Journal of Turkish

Operations Management

Airline customer satisfaction analysis using machine learning methods

Sonya Javadi[1*], Amine Hatun Ergin[2 *], Özge Yüksel [3]
[1]Department of Industrial Engineering, Işık University, Istanbul, Turkey
sonya.javadi@isikun.edu.tr, ORCID No: http://orcid.org/0000-0002-3726-6799
[2]Department of Management Information Systems, Işık University, Istanbul, Turkey
23itec5004@isik.edu.tr, ORCID No: https://orcid.org/0009-0002-4409-6369
[3]Department of Management Information Systems, Işık University, Istanbul, Turkey
23itec5003@isik.edu.tr, ORCID No: http://orcid.org/0009-0003-5801-0164
*Corresponding Author: Sonya Javadi

Article Info

Article History:
Received: 16.02.2025
Revised: 02.04.2026
Accepted: 21.04.2026

Keywords
Airline industries,
Customer satisfaction,
Machine learning,
Deep learning techniques

Abstract

The airline industry, a major component of global transportation, operates in a highly competitive environment where maintaining customer satisfaction is of critical importance. This study analyzes airline customer satisfaction using machine learning techniques in order to identify the key factors influencing passenger satisfaction and provide insights for service improvement. The dataset used in this study was obtained from a publicly available Kaggle dataset and contains 129,880 passenger records and multiple service-related features. Several machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN), were applied to analyze the dataset and predict customer satisfaction. The experimental results show that the Random Forest classifier achieved the highest prediction accuracy of 94.6%, outperforming the other models. The analysis further reveals that inflight entertainment has the strongest relationship with customer satisfaction, with a correlation value of approximately 0.52. In addition, the results indicate that customers who give high ratings to inflight entertainment are typically between the ages of 35 and 55, tend to travel in business class, and exhibit higher levels of customer loyalty. These findings provide valuable insights for airline companies seeking to improve service quality and enhance customer satisfaction. In particular, the results suggest that improving inflight entertainment services for loyal business-class passengers may significantly increase overall customer satisfaction. Therefore, airlines can use these insights to develop targeted service strategies and strengthen their competitive position in the airline industry.

1. Introduction

Customer satisfaction is crucial for the success of any business, and this holds true for the airline industry as well. The understanding and analysis of customer satisfaction play significant roles in enabling airlines to enhance their services, retain customers, and attain a competitive advantage. In recent years, significant progress has been made in utilizing machine learning techniques for the analysis and prediction of customer satisfaction. In the service-oriented airline industry, customer satisfaction stands out as a widely acknowledged key performance indicator. The satisfaction levels derived from passengers' travel experiences serve as direct reflections of the quality standards, service delivery, and customer-centric approaches employed by airline companies. For example, according to Hwang, Kim, Park, and Kwon (2020), the probability of customers' return visits to airline services can be predicted based on feedback comments and satisfaction ratings, achieving an accuracy of 83.42%. Their study also found that longer feedback comments improve prediction accuracy, although certain implications and

limitations remain in understanding customer preferences. Moreover, new machine learning models and deep learning methods have been applied in the literature to analyze customer satisfaction not only in the airline industry but also in other service sectors. For instance, a study by García et al. (2019) evaluates the effectiveness of ensemble models for predicting customer satisfaction using a real database consisting of 129,890 airline samples. Specifically, the study assesses the performance of the Bagging ensemble model with the k-nearest neighbor (k-NN) algorithm as the base learner. The results indicate that the Bagging ensemble model outperforms the single classifier in terms of RMSE and MAE, demonstrating its superior predictive accuracy.
This study examines the application of machine learning techniques for analyzing customer satisfaction in the airline industry through a case study of Turkish Airlines. The main objectives are to identify the key factors influencing customer satisfaction and to assess the effectiveness of machine learning methods in predicting and analyzing customer satisfaction using historical data. To achieve this objective, a comprehensive methodology was designed to analyze the satisfaction levels of customers who choose Turkish Airlines for their flights. This section includes several key steps, such as data collection from Kaggle, an open-source data platform, data preprocessing, the development of machine learning models, and a comparative analysis of different analytical methods. Finally, the study evaluates the predictive performance of the models using evaluation metrics such as accuracy, precision, and recall in forecasting customer satisfaction. The findings aim to contribute to both academic and managerial understanding of customer satisfaction analysis and demonstrate the potential of machine learning techniques to support strategic decision-making in the airline industry, particularly for Turkish Airlines.
The structure of the paper is organized as follows: Section 2 provides a comprehensive literature review on customer satisfaction analysis using machine learning techniques and different machine learning methods. Section 3 presents a brief introduction to the problem. Section 4 discusses the methodology and data preparation procedures used to analyze the problem. Section 5 presents the results of the different machine learning methods and discusses their findings. Section 6 introduces the prediction approach for future analysis of customer satisfaction in the airline industry. Finally, Section 7 concludes the paper and provides recommendations for future research in this field.

2. Literature Review

The existing research on customer satisfaction in the airline industry has primarily concentrated on traditional survey-based methods and statistical analysis. For example, Gures, Arslan, and Tun (2014), using a 5-point Likert questionnaire administered to 821 passengers in Turkish airports, found that reliability and facilities significantly enhance customer satisfaction, which in turn drives customer loyalty in the Turkish airline industry. While these approaches have yielded valuable insights, they often encounter challenges in capturing intricate patterns and trends in customer satisfaction, especially with the advent of social media platforms where customers can provide feedback. However, there exists a gap in the literature concerning the specific application of machine learning to analyze customer satisfaction within the airline industry.
For instance, Xu, Zhu, Metawa, and Zhou (2022) examine factors influencing brand equity from financial and customer perspectives, emphasizing the impact of ethical marketing on satisfaction and perception, and offering strategies to enhance customer satisfaction. Meanwhile, another study by Al-Mashraie, Chung, and Jeon (2020) analyzes customer churn in the telecommunications industry by comparing various prediction models, including logistic regression, support vector machines, random forests, and decision trees. Using the push-pull-mooring framework and partial least squares regression, the study identifies influential factors and finds that logistic regression has the highest prediction accuracy, while service quality emerges as a key factor in customer churn.
In contrast, machine learning techniques have the potential to analyze extensive volumes of customer data, reveal hidden patterns, and predict customer satisfaction with high accuracy. For instance, Hong, Khaw, Chew, and Yeong (2023) applied machine learning models to predict airline customer satisfaction, identifying key features such as online boarding, inflight entertainment, and seat comfort as highly correlated with satisfaction. Their model achieved 89.20% accuracy, 93.04% precision, and an 88.80% F1-score, providing valuable insights for improving airline service quality. Another study by Tayaba et al. (2023) uses machine learning to analyze tweets in order to improve customer experience in the airline industry. The findings show that a CNN model outperforms SVM and ANN models in sentiment classification, while association rule mining reveals connections between tweet categories and sentiments.
Similarly, Aktepe, Ersöz, and Toklu (2014) investigate customer satisfaction and loyalty in the white goods industry by categorizing customers into four groups based on 15 criteria. Using WEKA classification algorithms and structural equation modeling (SEM) with LISREL, the study analyzes the effects of satisfaction and loyalty. A survey of 200 customers supports the development of a method for identifying high-performance customer groups and relevant criteria. The findings are used to develop a tool for improving customer strategies and enhancing customer relationship management. Some studies focus more specifically on improving machine

learning techniques themselves. For example, Gou et al. (2019) propose two improved k-nearest neighbor (KNN) methods, namely the weighted representation-based KNN (WRKNN) and the weighted local mean representation-based KNN (WLMRKNN), to address the sensitivity of traditional KNN to neighborhood size and outliers. WRKNN uses the representation coefficients of k-nearest neighbors to calculate class-specific distances, while WLMRKNN uses k-local mean vectors. Extensive experiments on multiple datasets demonstrate that these methods outperform traditional KNN and are less sensitive to neighborhood size, particularly with small sample sizes.

In addition to considering product and service characteristics, customers also evaluate the attributes of firms when making their choices. Challenges have emerged not only due to the high level of competition but also because of increasing consumer expectations for better service. A study conducted in the United States found that consumers are increasingly seeking products produced according to ethical standards (Bockhorst, Yu, Polania, & Fung, 2017). This study describes a system implemented at a large U.S. insurance company that predicts customer satisfaction after call center interactions using data derived from speech-to-text systems, call metadata, customer profiles, and insurance policy information. The proposed workflow involves training a ranking model on call data and applying a convolutional fitting function to map rankings to survey scores, producing more accurate predictions than standard methods. This approach can be generalized to other customer satisfaction prediction problems.

The study by Ouf (2023) addresses passenger satisfaction in the highly competitive aviation industry by applying deep neural networks with the Adam optimization algorithm to improve classification performance. Unlike previous research that overlooked dataset quality, this approach was validated using the airline passenger satisfaction dataset from Kaggle and compared with artificial neural networks (ANNs), random forests, and support vector machines. The proposed method achieved an accuracy of 99.3%, outperforming previous studies. Similarly, Park, Kim, Kim, and Park (2022) apply deep learning techniques to analyze survey data from Korean airline customers in order to identify factors influencing customer churn risk and satisfaction. Their study uniquely focuses on the social servicescape, including interactions between cabin crew and passengers. The results show that incorporating human service factors into predictive models improves accuracy by up to 10% for customer churn risk and 9% for satisfaction prediction.

For example, Krishnan, Robinson, and Chilamkurti (2020) provide an overview of technological advancements in speech recognition using supervised learning, focusing on how deep neural networks can recognize speech from large datasets. The study highlights the importance of supervised learning in inferring functions from labeled data for speech recognition, a trend that has become prominent in automation technologies over the decades. A more general study by Thakur and Han (2021) contributes to fall detection for elderly individuals in IoT-based environments such as smart homes. The study compares 19 machine learning methods and finds that the k-NN classifier provides the highest accuracy for fall detection. It also introduces a framework capable of detecting both falls and fall-like motions, overcoming the limitations of binary classifiers. Furthermore, the study enhances the k-NN classifier's accuracy using k-fold cross-validation and the AdaBoost algorithm, achieving accuracies of 99.87% and 99.66% on two datasets.

Several studies specifically investigate customer satisfaction in the airline industry. For example, Chow (2014) analyzes the relationship between customer satisfaction and service quality among twelve Chinese carriers using quarterly panel data. Fixed-effect Tobit analysis reveals that customer complaints increase with damaged baggage but decrease at a diminishing rate, while on-time performance has no significant effect. Non-state carriers receive more complaints than state-owned ones, with the highest complaint levels occurring in the third quarter during summer holidays. Similarly, Hussain (2016) examines how customer satisfaction mediates the effects of service quality, corporate image, and perceived value on brand loyalty in the UAE airline industry. Based on 253 questionnaires, the results confirm that customer satisfaction plays a crucial role in converting passengers into loyal customers. This research represents one of the first attempts to examine these relationships in the UAE airline context.

Another study by Park et al. (2019) investigates the determinants of customer satisfaction in the airline industry by analyzing feedback from more than 133,000 customers using sentiment analysis and structural equation modeling. The findings show that customers' affective values significantly influence satisfaction and highlight important differences between low-cost and full-service carriers. Similarly, Leong, Hew, Lee, and Ooi (2015) examine the impact of SERVPERF on customer satisfaction and loyalty in both low-cost and full-service airlines using a combined SEM–artificial neural network approach. Unlike previous studies that relied on the GAP-5 SERVQUAL model, this research demonstrates that SERVPERF dimensions significantly influence satisfaction and loyalty, explaining 63.1% and 55.6% of the variance, respectively. The findings provide valuable insights for airline managers seeking to improve customer satisfaction and loyalty.

Customer satisfaction analysis using social media data has become an important research area in recent literature. Social media platforms provide vast amounts of real-time feedback and opinions from customers, enabling researchers to analyze sentiment, trends, and preferences at scale. For example, Hwang, Kim, Park, and Kwon

(2020) estimate the probability of customers'return visits to airline services using machine learning applied to feedback comments and satisfaction ratings. By analyzing sentiment features with seven classifiers, the model achieves an accuracy of 83.42%, and longer feedback comments improve prediction accuracy. Similarly, Kumar and Zymbler (2019) analyze airline customer feedback from Twitter using machine learning techniques such as word embedding and n-gram models. Their results show that CNN models outperform SVM and ANN models in classifying tweets as positive or negative, while association rule mining reveals useful insights for improving customer experience.

Nourbakhsh and Chelkasari (2023) also use machine learning techniques to analyze tweets and improve aviation customer experience. Their approach employs deep learning to classify tweets by sentiment, achieving an accuracy of 99.97% in two-class analysis and 88.83% in three-class analysis. The method extracts word vectors and constructs polarity features using the WordNet dictionary, enabling effective identification of passenger sentiments. Similarly, Nurdina and Puspita (2023) compare the effectiveness of Naïve Bayes and k-Nearest Neighbor (k-NN) algorithms for classifying airline passenger satisfaction. Their results show that Naïve Bayes outperforms k-NN, achieving an accuracy of 84.48% compared to 65.38% for k-NN. Additionally, Naïve Bayes achieved a precision of 82.25% and a recall of 82.43%, while k-NN obtained a precision of 67.35% and a recall of 74.33%.

Some studies have also applied alternative analytical techniques to analyze customer satisfaction. For example, Surpato and Oetama (2023) analyze key factors influencing airline customer satisfaction using decision tree algorithms, identifying important variables such as online boarding, inflight entertainment, Wi-Fi services, class, and travel type. The Naïve Bayes algorithm, with an accuracy of 87%, effectively predicts passenger satisfaction. Similarly, Taliah and Zervopoulos (2023) analyze 83 airlines between 2011 and 2019 using a Bayesian meta-frontier framework. Their findings indicate that there is no trade-off between customer satisfaction and airline efficiency, and that larger airlines are more likely to achieve both simultaneously. The study also shows that airlines offering premium services such as premium economy cabins and inflight amenities at reasonable prices can enhance customer satisfaction regardless of external shocks.

3. Problem Statement

Turkish Airlines is recognized as one of the leading and most prestigious airlines globally, serving millions of passengers from diverse nationalities each year. Therefore, understanding customer satisfaction and how passengers evaluate the airline's services is of critical importance for managers and strategic decision-makers within the company. This study focuses on analyzing customer satisfaction for Turkish Airlines. Unlike traditional approaches to evaluating customer satisfaction, this research employs machine learning techniques to analyze airline passenger satisfaction data obtained from a publicly available dataset. By leveraging these modern analytical methods, the study aims to gain deeper insights into customer sentiments and perceptions of Turkish Airlines.

Recent advances in machine learning and deep learning have significantly improved the ability of airline companies to analyze passenger behavior and satisfaction. Real-time data collected from various digital platforms allows airlines to better understand customer expectations and improve service quality (Ouf, 2023). Traditional approaches to customer satisfaction analysis often rely on structured surveys, which require substantial time and financial resources to conduct (Magsi et al., 2021).

In recent years, machine learning and data mining techniques have been increasingly applied to analyze large volumes of customer feedback and identify patterns in passenger behavior. These techniques allow researchers to extract meaningful insights from large datasets and improve prediction accuracy in customer satisfaction analysis. Among these approaches, supervised learning methods such as Support Vector Machines (SVM), Decision Trees, Random Forest, and k-Nearest Neighbor (k-NN) are commonly used for classification tasks. Motivated by these developments, this study investigates the application of machine learning techniques to analyze airline customer satisfaction using a publicly available airline passenger satisfaction dataset. The objective of this study is to identify the key factors that influence passenger satisfaction and evaluate the performance of different machine learning algorithms in predicting customer satisfaction.

4. Materials and Methods

This section presents the dataset used in the study, the preprocessing procedures applied to the data, and the machine learning models used to predict customer satisfaction.

4.1. Dataset Description

The dataset used in this study was obtained from the publicly available Airline Passenger Satisfaction dataset on Kaggle (Kaggle, 2020), which provides precompiled datasets for machine learning research. This particular dataset consists of customer ratings of airport services on a scale ranging from 1 to 5, along with additional information such as customers' gender, age, flight distance, and flight delay time. For this study, data from 129,880 customers were analyzed. The dataset includes the information presented in Table 1, which represents the 18 selected features used in this study.

Table 1. The structure of the dataset

Variable	Explanation	Data Type
Satisfaction	Satisfaction of customers	Object
Gender	Male or female	Object
Customer Type	Regular or non-regular airline customer	Object
Age	The actual age of the passenger	Int64
Type of Travel	The purpose of the passenger's flight (personal or business)	Object
Class	Business, economy, economy plus	Object
Flight Distance	Flight distance	Int64
Seat comfort	Seat satisfaction level	Int64
Departure/Arrival time convenient	Departure/arrival time satisfaction level	Int64
Food and drink	Food and drink satisfaction level	Int64
Gate location	Level of satisfaction with the gate location	Int64
Inflight wifi service	Satisfaction level with Wi-Fi service on board	Int64
Inflight entertainment	Satisfaction with inflight entertainment	Int64
Online support	Online support	Int64
Ease of Online booking	Online booking satisfaction rate	Int64
On-board service	Level of satisfaction with on-board service	Int64
Leg room service	Level of satisfaction with leg room service	Int64
Baggage handling	Level of satisfaction with baggage handling	Int64
Checkin service	Level of satisfaction with checkin service	Int64
Cleanliness	Level of satisfaction with cleanliness	Int64
Online boarding	Satisfaction level with online boarding	Int64
Departure Delay in Minutes	Departure delay in minutes	Int64
Arrival Delay in Minutes	Arrival delay in minutes	Float64

4.2. Data Preprocessing

Data preprocessing is an essential step before implementing machine learning models, as it involves preparing the raw dataset for analysis and improving data quality. During this phase, two key preprocessing steps were performed: handling missing values and eliminating redundant variables. First, missing values in the dataset were identified and addressed to ensure data consistency. Records containing missing values were removed from the dataset to prevent potential bias in the machine learning models. Second, redundant variables were examined using correlation analysis. The variables *Departure Delay Minutes* and *Arrival Delay Minutes* were found to contain highly similar information. To avoid redundancy and multicollinearity in the dataset, the *Arrival Delay Minutes* variable was removed. After completing these preprocessing steps, the cleaned dataset was used for subsequent analysis and machine learning modeling.

4.3. Machine Learning Models

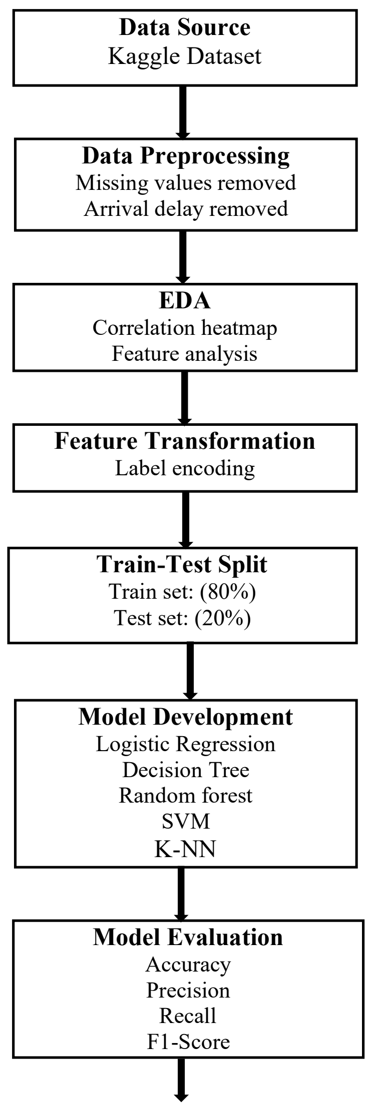
To predict customer satisfaction, several supervised machine learning algorithms were applied. These algorithms were selected due to their effectiveness in classification problems and their widespread use in customer analytics research. Logistic Regression is a statistical classification method that models the probability of a binary outcome.

Decision Tree classifiers divide the dataset into subsets based on feature values and construct a tree-like decision structure. Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to improve classification accuracy and reduce overfitting. Support Vector Machine (SVM) identifies an optimal hyperplane that separates different classes in the feature space. The k-Nearest Neighbor (k-NN) algorithm classifies observations based on the majority class among the nearest neighbors. These machine learning models were applied to the dataset in order to compare their performance in predicting customer satisfaction.

4.4. Model Training and Evaluation

The dataset was divided into training and testing subsets in order to evaluate the performance of the machine learning models. In this study, 80% of the data was used to train the models, while the remaining 20% was reserved for testing and validation. The models were trained using the training dataset and subsequently evaluated on the test dataset to assess their predictive performance.

Model performance was evaluated using several metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of classification performance by measuring different aspects of model prediction quality. While accuracy measures the overall correctness of predictions, precision and recall evaluate the model's ability to correctly identify satisfied and dissatisfied customers. The F1-score provides a balanced measure of precision and recall. The evaluation results of the applied machine learning models are presented and discussed in the following section. To provide a clear and reproducible overview of the methodology, the overall machine learning workflow of the proposed approach is illustrated in Figure 1.



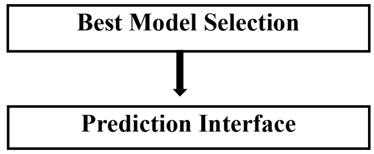


Figure 1. Machine learning pipeline of the proposed approach

5. Evaluation and Results

This section presents the results of the exploratory data analysis and the performance of the machine learning models used to predict customer satisfaction. First, a correlation analysis was conducted to examine the relationships among the variables in the dataset. A correlation matrix was generated and visualized using a heatmap, as shown in Figure 2.

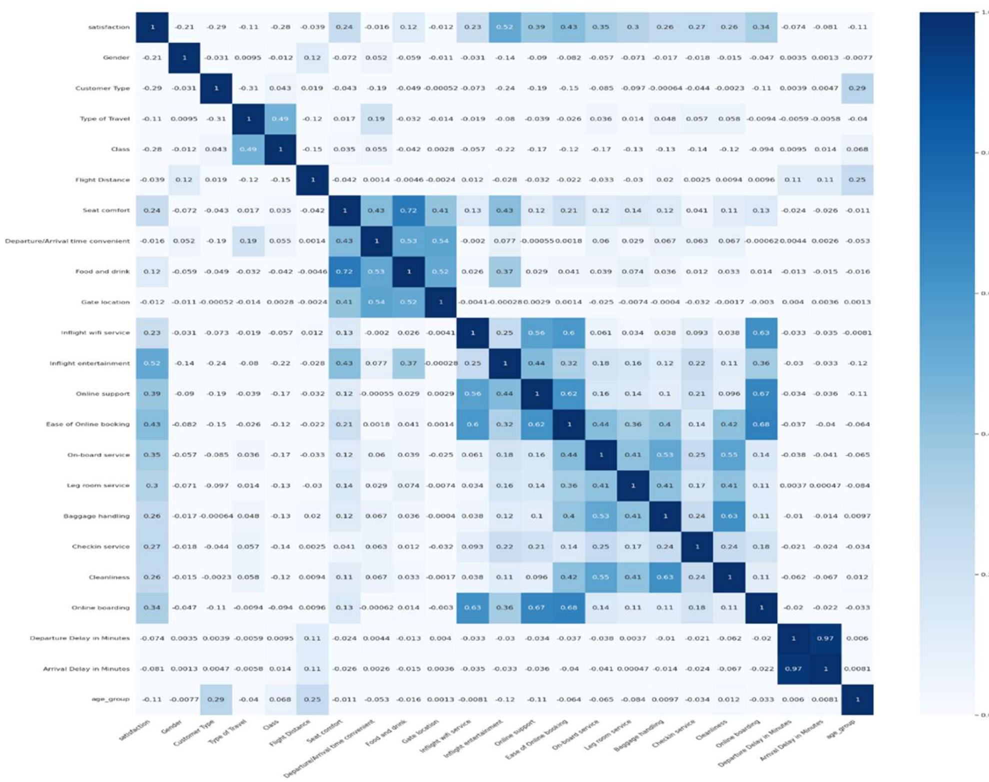


Figure 2. The correlation heatmap

According to the correlation matrix, as the correlation value approaches 1, the relationship between two variables becomes stronger. A negative value indicates an inverse relationship between the variables, meaning that as one variable increases, the other decreases. Based on this analysis, several data visualizations were generated to further explore the relationships between the variables and customer satisfaction levels. Some examples of these visualizations are presented below in Figures 3 and 4.

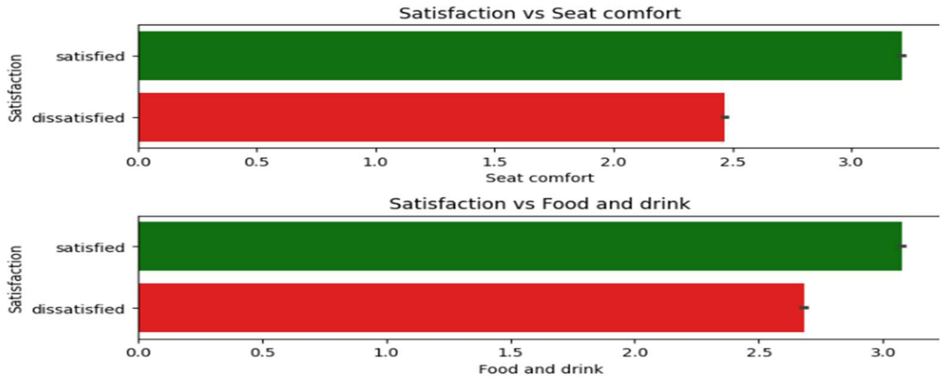


Figure 3. Relationships between customer satisfaction and other factors

An important observation from the correlation matrix is that *inflight entertainment* shows a correlation of approximately 52% with customer satisfaction. This indicates that inflight entertainment is one of the factors most strongly associated with passenger satisfaction. Therefore, a more detailed analysis of the inflight entertainment variable was conducted. This analysis helps identify which customer segments report higher satisfaction levels with inflight entertainment services. The distribution of inflight entertainment ratings by class and customer type is presented in the following figure.

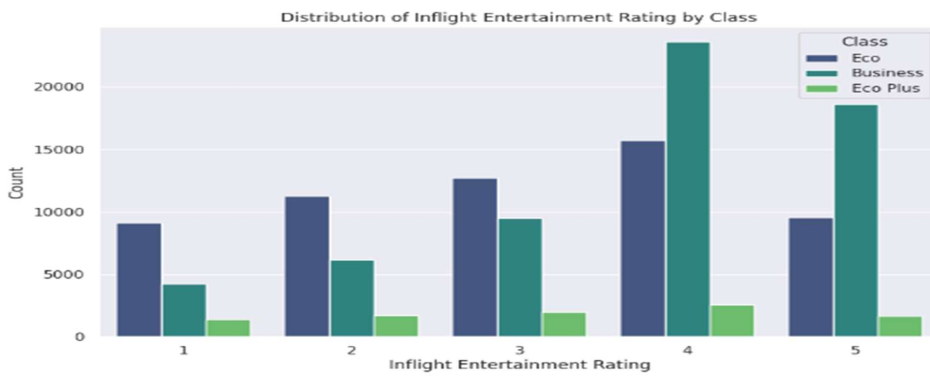


Figure 4. Distribution of inflight entertainment retings

According to the data presented in Table 1, Business Class passengers report the highest ratings for inflight entertainment. In addition, the visualization in Figure 4 shows that loyal customers tend to give higher ratings for inflight entertainment compared to disloyal customers. This finding suggests that inflight entertainment may play an important role in shaping satisfaction among loyal passengers. Furthermore, an analysis of the satisfaction levels of both loyal and disloyal customers is presented in Figure 5.

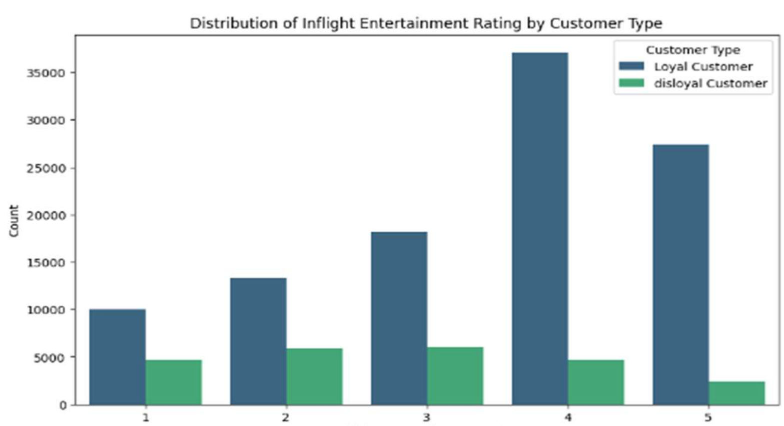


Figure 5. Distribution of customer satisfaction by loyalty status

According to this analysis, 61.6% of loyal customers are satisfied with airline services, whereas only 24% of disloyal customers report satisfaction. This finding supports the pattern illustrated in Figure 4 and indicates that loyal customers tend to report higher satisfaction levels. To further examine customer satisfaction patterns, passengers were categorized into different age groups, including children, teenagers, young adults, adults, and seniors. The results show that customers in the young adult age group exhibit higher satisfaction levels compared to other age groups. In this study, the young adult group includes individuals between 35 and 55 years old, as illustrated in Figure 6.

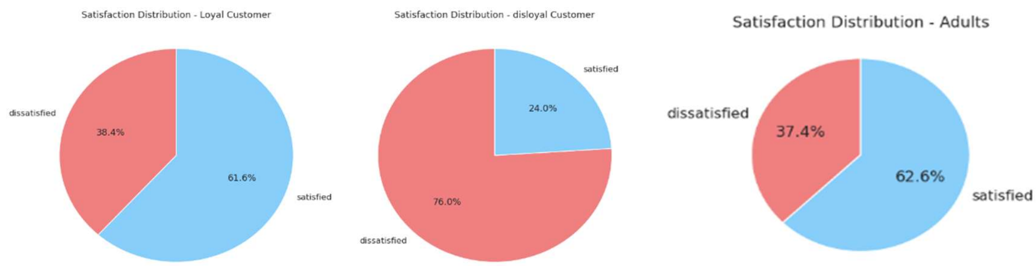


Figure 6. Distribution of customer satisfaction by age group

5.1 Machine Learning Results

Following the exploratory data analysis, the dataset was prepared for machine learning analysis. Since machine learning algorithms require numerical input, the categorical variables were transformed into numerical form to ensure compatibility with the models. To achieve this, categorical variables were encoded using a label encoding technique. For example, before encoding, the gender variable was represented as *male* or *female*. After applying the encoding process, these values were converted to 0 for male and 1 for female. The same encoding procedure was applied to other categorical variables, including customer type, travel class, and trip type.
Subsequently, several machine learning models were applied to the dataset, including Logistic Regression, Decision Tree Classification, Random Forest Classification, Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN). The performance of these models was evaluated using accuracy, precision, recall, and F1-score metrics. The accuracy values obtained from the models are as follows: Logistic Regression (0.8396), Decision Tree (0.9220), Random Forest (0.9461), SVM (0.9178), and k-NN (0.9113). The detailed performance comparison of the models is presented in Table 2.

Table 2. Performance comparison of machine learning models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.84	0.84	0.83	0.83
Decision Tree	0.92	0.92	0.92	0.92
Random Forest	0.95	0.95	0.95	0.94
SVM	0.92	0.91	0.91	0.92
K-NN	0.91	0.90	0.90	0.91

The results indicate that the Random Forest classifier achieves the best overall performance across all evaluation metrics, confirming its suitability for predicting airline customer satisfaction. In addition, the ensemble nature of the Random Forest algorithm helps reduce overfitting and improves prediction stability compared to single classifiers such as Decision Trees.

6. The Prediction Approach

This study provides valuable insights for airline companies by identifying the key factors that influence customer satisfaction and demonstrating how machine learning models can be used to predict passenger satisfaction levels. Understanding these factors enables airlines to improve their services, strengthen customer loyalty, and enhance their competitive position in the market. Within the scope of this study, a dataset containing airline customer satisfaction ratings was analyzed using several machine learning algorithms. Among the evaluated models, the Random Forest classifier achieved the highest prediction accuracy and was therefore selected as the most effective

model for the prediction task. Based on this model, a simple web-based interface was developed that allows users to enter their travel-related information and obtain a prediction indicating whether they are likely to belong to the satisfied or dissatisfied customer group.
Passengers rate the services they receive during the flight on a scale from 1 to 5 and provide additional information such as age, flight distance, and departure delay time. To enable prediction modeling, categorical variables such as gender, travel type, and travel class were converted into numerical format. In the gender variable, female is represented by 1 and male by 0. For travel type, business travel is denoted by 0 and personal travel by 1. In addition, business class is represented by 0, economy by 1, and economy plus by 2. Based on these input variables, the Random Forest model predicts customer satisfaction levels, as illustrated in Figure 7.

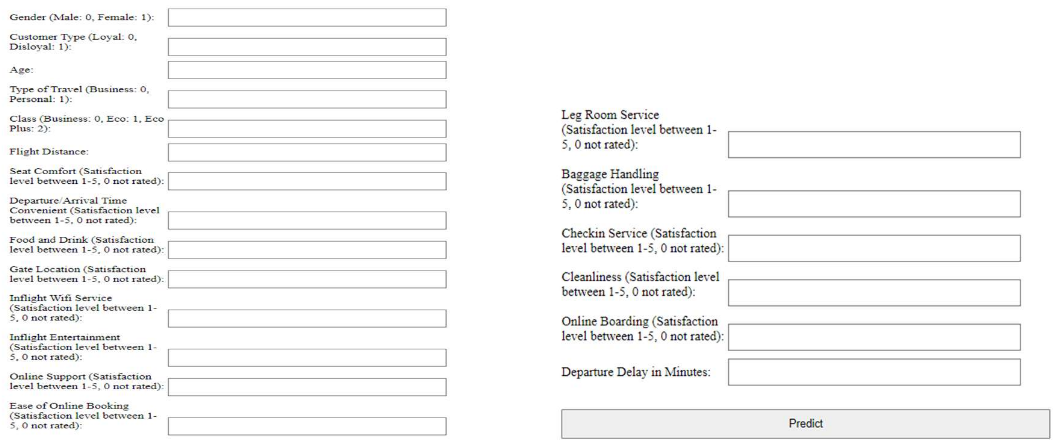


Figure 7. User interface for customer satisfaction prediction

The prediction results further indicate that inflight entertainment is one of the most influential factors affecting customer satisfaction. The analysis also suggests that passengers who give higher ratings to inflight entertainment are typically between the ages of 35 and 55, tend to travel in business class, and are generally loyal customers. These findings provide useful insights for airlines seeking to improve service quality and enhance passenger satisfaction as presented in Figure 8.

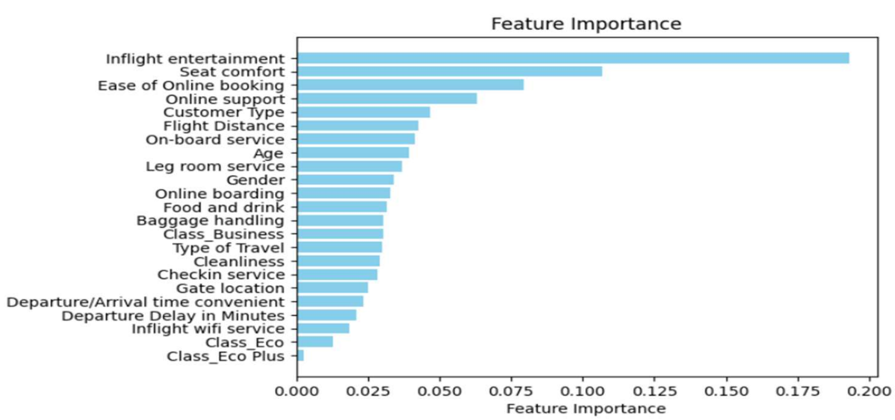


Figure 8. Prediction of customer satisfaction

7. Conclusion and Future Recommendations

This study is important for airline companies in identifying the factors that affect customer satisfaction and improving their services accordingly. Previous studies on this topic were also reviewed, highlighting the importance of customer satisfaction. Customer satisfaction is crucial for companies to enhance their success in the industry, strengthen their marketing strategies, and improve the services they provide. Within the scope of this study, a dataset containing airline customers' satisfaction levels with airline services was analyzed using various machine learning algorithms. Among the algorithms applied, the Random Forest model achieved the highest accuracy and was therefore selected as the most effective model. Additionally, a website was developed that allows customers to view their own information and determine whether they belong to the satisfied or dissatisfied group. The results indicate that inflight entertainment is one of the most influential factors affecting customer satisfaction. Furthermore, it was observed that customers who give high ratings to inflight entertainment are generally between the ages of 35 and 55, tend to fly in business class, and are loyal customers. Accordingly, it can be concluded that loyal customers in the 35–55 age group who travel in business class are generally more satisfied with airline services. Despite the valuable findings of this study, several limitations should be acknowledged. First, the dataset used in this research was obtained from a publicly available Kaggle dataset rather than directly from airline companies, which may limit the generalizability of the results. Second, the analysis focuses primarily on structured survey data and does not fully incorporate unstructured textual data from social media platforms. Future research could extend this work by integrating real-time social media sentiment analysis, applying deep learning models, and exploring larger datasets from multiple airlines to improve prediction performance and generalizability.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this study.

Contribution of Authors

In this study, Sonya Javadi contributed to the definition of the research problem, study supervision, and writing of the article. Amine Hatun Engin contributed to data collection and methodology development. Özge Yüksel contributed to the literature review and data collection process.

References

Aktepe, A., Ersöz, S., & Toklu, B. (2015). Customer satisfaction and loyalty analysis with classification algorithms and Structural Equation Modeling. *Computers & industrial engineering*, *86*, 95-106. https://doi.org/10.1016/j.cie.2014.09.031

Al-Mashraie, M., Chung, S. H., & Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, *144*, 106476. https://doi.org/10.1016/j.cie.2020.106476

Bockhorst, J., Yu, S., Polania, L., & Fung, G. (2017). Predicting self-reported customer satisfaction of interactions with a corporate call center. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia,* 179-190. 10.1007/978-3-319-71273-4_15

Chow, C. K. W. (2014). Customer satisfaction and service quality in the Chinese airline industry. *Journal of air transport management*, *35*, 102-107. https://doi.org/10.1016/j.jairtraman.2013.11.013

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

García, V., Florencia-Juárez, J. P., Sánchez-Solís., G. Rivera-Zarate., R. Contreras-Masse. (2019). Predicting airline customer satisfaction using k-nn ensemble regression models. *Instituto de Ingeniería y Tecnología*, 1870-4069. :10.13053/rcs-148-6-15

Gou, J., Qiu, W., Yi, Z., Shen, X., Zhan, Y., & Ou, W. (2019). Locality constrained representation-based K-nearest neighbor classification. *Knowledge-Based Systems*, *167*, 38-52. 10.1016/j.knosys.2019.01.016

Gures, N., Arslan, S., & Tun, S. Y. (2014). Customer expectation, satisfaction and loyalty relationship in Turkish airline industry. *International Journal of Marketing Studies*, *6*(1), 66. 10.5539/ijms.v6n1p66

Hayadi, B. H., Kim, J. M., Hulliyah, K., & Sukmana, H. T. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. *International Journal of Informatics and Information Systems*, *4*(1), 82-94. https://doi.org/10.47738/ijiis.v4i1.80

Hong, A. C. Y., Khaw, K. W., Chew, X., & Yeong, W. C. (2023). Prediction of US airline passenger satisfaction using machine learning algorithms. *Data Analytics and Applied Mathematics (DAAM)*, 7-22. https://doi.org/10.15282/daam.v4i1.9071

Hussain, R., Al Nasser, A., & Hussain, Y. K. (2015). Service quality and customer satisfaction of a UAE-based airline: An empirical investigation. *Journal of Air Transport Management*, *42*, 167-175. https://doi.org/10.1016/j.jairtraman.2014.10.001

Hwang, S., Kim, J., Park, E., & Kwon, S. J. (2020). Who will be your next customer: A machine learning approach to customer return visits in airline services. *Journal of Business Research*, *121*, 121-126.
Kaggle.(2020). *Airline Passenger Satisfaction Dataset*. Available at: https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

Krishnan, C. G., Robinson, Y. H., & Chilamkurti, N. (2020). Machine learning techniques for speech recognition using the magnitude. *Journal of Multimedia Information System*, *7*(1), 33-40. https://doi.org/10.33851/JMIS.2020.7.1.33

Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, *6*(1), 1-16. :10.1186/s40537-019-0224-1

Leong, L. Y., Hew, T. S., Lee, V. H., & Ooi, K. B. (2015). An SEM–artificial-neural-network analysis of the relationships between SERVPERF, customer satisfaction and loyalty among low-cost and full-service airline. *Expert systems with applications*, *42*(19), 6620-6634. 10.1016/j.eswa.2015.04.043

Magsi, H., Sodhro, A. H., Al-Rakhami, M. S., Zahid, N., Pirbhulal, S., & Wang, L. (2021). A novel adaptive battery-aware algorithm for data transmission in IoT-based healthcare applications. *Electronics*, *10*(4), 367. https://doi.org/10.3390/electronics10040367

Nourbakhsh, A., & Rezaei Chelkasari, M. (2023). Airline passenger's sentiment analysis for improving the quality of airline services by using a deep learning approach. *International Journal of Applied Operational Research-An Open Access Journal*, *11*(2), 77-97. https://doi.org/10.71885/ijorlu-2023-1-638

Nurdina, A., & Puspita, A. B. I. (2023). Naive Bayes and KNN for Airline Passenger Satisfaction Classification: Comparative Analysis. *Journal of Information System Exploration and Research*, *1*(2). https://doi.org/10.52465/joiser.v1i2.167

Park, E., Jang, Y., Kim, J., Jeong, N. J., Bae, K., & Del Pobil, A. P. (2019). Determinants of customer satisfaction with airline services: An analysis of customer feedback big data. *Journal of Retailing and Consumer Services*, *51*, 186-190. https://doi.org/10.1016/j.jretconser.2019.06.009

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

Suprapto, D. S., & Oetama, R. (2023). Analysis of Airline Passenger Satisfaction Using Decision Tree and Naïve Bayes Algorithms. *Jurnal Informatika Ekonomi Bisnis*, 1493-1500. https://doi.org/10.37034/infeb.v5i4.728

Taliah, A. R. B., & Zervopoulos, P. D. (2023). Using a Bayesian meta-frontier approach to evaluate airline performance: The nexus between airline efficiency and customer satisfaction. *Heliyon*, *9*(10). 10.1016/j.heliyon.2023.e20391

Tayaba, M., Ayon, E. H., Mia, M. T., Sarkar, M., Ray, R. K., Chowdhury, M. S., ... & Puja, A. R. (2023). Transforming Customer Experience in the Airline Industry: A Comprehensive Analysis of Twitter Sentiments Using Machine Learning and Association Rule Mining. *Journal of Computer Science and Technology Studies*, *5*(4), 194-202. https://doi.org/10.32996/jcsts.2023.5.4.20

Thakur, N., & Han, C. Y. (2021). A study of fall detection in assisted living: Identifying and improving the optimal machine learning method. *Journal of sensor and actuator networks*, *10*(3), 39. https://doi.org/10.3390/jsan10030039

Ouf, S. (2023). An optimized deep learning approach for improving airline services. *Comput. Mater. Contin*, *75*(1), 1213-1233. https://doi.org/10.32604/cmc.2023.034399

Xu, Z., Zhu, G., Metawa, N., & Zhou, Q. (2022). Machine learning based customer meta-combination brand equity analysis for marketing behavior evaluation. *Information Processing & Management*, *59*(1), 102800. 10.1016/j.ipm.2021.102800