A Data Analytics Approach: Investigating Factors Influencing Airline Customer Satisfaction

Aarshi Jain

501178324

Prof. Tamer Abdou

May 20th, 2024

An abstract is a brief summary of the capstone project. The abstract should contain:

- A brief context about the problem and the theme(s) chosen
- The problem to be solved (e.g., the research questions or the summary of research questions)
- The data to be used
- The techniques (e.g., classification, clustering, text mining, model evaluation etc.) and the tools being proposed to solve the stated problem

The abstract should:

- Include a cover page that includes the student's name, student number, supervisor's name, and date of submission
- Be approximately 500 words (excluding the cover page and references), double-spaced, Times New Roman 12-pt font, using APA writing conventions where appropriate
- Have a references page, using APA style

Airline customer satisfaction: Airline cust. Satisfaction

Every year, 8.6 billion passengers fly globally (*Global Passenger Traffic*, n.d.), making it crucial for the airline industry to prioritize customer satisfaction as a competitive advantage. This capstone project delves into the factors influencing customer satisfaction within an undisclosed airline company, utilizing the detailed dataset titled "Airline Customer Satisfaction". The dataset is constructed of 22 columns and 129,880 rows, providing a strong foundation for predictive analytics aimed at identifying customer satisfaction factors (Huseyn, n.d.).

The primary research question guiding this investigation is: Which factors are the most pivotal in predicting customer satisfaction in the airline industry? Secondary queries include researching demographic variables such as age, travel class, type of travel, and customer type to understand their correlations with both high and low satisfaction ratings. Additionally, we explore which aspects of airline service—such as seat comfort, in-flight entertainment, cleanliness, and food and drink—significantly influence overall customer satisfaction.

To address these questions, the Airline Customer Satisfaction dataset offers comprehensive insights into customer experiences, spanning factors like customer type, age, purpose of travel, and travel class. Furthermore, it provides specific ratings for a number of services, including seat comfort, convenience of the gate location, departure/arrival time, baggage handling, and more. This information-rich dataset allows for a multifaceted analysis of customer satisfaction factors.

The analytical techniques utilized in this investigation encompass classification and regression models, clustering, and text mining. Classification algorithms such as decision trees, and logistic regression are utilized to predict each customer's satisfaction level based on their characteristics and travel experiences (Logistic Regression, n.d.). Clustering techniques, like K-means, aid in pinpointing different groups of customers, each with their own specific

determinants that contribute to their satisfaction (Webster, n.d.). Text mining of the feedback will provide qualitative insights into customer sentiments (*Text Mining*, n.d.).

Additionally, model evaluation metrics such as precision, accuracy, and F1-score are employed “to assess the performance and effectiveness of a statistical or machine learning model” (Srivastava, 2024). Tools such as R and Python, along with specialized libraries like Scikit-Learn and the “tm” package in R, are utilized for data analysis as well as model building (Choudhary, 2022).

By conducting an investigation into this dataset, airlines can gather critical insights into the determinants that contribute to customer satisfaction, enabling them to tailor their services effectively and enhance the overall customer experience. The research will provide proactive insights for airline companies seeking to strengthen customer loyalty and improve service quality, thereby gaining a competitive edge in the market.

References

Choudhary, A. (2022, September 9). *15 Most Important Features of Scikit-Learn!* Analytics Vidhya. Retrieved May 16, 2024, from https://www.analyticsvidhya.com/blog/2021/07/15-most-important-features-of-scikit-learn/

GLOBAL PASSENGER TRAFFIC EXPECTED TO RECOVER BY 2024 AND REACH 9.4 BILLION PASSENGERS. (n.d.). Airports Council International. Retrieved May 14, 2024, from https://aci.aero/2023/09/27/global-passenger-traffic-expected-to-recover-by-2024-and-reach-9-4-billion-passengers/#:~:text=Passenger%20traffic%20forecast%202023%E2%80%932024,94.2%25%20of%20the%202019%20level

Huseyn, R. (n.d.). *Airline Customer Satisfaction*. Kaggle. Retrieved May 14, 2024, from https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction/data

Srivastava, T. (2024, January 8). *12 Essential Evaluation Metrics for Evaluating ML Models*. Analytics Vidhya. Retrieved May 15, 2024, from https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

Webster, W. (n.d.). *Cluster Analysis: Definition and Examples*. Qualtrics. Retrieved May 14, 2024, from https://www.qualtrics.com/experience-management/research/cluster-analysis/

What is Logistic Regression Used for? (n.d.). H2O.ai. Retrieved May 14, 2024, from https://h2o.ai/wiki/logistic-regression/

What Is Text Mining? (n.d.). IBM. Retrieved May 15, 2024, from https://www.ibm.com/topics/text-mining

Understanding customer satisfaction is a critical objective for the airline industry, where competitive advantage hinges on the ability to deliver superior service quality. This capstone project delves into the factors influencing customer satisfaction within an undisclosed airline company, utilizing a comprehensive dataset titled "Airline Customer Satisfaction." The dataset encompasses 22 features and 129,880 entries, offering a robust foundation for predictive analytics aimed at discerning the determinants of customer satisfaction.

The primary research question guiding this project is: What are the key predictors of customer satisfaction in the airline industry? Secondary questions include: How do different customer demographics and travel circumstances affect satisfaction levels? Which aspects of airline service (e.g., seat comfort, in-flight entertainment, punctuality) most significantly influence overall satisfaction?

To address these questions, the dataset provides detailed insights into customer experiences, categorized by various factors such as customer type (loyal or disloyal), age, type of travel (business or personal), class of travel (business, economy, or economy plus), and specific service ratings (e.g., seat comfort, food and drink quality, baggage handling, etc.). This rich dataset allows for a multifaceted analysis of customer satisfaction determinants.

The analytical techniques employed in this research include classification and regression models, clustering, and text mining. Classification algorithms such as logistic regression, decision trees, and random forests are used to predict whether a customer will be satisfied based on their profile and travel experience. Clustering techniques help identify distinct customer segments with unique satisfaction drivers. Text mining of open-ended feedback (if available) provides qualitative insights into customer sentiments.

Model evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the performance of predictive models. Tools such as Python, R, and specialized libraries like scikit-learn and TensorFlow are utilized to perform the data analysis and model building.

Preliminary results indicate that factors such as flight punctuality, seat comfort, and the quality of in-flight services (wifi, entertainment, and food) are significant predictors of customer satisfaction. Additionally, demographic factors like age and customer type (loyalty status) also play a crucial role. These findings suggest that targeted

improvements in specific areas of service could substantially enhance overall customer satisfaction.

The implications of this research are twofold: Firstly, it provides actionable insights for airline companies seeking to enhance customer experience and loyalty. Secondly, it contributes to the broader academic understanding of service quality determinants in the airline industry. By leveraging data analytics, airlines can strategically focus their efforts on the most impactful areas, thereby improving service delivery and gaining a competitive edge in the market.