Evaluating Passenger Segment Sensitivity and Reliability in Airline Satisfaction AI Systems

Miloš Stojanović
Department of ICT
The Academy of Applied Technical and Preschool Studies
Niš, Serbia
milos.stojanovic@akademijanis.edu.rs

Milena Nikolić
Department of ICT
The Academy of Applied Technical and Preschool Studies
Niš, Serbia
milena.nikolic@akademijanis.edu.rs

***Abstract*—Airline passenger satisfaction has been extensively examined using machine learning techniques, yet limited attention has been given to how artificial intelligence systems perform across different passenger profiles. The proposed approach is evaluated using a publicly available Airline Passenger Satisfaction dataset from Kaggle, comprising categorical, ordinal, and numerical features that support evaluation across passengers. After data preprocessing, normalization, and categorical encoding, supervised classification models are trained and assessed across passengers defined by travel purpose, service class, and flight distance. System behavior is evaluated using accuracy, macro F1-score, false negative rate, and feature importance stability. Feature sensitivity analysis is applied to examine whether feature contributions remain consistent or vary across passenger groups. Experimental results achieve overall accuracy above 85%, exceeding 90% for neural network models, while the analysis reveals substantial variation in reliability, particularly in false negative rates, together with corresponding shifts in dominant feature contributions across passenger groups. These findings indicate that models with strong aggregate performance may still exhibit uneven reliability across passenger populations. From a business perspective, the results highlight the value of evaluating passenger categories separately to support customer experience management, targeted service improvement, and informed planning within airline organizations.**

***Index Terms*—artificial intelligence, supervised machine learning, airline satisfaction, passenger segmentation, model reliability**

I. Introduction

The airline industry increasingly relies on artificial intelligence (AI) and machine learning (ML) techniques to analyze passenger satisfaction and improve operational decisions. Accurate satisfaction prediction plays an important role in service quality monitoring, customer retention, and operational planning in a highly competitive environment. Predictive analytics enable the combined analysis of passenger feedback and operational attributes to support informed decision-making and enhance customer experience [1].

Airline passenger satisfaction has been analyzed from multiple analytical perspectives. Multidimensional modeling has been utilized to capture the influence of factors including in-flight comfort, staff interaction, and operational reliability [2]. In parallel, sentiment analysis approaches successfully extract passenger opinions from unstructured sources, including social media and online platforms such as Twitter [3].

Despite progress in predictive modeling, an important limitation remains insufficiently explored. Most existing studies evaluate model performance using global metrics and implicitly assume uniform behavior across passengers. As a result, models with strong average performance may still behave inconsistently across passenger segments, leading to missed dissatisfaction signals in service monitoring contexts where early intervention is most valuable.

This paper addresses this challenge by proposing a reliability auditing framework for satisfaction AI systems, in which a single trained classifier is evaluated under realistic heterogeneous passenger conditions. By combining supervised learning with systematic performance and error analysis across passenger categories, the study provides deeper insight into model behavior and supports more informed use of AI for airline service quality assessment.

II. Related Work

A. Machine Learning for Airline Passenger Satisfaction

Iskandar *et al.* present a systematic approach where airline passenger satisfaction prediction is embedded in airline service and communication systems, demonstrating how supervised classifiers such as support vector machines and random forest models can be employed to identify dissatisfied passengers and highlight key service drivers [4]. However, model evaluation is primarily reported using aggregate performance measures, which may obscure differences in predictive behavior across distinct passenger categories.

Classical supervised learning models continue to play an important role in this domain. Decision tree and Naïve Bayes based studies show that interpretable classifiers can effectively identify key satisfaction drivers such as service quality indicators, delays, and onboard amenities [9]. These models are attractive due to their transparency and ease of deployment, but their reliance on relatively simple decision boundaries limits their ability to capture complex interactions among service factors. To mitigate this limitation, analyses based on decision tree structures have been used to rank influential attributes and support explanation of model outputs [10]. While such analyses improve interpretability, they typically rely on global importance measures, which may not fully reflect variation in passenger behavior.

More recent work explores deep learning models to improve predictive accuracy by learning richer representations of service characteristics. An optimized deep learning approach applied to airline satisfaction data demonstrates that neural networks can outperform traditional classifiers when combined with appropriate preprocessing and optimization strategies. These findings suggest improved prediction, but transparency and consistency remain limited by aggregate evaluation [7].

From a business intelligence (BI) perspective, fuzzy genetic systems have been proposed to balance prediction accuracy with interpretability. The fuzzy genetic approach combines rule-based reasoning with optimization mechanisms to develop predictive models characterized by interpretable rule sets and structured feature hierarchies [8]. This is considered relevant in managerial settings, where interpretability is as significant as predictive accuracy. However, explanations are generally evaluated in aggregated form, leaving uncertainty about their consistency across passenger populations.

B. Passenger Segmentation in Service Analytics

Passenger segmentation represents an important analytical approach in airline services, as passenger expectations and satisfaction drivers differ across travel contexts. A clustering approach using K-mode and XGBoost demonstrates how behavioral attributes can be used to identify distinct passenger groups and support targeted analysis [9]. This work highlights the value of acknowledging heterogeneity within airline customers and provides a structured representation of passenger populations beyond average trends.

Despite these advances, segmentation is often treated as a profiling or marketing tool rather than a foundation for evaluating predictive systems. In many studies, passenger groups are identified, but prediction performance and error behavior continue to be reported at an aggregate level. Consequently, the relationship between passenger grouping and model reliability remains insufficiently explored, particularly across different travel purposes, service classes, and flight characteristics.

The present study builds on segmentation research by examining model behavior within passenger groups rather than using segmentation solely for descriptive purposes. By extending segmentation from population description to system behavior analysis, the research supports a more realistic assessment of airline satisfaction prediction models.

C. Reliability and Explainability in AI Systems

Reliability and explainability are important considerations in airline satisfaction AI systems, especially when model outputs are used to inform managerial assessment and service improvement activities. In this paper, the focus is not on proposing new explanation methods but on analyzing whether feature influence patterns remain stable within passenger categories. Interpretable classifiers, including decision tree-based models, provide transparent decision logic, but their explanations are often derived from aggregated data and tend to reflect average behavior rather than variation across passenger contexts.

Recent studies have begun to integrate explainable AI techniques with classification models to enhance interpretability of satisfaction predictions, demonstrating the practical value of explanation methods in applied analytics [10]. Explanations in these studies are still examined in aggregated form, with limited analysis of whether explanation patterns remain consistent across different passenger categories. This limitation is important for false negatives, where dissatisfied passengers are classified as satisfied, potentially delaying identification of service issues and reducing the effectiveness of targeted improvement actions.

In contrast, the present study treats reliability and explainability as evaluation properties that must be examined across passenger categories. By examining variation in performance measures, error patterns, and explanation outcomes across passenger groups, the proposed methodology provides a more detailed view of system behavior. This perspective addresses a gap not adequately covered by existing machine learning, segmentation, or business intelligence solutions and supports more trustworthy application of airline satisfaction AI systems.

III. Data Selection and Passenger Segmentation

A. Airline Passenger Satisfaction Dataset

This study uses the public *Customer Satisfaction in Airline* dataset from Kaggle [11]. The dataset contains 129,880 passenger records with 22 variables, including categorical passenger descriptors, ordinal service quality ratings, and numerical operational factors like flight distance and delays. The target label is binary (`satisfied` vs. `dissatisfied`). Only the `Arrival Delay in Minutes` attribute contains missing values (393 records, about 0.30%), which are handled later during preprocessing.

TABLE I: Dataset summary and variable types.

Item	Value
Number of records	129,880
Number of variables (including target)	22
Categorical variables	3
Continuous numerical variables	4
Ordinal service rating variables (0–5)	14
Target distribution: satisfied	71,087 (54.73%)
Target distribution: dissatisfied	58,793 (45.27%)
Missing values: Arrival Delay in Minutes	393 (0.30%)

B. Feature Overview and Target Definition

The target variable is `satisfaction` with two classes: `satisfied` and `dissatisfied`. Input variables contain the following: (i) passenger descriptors (`Customer Type`, `Type of Travel`, `Class`); (ii) continuous operational variables (`Age`, `Flight Distance`, `Departure Delay in Minutes`, `Arrival Delay in Minutes`); and (iii) ordinal service ratings (e.g., seat comfort, inflight entertainment, online boarding, cleanliness). This structure makes the dataset suitable for supervised learning and interpretability analysis, since ML models can be evaluated by predictive performance and by how service contributions change under different passenger conditions.

To motivate the segment analysis, Fig. 1 presents empirical satisfaction rates by service class. Business class passengers exhibit substantially higher satisfaction than economy categories, indicating that aggregate performance measures may mask behavior differences across passenger groups.

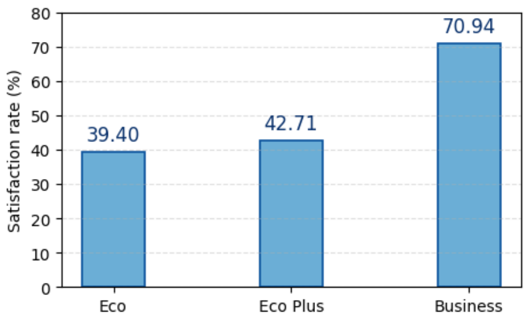

Fig. 1: Observed satisfaction rate by service class. *Eco* denotes standard economy, *Eco Plus* is the economy with more comfort or benefits, and *Business* represents premium service offers.

C. Passenger Segment Definition

Passenger groups are defined to enable structured evaluation of model behavior across heterogeneous profiles. Following common operational distinctions in airline analytics, three segmentation dimensions are used:

- **Travel purpose:** derived from `Type of Travel` field (`Business travel` vs. `Personal travel`).
- **Service class:** derived from `Class` attribute (`Eco`, `Eco Plus`, `Business`).
- **Flight distance range:** directly obtained from `Flight Distance` using three ranges: short (< 1000), medium (1000–2499), and long (≥ 2500) miles.

These groupings are used in later sections to compare predictive performance, false negative rates, and feature patterns across passenger categories, supporting a more detailed assessment than aggregate evaluation alone.

IV. Methodology

A. Data Preprocessing and Normalization

The chosen dataset contains a mix of categorical descriptors, ordinal service ratings, and continuous numerical variables. Prior to model training, a preprocessing strategy consistent with established airline satisfaction studies is applied to ensure data consistency and numerical stability [12]. Missing values are limited to the `Arrival Delay in Minutes` and are handled using median imputation, which reduces sensitivity to extreme delay values and reflects the skewed nature of delay distributions commonly observed in airline data.

Continuous numerical features, including passenger age, flight distance, and delay variables, are scaled using min–max normalization. Ordinal service quality ratings are retained in their original discrete form, as they represent passenger assessments on a consistent scale (0–5). The same preprocessing pipeline is applied across all experiments to ensure reliable comparison of model behavior across passenger categories.

B. Categorical Encoding Strategy

Categorical variables describing passenger characteristics and travel conditions are transformed into numerical representations for supervised learning. Low cardinality categorical attributes such as `Gender`, `Customer Type`, `Type of Travel`, and `Class` are encoded using one-hot encoding. This strategy avoids imposing artificial ordering on nominal categories and allows models to learn independent contributions for each category. One-hot encoded variables support subsequent analysis of model behavior, since categories are represented explicitly and can be examined independently in feature importance and reliability assessments. Encoded features are concatenated with normalized numerical variables to form the final input feature matrix for supervised models.

C. Supervised Learning Models

The supervised learning task is formulated as a binary classification problem, where the objective is to predict passenger satisfaction status. Given an input feature vector $\mathbf{x}_i \in \mathbb{R}^d$ describing the i-th passenger, each model learns a mapping

$$f(\mathbf{x}_i) \rightarrow \hat{y}_i, \quad \hat{y}_i \in \{0, 1\}, \tag{1}$$

where $\hat{y}_i = 1$ denotes a satisfied passenger and $\hat{y}_i = 0$ denotes a dissatisfied passenger.

To examine model behavior across heterogeneous passenger populations, predictions are analyzed with respect to passenger categories $g \in \mathcal{G}$ defined by travel purpose, service class, and flight distance. The aim is not to tailor separate models to individual passenger groups but to analyze the behavior of deployed models under realistic heterogeneous conditions. Performance and error characteristics are evaluated both on the full test set and on category subsets using identically trained models, enabling consistent and reliable analysis.

Multiple supervised models are employed to capture different characteristics, following comparative findings in airline satisfaction analytics [13]. Tree-based classifiers, including decision trees and random forests, are used due to their transparency and ability to capture complex feature interactions. Ensemble learning is explored through a Gradient Boosting classifier, while a feedforward Neural Network is introduced to examine performance under a non-linear learning setting.

The Decision Tree model is used as a baseline, with depth and minimum sample constraints applied to limit overfitting while preserving interpretability. The Random Forest aggregates multiple decision trees trained on bootstrap samples with randomized feature selection to improve prediction stability. Gradient Boosting builds an ensemble of shallow trees iteratively, with learning rate and tree depth selected to control model complexity. The feedforward Neural Network consists of two fully connected hidden layers with rectified linear unit activations and is optimized using the Adam optimizer, with early stopping based on validation loss applied to mitigate overfitting. All models are trained using the same input features and preprocessing pipeline to ensure a fair comparison of predictive accuracy, error behavior, and reliability across passenger categories.

D. Training and Validation Protocol

Model training uses a stratified train–test split to preserve the original satisfaction distribution, with an 80:20 division between training and testing data. Hyperparameter tuning is performed on the training set using five-fold cross-validation to avoid information leakage. Cross-validation is used to reduce variance in model selection, while formal statistical significance testing of inter-segment performance differences is deferred to future work.

For each model, key hyperparameters (such as tree depth and minimum sample size for tree-based models, learning rate for Gradient Boosting, and hidden layer size for the neural network) are selected by maximizing validation performance over predefined parameter ranges, ensuring balanced model complexity and generalization.

E. Evaluation Metrics

Model performance is evaluated through accuracy, macro-averaged F1-score and false negative rate. Accuracy measures overall classification correctness, while the macro-averaged F1-score ensures equal weighting of satisfied and dissatisfied classes, preventing dominance of the majority class and enabling balanced performance assessment. Particular attention is given to the false negative rate, as misclassifying dissatisfied passengers as satisfied can obscure service deficiencies and reduce the effectiveness of satisfaction monitoring.

In addition to numerical performance metrics, feature importance and feature sensitivity outputs are analyzed to identify influential service attributes and then to examine how decision drivers vary across passenger categories. Combining predictive evaluation with feature sensitivity analysis provides a more comprehensive understanding of system behavior compared to aggregate metrics alone, especially when reliability differs across passenger groups.

V. Experimental Results and Segment Analysis

A. Overall Model Performance

Across all trained models, classification accuracy exceeds 85%, indicating that passenger satisfaction can be predicted reliably using the selected feature set and preprocessing strategy. Performance improves with model complexity, increasing from 85.42% for the Decision Tree to 90.12% for the Neural Network, suggesting that non-linear modeling captures additional structure in the data.

Macro averaged F1 scores follow a similar trend, rising from 0.84 to 0.90, which confirms balanced predictive performance across satisfied and dissatisfied classes rather than reliance on majority class prediction. At the same time, false negative rates decrease steadily across model families, from 18.73% for the Decision Tree to 12.94% for the Neural Network, indicating improved sensitivity to dissatisfied passengers and stronger support for operational monitoring.

While these aggregate results are comparable to or exceed those reported in prior airline satisfaction studies, they provide only a partial view of system behavior. High overall accuracy and reduced error rates may still mask systematic differences in prediction quality across passenger groups, particularly in heterogeneous populations. Table II summarizes these metrics and provides a reference summary of model performance.

TABLE II: Overall model performance on the full test set.

Model	Accuracy (%)	Macro F1	False Negative Rate (%)
Decision Tree	85.42	0.84	18.73
Random Forest	88.91	0.88	14.26
Gradient Boosting	89.67	0.89	13.58
Neural Network	90.12	0.90	12.94

B. Performance Across Passenger Segments

To assess how predictive performance varies across passenger profiles, models are evaluated separately within groups defined by travel purpose, service class, and flight distance. This analysis reveals variation in predictive behavior that is not visible in aggregate results, with marked differences in error distribution and reliability across passenger categories. Business class passengers exhibit higher prediction accuracy, while economy class and long-distance travel segments show decreased performance for certain models. These differences suggest that satisfaction drivers and their statistical relationships are not uniform across passenger populations.

C. False Negative Analysis and Reliability Discussion

Beyond accuracy and F1-score, the false negative rate offers important insight into model behavior under practical deployment conditions. In airline satisfaction analysis, false negatives correspond to dissatisfied passengers that are incorrectly classified as satisfied. For a passenger category $g \in \mathcal{G}$, the false negative rate is defined as

$$\mathrm{FNR}_g = \frac{\sum_{i \in g} \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 1)}{\sum_{i \in g} \mathbb{I}(y_i = 0)}, \tag{2}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, enabling direct comparison of error behavior across passenger categories under identical model configurations.

The results showcase that false negative rates are unevenly distributed across passenger groups. As shown in Fig. 2, economy and economy-plus passengers exhibit higher false negative rates in several contexts, especially for personal travel. For example, false negative rates above 45% are observed for economy and economy-plus personal travel on short-distance flights, indicating that almost half of dissatisfied passengers in these categories are not identified by the model, whereas values below 10% correspond to more reliable detection.

These elevated false negative rates occur in specific passenger categories and do not contradict the aggregate performance of the models. Instead, they reflect a concentration of prediction errors in heterogeneous contexts where passenger expectations and dissatisfaction cues are more diverse. A clear asymmetry is observed between travel purposes, with business travel showing slightly more consistent dissatisfaction detection, while personal travel exhibits higher error concentration, particularly among economy passengers.

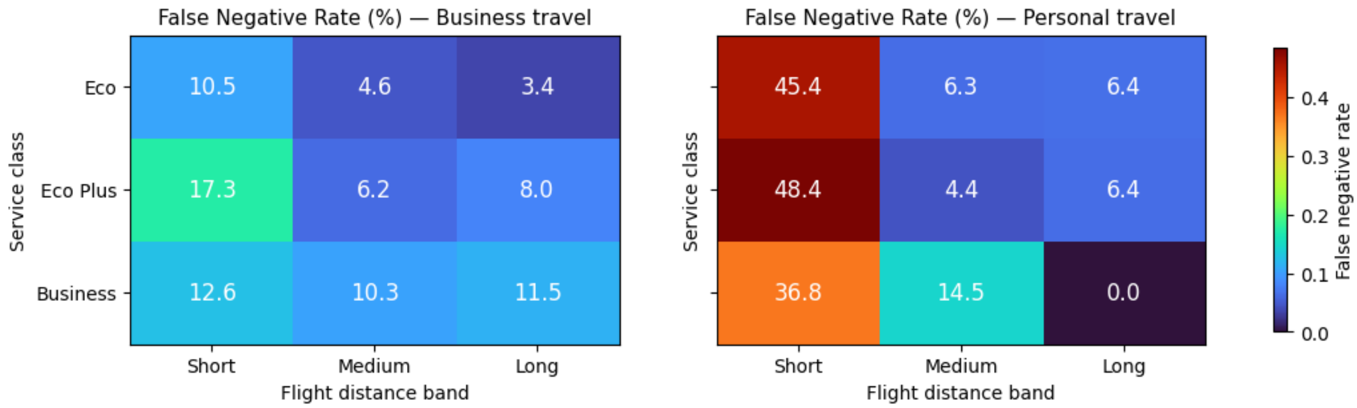

Fig. 2: False negative rate (FNR) across passenger categories, split by travel purpose. Rows denote service classes, and columns denote flight distance bands. Each cell reports the percentage of dissatisfied passengers predicted as satisfied.

VI. Feature Sensitivity and Importance Analysis

To complement performance evaluation, feature sensitivity analysis is employed to explore how input variables contribute to satisfaction predictions and whether learned decision patterns remain consistent in different passenger profiles. Feature influence is assessed using permutation importance computed on the test split: for each feature, values are randomly permuted, and the resulting drop within evaluation metrics is recorded. Importance is computed separately for each trained model, and segment importance is computed by applying the same procedure within each passenger subset using the identical trained model. To reduce dependence on a single random permutation, each importance value is averaged over multiple permutation repeats to ensure for comparability.

To support comparison among passengers defined by travel purpose, service class, and flight distance, importance values are normalized within each category, allowing relative feature influence analysis. Figure 3 illustrates normalized feature importance in service classes. Across all categories, seat comfort emerges as the dominant predictor, representing approximately 46% of importance in economy, 57% in economy-plus, and 55% in business class, confirming its central role in satisfaction assessments. Inflight entertainment consistently ranks second, contributing roughly 24–27%, indicating a shared core structure of satisfaction drivers across service tiers.

Beyond these primary factors, notable variation appears in secondary attributes. Customer type contributes substantially more to business class predictions (18.6%) than to economy (10.0%) or economy-plus (6.8%), suggesting that loyalty status and usage patterns carry greater informational value within higher service levels. In contrast, economy and economy-plus predictions depend more evenly on booking convenience and timing features (e.g., departure delay), reflecting greater sensitivity to service accessibility and process quality.

Overall, the results indicate that a small set of core attributes shapes passenger satisfaction, while the relative importance of supporting factors differs meaningfully by service class. This demonstrates that aggregate feature importance alone may mask important variation in decision patterns and underscores the value of feature sensitivity in passenger categories for reliable deployment of airline satisfaction prediction systems.

VII. Discussion and Industrial Implications

The results show that airline satisfaction prediction systems with strong aggregate accuracy may still perform unevenly within passenger populations. Although all evaluated models achieve high accuracy and balanced macro F1-scores, false negative rates are not uniformly distributed. Specifically, dissatisfaction is more frequently overlooked among economy and economy-plus passengers on personal trips, revealing a gap between average performance and operational usefulness.

This uneven error distribution has direct implications for customer experience management. False negatives correspond to dissatisfied passengers incorrectly classified as satisfied, which can delay the detection of service problems and weaken corrective actions. The findings indicate that uniform monitoring rules and global decision thresholds may fail to capture dissatisfaction signals in passenger contexts where dissatisfaction cues are less consistent.

Feature sensitivity analysis further shows that while core service attributes such as seat comfort dominate satisfaction prediction in general, secondary factors vary by service class. Business passenger predictions emphasize passenger characteristics, whereas economy passengers are more sensitive to service accessibility and operational factors. This suggests that managerial insights based solely on aggregate feature rankings may overlook important drivers of dissatisfaction.

In practical airline operations, these findings support the use of category analytics for satisfaction monitoring, enabling targeted quality interventions, differentiated alerting strategies, and more effective allocation of service improvement resources. Incorporating passenger context into model evaluation and interpretation can therefore enhance the practical value of AI-driven satisfaction systems and improve their alignment with operational decision-making processes.

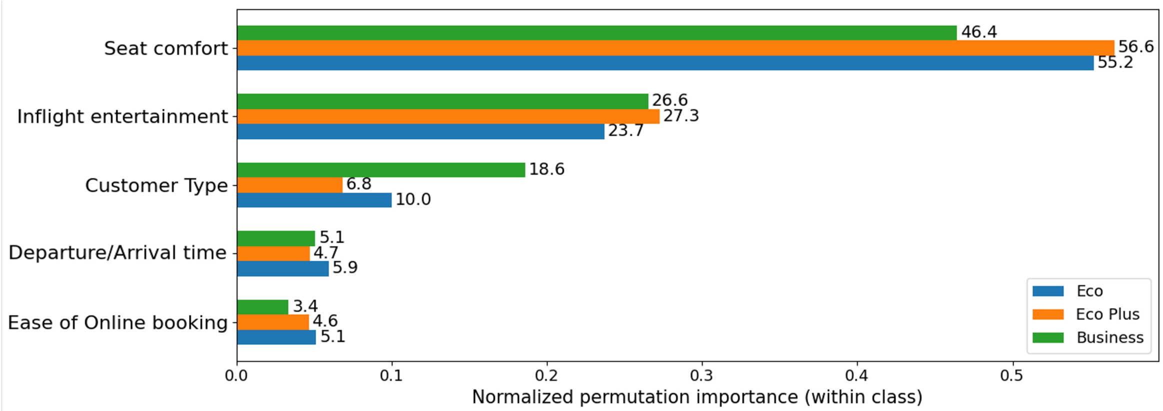

Fig. 3: Normalized permutation feature importance across passenger service classes. Each bar represents the relative contribution of a feature to satisfaction prediction within a given service class, normalized so that importance sum to one per class.

VIII. Conclusion and Future Work

This study evaluated airline satisfaction prediction systems across passengers defined by travel purpose, service class, and flight distance. Although the supervised learning models achieve high accuracy above 85%, and exceed 90% for neural network models, the results show that strong aggregate performance does not ensure consistent behavior across passenger populations. Significant variation is observed in false negative rates and feature influence, underscoring uneven sensitivity to dissatisfaction across travel contexts. Feature sensitivity results indicate that while seat comfort remains the key driver of satisfaction across all passengers, secondary factors vary by service class, showing differences in how service quality is evaluated across travel contexts.

Several limitations should be acknowledged. This analysis relies on a single public dataset and focuses on structured service ratings and service delivery indicators, which may not fully capture behavioral, temporal, or situational aspects of passenger experience. In addition, passenger groupings are examined after model training rather than being incorporated directly into the learning process.

Future research will extend this work by introducing modeling approaches that integrate passenger attributes into training, enabling prediction behavior to respond more effectively across diverse travel contexts. Planned directions include temporal models to capture satisfaction changes in repeated journeys, the integration of behavioral indicators like booking patterns and travel frequency, and the development of ensemble systems that adjust sensitivity to dissatisfaction in higher-risk travel scenarios. These efforts aim to translate satisfaction prediction models into decision support tools that help airlines prioritize interventions, allocate resources, and address issues.

Acknowledgment

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7502, *Intelligent Multi-Agent Control and Optimization Applied to Green Buildings and Environmental Monitoring Drone Swarms (ECOSwarm).*

References

[1] "Predictive analytics for enhanced passenger satisfaction in the airline industry: Leveraging machine learning to drive strategic decision-making," in *Proc. IEEE Int. Conf. on Optimization and Applications (ICOA)*, 2024. doi: 10.1109/ICOA62581.2024.10753807.

[2] A. V. Keerthana, A. S. Surabhi, and S. Supreeth, "Analysis of airline passengers' satisfaction in multiple dimensions," in *Proc. 2024 IEEE North Karnataka Subsection Flagship Int. Conf. (NKCon)*, Sep. 2024, pp. 1–7. doi: 10.1109/NKCon62728.2024.10774776.

[3] S. P. Wu and Y. Gao, "Happy or grumpy? A machine learning approach to analyze the sentiment of airline passengers' tweets," *Transportation Research Record*, 2022. doi: 10.48550/arXiv.2209.14363.

[4] R. Iskandar, O. R. A. Rasyid Anies, R. Iskandar, M. E.-K. Kesuma, and M. Konecki, "Analyzing airline services and communication systems by designing machine learning model to predict passenger satisfaction," *Int. J. Electronics and Communications Systems*, vol. 3, no. 2, 2023. doi: 10.24042/ijecs.v3i2.19782.

[5] "Analysis of airline passenger satisfaction using decision tree and naïve Bayes algorithms," *Information Engineering and Business*, vol. 5, no. 4, 2023. doi: 10.37034/infeb.v5i4.728.

[6] "Feature analysis and evaluation of airline passenger satisfaction based on decision tree," in *Proc. IEEE Int. Conf. on Industrial Internet of Things, Big Data and Supply Chain (IIoTBDSC)*, 2023. doi: 10.1109/IIoTBDSC60298.2023.00014.

[7] "An optimized deep learning approach for improving airline services," *Computers, Materials & Continua*, 2023. doi: 10.32604/cmc.2023.034399.

[8] "Business intelligence in airline passenger satisfaction study—A fuzzy-genetic approach with optimized interpretability-accuracy trade-off," *Applied Sciences*, vol. 11, no. 15, 2021. doi: 10.3390/APP11115098.

[9] "Airline customer segmentation based on complex behavioral approach using K-mode and XG-Boost algorithm," in *Proc. IEEE Int. Conf. on Digital Technologies (ICDT)*, 2023. doi: 10.1109/ICDT57929.2023.10151011.

[10] C. C. E. van Geest, Y. W. Yit, Z. T. Gouliev, and K. Quille, "Predictive models with XAI: A comparative study of enhancing airline customer satisfaction," in *Proc. 2023 Conf. on Human-Centered Artificial Intelligence: Education and Practice (HCAIep)*, New York, NY, USA, 2023, pp. 36–41. doi: 10.1145/3633083.3633189.

[11] Y. Akhyojon, "Customer satisfaction in airline," Kaggle dataset. [Online]. Available: https://www.kaggle.com/datasets/yakhyojon/customer-satisfaction-in-airline

[12] M. Salah-Ud-Din, B. T. L. S. S., and H. Al Ali, "Exploratory data analysis and prediction of passenger satisfaction with airline services," in *Proc. IEEE National Technology Conf. on Automation (NTCA)*, Apr. 2024, pp. 295–302. doi: 10.23919/ntca60572.2024.10517814.

[13] A. Ashwika, D. G. K. Dishali, and N. Hemalatha, "Airline passenger satisfaction prediction using machine learning algorithms," *Redshi Arch*, vol. 4, no. 1, Jul. 2023. doi: 10.25215/8119070682.24