Alex Parson
Ammar Mustafa
Isha Verma
Navya Singhal
Theresa Sushil

Airline Passenger Satisfaction Project Report

Group 1 (MIS S381N)

1. Project Goals

1.1 Description

This project investigates the dataset from a US airline passenger satisfaction survey aiming to answer the question: What factors are highly correlated with a satisfied or dissatisfied passenger?

1.2 Significance of the problem

Customer satisfaction is a critical metric for airlines, directly influencing customer loyalty and repurchase intentions. The aviation sector is a significant contributor to the US economy, with a forecasted growth rate of 1.7% CAGR. Retaining satisfied customers is far more cost-effective than acquiring new ones, making the insights from this study valuable not only for the airline in question but also for the broader airline industry, which is poised to grow at a 4.1% CAGR over the next 20 years. The findings of this project could help airlines enhance customer experiences, improve satisfaction rates, and ultimately boost their profitability.

1.3 About the Dataset

The dataset includes a range of demographic and satisfaction-related variables, such as age, gender, type of travel, travel class, and ratings for various services like inflight Wi-Fi, seat comfort, and online booking with ratings collected on a Likert scale of 1 (lowest) to 5 (highest). If a particular service was missing on a particular flight, it was marked as 0 by the customer. Our dataset was sourced from Kaggle with data from almost 130k customers across 22 variables.

2. Exploratory Analysis

Through exploratory data analysis, several patterns were uncovered within the dataset. A new "average delay" feature was created by calculating the mean of arrival and departure delays, driven by a strong correlation (0.96) between these variables (Fig 1). We tested several hypotheses based on trends in the industry and found the following patterns:

H1: All service ratings are important for predicting customer satisfaction

Finding: Only some factors like "Online Booking", "In-Flight WiFi services" & "Online Boarding" have a high impact on dissatisfaction

H2: Flight delays have a negative impact on passenger satisfaction

Finding: Flight delays do not have a strong relationship with satisfaction, and are highly dependent upon seating class. Travelers in 'Eco' and 'Eco Plus' experience greater dissatisfaction as the length of delay increases, while 'Business' travelers' satisfaction remains fairly constant (Fig 2,3)

H3: Purpose of travel has no impact on satisfaction in the same travel class

Finding: Personal travelers were far more likely to be dissatisfied with their trip, which was consistent across the seating class. Additionally, personal travelers and business travelers also differed in their ratings of amenities, even when conditioning on flight length and seating class (Fig 4).

H4: People are more likely to be dissatisfied on a longer flight

Finding: Class of travel impacted the satisfaction level across all distance ranges of a flight

H5: Loyalty to the airline is a strong indicator of satisfaction

Finding: Loyalty to an airline plays a significant role in overall passenger satisfaction, which was expected (Fig 5).

3. Model Techniques & Comparison

The most impactful features identified were online booking, inflight Wi-Fi, and online boarding. These features were used across all the following models to predict satisfaction.

- **Naive Bayes**: This model highlighted that personal travel, economy class, and disloyal customers are strong indicators of dissatisfaction. (Fig 11)
- **K-Nearest Neighbors (KNN)**: With k=5, this model achieved an accuracy of 92%, significantly improving over the baseline. (Fig 10)
- **Logistic Regression**: Achieved an accuracy of 87%, indicating that personal travel is a major dissatisfaction driver, while online boarding is a significant satisfaction indicator. (Fig 9)
- **Ensemble Methods (Random Forest & Gradient Boosting)**: These models performed the best, with Random Forest achieving 96% accuracy, emphasizing the importance of the service ratings in predicting satisfaction.

Among the models, **Ensemble Methods delivered the highest performance**, achieving an accuracy of 96%, precision of 95%, sensitivity recall of 97%, and specificity of 94%. (Fig 6,7). Logistic Regression also performed well, with an accuracy of 87% and a 43% improvement compared to the baseline (Table 1).

We identified the following top five predictors based on their importance from the models (Fig 8):

1. Online Boarding
2. Class of Travel - Business
3. In-flight Wi-Fi Service
4. In-flight Entertainment
5. Type of Travel - Business/Personal

4. Recommendations & Findings

Based on our findings, we recommend that airlines prioritize the following areas:

- **Enhance Digital Touchpoints**: Focus on improving online boarding and inflight Wi-Fi services. For example, simplifying the web check-in process and introducing tiered pricing for Wi-Fi could significantly improve customer satisfaction, especially for business travelers.
- **Segmented Loyalty Programs**: Implement tailored loyalty programs, offering periodic upgrades and personalized rewards for frequent economy passengers. This could incentivize customers to upgrade to business class, with higher satisfaction rates.

- **Maintain Standards Across Services**: Ensure that basic services like cleanliness and food & drink meet customer expectations. These non-critical factors, if poorly managed, could lead to dissatisfaction.

  Some surprising findings contrary to our initial assumption were that factors like flight delay and food & drink did not significantly impact overall satisfaction. We hypothesize that in cases of delays airlines offer compensation to passengers which might negate the dissatisfaction.

5. Future Scope

- **Flight Delay Analysis**: Conduct a detailed study on the impact of flight delays for each class of travel, and assess the effectiveness of compensation strategies like vouchers, miles, or discounts.
- **Expand Demographic Data**: Collect additional socio-demographic data (e.g., income, education, occupation) to uncover more accurate trends in satisfaction.
- **Price Perception**: Include variables like flight price to study the impact of perceived price fairness on satisfaction (Fig. 12).
- **Post-COVID Analysis**: Perform the analysis on post-COVID customer data to understand shifts in passenger priorities, such as the increased importance of cleanliness and add-on services like seat selection.

5. Appendix

Fig 1: Correlation between arrival and departure delay in minutes

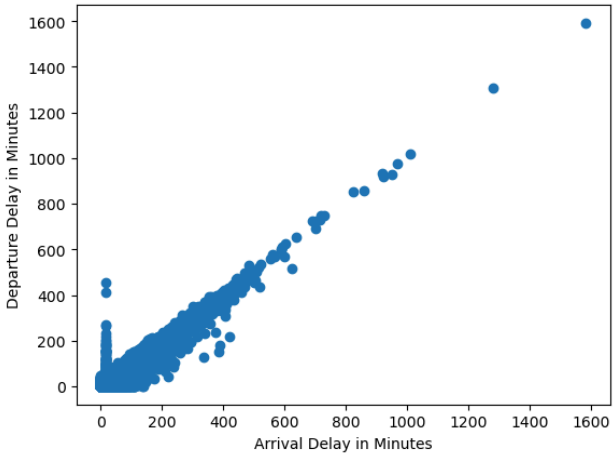


Fig 2: Satisfaction based on average flight delay in hours

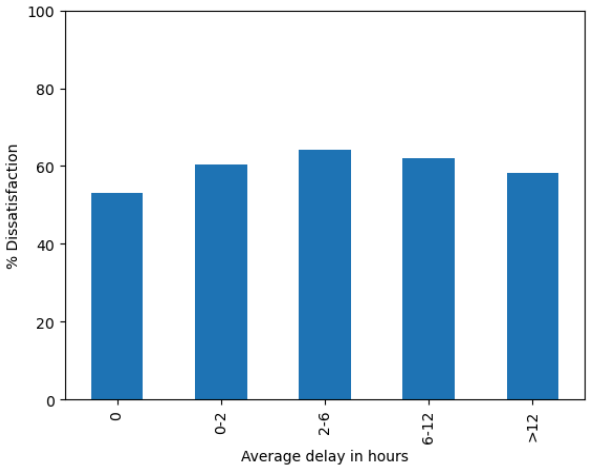


Fig 3: Satisfaction based on average flight delay in hours at a class level

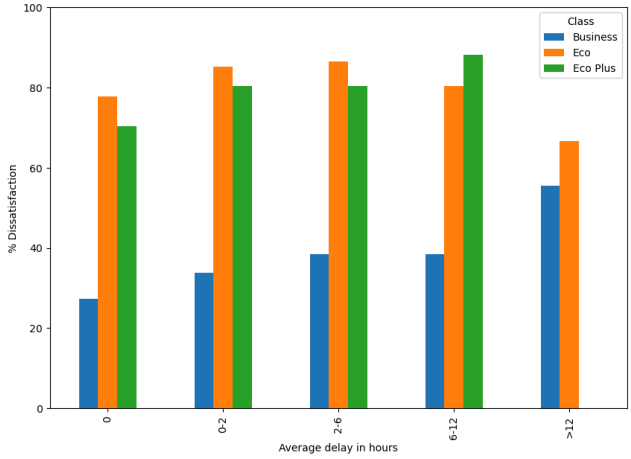


Fig 4: Satisfaction rating based on travel type

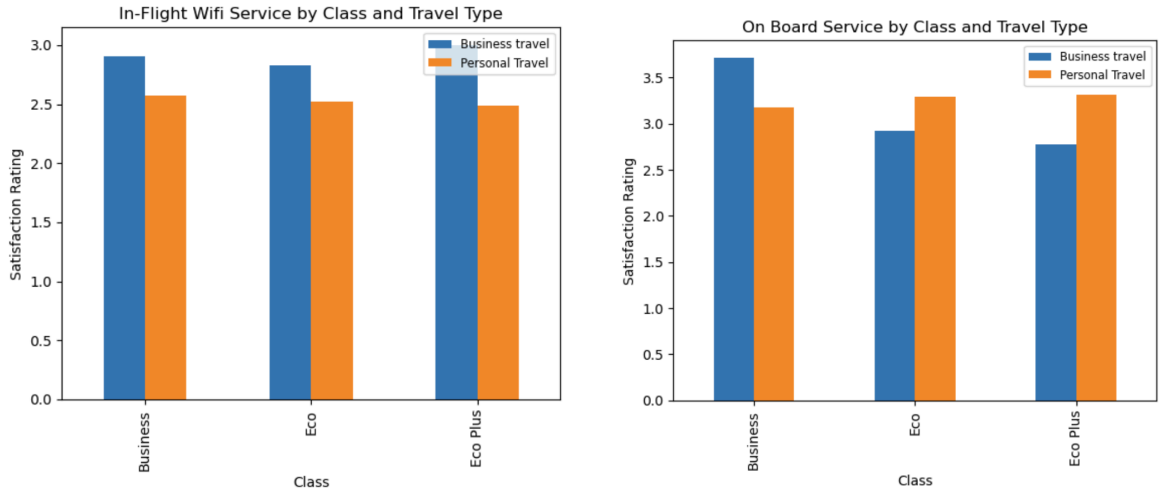


Fig 5: Satisfaction based on loyalty

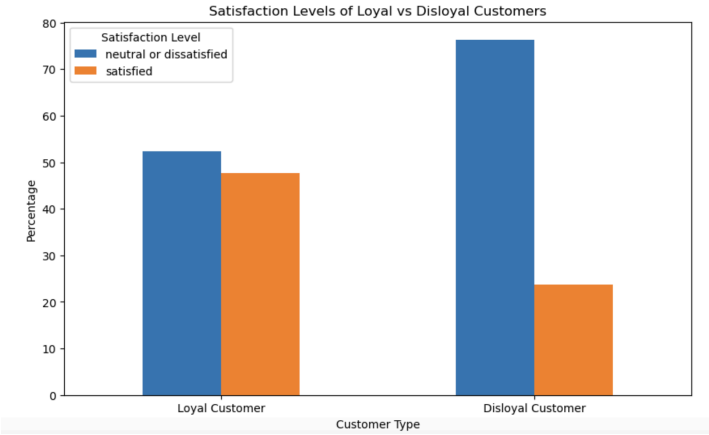
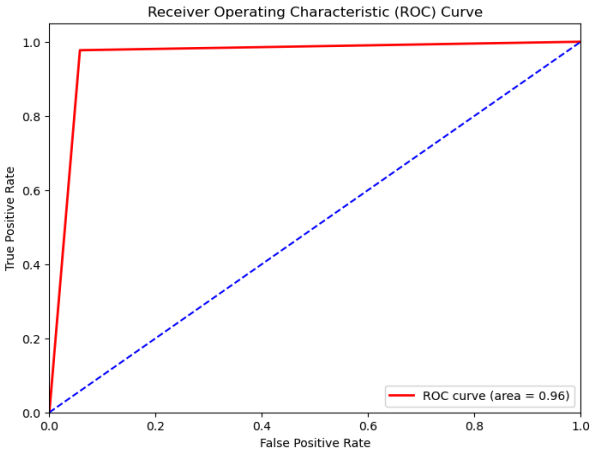


Fig 6: ROC curve of the ensemble classifiers

Random Forest



Boosting

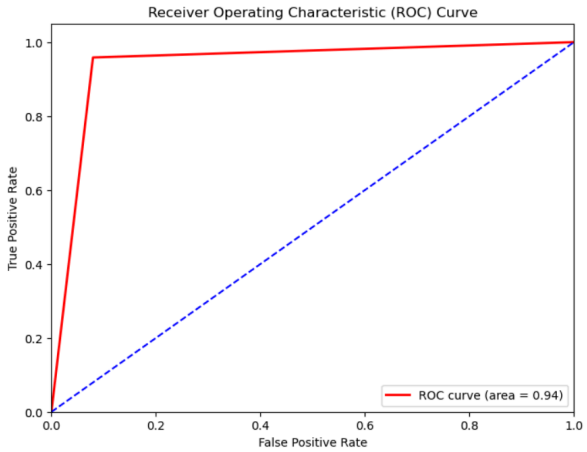
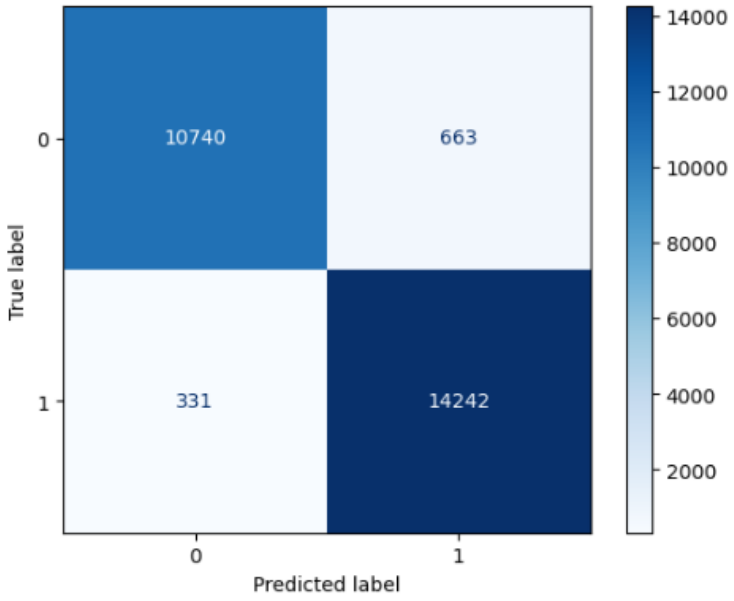

Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.94)

Fig 7: Confusion matrix of ensemble classifiers

Random Forest



Boosting

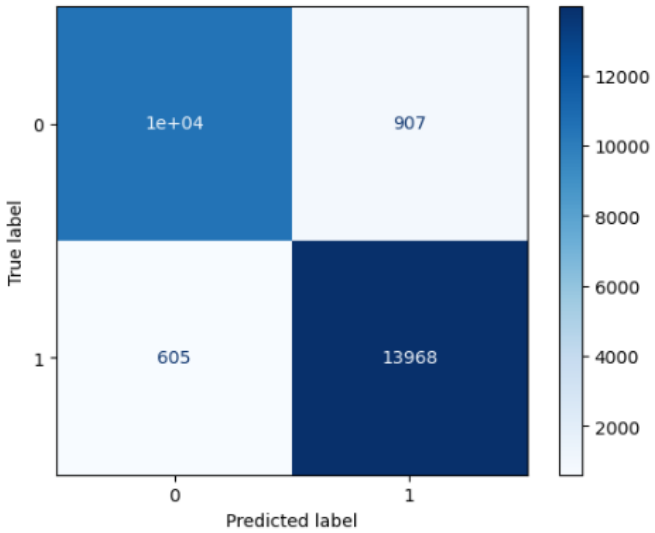


Fig 8: Variable Importance of the Random Forest classifier

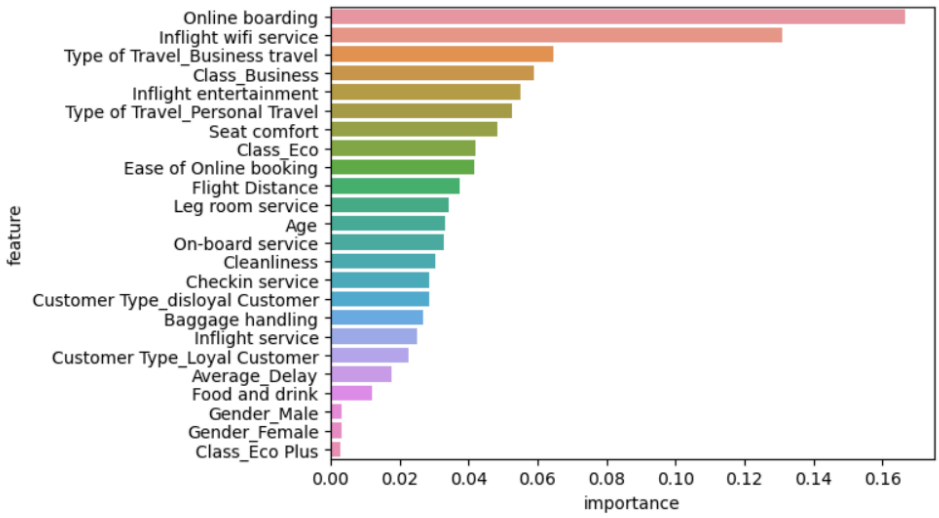
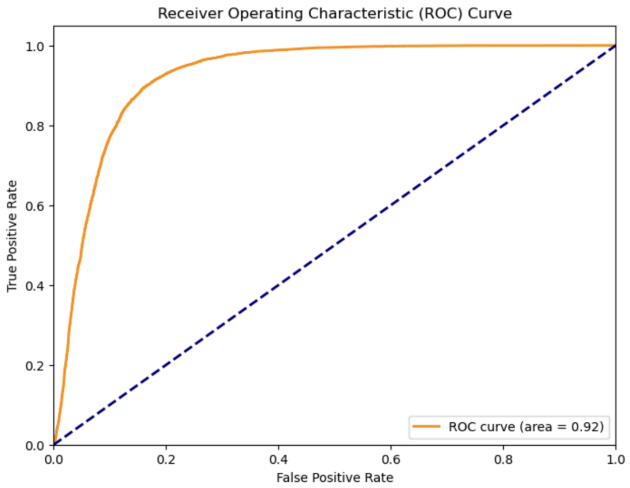
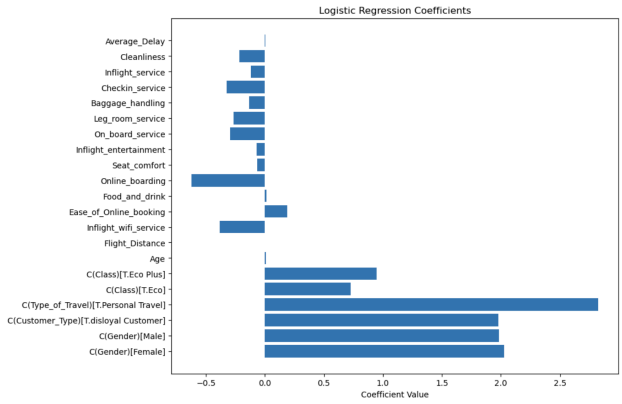


Fig 9: Logistic Regression

ROC Curve



Logistic Regression Coefficients



Confusion Matrix

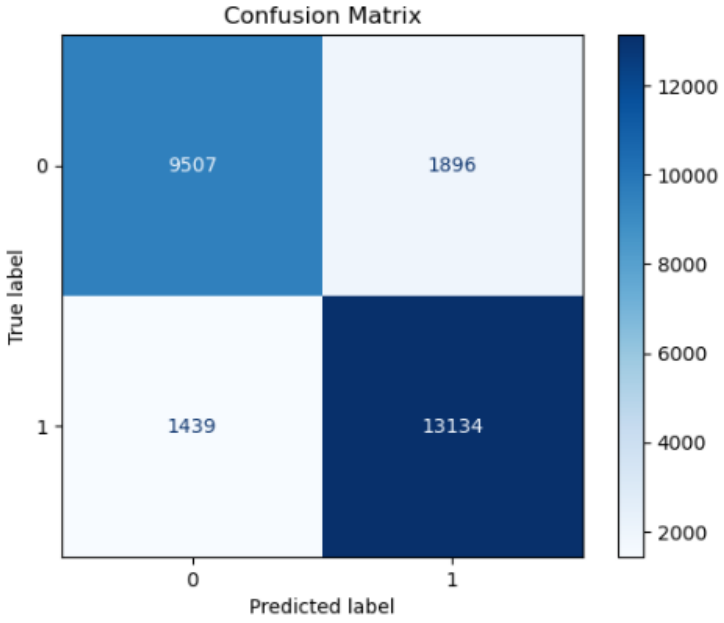
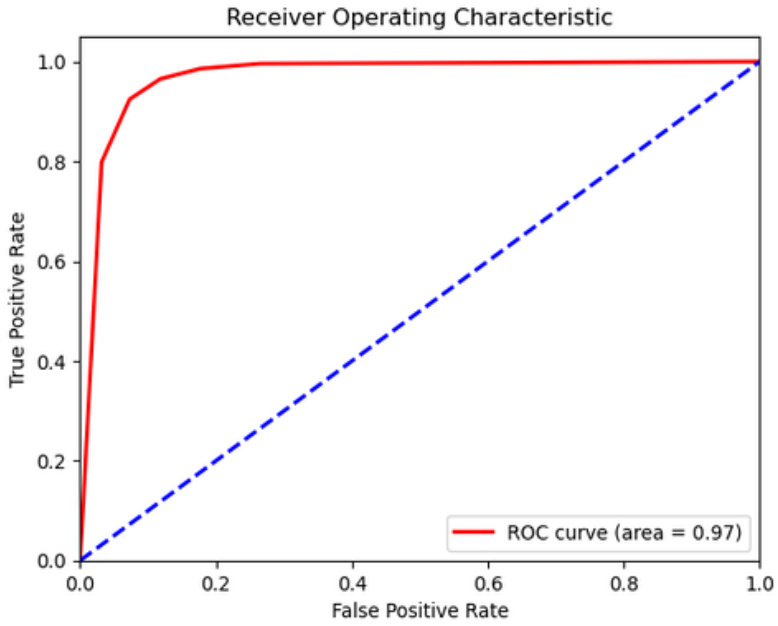


Fig 10: K-Nearest Neighbor

ROC Curve



Confusion Matrix

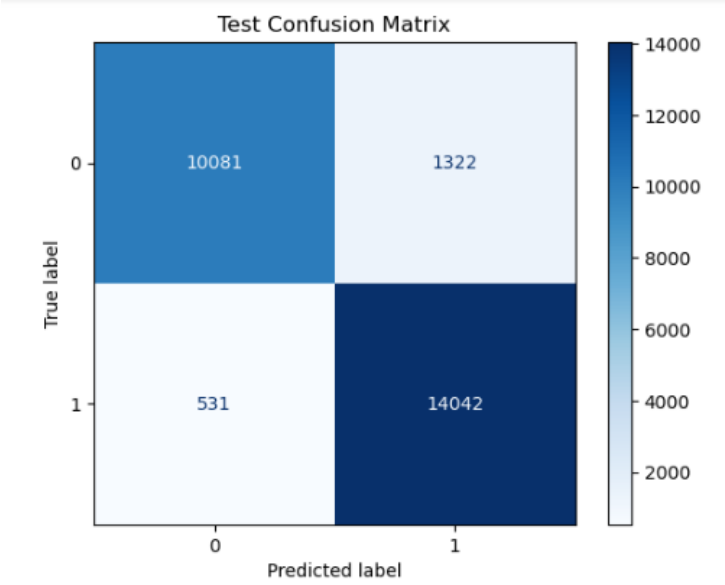
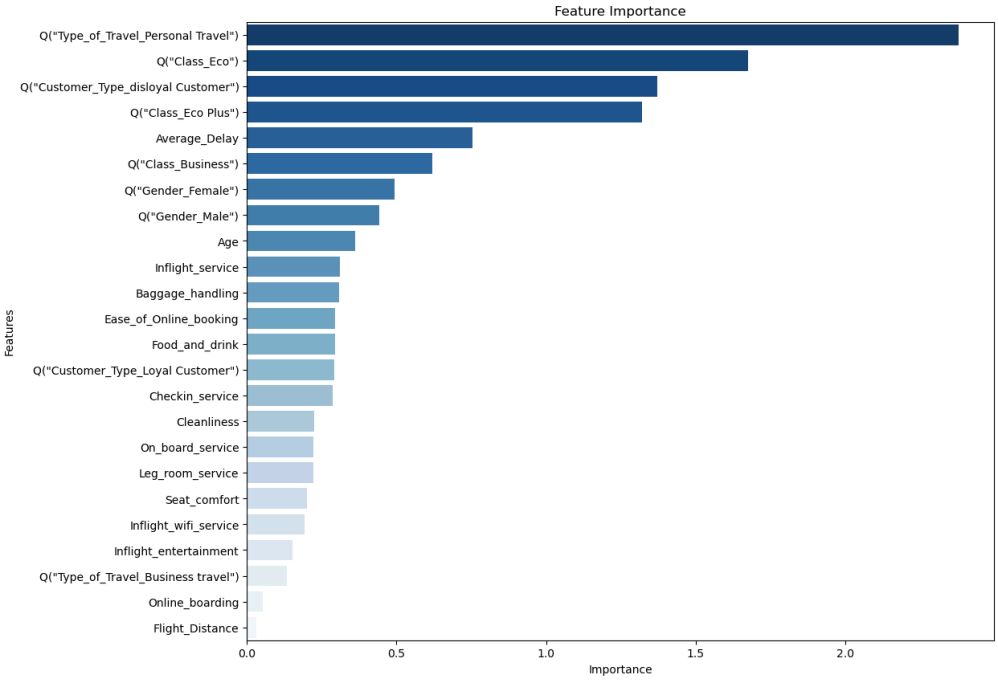


Fig 11: Naive Bayes



	Positive class	Negative class	Positive/Negative Ratio	Importance
Q("Type_of_Travel_Personal Travel")	0.000481	0.000045	10.787933	2.378428
Q("Class_Eco")	0.000632	0.000118	5.339826	1.675193
Q("Customer_Type_disloyal Customer")	0.000241	0.000061	3.938951	1.370914
Q("Class_Eco Plus")	0.000094	0.000025	3.740912	1.319330
Average_Delay	0.016442	0.007729	2.127285	0.754846
Q("Class_Business")	0.000252	0.000469	0.537913	0.620058
Q("Gender_Female")	0.000502	0.000306	1.636484	0.492550
Q("Gender_Male")	0.000476	0.000306	1.557778	0.443261
Age	0.036738	0.025564	1.437129	0.362647
Inflight_service	0.003314	0.002431	1.363551	0.310092
Baggage_handling	0.003302	0.002429	1.359441	0.307073
Ease_of_Online_booking	0.002491	0.001856	1.341804	0.294015
Food_and_drink	0.002893	0.002156	1.341702	0.293939
Q("Customer_Type_Loyal Customer")	0.000737	0.000551	1.337593	0.290872
Checkin_service	0.002976	0.002232	1.332996	0.287429
Cleanliness	0.002871	0.002293	1.252432	0.225087
On_board_service	0.002953	0.002362	1.250129	0.223247
Leg_room_service	0.002925	0.002340	1.249791	0.222976
Seat_comfort	0.002969	0.002429	1.222612	0.200989

Fig 12: Customer satisfaction conceptual model in the air travel industry

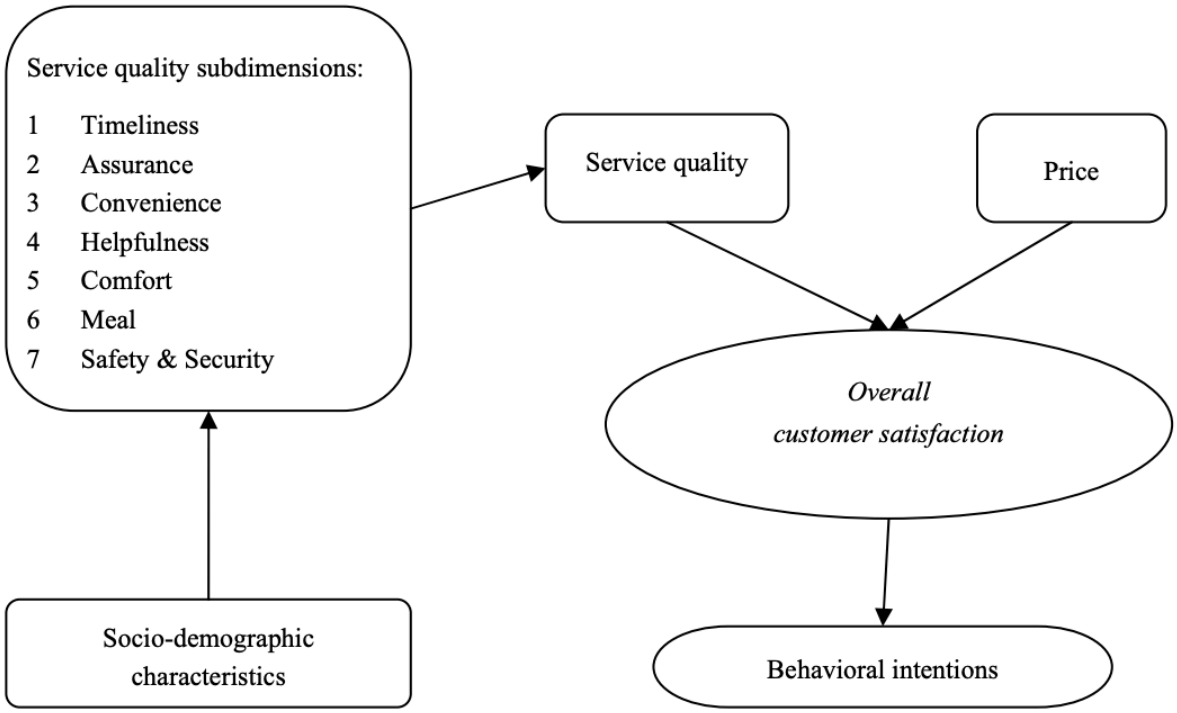


Table 1

Metrics	**Gradient Boosting**	**Random Forest**	Logistic Regression	KNN
ROC AUC	**0.94**	**0.96**	0.92	0.97
Accuracy	**0.94**	**0.96**	0.87	0.92
Sensitivity	**0.93**	**0.95**	0.87	0.91
Specificity	**0.92**	**0.94**	0.83	0.88
Precision	**0.95**	**0.97**	0.90	0.96

6. Citations

- Cronin, J.J. and Taylor, S.A. (1992), “Measuring service quality: a reexamination and extension”, Journal of Marketing, Vol. 56, July, pp. 55-68
- Fornell, C. (1992), “A national customer satisfaction barometer: The Swedish Experience”, Journal of Marketing, 56, pp. 6-21
- Blodgett, J.G., Wakefield, K.L. and Barnes, J.H. (1995), “The Effects of Customer Service on Consumer Complaining
- Innovative Marketing, Volume 4, Issue 2, 2008 Behavior”, Journal of Service Marketing, Vol. 9 (4), pp. 31-42
- “20 Year Passenger Forecast.” IATA, www.iata.org/en/services/data/market-data/20-year-passenger-forecast/
- “Future of Aviation.” ICAO, https://www.icao.int/Meetings/FutureOfAviation/Pages/default.aspx
- Total operating revenues of the U.S. airline industry 2023 | Statista
- Passenger revenues in U.S. airline industry 2023 | Statista
- “The Importance of Air Transport to United States.” IATA https://www.iata.org/en/iata-repository/publications/economic-reports/the-united-states--value-of-aviation
- US Aviation Market Size & Share Analysis - Industry Research Report - Growth Trends (mordorintelligence.com)