

CIND 820 Literature Review

Name: Nelly Grillo

Student number: 501144764

Supervisor: Ceni Babaoglu, Ph.D

Date: October 31st, 2022

Table of Contents

[Abstract 3](#_Toc118109232)

[Literature Review 6](#_Toc118109233)

[Descriptive Statistics of Selected Dataset 12](#_Toc118109234)

[Methodology 13](#_Toc118109235)

[GitHub Repository Link 14](#_Toc118109236)

[References 15](#_Toc118109237)

# Abstract

The Airline sector has played a unique role within the global transportation system providing economic growth and development in several countries around the world (Heiets & Xie, 2021).

Since the COVID-19 outbreak made its first appearance in China, it is well known that this pandemic has severely impacted in the world’s airline sector during the 2020-2021 period (Heiets & Xie, 2021). Regardless worldwide society is trying to recover from this COVID-19 pandemic, it is still ongoing and posing a continuous threat to both: global economy and public health (Olaganathan, 2021). Reports show that a recovery path is expected between 2022 and 2026 in the most pessimistic scenario (Sulu et al., 2021). Therefore, to accelerate the recovery of this business sector, governments around the world are currently lifting the restrictions that were meant to contain the spread of COVID-19 (Sonntag, 2021). At the same time, private sector is managing ways to enhance and expedite the sector revival. New behavioral passenger patterns have been developed and airline companies were in the need to adjust services at a whole different level (Heiets & Xie, 2021).

One main factor to help with the airline sector recovery, is to determine the level of passenger satisfaction (Sulu et al., 2021). The degree of content shown by costumers with airline services defined the passenger satisfaction (Suhartanto, 2014 and Khudhair et al., 2021). It is important for the airline sector to provide adequate and effective strategies to meet their customer requirements (Sulu et al., 2021 and Khudhair et al., 2021) and to encourage positive post-purchase behaviours which promotes the development of customer loyalty (Razafimanjary, 2019).

The purpose of this project is to:

* Determine the most important factors that impact the airline customer satisfaction
* Determine if customer loyalty have an effect on airline costumer satisfaction
* Determine if customer class play a key role in airline costumer satisfaction
* Use visualizations to understand the correlation between the features which impact the airline costumer satisfaction
* Predict customer satisfaction using Machine learning (ML) algorithms
* Determine which ML algorithm is the most accurate

The dataset used to solve the problem of this project is *Airline passenger satisfaction measurement and service quality improvement* obtained from *Kaggle*. The dataset has both categorical and numerical attributes and includes 24 features. Moreover, the dataset has more than 100, 000 instances or rows.

Python is the tool planned to use for this project. The techniques of Supervised Machine Learning intended to use in this project are Classification and Regression. The algorithms proposed are tree and non-tree based used for model building and prediction:

* Decision Tree: easy to interpret, visualization ability, works with both categorical and

numerical data (Liberman, 2020)

* Random Forest: strong modeling technique, performs feature selection, works with both

categorical and numerical data, high accuracy (Singh, 2020)

* Naïve Bayes: simple and easy to implement, can be used to make real-time predictions

(Soni, 2020)

* K-Nearest Neighbors: no training period, easy implementation (Soni, 2020)
* Logistic Regression: easy to implement, interpret and efficient to train, good accuracy

(Pareek, 2021)

* XGBoost: highly flexible, fast to interpret (Abhishek, 2020)

Furthermore, this project will show the Exploratory data analysis (EDA) and feature visualizations.

# Literature Review

Airline customer satisfaction and customer loyalty are two important variables that the airline service industry must consider providing adequate and effective strategies to meet their customer requirements, and to promote customer return (Maminiaina, 2019).

**Airline customer satisfaction**

Considering that an airline passenger is a customer in the airline industry, satisfaction can be seen as a complex level of the customer knowledge and experience when comparing the service performance with their prior expectations. (Noviantoro & Huang, 2022; Jiang & Zhang, 2016)

There are many factors that make customer satisfaction such as: product quality, price, service quality, emotional factor, and cost and convenience (Maminiaina, 2019).

In terms of marketing, the customer satisfaction highly influences the future consumer purchase behavior, profitability, and shareholder value (Suhartanto & Noor, 2012).

**Customer loyalty**

One important factor for the airline industry to preserve business stability is the customer loyalty.

Customer loyalty is described as customers who return to the same organization because they are very satisfied with the product or service (Maminiaina, 2019)

**Airline service quality**

Service quality is defined as an indispensable approach to achieve business survival and success because it can persuade customer purchase behaviour and business performance (Suhartanto & Noor, 2012).

To achieve and provide high-quality service, it is necessary for the airline industry to understand the customer needs (Jiang & Zhang, 2016). The airline customer is the one who defines the service quality by comparing their expectations and experiences (Noviantoro & Huang, 2022). When the airlines offer better quality services, it is possible to achieve customer loyalty (Jiang & Zhang, 2016).

The airline service can be seen as a group of services which can be divided on pre-flight services and in-flight services (Gao et al., 2021; Etemad-Sajadi, et al., 2016).). The pre-flight service comprises attributes such as online boarding, gate location, ease of online booking and on-board service. In addition, the in-flight service includes attributes such as seat comfort, in-flight wi-fi service and entertainment, legroom, and cleanliness.

There are several investigations on service quality that explore the problems related to the service quality in the airline industry. Some of these studies that assess the airline customer satisfaction levels uses traditional statistical testing or multiple-criteria methods (Noviantoro & Huang, 2022).

Etemad-Sajadi, et al., (2016) focus on the customer perception of the pre-flight and in-flight service quality. This study shows that there is a positive direct effect between the pre-flight service quality perceived by the airline customer and the airline customer satisfaction. Moreover, there is a positive direct and indirect effect on the airline customer loyalty which can be improved by enhancing the pre-flight quality service.

Suhartanto and Noor (2012) aims to aims to investigates how service quality and price affect customer satisfaction in low-cost airlines and full-service airlines. By applying SERVPERF, this study concludes that the low-cost airlines use price as competitive strategy, while full-service airlines depend on quality of services. Moreover, this study highlights that both types of airlines should consider as factors: the attitude of employees when delivering the service and the price.

There are other studies that investigate the role of the customer class on the airline customer satisfaction. In this study, An and Noh (2009) concludes that there is a different perception of the service quality according to the customer seat class which can be explain in terms of a difference in income and professional status. Moreover, this study shows that the in-flight service quality impacts the customer loyalty, and the extent of the impact depends on the seat class (An & Noh, 2009). On the other hand, Jiang and Zhang (2016) suggests that, although, there is a positively association between the customer satisfaction and the customer loyalty, this is not greatly reflected among business travellers.

Using the machine learning approach, it is common to study the airline customer satisfaction by applying the sentiment analysis, which analyze tweets to detect positive or negative customer satisfaction (Khan & Urolagin, 2018).

For example, Khan and Urolagin, 2018 analyze tweets using sentiment analysis and provides a general opinion of airline customer satisfaction. Furthermore, Khan & Urolagin, 2018 use three classifiers to perform consumer loyalty prediction.

Kumar and Zymbler (2019) uses sentiment classification approaches such as SVM, ANN, and CNN. Also, this study recommends applying association analysis to provide suggestions to improve the customer satisfaction.

There are other studies that apart from applying sentiment analysis, also use more advanced methods of machine learning approach.

Tan, C. (2021) aims to classify customers satisfaction level by applying multiples machine learning models such as KNN, Logistic Regression, Decision Tree, Random Forest, and several boosting models including Gradient Boosting, Adaboost, Xgboost, LGBM and Catboost. In addition, it applies sentiment analysis on tweets and Bidirectional LSTM model to identify emotional tendency of tweets.

It concludes that the most important factors that significantly affect the airline customer satisfaction includes online boarding, in-flight wi-fi service, in-flight entertainment, and seat comfort.

Nowadays, it is usually to apply machine learning methods for customer analysis and marketing.

Machine learning and deep learning analysis allows to work with complex multi-dimensional data and enables to determine hidden correlations and important insights from the mentioned data (Park et al., 2022).

Noviantoro & Huang (2022) propose a study that uses Data mining method to examine the airline passenger satisfaction focusing on investigating the passenger most desired airline services by using feature selection. This study runs and compares classification algorithms such as decision tree, random forest, gradient boosted tree, k-NN, Naïve Bayes, rule induction, logistic regression, neural net, deep learning, and support vector machine. This study concludes that the most critical features that the airline should improve to obtain passenger satisfaction are online boarding, inflight wi-fi service, baggage handling, and inflight entertainment.

Park et al. (2022) aim to demonstrates the prediction of customer churn risk and satisfaction by using deep learning models. Furthermore, it compares deep learning model with machine learning models concluding that, in terms of accuracy, CNN-LSTM outperformed the machine learning models.

Research from Gao et al. (2021) uses various machine learning algorithms and interpretation techniques to understand the different factors that impact the airline travel satisfaction. It compares machine learning algorithms and conventional logistic regression methods and determines that machine algorithms such as Multi-layer Perceptron, Random Forest and Support Vector Machine demonstrate superior modeling and predictive performances for modeling airline travel satisfaction. Moreover, it determines the key factors that have greater impacts on airline travel satisfaction such as type of travel, class and customer type, Wi-Fi service, online boarding, and baggage handling.

García et at. (2019) predicts airline customer satisfaction by using a combination of k-NN and ensemble regression model (BAGGING).

Hayadi et al. (2021) applies various classification algorithms such as k-NN, Logistics Regression, Gaussian Naïve, Decision Tree, Random Forest to determine airline customer satisfaction.

This study recommends optimizing the in-flight wi-fi service experience by making it more accessible or by lowering its cost.

Bellizzi et al. (2022) aims to analyse the highly educated people’s satisfaction with airline services by applying the Classification and Regression Tree (CART) approach. In addition, this study uses an Importance Performance Analysis (IPA) which permits to identify the most critical factors to be considered by the airline company.

Although there are some studies that applies various machine learning algorithms to determine the level of airline customer satisfaction. This study uses a different airline customer dataset and aims to determine the most important factors that impact the airline customer satisfaction, focusing in establish if customer loyalty and customer class have a particular effect on customer satisfaction. Moreover, this study intends to apply and compare various Classification and Regression algorithms to determine which is the most accurate.

# Descriptive Statistics of Selected Dataset

The dataset has 24 features for 103,904 instances. There is an “Unnamed” column that has no relevant information; therefore, it has been dropped.

There are 18 quantitative variables and 5 categorical variables including the target variable “Satisfaction”, that shows the level of satisfaction of the airline passenger: satisfied, neutral or dissatisfied.

Table 1. Descriptive Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Type | Mean | Std | Min | Max |
| Gender | Categorical |  |  |  |  |
| Customer Type | Categorical |  |  |  |  |
| Age | Numerical | 39.380466 | 15.113125 | 7.0 | 85.0 |
| Type of Travel | Categorical |  |  |  |  |
| Class | Categorical |  |  |  |  |
| Flight Distance | Numerical | 1189.325202 | 997.297235 | 31.0 | 4983.0 |
| Inflight Wi-Fi service | Numerical | 2.729753 | 1.327866 | 0.0 | 5.0 |
| Departure/Arrival time convenient | Numerical | 3.060081 | 1.525233 | 0.0 | 5.0 |
| Ease of Online booking | Numerical | 2.756984 | 1.398934 | 0.0 | 5.0 |
| Gate location | Numerical | 2.977026 | 1.277723 | 0.0 | 5.0 |
| Food and drink | Numerical | 3.202126 | 1.329401 | 0.0 | 5.0 |
| Online boarding | Numerical | 3.250497 | 1.349433 | 0.0 | 5.0 |
| Seat comfort | Numerical | 3.439765 | 1.318896 | 0.0 | 5.0 |
| Inflight entertainment | Numerical | 3.358341 | 1.333030 | 0.0 | 5.0 |
| On-board service | Numerical | 3.382609 | 1.288284 | 0.0 | 5.0 |
| Leg room service | Numerical | 3.351401 | 1.315409 | 0.0 | 5.0 |
| Baggage handling | Numerical | 3.631687 | 1.181051 | 1.0 | 5.0 |
| Check-in service | Numerical | 3.304323 | 1.265396 | 0.0 | 5.0 |
| Inflight service | Numerical | 3.640761 | 1.175603 | 0.0 | 5.0 |
| Cleanliness | Numerical | 3.286397 | 1.312194 | 0.0 | 5.0 |
| Departure Delay in Minutes | Numerical | 14.747939 | 38.116737 | 0.0 | 1592.0 |
| Arrival Delay in Minutes | Numerical | 15.178678 | 38.698682 | 0.0 | 1584.0 |
| Satisfaction | Categorical |  |  |  |  |

# Methodology

Table 2 shows a flow chart of the methodology that will be done throughout this capstone project.

Table 2. Methodology

Load the Data

Data Preparation/Exploration

Data Cleaning

Exploratory Data Analysis (EDA)

Dataset for Analysis

Data Visualization

Data Splitting

Training Models

* Decision Tree
* Random Forest
* Naïve Bayes
* K-Nearest Neighbors
* Logistic Regression
* XGBoost

Models Training and Evaluation

Test Data

Prediction

Evaluation

Model

Training Models

Train Data

# GitHub Repository Link

https://github.com/nvgril/CIND820-Project\_2022

# References

1. Abhishek, M. (2020, February 27). *XGBoost an efficient implementation of gradient boosting*. Agency.

https://datascience.foundation/datatalk/xgboost-an-efficient-implementation-of-gradient-boosting

1. An, M., & Noh, Y. (2009). Airline Customer Satisfaction and loyalty: Impact of in-flight service quality. *Service Business*, *3*(3), 293–307.

https://doi.org/10.1007/s11628-009-0068-4

1. Bellizzi, M. G., Eboli, L., Mazzulla, G., & Postorino, M. N. (2022). Classification trees for analysing highly educated people satisfaction with airlines’ services. *Transport Policy*, *116*, 199–211.

https://doi.org/10.1016/j.tranpol.2021.12.008

1. Etemad-Sajadi, R., Way, S. A., & Bohrer, L. (2016). Airline passenger loyalty: the distinct effects of airline passenger perceived pre-flight and in-flight service quality. *Cornell Hospitality Quarterly*, *57*(2), 219–225.

https://doi.org/10.1177/1938965516630622

1. Gao, K., Yang, Y., & Qu, X. (2021). Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. *Transportation Research Part D: Transport and Environment*, *97*, 102957.

https://doi.org/10.1016/j.trd.2021.102957

1. Garcia, V., Florencia-Juarez, R., Sanchez-Solis, J. P., Rivera-Zarate, G., & Contreras-Masse, R. (2019). Predicting airline customer satisfaction using K-NN Ensemble regression models. *Research in Computing Science*, *148*(6), 205–215.

https://doi.org/10.13053/rcs-148-6-15

1. Hayadi, B., Kim, J., Hulliyah, K., & Sukmana, H. (2021). Predicting airline passenger satisfaction with classification algorithms. IJIIS: International Journal of Informatics and Information Systems, 4(1), 82–94.

https://doi.org/10.47738/ijiis.v4i1.80

1. Heiets, I., Xie, Y. (2021). The impact of the COVID-19 pandemic on the Aviation Industry. *Journal of Aviation*.

https://doi.org/10.30518/jav.933296

1. Jiang, H., & Zhang, Y. (2016). An investigation of service quality, customer satisfaction and loyalty in China's airline market. *Journal of Air Transport Management*, *57*, 80–88. https://doi.org/10.1016/j.jairtraman.2016.07.008
2. Khan, R., & Urolagin, S. (2018). Airline sentiment visualization, consumer loyalty measurement and prediction using Twitter data. *International Journal of Advanced Computer Science and Applications*, *9*(6).

https://doi.org/10.14569/ijacsa.2018.090652

1. Khudhair, H.Y., Jusoh, A., Nor, K.M. and Mardani, A. (2021) Price sensitivity as 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*, Vol. 11, Nos. 2/3, pp.114–125. <https://doi.org/10.1504/ijbcrm.2021.116274>
2. Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, *6*(1).

https://doi.org/10.1186/s40537-019-0224-1

1. Liberman, N. (2020, May 21). *Decision trees and random forests*. Medium. https://towardsdatascience.com/decision-trees-and-random-forests-df0c3123f991
2. Maminiaina, R. (2019). A thorough literature review of customer satisfaction definition, factors affecting customer satisfaction and measuring customer satisfaction. *International Journal of Advanced Research*, 7(9), 828–843.

<https://doi.org/10.21474/ijar01/9733>

1. Noviantoro, T., & Huang, J.-P. (2022). Investigating airline passenger satisfaction: Data Mining Method. *Research in Transportation Business &* *Management*, 43, 100726. <https://doi.org/10.1016/j.rtbm.2021.100726>
2. Olaganathan, R. (2021). Impact of covid-19 on airline industry and strategic plan for its recovery with special reference to data analytics technology. *Global Journal of Engineering and Technology Advances*, 7(1), 033–046. https://doi.org/10.30574/gjeta.2021.7.1.0050
3. Pareek, P. (2021, September 3). *Logistic regression: Essential things to know*. Medium. https://medium.datadriveninvestor.com/logistic-regression-essential-things-to-know-a4fe0bb8d10a
4. Park, S.-H., Kim, M.-Y., Kim, Y.-J., & Park, Y.-H. (2022). A deep learning approach to analyze airline customer propensities: The case of South Korea. *Applied Sciences*, *12*(4), 1916.

https://doi.org/10.3390/app12041916

1. Razafimanjary M. A. (2019). A thorough literature review of customer satisfaction definition, factors affecting customer satisfaction and measuring customer satisfaction. *International Journal of Advanced Research*, 7(9), 828–843. <https://doi.org/10.21474/ijar01/9733>
2. Singh, J. (2020, December 26). *Random Forest: Pros and cons*. Medium. <https://medium.datadriveninvestor.com/random-forest-pros-and-cons-c1c42fb64f04>
3. Soni, A. (2020, July 3). *Advantages and disadvantages of KNN*. Medium. <https://medium.com/@anuuz.soni/advantages-and-disadvantages-of-knn-ee06599b9336>
4. Soni, A. (2020, July 4). *Pros and cons of naive bayes classifier*. Medium. <https://medium.com/@anuuz.soni/pros-and-cons-of-naive-bayes-classifier-40b67249ae8>
5. Sonntag, T. (2021, December 15). *How did covid-19 impact the aviation industry*. CUNY Academic Works. <https://academicworks.cuny.edu/bb_etds/122/>
6. Suhartanto, D. (2014, June 8). Customer satisfaction in the airline industry: *The role of service quality and Price*. Academia.edu. <https://www.academia.edu/3215480/Customer_Satisfaction_in_the_Airline_Industry_The_Role_of_Service_Quality_and_Price>
7. Suhartanto, D. & Noor, A.A. (2012). Customer Satisfaction in the Airline Industry: The Role of Service Quality and Price. *Asia Tourism Forum Conference*. 1–9.
8. Sulu, D., Arasli, H., Saydam, M. B. (2021). Air-travelers’ perceptions of service quality during the COVID-19 pandemic: Evidence from Tripadvisor Sites. *Sustainability*, 14(1), 435. <https://doi.org/10.3390/su14010435>
9. Tan, C. (2021). Bidirectional LSTM model in predicting satisfaction level of passengers on airline service. *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*.

<https://doi.org/10.1109/icaice54393.2021.00107>

1. Vishalmendekarhere. (2021, January 22). It's all about assumptions, Pros &amp; Cons. Medium. https://medium.com/swlh/its-all-about-assumptions-pros-cons-497783cfed2d