



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

Air Insight AI: Passenger Satisfaction Analysis for Airline Service Optimization

¹Neha Punnam,²Ch.Mahesh,

¹M.Tech Scholar, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India. Mail id :nehareddy3636@gmail.com

²Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India. Mail id: chekurimahesh364@gmail.com

Abstract

In this research, we use Python-based tools including Pandas, NumPy, Seaborn, Matplotlib, and a few more to conduct an exploratory data analysis (EDA) and machine learning model on a dataset representing airline customer satisfaction surveys. We use EDA to look at all the different parts of the information, including flight details, delays, and service reviews. Finding patterns and predicting passenger satisfaction levels is the next goal of using machine learning techniques like ensemble methods, regression, or classification. By combining EDA with machine learning, we can have a better grasp of the aspects that affect passenger pleasure in the aviation industry, as well as confirm our results and get deeper insights.

Problem Statement

In order to get a better knowledge of the variables that impact airline customer pleasure, this study primarily aims to extract and analyze significant insights from the 'Airline customer pleasure' dataset. Finding these elements requires sifting through data, checking it for quality, and doing a thorough statistical and visual analysis. In order to tackle this issue, the research utilizes many machine learning models, each with its own set of skills and measures for assessment. Utilizing its inherent simplicity, Linear Regression establishes a linear connection between characteristics and the target variable, serving as a baseline for further comparisons. For jobs requiring binary classification, including determining the likelihood of passenger satisfaction ratings, Logistic Regression is used. When it comes to robust prediction performance, two popular models are the Random Forest Regressor and the AdaBoost Regressor. The former is renowned for its ability to handle imbalanced datasets accurately via its many decision trees, while the latter combines several weak learners into a stronger model. The regression models are assessed using the Mean Squared Error (MSE) and R^2 Score, while the classification model is assessed using the Accuracy Score, Precision, Recall, and F1-score. The study aims to use these models to determine the main factors that affect customer pleasure, which will help the airline sector. This is highlighted in the issue statement.

Introduction

In the ever-changing airline industry, maintaining and increasing customer happiness is not just a goal, but a must for survival and growth. The advent of data mining technology has allowed for the creation of novel approaches to the study of customer preferences, habits, and variables influencing satisfaction. Airline passenger happiness is influenced by a complex web of variables; our study aims to utilize a large dataset (129,880 items) to investigate this web in detail. Once thought of as a

luxury, air travel is now a common means of transportation for both individuals and large organizations. Client expectations have evolved in tandem with this development, becoming more intricate and demanding. In addition to punctuality and safety, airlines now have an additional responsibility to provide customers with an enjoyable, tailored, and dependable travel experience. It is crucial to have a good grasp of the many factors that influence how passengers perceive and are satisfied in order to do this. At its core, this project's

dataset is a treasure trove of information. Passenger demographics like age and gender are included with travel-related data like customer type (first-time or returning), trip type (personal or business), flight class, and flight distance. From the services offered before the flight, such as online booking and check-in, to those offered during the flight, such as leg room service, comfortable seats, a clean cabin, tasty food and drink, in-flight entertainment, and Wi-Fi, it provides a comprehensive analysis of the travelers' experiences. When planning a trip, it's important to pay attention to the statistics on arrival and departure delays. The project's goal is to answer critical questions such, "Which variables make the biggest effects on how satisfied passengers are?" by using sophisticated data mining methods and machine learning algorithms to this information in order to find patterns and connections. Can we generalize about the expectations and satisfaction levels of various flying classes or populations? When all the parts of the flying experience work together, how satisfied are you with it overall? An important first step in comprehending the contemporary airline passenger's mental state, this research will use a number of data mining techniques—including clustering, classification, regression analysis, and association rule mining—to address the fundamental questions. Airline companies may use our study's findings to inform service redesigns that boost customer happiness and loyalty and, in the long run, help them thrive in a cutthroat industry.

Literature Survey

Service Quality

According to Ganiyu (2017) and Ishutkina and Hansman (2008), the airline industry plays a pivotal role in the global economy since it greatly impacts tourism, international commerce, and economic growth. In order to retain the market and provide future development, assessing and improving customer satisfaction is a crucial feature of this sector. Regarding the quality of airline service, there have been a lot of research. Parasuraman et al. (1988) developed the SERVQUAL model, which is widely used to assess the quality of services in many different sectors, including aviation. Two more qualities, certainty and empathy, were added to this model's assessed 10 traits in order to emphasize the significance of comprehending customers' requirements and offering them competent and thoughtful service (Parasuraman et al. 1988). The impact of service quality on a passenger's decision to

choose an airline has been shown in studies conducted in this field (Ostrowski et al., 1993). Further research by Tsauro et al. (2002) and others has shown that responsiveness and tangibility are critical indicators of service quality (Gilbert, Chen & Chang et al., 2005).

Methodology

Data Collection

Our research relies on a dataset called "Airline Passenger Satisfaction," which we acquired specifically for this study. The data offered by Kaggle includes a variety of variables that capture the flight patterns and levels of satisfaction of passengers. The data pre-processing phase included cleansing the dataset and making sure it was relevant to the intended analysis. The flowchart is a visual depiction of the standard data analysis cycle, which begins with data acquisition and ends with data extraction for processing. Data is handled in a way that ensures it is cleaned up if issues like duplicate or missing values are detected. After that, you may go on to exploratory data analysis, which is great for building data visualizations and looking for important trends and patterns. The goals were tackled using a variety of machine learning techniques, such as classification to organize ordered data and regression to predict values. Before the project ends, the model results are created, checked for correctness, and the right communication techniques are used.

Data Preprocessing

We will now move on to the cleaning and preprocessing steps: Since we discovered certain entries to be duplicated and may impact the study, we tackled the issue of duplicate data first and foremost. Addressing Data Missing: The complete dataset and the "Arrival Delay" parameter were examined for missing values. To ensure the dataset's quality, records with missing values will be removed since the proportion of missing values is rather low. To facilitate their use in quantitative analysis, categorical variables were "encoded" using suitable encoding procedures, which converted them into a numerical representation. To reduce the amount of data needed for the study, we have eliminated superfluous variables from the database, such as passenger IDs, which were unique for each case.

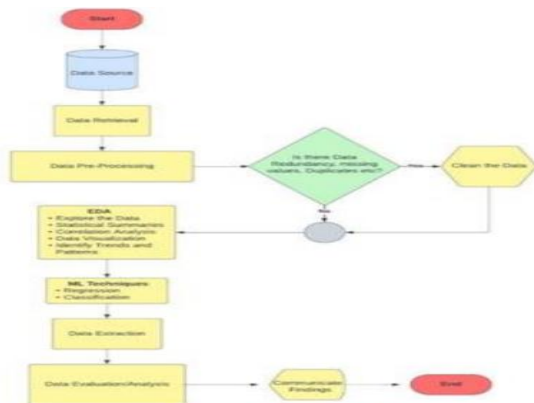


Fig.1 Flow Chart: Methodology

Exploratory Data Analysis (EDA)

The numerical variables' central tendency, dispersion, and distribution shapes were computed and assessed in order to understand the descriptive statistics. For continuous variables such as 'Age' and 'Flight Distance,' we created histograms to show how they were distributed. Boxplots: We compared passenger happiness with factors like 'Age' and 'Flight Distance' to understand the distribution of satisfaction across various categories.

Third, we ran a correlation analysis to see how all the numerical variables were related to one another; this helped us find factors that could be a predictor of passenger happiness. #4. Scatter Plots: We looked at the correlation between 'Departure Delay' and 'Arrival Delay,' among other pairs of variables, using scatter plots.

Step four of our process involves training, evaluating, and validating models using the various machine learning approaches that were chosen for the research. Training and testing the algorithms/models and then assessing their prediction and selection performance are the two main components of this process. After much work, four ML models were finally approved: One simple model is "Modelled Linearly," which states that the characteristics you choose should have a direct correlation with the result. Here, it served just as a reference point. 2. A model that describes the odds of class membership is logistic regression. Predicting the level of satisfaction is an example of a two-class classification issue that may be solved using this kind of model.

3. The Random Forest Regressor is a multiple model that is known to provide excellent predictions, even with imbalanced datasets. It consists of several decision trees. Fourth, the AdaBoost Regressor is an ensemble approach that combines several weak

classifiers into a single strong one. Decision trees often make use of it to improve their results. Currently, the models were assessed using appropriate metrics: Linear Regression, Random Forest, and AdaBoost were the regression models that used Mean Squared Error (MSE) and R² Score. Accuracy Score, Classification Report (including precision, recall, and F1-score), and Logistic Regression were generated for the classification model.

Data Mining Techniques and Analysis

Several software tools and libraries were used to do the analysis, such as: Python modules for data manipulation (Pandas) and machine learning (Skit-Learn) • Prototyping and data visualization software that is frequently used: spreadsheets • A Jupyter Notebook to record and execute coding operations while doing analysis Our main objective was to establish the predictors' proxies and their relationship with the customer satisfaction level using regression and correlation analysis. This helped us understand the strength and direction of the various predictor variables in relation to the satisfaction level, as well as which variables were more reliably and strongly associated with the level of satisfaction within the assessment model. We received a wealth of information on the impacts on customer satisfaction inside the data set that was the subject of our research. Building and assessing models to predict consumer satisfaction level was the next step, all based on the previous investigation. The fact that there is a strong relationship between the variables "Arrival Delay" and "Departure Delay" has generated questions about linearity, which should be highlighted. To address the issue, we decided to use only one of them in the modeling process, ensuring that it would not be unnecessary or negatively impact the model's performance. The next part of the series covers how our strategy's categorization algorithms began to think about outline. Based on the parameters identified as relevant during the preliminary exploratory data analysis, these algorithms aimed to identify the characteristics of pleased and unsatisfied travelers. We aimed to implement categorization algorithms that could handle all levels in relation to the factors taken into account and the anticipated effect on passengers' levels of satisfaction. This strategy to understand the levels of passenger satisfaction depending on the established determinants of satisfaction was formed by first conducting correlated analysis to determine if there

were any associations, then using models to address the multicollinearity problem, and finally considering the use of classification algorithms.

Discussion

In order to analyze and learn from the dataset, the right procedures were followed in a sequential fashion that guarantees the data is consistent and reliable. Detailed explanations of these stages are provided below: Information Gathering and Initial Investigation From the First Load of "Airline The dataset used for this research, known as "Passenger Satisfaction," is given to a Data Frame using Panda's library. Dataset loading, which assigns the data frame, is the first step in the analysis. By extracting a few of preliminary records, we can get a bird's-eye perspective of the data. Important for setting the stage for further in-depth study, it is vital. There are 129880 rows and 24 columns in the dataset, making it fairly huge. Data Preparation and Cleaning Perhaps the most crucial part of doing the study is the cleaning and preparation of the data sets. This procedure includes: Finding Unique Values—A crucial part of assessing the breadth and depth of a database across various parameters is identifying the unique values included inside each column. To ensure the data remains intact, duplicate record management involves checking the dataset for instances of duplicate records and resolving the issue of duplicate entries. If there are duplicate records, it could skew the results and lead to bad conclusions. Dealing with Values That Aren't There: Some columns, such "Arrival Delay," which is lacking values, are given extra attention. It has been decided that 393 instances that do not have a "Arrival Delay" will be removed. This procedure guarantees a high-quality dataset and its dependability. Examining the Data Frame: The info() method gives a quick overview of the dataset's structure by detailing the categories of variables and the total number of non-null variables. For the purpose of planning future data processing and analysis, this information is vital. Statistics for Characteristics For a comprehensive statistical review of the data set, use the describe() function. Every numerical variable in the dataset has its range, median, standard deviation, and average included in this summary. Before implementing complex statistical processes, it is vital to evaluate elements such as the overall distribution and measures of central trends of the dataset. This phase is crucial. Python Library Utilization 1. Pandas: One of the most important Python libraries for a variety of

data processing and annual tourist analyses. The primary usage of pandas in this research is to load the dataset into a DataFrame from a csv file. Cleanup, transformation, and analysis of tabular data sets are made easy and straightforward by pandas's DataFrame structure. When dealing with basic data exploration, summarization, outlier handling, and missing value difficulties, functions like head(), info(), describe(), and dropna() come in handy. Python scientific calculations rely on the NumPy library. Its primary emphasis is on numerical computation, an essential tool for any data processing. For this research, NumPy is a great complement to pandas, which is useful for things like doing mathematical calculations on DataFrame columns. When it comes to data visualization, two libraries are crucial: Matplotlib.pyplot and Seaborn. Matplotlib.pyplot enables the creation of detailed static charts in a variety of formats.

EDA Visualization

Following extensive investigation, we determined the links between the datasets in order to have a better grasp of them and to aid in summarizing our results for the readers. Take the histogram for Age Distribution as an example; it shows the range of ages of passengers. Although it is not entirely symmetrical, the distribution of ages among passengers follows a bell-shaped curve, which is considered normal. Passengers in the Middle Ages: The distribution's highest point is around the 30–40 age group, suggesting that a considerable percentage of the airline's customers are in the middle age range. There is a noteworthy frequency of passengers up to the age of 80 in the distribution, which spans a broad range of ages from young to old. The frequency increases gradually from the youngest ages up to the peak, then declines more gradually. This indicates that there are less extremely young (children) and older (senior) passengers than middle-aged passengers.

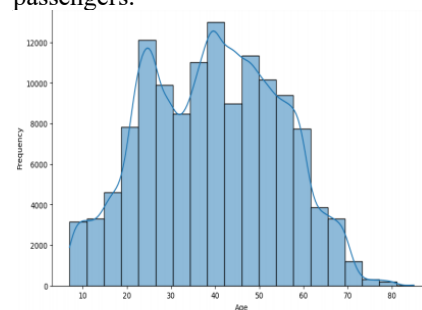


Fig 2: Age distribution in airline passenger satisfaction

When making changes to marketing or services, these age demographics might be helpful. For example, if the airline is looking to attract more customers in the younger or older demographics, they may think about offering certain services or facilities that are popular with that age groups.

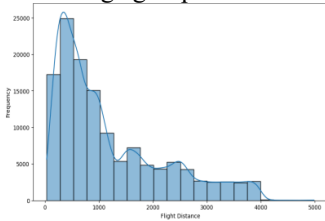


Fig 3: Flight Distance Distribution in airline passenger satisfaction

The frequency of flights over various distances is shown in this histogram for Flight Distance Distribution. Our interpretation of the graph is as follows: The graph shows that there are more flights covering small distances than large distances, which is known as a right-skewed distribution. The bulk of flights are short-distance, as shown by the increased frequency of flights within the 0-1000 km range. Long-Distance Flights: This dataset seems to include fewer instances of long-distance flights since the frequency of flights declines with increasing distance. The histogram shows a few very long-distance flights, which may be seen as outliers, since there are bars at the far right of it. A concentration on short-haul flights or a client base that mostly travels short distances are two possible interpretations of this distribution regarding the airline's operations. It could also guide decisions about how to divide up scarce resources, such crew schedules and fleet management, in ways that are unique for flights of varying lengths.

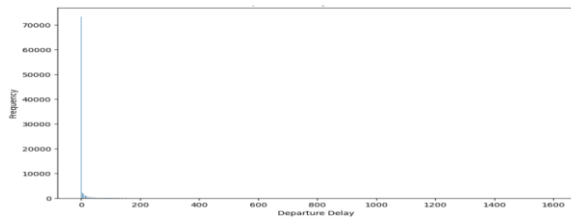


Fig 4: Departure Delay distribution in airline passenger satisfaction

Most flights depart on time or with a little delay, as seen by the right-skewed Departure Delay Distribution histogram. There may not be many planes with very lengthy departure delays, as shown by the long right tail, but these outliers may have a major influence on customer happiness. In order to identify and solve the reasons of excessive delays, it

is sometimes necessary to do particular analysis of outliers when dealing with such a distribution.

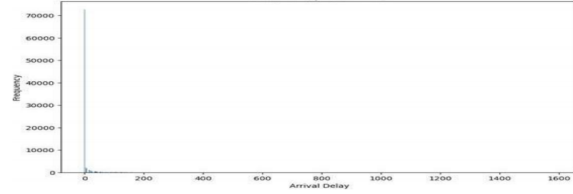


Fig 5: Arrival Delay distribution in airline passenger satisfaction

Just like the Arrival Delay Distribution histogram, it is biased to the right. This indicates that while there are cases of significant delays, the majority of flights arrive on time or with just little delays. These anomalies, like departure delays, must be thoroughly investigated as they might lead to general consumer discontent and could cost the airline money in compensations and scheduling disruptions.

Data-Driven Decisions

Transforming or segmenting the delay variables (e.g., placing delays into bins) may be necessary for predictive modeling in order to capture the link between delays and passenger satisfaction more accurately. Customer service interventions, such as proactive communication on delays, may be better planned with knowledge of the distribution of delays.

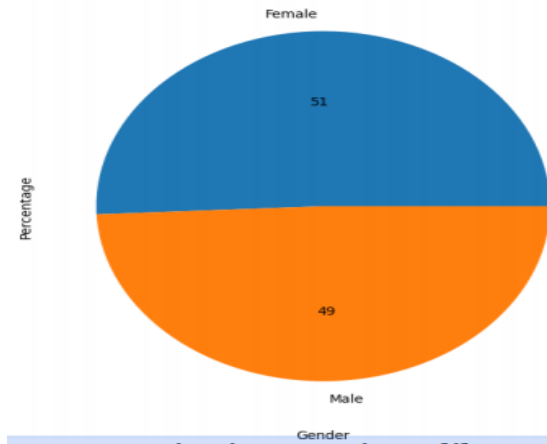


Fig 6: Gender distribution in airline passenger satisfaction

The gender breakdown of the passengers is almost balanced in the pie chart, with slightly more females than men (51% to 49%). Given this parity, it's safe to assume that marketing and service adjustments will have an equal impact on passengers of both sexes and that gender won't play a role in any study of passenger happiness. In any customer satisfaction

study, having a balanced gender distribution ensures that all opinions and preferences are represented.

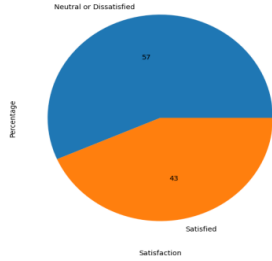


Fig 7: Satisfaction level of all the passengers in airline in percentages

In terms of satisfaction, 43% of passengers were pleased while 57% were either indifferent or unhappy, as seen in the pie chart. The fact that the majority of customers are dissatisfied suggests that the airline still has room to grow in terms of customer satisfaction. Passenger services or experiences might be enhanced according to the satisfaction statistics. As a means of better serving its customers, the airline may look into what exactly is making them unhappy.

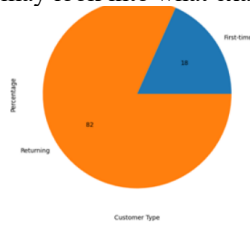


Fig 8: Customers Type Distribution Pie Chart in airline passenger satisfaction

Based on the data in the graphic, it can be seen that 82% of passengers are returning customers and just 18% are new consumers. This may indicate a very devoted clientele or an efficient customer retention campaign. On the other hand, it suggests that there might be opportunities to expand the clientele by appealing to first-time travelers. The airline seems to have a devoted client base, as seen by the high rate of repeat business, which bodes well for consistent income. On the other hand, increasing market share could need plans to draw in new consumers.

Implications for Airline Passenger Satisfaction

If we want to make passengers happier, we need to know these relationships. If we could improve the on-board service, seat comfort, and leg space all at once, it would improve the entire experience for passengers since these three factors are highly connected. Bear in mind, nevertheless, that a correlation does not prove a causal relationship. These connections could help with research questions, but they don't prove that one

variable causes changes in another. To further investigate potential correlations, other statistical analyses like regression modeling might be performed. In order to maximize customer happiness, the airline may focus on fixing problems that have strong relationships to that outcome. By incorporating customer happiness data into this heat map, the airline might pinpoint the factors that have the greatest impact on customer satisfaction and work to enhance those areas. As an example, if there was a strong positive association between 'In-flight Entertainment' and 'Satisfaction,' the airline should prioritize improving entertainment offerings to boost overall satisfaction levels.

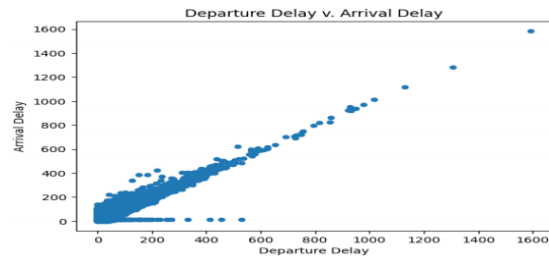


Fig 9: Relationship between 'Departure Delay' and 'Arrival Delay' in airline

Machine Learning Model Evaluation

The following four machine learning models were tested on a dataset consisting of airline customer satisfaction surveys: Logistic Regression, Linear Regression, Random Forest Regressor, and AdaBoost Regressor. Among them, Logistic Regression had an accuracy score of around 66.07%, which is considered moderate. But as we dug more, we saw that the accuracy and recall numbers for the two satisfaction groups were skewed, so it was clear that there were problems with accurately identifying "satisfied" consumers. The model's recall was 0.43 and accuracy was 0.67 for "satisfied" instances, indicating that it had difficulty capturing a substantial number of real "satisfied" occurrences.

```

Logistic Regression Metrics:
Accuracy Score: 0.660686788041269
Classification Report:

```

	precision	recall	f1-score	support
0	0.66	0.83	0.74	14723
1	0.67	0.43	0.53	11253
accuracy			0.66	25976
macro avg	0.66	0.63	0.63	25976
weighted avg	0.66	0.66	0.64	25976

```

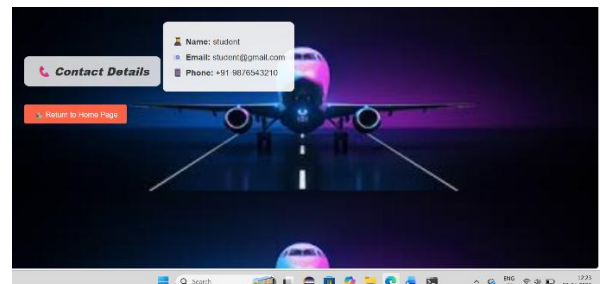
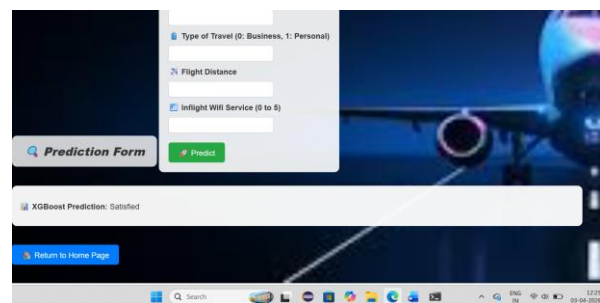
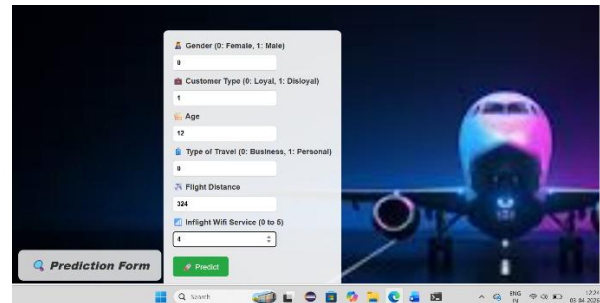
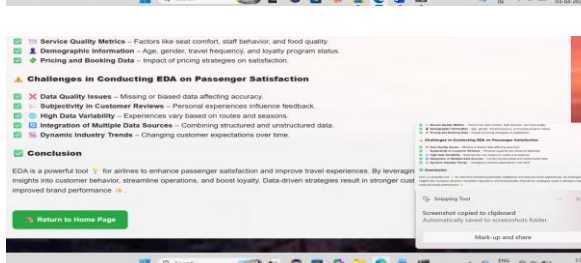
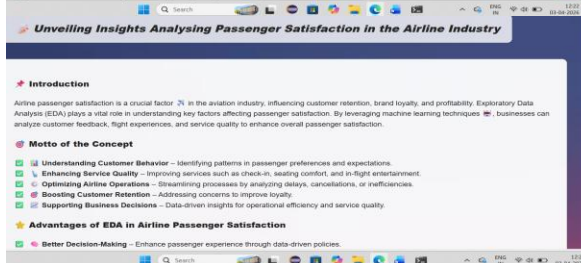
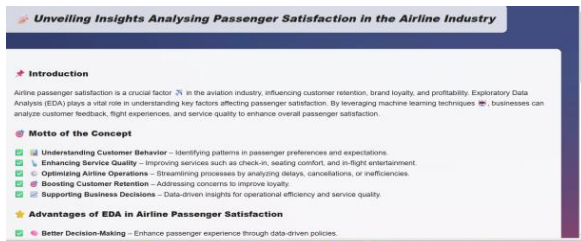
Random Forest Regressor Metrics:
Mean Squared Error: 0.24639502730153187, R2 Score: -0.003487250375826756
AdaBoost Regressor Metrics:
Mean Squared Error: 0.2119083404094793, R2 Score: 0.136965870707363

```

Fig 10: Performance metrics for model

However, the Random Forest Regressor and the ensemble approaches were clearly not up to the task of dealing with the categorical focus variable. A negative R^2 value, which indicates poor model performance, provided proof of this. Several suggestions were made to enhance the accuracy and reliability of the model. These included looking into other pertinent features, fine-tuning the model's hyperparameters, preventing possible class imbalances, and considering other models such as neural networks or Support Vector Machines. Refining the predictive models and obtaining greater accuracy related to the aims of the airlines requires aligning the model selections with the project's objectives and undertaking in-depth study.

Results



Conclusion

Finally, this paper's predictive modeling and exploratory data analysis have provided a thorough comprehension of the elements impacting airline customer pleasure. A number of conclusions have been drawn from the investigation; one of the most important is that logistic regression for classification produced an accuracy of around 66.07%, suggesting a rather reliable level of prediction. When it comes to recognizing "not satisfied" consumers, this model shows acceptable accuracy and recall. However, there is a clear need for improvement when it comes to identifying "satisfied" passengers. The AdaBoost regression model had a poor fit to the data and the Random Forest model had a negative R^2 score, hence its performance was subpar. A categorical target variable, such as "Satisfaction," may not be the ideal candidate for regression models, according to these results. The article emphasizes the possibility of

better model performance via feature engineering, modifying the model, balancing the classes, and investigating other modeling approaches like neural networks or Support Vector Machines. These results highlight how critical it is to match model selection with analytic goals and data characteristics. The study makes a substantial contribution to the field of customer service in the airline sector with these predicted insights. In particular, it encourages research into areas with the potential to greatly improve the passenger experience. Industry stakeholders seeking to engage in data-driven initiatives to enhance customer happiness and loyalty may find a complete resource in the outcomes and methodology utilized. We can improve our knowledge of passenger requirements and preferences by including a larger spectrum of data and refining modeling tools. This study will lead the way in this regard.

References

- [1]Ashwika, Dishali G. K., and Hemalatha N., "Airline Passenger Satisfaction Prediction Using Machine Learning Algorithms," Redshine Archive, 2020.
- [2]R. Murugesan, A. P. Rekha, N. Nitish, and B. Raghavan, "Forecasting Airline Passengers' Satisfaction Based on Sentiments and Ratings Using Machine Learning Techniques," *Journal of Air Transport Management*, vol. 120, 2024.
- [3]K. Q. Maqbool, M. F. Kayani, M. Z. Hasan, M. Z. Hussain, and M. A. Yaqub, "Unlocking Insights into Airline Passenger Satisfaction with Data-Driven Exploratory Mining Approach," *Dialogue Social Science Review*, 2025.
- [4]X. Jiang, Y. Zhang, Y. Li, and B. Zhang, "Forecast and Analysis of Aircraft Passenger Satisfaction Based on RF-RFE-LR Model," *Scientific Reports*, vol. 12, 2022.
- [5]S. Jethwa and N. Vora, "A Comprehensive Analysis of Airline Passenger's Satisfaction through Classification Model," *Journal of Trends in Computer Science and Smart Technology*, vol. 6, no. 4, 2024.
- [6]A. C. Y. Hong, K. W. Khaw, X. Y. Chew, and W. C. Yeong, "Prediction of US Airline Passenger Satisfaction Using Machine Learning Algorithms," *Data Analytics and Applied Mathematics*, 2022.
- [7]S. Kumar and M. Zymbler, "A Machine Learning Approach to Analyze Customer Satisfaction from Airline Tweets," *Journal of Big Data*, vol. 6, no. 62, 2019.
- [8]"Investigating Airline Passenger Satisfaction: Data Mining Method," *Research in Transportation Business & Management*, vol. 43, 2022.
- [9]T. Mirthipati, "Enhancing Airline Customer Satisfaction: A Machine Learning and Causal Analysis Approach," arXiv preprint, 2024.
- [10]S. Wu and Y. Gao, "Happy or Grumpy? A Machine Learning Approach to Analyze the Sentiment of Airline Passengers' Tweets," arXiv preprint, 2022.
- [11]Chang, EC., Huang, SC. & Wu, HH. Using K-means method and spectral clustering technique in an outfitter's value analysis. *Qual Quant* 44, 807–815 (2010). <https://doi.org/10.1007/s11135-009-9240-0>
- [12]Bellizzi, M. G., Eboli, L., & Mazzulla, G. (2020). An online survey for the quality assessment of airlines' services. *Research in Transportation Business & Management*, 100515. <https://doi.org/10.1016/j.rtbm.2020.100515>
- [13]Noriza, M. S. (2013). Customer satisfaction with service delivery in the life insurance industry: An empirical study. *Jurnal Pengurusan*, 38, 101–109.
- [14]Malaysia, Tourism (2013). Malaysia receives a record 25 million tourist arrivals .http://corporate.tourism.gov.my/mediacentre.asp?page=news_desk&news_id=763(accessed 25-09-2012) [
- 15] Liou, J. J. H., Hsu, C.-C., Yeh, W.-C., & Lin, R.-H. (2011). Using a modified grey relation method for improving airline service quality. *Tourism Management*, 32(6), 788
- [16]Ganiyu, R. (2017). Customer Satisfaction and Loyalty: A Study of Interrelationships And In Nigerian Domestic Airline Industry. http://ojbe.steconomicsuoradea.ro/wp-content/uploads/2017/03/OJBE_vol-21_01_Ganiyu.pdf
- [17]Abdullah Al Noman, Md Tanvir Rahman Tarafder, S. M. Tamim Hossain Rimon, Asif Ahamed, Shahriar Ahmed, and Abdullah Al Sakib, "Discoverable Hidden Patterns in Water Quality through AI, LLMs, and Transparent Remote Sensing," *The 17th International Conference on Security of Information and Networks (SIN 2024)*, Sydney, Australia, 2024, pp. 259–264.
- [18]S. M. T. H. Rimon, Mohammad A. Sufian, Zenith M. Guria, Niaz Morshed, Ahmed I. Mosaddeque, and Asif Ahamed, "Impact of AI-Powered Business Intelligence on Smart City Policy-Making and Data-Driven Governance," *International Conference on Green Energy, Computing and*

Intelligent Technology (GEn-CITy 2024), Johor, Malaysia, 2024.

[19]A. Nuthalapati, "Building Scalable Data Lakes for Internet Of Things (IoT) Data Management," Educational Administration: Theory and Practice, vol. 29, no. 1, pp. 412–424, Jan. 2023, doi:10.53555/kuely.v29i1.7323.

[20]M. A. Sufian, Z. M. Guria, N. Morshed, S. M. T. H. Rimon, A. I. Mosaddeque, and A. Ahamed, "Leveraging Machine Learning for Strategic Business Gains in the Healthcare Sector," 2024 International Conference on TVET Excellence & Development (ICTeD-2024), Melaka, Malaysia, 2024.