Determinants of Customer Satisfaction at the San Francisco International Airport

Ashok K. Singh\(^1\)*, Myongjee Yoo\(^2\) and Rohan J. Dalpatadu\(^3\)

\(^1\)William F. Harrah College of Hospitality, University of Nevada, Las Vegas, USA; \(^2\)Cal Poly Pomona, Pomona, California, USA; \(^3\)Department of Mathematical Sciences, University of Nevada, Las Vegas, USA

**ABSTRACT**

This study attempts to determine the overall satisfaction factors from airline passengers at the San Francisco International Airport (SFO), using the classification method of random forest. The analysis is based on the 2014 annual survey conducted by SFO that collects data on passenger demographics and satisfaction with airport facilities and services. Results of this study indicate that some service attributes are more important than others for passengers’ overall satisfaction at SFO. Study results are expected to provide practical insights to the airport industry. This study, in addition, introduces the machine learning method of random forest to tourism research.

**Keywords:** Airport; Customer satisfaction; Predictive modeling; Random forest; Service attributes

**INTRODUCTION**

Researchers in tourism have generally used (i) multiple linear regression [1,2] which ignores the fact that the response is ordinal and not interval scale data, (ii) multinomial or ordinal logistic regression [3,4] or (iii) transformation to convert a 5-point Likert scale response to a binary response that is modeled by the binary logistic regression method [5], which is not necessary. The method of random forest [6] is a machine learning tool for classification and regression problems; the method uses decision trees and bootstrapping to predict a multinomial response (classification) or a continuous response (regression). This study attempts to determine the overall satisfaction factors from airline passengers at the San Francisco International Airport (SFO) (hereon called “SFO”) by using the method of random forest.

**LITERATURE REVIEW**

Airports are complex service settings where passenger satisfaction is influenced by a variety of attributes [7]. Some of the known factors that influence passengers’ satisfaction are: security check, art display, accessibility, airport parking, baggage, cleanliness, information availability, restrooms, restaurants, shops, staff, signage, and Wi-Fi [8-12]. In study of the service quality at Melbourne airport [11], significant discrepancies between passengers’ expectations and their perceptions of service quality at the airport were found, indicating room for improvement in service quality at the Melbourne Airport. Another study [13] used observations and information collected from a focus group study, and in-depth interviews to determine reasons for delays in baggage access. Researchers in hospitality and tourism have also investigated the problems related to determinants of customer satisfaction [14-16].

**METHODOLOGY**

**Data collection and description of variables**

SFO conducts an annual survey and collects data on passenger demographics and satisfaction with airport facilities and services from stratified random samples [15]. This study uses secondary data from the 2014 SFO annual survey, which provided a random sample of 2820 responses on 95 questions, with a number of missing responses ranging from 0 to 2820. A total number of 23 variables are selected for the analysis based on existing literature. The method of multivariate imputation by chained equations (MICE) yields a complete data set and results in estimates with smaller standard errors and narrower confidence intervals [16]. The R-package mice is therefore used to replace missing values [17].

In this study, three types of predictor variables are selected to determine the key drivers of overall satisfaction at SFO: ratings, cleanliness, and demographics. Ratings include a total of 15 items (artwork exhibitions, restaurants, retail shops and concessions, signs and directions inside SFO, escalators/elevators/moving walkways, information on screens/monitors, information booths (lower level - near baggage claim), information booths (upper level - departure area), accessing and using free Wi-Fi at SFO, signs and...
directions on SFO airport roadways, airport parking facilities, AirTrain, long term parking lot shuttle (bus ride), airport rental car center, and SFO Airport as a whole. A 5-point Likert scale, with 1 as “Unacceptable” to 5 as “Outstanding” is used to measure ratings.

Cleanliness includes a total of 6 items (Boarding areas, AirTrain, airport rental car center, airport restaurants, restrooms, and overall SFO cleanliness). A 5-point Likert scale, with 1 as “Dirty”, 3 as “Average” to 5 as “Clean” is used to measure cleanliness. Age, gender, and income are the demographic variables, with age categorized into 7 levels, gender categorized into 3 levels, and income categorized into 5 levels. Table 1 summarizes the variables selected in this study.

Method of random forest

The analyses are performed using the statistical software environment R [18]. The method of random forest is utilized to build a predictive model for overall satisfaction as a function of the 22 selected predictors. Random forest is a highly accurate ensemble machine learning method for classification or regression, which involves building a large number of decision trees in the training step, and outputs the model of the classes predicted by individual trees [4,19,20].

This study uses the R-package randomForest [21] to perform the method of random forest. The package randomForest outputs ’Out of Bag’ (OOB) (i.e., out of the training sample) estimates of prediction accuracy as well as a plot showing the importance of predictors in the model. The package is iteratively used by adding and dropping predictors until a final model with good prediction accuracy is obtained. The association between the response variable and each individual predictor is further tested by the chi-square test of independence; in majority of the cases, the expected frequencies of several cells turn out to be less than 5, and the p-values for the chi-square test are evaluated by bootstrap [22].

Performance measures for prediction

A large number of performance measures for multi-level classifiers exist in machine learning literature [23]. Accuracy, precision, recall and the geometric mean F1 of precision, and recall are commonly used [24,25]. To compute these measures, the confusion matrix is first calculated. Since the response has five categories, the confusion matrix is comprised of a 5x5 matrix of cell frequencies

| Variable label | Attribute |
|----------------|-----------|
| Q7ART          | Artwork and exhibitions |
| Q7FOOD         | Restaurants |
| Q7STORE        | Retail shops and concessions |
| Q7SIGN         | Signs and directions inside SFO |
| Q7WALKWAYS     | Escalators/Elevators/Moving walkways |
| Q7SCREENS      | Information on screens/monitors |
| Q7INFODOWN     | Information booths (lower level - near baggage claim) |
| Q7INFOUP       | Information booths (upper level - departure area) |
| Q7WIFI         | Accessing and using free WiFi at SFO |
| Q7ROADS        | Signs and directions on SFO airport roadways |
| Q7PARK         | Airport parking facilities |
| Q7AIRTRAIN     | AirTrain |
| Q7LTPARKING    | Long term parking lot shuttle (bus ride) |
| Q7RENTAL       | Airport Rental Car Center |
| Q7ALL          | SFO Airport as a whole |
| Q9BOARDING     | Boarding areas |
| Q9AIRTRAIN     | Airtrain |
| Q9RENTAL       | Airport Rental Car Center |
| Q9FOOD         | Airport restaurants |
| Q9RESTROOM     | Restrooms |
| Q9ALL          | Overall cleanliness |

Demographic variables

| Variable label | Attribute |
|----------------|-----------|
| Q18AGE         | Age (1 = Under 18, 2 = 18-24, 3=25-34, 4=35-44, 5=45-54, 6=55-64, 7 = 65 and over, 8=don’t know/refused, 9= multiple responses, 10 = Blank) |
| Q19GENDER      | Gender (1 = Male, 2 = Female, 3 = Other) |
| Q20INCOME      | Income (1 = Under 50K, 2 = 50K-100K, 3=101K-150K, 4=Over 150K, 5 = Other currency, 0 = Blank ) |
| LANG           | Language of questionnaire (1 = English, 2 = Spanish, 3 = Chinese, 4 = Japanese) |
Peinlang BL

$C_{i,j}$ where $C_{i,j}$ is the number of times true response of $j$ get predicted as $i$ ($i, j = 1, 2, ..., 5$) (Table 2).

The performance measures accuracy, precision, recall and $F_1$ were calculated for each category from the following formulas [24]:

$$\text{Accuracy} = \frac{\sum_{j=1}^{5} C_{j,j}}{\sum_{j=1}^{5} \sum_{i=1}^{5} C_{i,j}}$$

$$\text{Precision}_{j} = \frac{C_{j,j}}{\sum_{k=1}^{5} C_{j,k}}$$

$$\text{Recall}_{j} = \frac{C_{j,j}}{\sum_{i=1}^{5} C_{i,j}}$$

$$F_1_{j} = \frac{2 \times \text{Precision}_{j} \times \text{Recall}_{j}}{\text{Precision}_{j} + \text{Recall}_{j}}$$

There are examples in the literature when a multi-level classification or prediction problem is transformed into a binary classification so that the binary logistic regression can be used [3]; for this reason, the overall ratings are transformed as follows:

“Unacceptable (1)”, “Below Average (2)”, “Average (3)” = 0, “Good (4)”, and

“Outstanding (5)” = 1 and the performance measures are recalculated; these are referred to as binary accuracy, precision, recall, and $F_1$ in this study.

RESULTS

Data imputation

Table 3 shows that the method of multivariate imputation by chained equations (MICE) has performed quite well for the data set; the five-point summaries of data before and after imputation are very close to each other.

The stacked bar chart of Wi-Fi service (RATE_WIFI) (Figure 1) shows that the majority of SFO passengers give a rating of 4 or 5 to the Wi-Fi service at SFO; Figure 1 further suggests that the proportions of Wi-Fi service ratings of 1, 2,..., 5 are similar across the gates, i.e., there is no association between Wi-Fi service rating and gate; this is confirmed by the chi-square test of association between gate and Wi-Fi service (p = 0.18), which implies that the quality of Wi-Fi is similar at each gate.

All of the bar charts suggest the presence of association between the response and the predictor; the chi-square test of independence

| Predicted Overall Satisfaction | True Overall Satisfaction |
|-------------------------------|---------------------------|
| 1                             | C_{1,1} C_{1,2} C_{1,3} C_{1,4} C_{1,5} |
| 2                             | C_{2,1} C_{2,2} C_{2,3} C_{2,4} C_{2,5} |
| 3                             | C_{3,1} C_{3,2} C_{3,3} C_{3,4} C_{3,5} |
| 4                             | C_{4,1} C_{4,2} C_{4,3} C_{4,4} C_{4,5} |
| 5                             | C_{5,1} C_{5,2} C_{5,3} C_{5,4} C_{5,5} |

Table 2: Performance measures for prediction.

Figure 1: Stacked bar charts of WIFI ratings by Gate.
confirms this association; Table 4 shows that strong association exists between the response variable and each of the potential predictors.

Figure 4 shows the stacked bar charts of age (AGE) and gender (GENDER) by the response variable overall satisfaction with SFO (RATE_ALL). Figure 4 suggests that overall satisfaction with SFO is not affected by age or gender. Table 5 shows the results of the chi-square test of independence between the response variable and the two demographic variables age and gender. The associations between the response and these two potential predictors are insignificant (p > 0.05).

The random forest model
The backward selection procedure is used to find the important predictors of the response variable overall satisfaction with SFO (RATE_ALL). Table 6 shows the multi-level confusion matrix of the full random forest model for the response as a function of all of the 22 potential predictors, and Table 6 shows the binary confusion matrix of prediction obtained from Table 6. Tables 6 and 7 show that the random forest model has high accuracy (75%) and very high binary accuracy (98.5%).

Figure 5 shows the plot of variable importance measures for the full random forest model; gender (GENDER), language (LANG), age
Table 3: Results of data imputation by MICE - number of missing values, and five-point summary of data before (B) and after (A) data imputation.

| Variable          | N    | Mean   | Median | sd  | Min | Max |
|-------------------|------|--------|--------|-----|-----|-----|
| RATE_ART          | 675  | 3.92   | 3.91   | 0.88| 1   | 5   |
| RATE_FOOD         | 481  | 3.58   | 3.59   | 0.89| 1   | 5   |
| RATE_STORE        | 512  | 3.6    | 3.63   | 0.86| 1   | 5   |
| RATE_SIGN         | 122  | 4.01   | 4.02   | 0.87| 1   | 5   |
| RATE_WALKWAYS     | 274  | 4.09   | 4.08   | 0.81| 1   | 5   |
| RATE_SCREEN       | 174  | 4.05   | 4.05   | 0.82| 1   | 5   |
| RATE_INFODOWN     | 1298 | 3.84   | 3.86   | 0.88| 1   | 5   |
| RATE_INFOUP       | 1298 | 3.86   | 3.86   | 0.86| 1   | 5   |
| RATE_WIFI         | 864  | 3.91   | 3.9    | 1.13| 1   | 5   |
| RATE_ROADS        | 964  | 3.95   | 3.95   | 0.88| 1   | 5   |
| RATE_PARK         | 1710 | 3.79   | 3.8    | 0.95| 1   | 5   |
| RATE_AIRTRAIN     | 1410 | 4.08   | 4.12   | 0.85| 1   | 5   |
| RATE_LTPARKING    | 2123 | 3.79   | 3.8    | 0.95| 1   | 5   |
| RATE_RENTAL       | 1824 | 3.72   | 3.73   | 1.02| 1   | 5   |
| RATE_ALL          | 143  | 4.11   | 4.13   | 0.71| 1   | 5   |
| CLEANLINESS_BOARDING | 56   | 4.34   | 4.34   | 0.76| 1   | 5   |
| CLEANLINESS_FOOD  | 661  | 4.13   | 4.11   | 0.82| 1   | 5   |
| CLEANLINESS_RESTROOM | 213  | 4.08   | 4.06   | 0.86| 1   | 5   |
| CLEANLINESS_ALL   | 81   | 4.2    | 4.19   | 0.75| 1   | 5   |
| AGE               | 516  | 4.11   | 4.13   | 1.56| 1   | 7   |
| GENDER            | 114  | 1.5    | 1.5    | 0.52| 1   | 3   |
| INCOME            | 494  | 2.62   | 2.62   | 1.12| 1   | 5   |
| LANG              | 2    | 1.06   | 1.06   | 0.35| 1   | 4   |

Figure 4: Stacked bar charts of demographic variables AGE and GENDER by the response variable overall satisfaction with SFO (RATE_ALL).
(AGE), and income (INCOME) are the least important predictors in this model, and overall SFO cleanliness (CLEANLINESS_ALL), signs and directions inside SFO (RATE_SIGN), artwork exhibitions (RATE_ART), and restaurants (RATE_FOOD) are the most important ones. Key drivers of overall satisfaction were obtained by successively removing predictors from the bottom of Figure 5: signs and directions inside SFO (RATE_SIGN), overall SFO cleanliness (CLEANLINESS_ALL), signs and directions on SFO airport roadways (RATE_ROADS), artwork exhibitions (RATE_ART), retail shops and concessions (RATE_STORE), restaurants (RATE_FOOD), airport rental car center (RATE_RENTAL) and accessing and using free Wi-Fi at SFO (RATE_WIFI).

Table 8 shows the multi-level confusion matrix, and Table 9 shows the binary confusion matrix for the final random forest model. The OOB accuracy of the final random forest model (74.6%) is very close to that of the full random forest model (75.5%). Figure 6 shows the variable importance of the predictors in the final random forest model.

**DISCUSSIONS AND IMPLICATIONS**

This study introduces the machine learning tool of random forest to tourism literature, and shows the applicability of this approach in determining drivers of passenger satisfaction using data from the 2014 SFO customer satisfaction survey. The methods used in this study (data imputation, random forest predictive model) and performance measures computed for multi-level response (precision, recall, F1) are taken from the machine learning literature and applied to analysis of SFO customer satisfaction data. These methods can be clearly applied to any modeling situation in which the response variable is multi-level, without transforming it to binary response, or using methods such as multiple linear regression which should not be used for ordinal data.

Generally, this study suggests that the key drivers of overall satisfaction at the SFO airport are artwork and exhibitions, restaurants, retail shops and concessions, signs and directions inside SFO, signs and directions inside SFO airport roadways, airport rental car center, accessing and using free Wi-Fi at SFO, and overall cleanliness of SFO. Among these key drivers, overall cleanliness of SFO, signs and directions inside SFO, artwork and exhibitions, and restaurants are regarded most important. Several limitations exist in this study. Study results cannot be generalized as data is from a single airport and from 2014 only. Moreover, there is no ‘typical’ airport in terms of services and facilities provided [26-29]: airports differ in size, infrastructure, service facilities etc., so not all airports may have all of the features at SFO. It is recommended to

| Predictor | Chi-square statistic | P-value |
|-----------|----------------------|---------|
| RATE_SIGN | 1458.1               | 0       |
| RATE_FOOD | 1347.0               | 0       |
| RATE_ROADS| 1303.4               | 0       |
| CLEANLINESS_ALL | 1376.3 | 0       |
| RATE_ART | 1224.5               | 0       |
| RATE_STORE| 1288.7               | 0       |
| RATE_RENTAL| 925.17              | 0       |
| RATE_WIFI| 729.34               | 0       |

Table 4: Results of the chi-square test of independence between the response and the potential predictors.

| Predictor | Chi-square statistic | P-value |
|-----------|----------------------|---------|
| AGE       | 34.98                | 0.096   |
| GENDER    | 6.81                 | 0.441   |

Table 5: Results of the chi-square test of independence between the response and the demographic variables AGE and GENDER.

| Predictor | Predictor Chi-square statistic | P-value |
|-----------|--------------------------------|---------|
| Observed Overall satisfaction | Unacceptable | Below Average | Average | Good | Outstanding |
| Unacceptable | 3 | 0 | 2 | 1 | 0 |
| Below Average | 0 | 5 | 29 | 5 | 0 |
| Average | 0 | 4 | 308 | 232 | 4 |
| Good | 0 | 0 | 87 | 1400 | 102 |
| Outstanding | 0 | 0 | 0 | 225 | 413 |

Table 6: Confusion matrix of the random forest model for 5-point Likert scale response RATE_ALL with all potential predictors.

| True Binary Overall Satisfaction | Observed Binary Overall Satisfaction |
|----------------------------------|-------------------------------------|
| 0                  | 0                        | 1 |
| 351                | 36                      | 6 |
| 87                 | 90                     | 2140 |

Table 7: Confusion matrix of the random forest model for RATE_ALL with all potential predictors for binary response (Unacceptable and Below Average = 0, Average, Good, or Outstanding = 1) obtained by collapsing rows and columns of Table 6.
Table 8: Confusion matrix of the random forest model for 5-point Likert scale response RATE_ALL using the final predictors.

| Observed Overall satisfaction | Predicted Overall satisfaction | Unacceptable | Below Average | Average | Good | Outstanding |
|------------------------------|--------------------------------|--------------|---------------|---------|------|-------------|
| Unacceptable                 | 3                              | 0            | 2             | 1       | 0    |             |
| Below Average                | 0                              | 8            | 25            | 6       | 0    |             |
| Average                      | 0                              | 3            | 308           | 232     | 5    |             |
| Good                         | 0                              | 1            | 112           | 1343    | 133  |             |
| Outstanding                  | 0                              | 0            | 0             | 196     | 442  |             |

Figure 5: Variable importance plot of the full random forest model, i.e., from random forest model for overall satisfaction as a function of all of the 22 selected predictors.

Figure 6: Variable importance plot from the final random forest model for overall satisfaction.
Table 9: Confusion matrix of the random forest model for RATE_ALL using the final predictors for binary response (Unacceptable and Below Average = 0, Average, Good, or Outstanding = 1) obtained by collapsing rows and columns of Table 8.

| Observed Binary Overall Satisfaction | True Binary Overall Satisfaction |
|-------------------------------------|----------------------------------|
|                                     | 0  | 1  |
| 0                                   | 349 | 7  |
| 1                                   | 113 | 2114 |

replicate this study for different years and different sizes of airports. Additionally, this study did not use the entire list of variables from the SFO survey. Future studies are encouraged to include a broader variety of predictor variables to determine the drivers of passengers’ overall satisfaction.

ACKNOWLEDGMENTS

Funding for this project was provided by the Caesars Foundation.

REFERENCES

1. Wang X, Hong M, Berger PD. Customer-satisfaction analysis at San Francisco international airport. International Journal of Management Studies. 2015;2(1):1-12.

2. Kwanisai, G, Vengesayi S. Destination attributes and overall destination satisfaction in Zimbabwe. Tourism Analysis. 2016;21(1):17-25.

3. Hess S, Polak JW. Mixed logit modelling of airport choice in multi-airport regions. Journal of Air Transport Management. 2005;11(2):59-68.

4. Eboli L, Mazzulla G. An ordinal logistic regression model for analysing airport passenger satisfaction. EuroMed Journal of Business. 2009;4(1):40-57.

5. Chen P, Gryschek O, Harshman G, McDonald J, Olson N. San Francisco International Airport – Understanding Customer Service, 2011.

6. Breiman L. Random Forest. Machine Learning. 2001;45:5-32.

7. Bogicevic V, Yang W, Bilgihan A, Bujisic M. Airport service quality drivers of passenger satisfaction. Tourism Review. 2013;68(4):3-18.

8. Bester BGL, Gomes CF. The effects of service quality dimensions and passenger characteristics on passenger’s overall satisfaction with an airport. Journal of Air Transport Management. 2015;44; 77-81.

9. Arif M, Gupta A, Williams A. Customer service in the aviation industry: An exploratory analysis of UAE airports. Journal of Air Transport Management. 2013;32: 1-7.

10. Chen J, Yu Y, Batnasan J. Services innovation impact to customer satisfaction and customer value enhancement in airport. Proceedings of PICMET ‘14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration. 2014;27-31.

11. Jiang H, Zhang Y. An assessment of passenger experience at Melbourne Airport. Journal of Air Transport Management. 2016;54:88-92.

12. Yeh C, Kuo Y. Evaluating passenger services of Asia-Pacific international airports. Transportation Research Part E: Logistics and Transportation Review. 2003;39(1):35-48.

13. Ofiac BS, Yumurtaci IO. Improving passenger satisfaction at airports: an analysis for Shortening baggage access time. Journal of Management, Marketing and Logistics. 2014;1(4): 339-347.

14. Rana MW, Lodhi RN, Butt GR, Dar WU. How Determinants of Customer Satisfaction are Affecting the Brand Image and Behavioral Intention in Fast Food Industry of Pakistan? J Tourism Hospit. 2017;6(6):316.

15. Yaru L, Liu X, Jing M. Study on the Quality of Service in Rural Homestay - Taking Shanli Lohas as an Example. J Tourism Hospit. 2018;7(4):370.

16. Yenida, Saad ZI, Chandra AR. Satisfaction Study of Padang Air Manis Beach Visitors Using Importance Performance Analysis. J Tourism Hospit. 2018;7(5):391.

17. San Francisco International Airport. Services & Amenities, 2017.

18. Boughiha DS, Sellaouti F. Multiple imputation using chained equations for missing data in TIMSS: a case study. Large-scale Assessments in Education. 2013;1(4):1-14.

19. van Buuren S, Groothuis-Oudshoorn K. Multivariate imputation by chained equations in R. Journal of Statistical Software. 2011;45(3): 1-67.

20. The R Project for Statistical Computing. What is R? 2017.

21. Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing. 2005;26(1):217-222.

22. Shi T, Seligson D, Belledegrun AS, Palotie A, Horvath S. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Modern Pathology. 2005;18(4):547-557.

23. Liaw A, Wiener M. Classification and Regression by pa. R News. 2002;2:18-22.

24. Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing. 2005;26(1):217-222.

25. Habermans SS. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. Journal of the American Statistical Association. 1988;83(402): 555-560.

26. Sokolova M, Lavalme G. A systematic analysis of performance measures for classification tasks. Information Processing and Management. 2009;45(4):427-437.

27. Guillem F, Hamilton HJ. Quality measures in data mining. Springer, New York, 2007.

28. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer, New York, 2013.

29. Graham A. Managing Airports: An International Perspective. (4th edn), Routledge, New York, 2013.