complaints: Service feedback and complaint data provide direct evaluations of hotel services from customers. Frequent negative feedback and complaints may indicate a decline in customer satisfaction, increasing the risk of customer churn. Sentiment analysis and topic modeling of these data can help hotels quickly identify weak links in services and take improvement measures.

4.3 Feature Selection and Dimensionality Reduction Techniques

Correlation Analysis: Correlation analysis is used to assess the strength of the relationship between individual features and customer churn. By calculating the correlation coefficient between features and the target variable, the most influential features can be identified. This helps simplify the model, improving the accuracy and efficiency of predictions.

Principal Component Analysis (PCA): Principal Component Analysis is a common dimensionality reduction technique that maps the original feature space to a new coordinate system through linear transformation, while retaining the most important variation information in the data. In customer churn prediction models, PCA can help reduce the number of features, eliminate multicollinearity between features, and improve the training speed and generalization ability of the model.

Through in-depth analysis and processing of data features, we can build a more accurate and reliable customer churn prediction model, providing scientific decision support for hotels. (Buhalis, D., & Law, R., 2008)

5. Model Training and Validation

In the process of building a customer churn prediction model, the training and validation of the model are crucial steps that directly affect the model's predictive performance and practical application effectiveness.

5.1 Preprocessing and Splitting of Historical Data

Data Cleaning: Data cleaning is the first step in preprocessing, aimed at ensuring data quality and laying a solid foundation for model training. Data cleaning includes removing duplicate records, handling missing values, identifying and correcting outliers, etc. In the context of customer churn prediction, for example, we need to ensure that the customer's transaction history data is accurate, and the service interaction records are complete. Data cleaning not only improves the usability of data but also helps to enhance the accuracy and robustness of the model.

Division of Training and Test Sets: After cleaning, the data needs to be divided into training and test sets. The training set is used for the model's learning process, while the test set is used to evaluate the model's predictive performance. Typically, data is randomly divided according to a certain ratio (such as 70% for training and 30% for testing). This division ensures that the model can also perform well on unseen data, demonstrating good generalization ability. In some cases, time series splitting is also used, especially when the data has temporal dependencies, to maintain the integrity of the data's timeline.

5.2 Steps and Techniques for Model Training

Hyperparameter Tuning: A key aspect of model training is hyperparameter tuning. Hyperparameters are parameters that are set before the learning process begins and have a significant impact on the model's performance. For example, in a random forest model, the number of trees and depth, and in a neural network, the number of layers and learning rate are hyperparameters. Tuning can be done through grid search, random search, or Bayesian optimization, with the goal of finding the best combination of hyperparameters to improve model performance.

Model Fitting: Model fitting is the process of training the model using training data. In this step, the model adjusts its parameters by learning patterns and relationships in the data. For different models, the fitting process may involve different algorithms and strategies. For example, in logistic regression, the model estimates parameters by maximizing the likelihood function; while in neural networks, the model updates weights through backpropagation. The goal of model fitting is to achieve good predictive results on the training data while avoiding overfitting.

5.3 Cross-Validation and Model Performance Evaluation

K-Fold Cross-Validation: K-fold cross-validation is a technique for evaluating model performance. It divides the dataset into K equal-sized subsets. In each round, one subset is used as the test set, and the remaining K-1 subsets are combined as the training set. This process is repeated K times, with a different subset used as the test set each time. K-fold cross-validation provides a more stable and reliable estimate of model performance because it reduces the dependence of model evaluation results on specific data divisions. (Xiang, Z., & Gretzel, U., 2010)

Model Stability and Generalization Ability: Model stability and generalization ability are two important indicators for evaluating model performance. Stability refers to the consistency of the model's predictive results across different datasets, while generalization ability refers to the model's predictive capability on unseen data. A model with good stability and generalization ability can maintain stable predictive performance on new and different data. This is usually assessed by comparing the model's performance on the training set and an independent test set. Additionally, statistical tests can be used to evaluate whether the differences in model performance are significant.

Through these training and validation steps, we can ensure the reliability and effectiveness of the customer churn prediction model, providing accurate decision support for hotels.

6. Intervention Strategies

After establishing and validating the customer churn prediction model, the key lies in how to transform the prediction results into practical actions to reduce customer churn. This chapter will discuss the design, implementation, and effectiveness evaluation of intervention strategies based on model predictions, as well as long-term customer relationship management strategies.

6.1 Design of Intervention Strategies Based on Model Predictions

Personalized Marketing Campaigns: Based on the high-risk customer groups identified by the model predictions, hotels can design personalized marketing campaigns to improve customer retention rates. These campaigns may include customized offers, special event invitations, or loyalty rewards. For example, for customers identified as at risk of churn by the prediction model, hotels can offer exclusive discounts or free upgrades to increase customer satisfaction and loyalty. The goal of personalized marketing campaigns is to enhance brand loyalty by providing products and services that match customer needs and preferences.

Customer Loyalty Enhancement Programs: Customer loyalty enhancement programs aim to encourage repeat purchase behavior through rewards and incentives. These programs can include points systems, membership levels, or exclusive offers. By analyzing the output of the customer churn prediction model, hotels can identify customer groups that require special attention and provide them with customized loyalty enhancement programs. These programs not only help prevent customer churn but also increase the customer’s lifetime value.

6.2 Implementation and Effectiveness Evaluation of Intervention Strategies

A/B Testing: To evaluate the effectiveness of intervention strategies, hotels can conduct A/B testing. In A/B testing, different customer groups are randomly assigned to different intervention groups and control groups. The intervention group receives specific retention strategies, while the control group receives standard service or no intervention. By comparing key indicators such as retention rates and satisfaction between the two groups, hotels can quantify the actual effects of intervention strategies and adjust strategies accordingly. A/B testing provides a scientific method for hotels to optimize customer retention strategies.

Customer Feedback Loop: After implementing intervention strategies, collecting and analyzing customer feedback is crucial. Through customer satisfaction surveys, online reviews, and direct feedback, hotels can assess the acceptance and effectiveness of intervention measures. The customer feedback loop not only helps hotels understand the actual impact of intervention strategies but also provides valuable data support for continuous improvement. By responding promptly to customer feedback, hotels can further adjust and optimize intervention measures to improve customer satisfaction and loyalty.

6.3 Long-Term Customer Relationship Management Strategies

Customer Journey Mapping: The key to long-term customer relationship management strategies is understanding the customer’s entire journey experience. Customer journey mapping is a visual tool used to track every touchpoint of customer interaction with the brand, from the first contact to repeat purchases. By analyzing the customer journey, hotels can identify opportunities to enhance customer experience and satisfaction, thereby reducing customer churn. Customer journey mapping also helps hotels design more coherent and personalized customer interaction strategies, strengthening brand loyalty. (Sigala, M., 2012)

Continuous Customer Value Enhancement: Hotels should be committed to continuously enhancing customer value, which includes providing high-quality products and services, optimizing customer experience, and building strong customer relationships. By regularly evaluating customer value and analyzing market trends, hotels can adjust their products and services to meet the changing needs of customers. Additionally, hotels can enhance customer experience and operational efficiency through technological innovations such as artificial intelligence and big data analysis. Continuous customer value enhancement not only helps reduce customer churn but also strengthens the hotel's market competitiveness.

Through these intervention strategies and long-term customer relationship management strategies, hotels can effectively address the challenges of customer churn, improve customer satisfaction and loyalty, and achieve sustainable business growth.

7. Conclusion

In this study, we explored the application of artificial intelligence technology in hotel customer churn prediction and summarized its effectiveness. The following is a review of the research results and a look forward to future research directions.

Summary of the Effectiveness of Artificial Intelligence in Hotel Customer Churn Prediction: By comparing various statistical and machine learning models, including logistic regression, random forests, neural networks, and support vector machines, we found that artificial intelligence technology is significantly effective in predicting hotel customer churn. These models can identify features associated with a high risk of customer churn, such as transaction frequency, service interactions, and customer feedback patterns. Through model training and validation, we were able to build models that accurately predict customer churn and provide a basis for hotels to implement intervention measures. Additionally, this study demonstrated how to evaluate and optimize intervention strategies through A/B testing and customer feedback loops, thereby increasing customer retention rates and satisfaction.

Implications for Hospitality Industry Practice: This study provides empirical support for the implementation of customer churn prediction and intervention strategies in the hospitality industry. By applying artificial intelligence technology, hotels can more accurately identify and understand the causes of customer churn and take timely personalized marketing and customer loyalty enhancement plans. Moreover, customer journey mapping and continuous customer value enhancement provide strategies for long-term customer relationship management. These practices not only help reduce customer churn but also enhance customer loyalty, brand image, and market competitiveness.

Future Research Directions and Technical Challenges: Although this study has achieved positive results, there are still some challenges and future research directions. First, with the rapid development of big data and artificial intelligence technology, exploring new algorithms and models, such as deep learning and reinforcement learning, in the application of customer churn prediction is an important research direction. Second, how to improve the interpretability and transparency of models so that hotel managers can better understand and trust the model's prediction results is also a key technical challenge. In addition, data privacy and security issues are becoming increasingly prominent in the collection and processing of customer data. Researching how to effectively use data while protecting customer privacy is also an urgent issue that needs to be addressed.

References

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—the state of eTourism research. *Tourism Management*, *29*(4), 609-623.

Deloitte. (2018). The future of hospitality: A look at the technologies shaping the travel and hospitality industry. Deloitte Insights.

IBM. (2020). The hospitality industry in the era of cognitive systems. IBM Institute for Business Value.

PwC. (2019). Outperformers in hospitality: Digitally reinventing the guest experience. PricewaterhouseCoopers.

Sigala, M. (2012). Integrating web 2.0 applications for co-creating value in the travel industry. *International Journal of Contemporary Hospitality Management*, *24*(6), 937-956.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179-188.