KDD-Based Predictive Modeling of Air Passenger Satisfaction using AutoML

Keerthana Parsa *Software Engineering Dept.*
San Jose State University
CA, USA
keerthana.parsa@sjsu.edu

***Abstract*—The ever-evolving aviation industry has positioned passenger satisfaction as a cornerstone for achieving operational excellence and sustaining market competitiveness. This intricate and comprehensive research paper embarks on a journey to explore, analyze, and predict passenger satisfaction using cutting-edge Automated Machine Learning (AutoML) with the PyCaret library, meticulously adhering to the Knowledge Discovery in Databases (KDD) methodology. The insights garnered from this elaborate exploration aim to act as a catalyst for reshaping and innovating customer service strategies, thereby elevating the overall passenger experience in air travel.**

***Index Terms*—Automated Machine Learning, PyCaret, Air Passenger Satisfaction, Predictive Modeling, Knowledge Discovery in Databases, Customer Experience, Aviation Industry Innovation.**

I. Introduction

THE aviation industry, a dynamic and rapidly evolving sector, has amplified the importance of passenger satisfaction, making it a critical factor influencing the success trajectories of airline companies. This expansive research is conceived with the intent to provide a profound and intricate understanding of the myriad factors influencing passenger satisfaction.

A. Motivation and Objective

The primary motivation behind this research is to unravel the complexities surrounding passenger satisfaction and to contribute substantial insights and reliable predictive models that can guide airlines in optimizing their service delivery mechanisms, thus enhancing overall passenger experiences and satisfaction levels.

II. Comprehensive Literature Review and Background

The pursuit of understanding passenger satisfaction has spurred numerous research studies, each contributing a wealth of insights into various aspects of airline services and passenger perceptions.

1. [Reference Paper 1] offered a comprehensive exploration of service quality dimensions, correlating them with passenger satisfaction and providing a detailed understanding of passenger preferences and expectations.

2. [Reference Paper 2] employed advanced machine learning models to predict passenger satisfaction and highlighted the critical role of meticulous feature selection, engineering, and optimization in attaining high-performing predictive models.

3. [Reference Paper 3] conducted a detailed investigation into the relationship between passenger satisfaction and brand loyalty, revealing intricate dynamics and offering valuable strategies for enhancing customer retention and brand loyalty.

4. [Reference Paper 4] deployed cutting-edge deep learning methodologies, demonstrating the capabilities of neural networks in deciphering complex, non-linear relationships within extensive and multifaceted datasets.

III. Detailed Methodological Framework

This research adopts a rigorous and systematic approach, based on the KDD methodology, ensuring that each phase of the study is conducted with utmost precision and attention to detail, to extract reliable and insightful knowledge from the dataset.

A. In-depth Data Understanding and Exploration

An extensive examination and exploration of the dataset were conducted to uncover underlying patterns, assess the distribution of variables, and identify potential anomalies, providing a foundational understanding for subsequent phases.

B. Meticulous Data Preprocessing and Transformation

Comprehensive strategies were employed to handle missing values, potential outliers were scrutinized and treated appropriately, and PyCaret was leveraged for efficient data transformations, ensuring the development of a solid groundwork for model development.

C. Exhaustive Model Development and Optimization

Various models were evaluated, and the most promising candidate underwent further optimization through hyperparameter tuning to enhance its predictive accuracy, reliability, and generalization capability.

D. Rigorous Model Evaluation and Interpretation

The final model was subjected to a series of rigorous evaluations and interpretations using a plethora of metrics and visual tools, allowing for a thorough understanding of its performance, reliability, and limitations.

IV. Results, Insights, and Comprehensive Discussions

The use of AutoML and PyCaret facilitated a streamlined and efficient modeling process. The optimized model, after undergoing extensive evaluations, exhibited high levels of accuracy and reliability.

A. Detailed Analysis of Feature Importance

Detailed analyses of feature importance revealed critical areas influencing passenger satisfaction, such as 'Inflight entertainment' and 'Seat comfort,' providing crucial insights for airlines to allocate resources and refine services strategically.

V. Challenges and Limitations

While this research provides substantial insights and a reliable predictive model, it is crucial to acknowledge the challenges and limitations encountered during the study, such as the handling of imbalanced datasets, the treatment of outliers, and the generalizability of the model to diverse airline operational settings.

VI. Potential Applications

The insights and predictive models developed in this research have vast potential applications, including real-time passenger feedback systems, dynamic allocation of airline resources, and the development of personalized passenger experiences based on individual preferences and expectations.

VII. Conclusion, Implications, and Future Directions

This elaborate study underscores the transformative potential of Automated Machine Learning in air passenger satisfaction. The insights derived have significant implications for airlines, offering a strategic direction for service improvement and customer experience enhancement.

A. Implications

The insights obtained from this research have the potential to revolutionize customer service strategies within the airline industry, offering a comprehensive roadmap for enhancing passenger experiences and satisfaction levels.

B. Future Directions

Future research avenues could explore more advanced models and methodologies, incorporating additional data sources like passenger feedback and sentiment analysis, and exploring the potential of ensemble and deep learning models for a more comprehensive understanding of passenger satisfaction dynamics.

Acknowledgment

The authors express their profound gratitude to [Your Institution or Collaborators] for their unwavering support, insightful discussions, and substantial contributions to this extensive research endeavor.