WEP Warwick Evans Publishing

Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 10
MEEA 2024

Optimize Airline Service Policy Based on Airline Passenger Satisfaction Data Analysis

Shiyuan Tian

College of Statistics, University of International Business and Economics, Beijing, 100000, China

202192018@uibe.edu.cn

Abstract. As the preferred means of transportation for long-distance travel after the epidemic, whether an airline can stand out among its peers is deeply related to whether it can satisfy passengers, and the service quality of an airline can significantly affect passenger satisfaction. In this paper, random Forest is applied to judge the correlation between satisfaction and different service factors, and several factors that have the most significant impact on satisfaction are selected. The results show that the influence degree of Online. boarding, Inflight. wifi. service is significantly higher than other factors. Then, Legroom service, Inflight entertainment, departure-arrival time convenience, and Seat comfort are four factors with significant influence. Therefore, this paper believes that airlines should provide convenient online registration, rich flight entertainment activities, and comfortable seat space to improve passenger satisfaction under the premise of ensuring that their costs do not change much. In this way, airlines can greatly improve their competitiveness to attract and retain more loyal passengers.

Keywords: Service quality; customer satisfaction; random forests; important value.

1. Introduction

After the end of the management of the journey, tourists become the most popular way for relaxation and entertainment. The convenient and fast features of aviation make it the best choice for long-distance travel. Whether the airline can stand out in the fierce competition of customers is closely related to whether it can satisfy customers. Aghaei proposed that customer satisfaction and customer loyalty are significantly correlated, at the same time, customer satisfaction can also enhance the reputation of the airline company [1]. Therefore, enhancing customer satisfaction is a key assignment for airlines to increase revenue. Rizvi believed that higher service quality can improve customer satisfaction [2]. However, the services offered by airlines are diverse, and this paper needs to explore which specific factors have a significant correlation with customer satisfaction.

According to Tri and Jenpeng, online/mobile boarding, in-flight wi-fi services, baggage handling, and in-flight entertainment are the main concerns for airlines to improve passenger satisfaction [3]. Masorgo suggests that both flight delays and involuntarily denied boarding can harm customer satisfaction [4]. Khudhair, Jusoh, and others believed that passengers' price sensitivity is a key factor affecting service quality and passenger satisfaction [5]. Xiong suggested that placing equipment maintenance time during periods with low passenger flow would help improve airport services [6]. Liao and Cao applied the Structural Equation Model-Logit model to investigate how excessively long mandatory processes, such as check-in and security checks, can decrease passenger satisfaction, while time costs like the time of arrival at the airport and the duration of the flight are key influencing factors, however, the model requires data to satisfy the assumption of multivariate normality, which is too strict [7]. Based on the HistGBDT algorithm, Huang proposed that online aboard and in-flight environments significantly impact passenger satisfaction. The performance of the HistGBDT algorithm largely depends on the parameter Settings, but finding the optimal parameters may require a lot of adjustment and verification work, and the results are not good if the appropriate parameters are not found [8]. Mulyadi believes that supporting facilities such as the Check-in counter, aviation security, flight security, support, and baggage have a significant effect on passenger satisfaction. They introduced the multiple linear regression method to see the significance of the influence of

explanatory variables through F-test results and determined the size of the influence through partial determination coefficient analysis. However, it is not appropriate that the dependent variable is qualitative in this article [9]. Zheng and Liu studied the impact on passenger satisfaction from new perspectives such as indication direction signs (identity identification), airport environment, security, and entry procedures. The study designed a passenger satisfaction questionnaire and used SPSS for statistical analysis. The angle is relatively new but the depth of data analysis is not enough [10].

Based on the dataset of Airline Passenger Satisfaction, this paper will use multiple logistic regression to determine whether ten factors, including gender, flight distance, in-flight entertainment, catering service, in-flight wifi service, convenience of online booking, and gate location, have significant impacts on passenger satisfaction. In turn, some good suggestions are given to airlines to improve the flying experience of passengers

2. Methods

2.1. Data Sources

The research is based on a dataset of airline passenger satisfaction from the Kaggle platform. The dataset contains flight information collected from multiple airlines and passenger reviews of service, totaling 103,904 records. The data set includes various characteristics such as flight attributes, service quality, passenger attributes, and delay time. The response variable was passenger satisfaction, which was classified as "satisfied" and "dissatisfied or neutral."

2.2. Data Preprocessing

Before training the model, this paper implemented several data preprocessing steps. Initially, this paper removed columns irrelevant to the analysis, such as ID and age, along with uncontrollable factors like flight distance and delay times. Subsequently, this paper addressed missing values by filling in the 'Arrival Delay in Minutes' with the median value of the column. Finally, this paper encoded categorical variables into numeric formats suitable for model processing.

Table 1. List of Variables

Variable	Description
const	The model intercept is a constant term in the model.
Inflight wifi service	Quality of the inflight wifi service (rated by passengers).
Departure/Arrival time convenient	Convenience of the flight's departure and arrival times.
Ease of Online booking	Ease and usability of the online booking system.
Online boarding	Efficiency and convenience of the online boarding process.
Seat comfort	The comfort of the airplane seating.
Inflight entertainment	Quality of the inflight entertainment options provided.
On-board service	Quality of the service provided by the onboard crew.
Legroom service	Amount of legroom available in seating, and service associated with it.
Baggage handling	Efficiency of the airline's baggage handling process.
Check-in service	Efficiency and convenience of the airline's check-in service.
Inflight service	The overall quality of service provided inflight.
Cleanliness	Cleanliness of the aircraft environment.
Departure Delay in Minutes	Minutes by which the departure was delayed.
Arrival Delay in Minutes	Minutes by which the arrival was delayed.

First of all, too many independent variables lead to too scattered influence. Secondly, this essay is interested in which airline services satisfy or dissatisfy customers. Therefore, this paper will not consider information about the Customer: 'Age', 'Gender Male', 'Customer Type Loyal Customer', Class: 'Class', 'Type of Travel Business travel'.Meanwhile, 'Flight_Distance'and 'Departure Delay in

Minutes' are mostly beyond the control. Finally, according to multiple logistic regression, it is verified whether the satisfaction of a single variable has a significant impact on satisfaction. The P-values of Inflight service and Cleanliness are 0.620 and 0.046, indicating insignificance and low significance, respectively. Therefore, these two factors are excluded. The specific description of variables is shown in Table 1.

2.3. Model Introduction

Random forests are an ensemble learning method that constructs multiple decision trees and aggregates their outcomes to make a final decision. This method combines the predictive power of multiple decision trees, enhancing overall prediction accuracy and stability through averaging or majority voting. Random forests exhibit several unique characteristics that make them powerful tools for machine learning. Firstly, they are robust, reducing the risk of overfitting through the construction of multiple decision trees, thus enhancing the model's generalization capabilities. Secondly, they possess strong processing capabilities, allowing them to manage and analyze a large set of input variables effectively and perform feature selection efficiently. Finally, as a non-parametric model, random forests do not assume a specific underlying data distribution, which gives them the flexibility to adapt to the various types of data encountered in real-world scenarios.

3. Results and Discussion

3.1. Random Forest Model Results

A Random Forest classifier was built to predict customer satisfaction based on the selected variables. The model was trained on the training data, with a total of 500 trees and 3 variables tried at each split. The model's performance was evaluated using the test dataset. The model is estimated in many ways and related images such as Figures 1, 2, and 3 are demonstrated below.

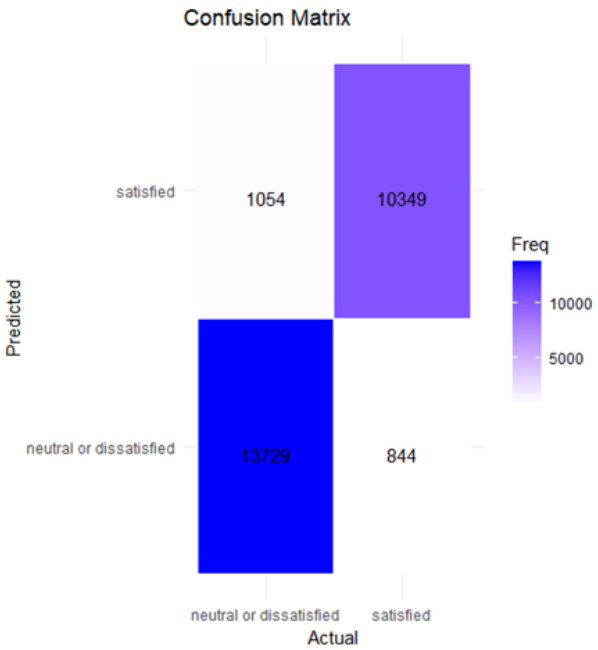

Figure 1. Confusion Matrix

From the confusion matrix, this paper observes that the number of correctly classified samples in the "neutral or dissatisfied" category is 13,729, while the number of misclassified samples in this category, predicted as "satisfied," is 1,054. Conversely, for the "satisfied" category, the model correctly classified 10,349 samples, with 844 samples misclassified as "neutral or dissatisfied". Overall, the model demonstrates a relatively balanced predictive ability across the two categories. The majority of "neutral or dissatisfied" samples and "satisfied" samples were correctly classified, indicating that the model performs well in distinguishing between these two classes (Figure 1).

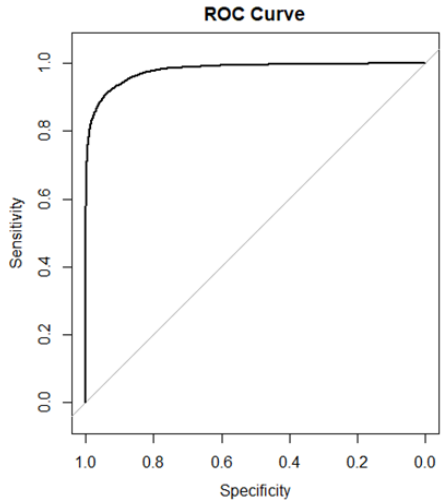

Figure 2. ROC

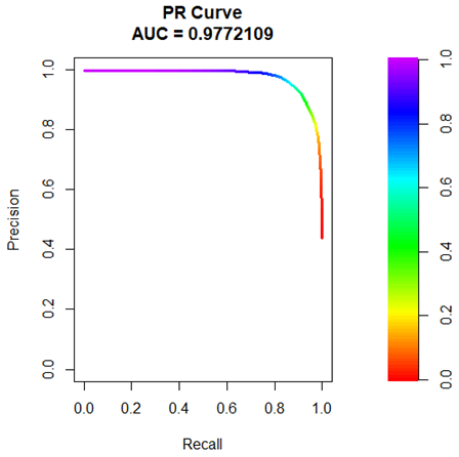

Figure 3. PR Curve&AUG

From the ROC curve and AUC value (0.977), the model performs very well on the test set. The AUC value is close to 1, indicating that the model has a strong ability to distinguish. The ROC curve shows high sensitivity and specificity, indicating that the model can maintain high performance at different thresholds. From the PR curve, the model performs well in dealing with class imbalance. High precision and high recall indicate that the model can maintain high accuracy when predicting 'satisfied' samples, and can correctly identify most positive samples (Figure 2).

By calculating the following comprehensive evaluation indicators, the Kappa value of 0.8514 indicates that the prediction results of the model are significantly better than a random guess, demonstrating the model's reliability. Additionally, the sensitivity, or recall, of 0.9421 indicates that the model can effectively identify positive samples, such as satisfied customers, showcasing its excellent performance in recognizing satisfied customers. Furthermore, the precision of 0.9287 indicates that most of the positive samples predicted by the model are indeed positive, highlighting the model's accuracy in predicting satisfied customers. This supports the conclusion that the model and its results significantly aid in exploring which airline service factors are related to passenger satisfaction (Figure 3).

3.2. Important Value

In this study, this paper uses the random forest model to predict the satisfaction of air passengers. The model itself has been evaluated above. Next, this paper evaluates the impact of various factors on passenger 'satisfaction' evaluation through the importance of features. The importance of features is evaluated by the Overall Importance index, which is an important indicator that combines the Mean Decrease Accuracy and the Mean Decrease Gini coefficient. The following is a detailed analysis of

the key influencing factors of various services on passenger satisfaction and suggestions for improvement (Including pictures in Figure 4, Figure 5, and Figure 6).

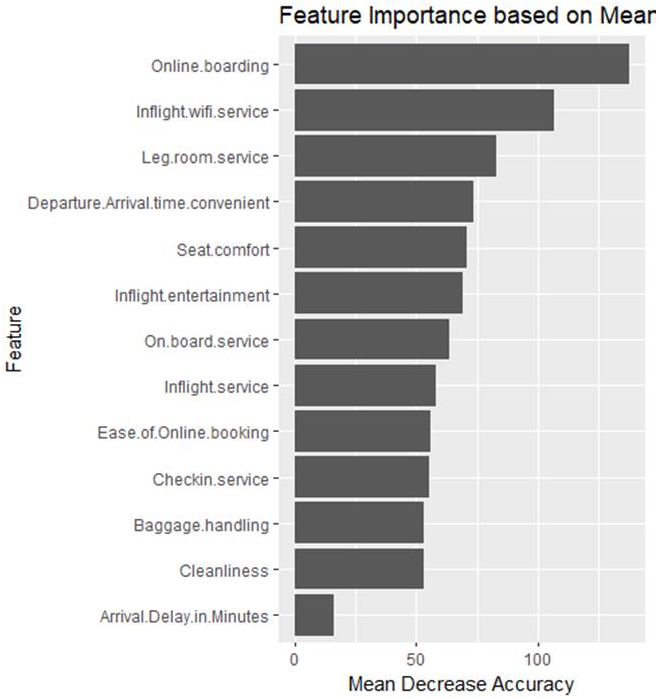

Figure 4. MDA results

The importance of features is sorted by MeanDecreaseAccuracy, which indicates the degree of decline in the overall model after deleting a feature. Online boarding is also the most important feature that significantly affects the Gini coefficient. Inflight wifi service and inflight entertainment also have a significant impact on the Gini coefficient of the model. Legroom service, inflight entertainment, and departure-arrival time convenience affect the Gini coefficient and the accuracy of the model to some extent (Figure 4).

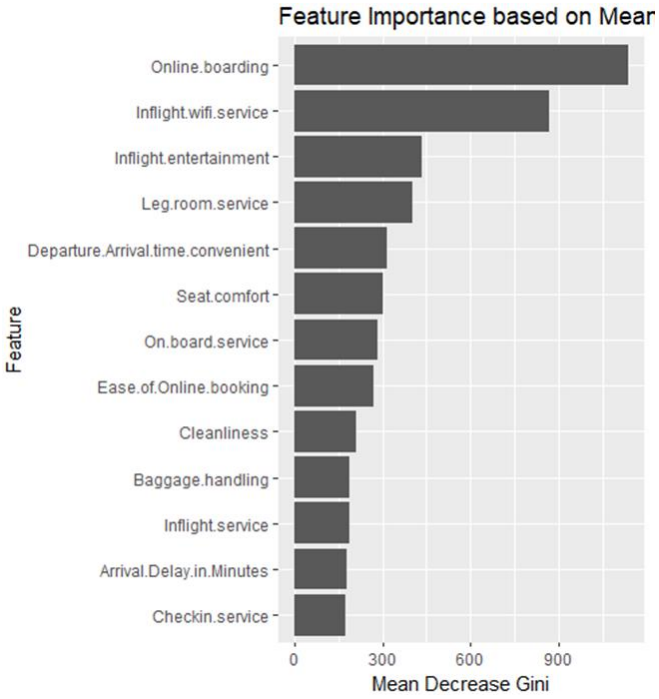

Figure 5. MDG results

The importance of features is sorted by MeanDecreaseGini, which indicates the change in the Gini coefficient after deleting a feature. The Gini coefficient represents the purity of nodes. Random forest

is committed to finding the nodes that make the Gini coefficient reduce the most. Online boarding is also the most important feature that significantly affects the Gini coefficient. Inflight wifi service and inflight entertainment also have a significant impact on the Gini coefficient of the model. Legroom service and departure arrival time convenience affect the Gini coefficient of the model to a certain extent (Figure 5).

Online boarding is also the most important feature that significantly affects the Gini coefficient.Inflight.wifi.service and Inflight. entertainment: also has a significant impact on the Gini coefficient of the model. Leg.room.service and Departure Arrival time convenient: These characteristics affect the Gini coefficient of the model to a certain extent.

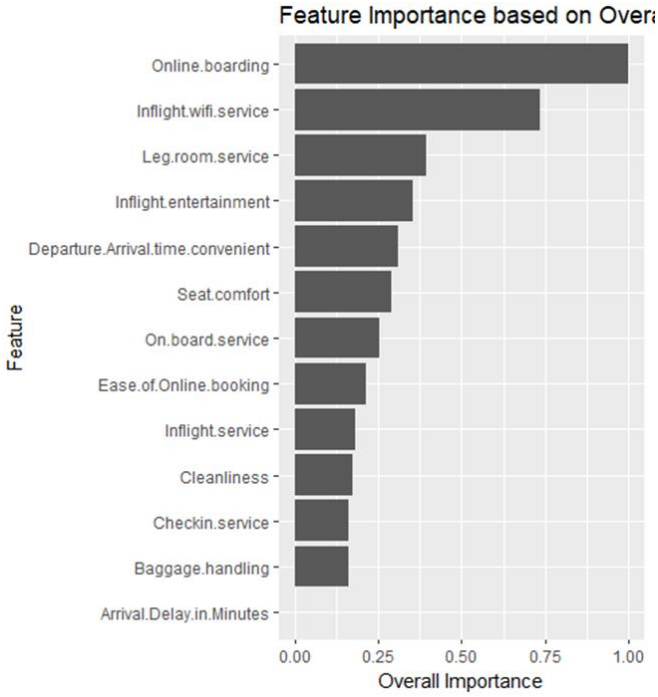

Figure 6. Overall Important

To sort the importance of each factor, this paper synthesizes the above two importance indicators, progress standardization summation obtains comprehensive importance indicators, and sets the overall importance value of the most important factor to 1. To identify the influence of a single factor on the random forest model (Figure 6).

Online boarding ranked first, which is the most important factor affecting passenger satisfaction. Good boarding convenience is the beginning of passengers ' satisfaction. Inflight. wifi. service, overall importance value is 0.733754079. These two factors are faults ahead of other factors, and airlines should focus on improving this aspect. Leg. room. service, Inflight.entertainment, Departure.Arrival. time. convenient, Seat.com's four indicators have an overall importance value close to or more than 0.3, they also have a certain degree of important influence on passenger satisfaction.

Other factors also have a certain degree of influence, but to better significantly improve passenger satisfaction, this paper only focuses on the particularly significant impact and significant impact of easy factors.

3.3. Suggestions

Through the above analysis, this paper concludes six factors that have the greatest correlation with passenger satisfaction, from high to low are Online. boarding, Inflight. wifi.service, Legroom service, Inflight entertainment, departure-arrival time convenience, and Seat comfort, while other factors have a lower degree of influence compared with these six factors. Therefore, this article will refer to these aspects to help airlines to improve service to obtain higher passenger satisfaction. First of all, the

online boarding system can improve the boarding experience by reducing the waiting time of passengers at the airport and allowing passengers to choose their seats in advance. Airlines should not only pay attention to the optimization of the system, such as a better interface and easier operation but also need to regularly maintain and update the system to ensure its stability and user experience. Secondly, because the time on the plane is mostly long, passengers are very concerned about leg space service and seat comfort, because this means whether passengers get good rest in flight. However, the increase in individual average space means a decrease in single-flight revenue, and airlines need to find a balance between passenger comfort and revenue. More importantly, passengers are very concerned about Wifi and entertainment facilities. Similarly, because passengers need to relax and enjoy the spirit when flying, airlines should improve faster Wifi and more diversified entertainment activities. Another important factor is Departure Arrival time because passengers want flights to take off and land at the right time to facilitate their travel arrangements. Airlines should provide more flexible change and refund services when optimizing flight schedules.

4. Conclusion

This study provides valuable insights into the factors affecting air passenger satisfaction. By optimizing online boarding, enhancing onboard Wi-Fi services, increasing leg space, expanding onboard entertainment options, and improving flight scheduling convenience and seat comfort, airlines can significantly improve passenger satisfaction and loyalty. These improvements not only improve the experience of passengers but also strengthen the competitive position of airlines in the fiercely competitive market. Although the research identifies these key factors and their importance, this paper must recognize the limitations of the data and methods used. External factors, such as economic conditions, technological progress, and passengers' personal preferences, also affect satisfaction, but they are not fully considered in this study. Therefore, airlines should consider these broader factors when implementing improvement measures.

References

[1] A. Aghaei, A.M. Mostafapour, H. Rezaei, Investigating the relationship between service quality and customer loyalty: A case study of the banking industry, Management Science Letters (2013) 2147-2154.

[2] F.A.D. Rizvi, A.D. Sabir, The Relationship between Customer Satisfaction and Service Quality with Special Reference to the Hypermarkets in Saudi Arabia, Research in Economics and Management (2018) 160.

[3] N. Tri, H. Jen-Peng, Investigating airline passenger satisfaction: Data mining method, Research in Transportation Business & amp; Management (2022) 43.

[4] M. Nicolò, M. Saif, Rossiter A H Expectations vs Experience: Managing the Adverse Effects of Service Failures on Customer Satisfaction in the Airline Industry, TRANSPORTATION JOURNAL (2022) 231-262.

[5] Y.H. Khudhair, A. Jusoh, M.K. Nor, et al, Price sensitivity is a moderating factor between the effects of airline service quality and passenger satisfaction on passenger loyalty in the airline industry, International Journal of Business Continuity and Risk Management (2021) 114-125.

[6] X. Honglin, Research on key technologies of air passenger travel based on machine learning method, University of Shanghai for Science and Technology (2021).

[7] L. Wang, C. Xiaoshu, L. Tao, et al, The impact of airport service quality on air travel choice behavior: A case study of Guangdong-Hong Kong-Macao Greater Bay Area, Tropical Geography (2024) 195-211.

[8] H. Xinya, Research on Airline Passenger Satisfaction Prediction Model Based on HistGBDT Algorithm, Shandong University (2023).

[9] M. Nur, M. Yahya, D. Sidiq, et al, Service Quality Model Towards Passenger Satisfaction at Sultan Hasanuddin Makassar International Airport, Asian Research Journal of Arts & Social Sciences (2023) 18-25.

[10] Z. Wang, L. Xiaoping, Study on Passenger Satisfaction with Service Quality at Terminals of Wuhan Tianhe International Airport, ACM International Conference Proceeding Series (2021) 489-493.