University of California, Irvine
M.S. Business Analytics

Airline Reviews

Machine Learning Fall 2024
Section B, Team 12

Kenjee Koh | Ivy Nguyen
Hsiang-Han Huang | Becky Wang
Dennis Wu

Index

Table of Contents

"How do the various aspects of a flight experience influence a customer's overall satisfaction and their likelihood to recommend the airline?"

Airline Reviews

Executive Summary

The primary goal of this project was to analyze customer reviews from a Kaggle dataset to:

1. Predict whether a customer would recommend a flight based on various aspects of their flight experience.
2. Understand how each aspect of the flight experience influences overall satisfaction and customer recommendations.

To achieve this, we employed Decision Tree and Random Forest models. Preprocessing techniques included feature engineering, feature selection, SMOTE for handling class imbalance, and hyperparameter tuning. By building both benchmark and fully preprocessed models, we aimed to ensure fairness in evaluation and maximize predictive accuracy.

Key Findings

1. **Most Influential Features:**
    - "Value for Money" emerged as the most significant factor influencing customer satisfaction and recommendations.
    - "Ground Service", "Cabin Staff Service", and “Seat Comfort” were the next most impactful factors.
2. **Model Performance:**
    - The Random Forest model with full preprocessing demonstrated the most robust performance, offering superior predictive power compared to the Decision Tree model.

Recommendations

Based on the findings, the following strategies are recommended:

1. **Focus on Price-Performance Optimization:** Airlines should prioritize initiatives such as promotional pricing, bundling, or loyalty programs to enhance the perceived value of flights.
2. **Enhance Ground Service:** Improving the quality of ground services will help deliver a more seamless end-to-end customer experience, complementing other satisfaction drivers.
3. **Maintain Adequate Standards:** While "Cabin Staff Service," "Seat Comfort," and "Inflight Entertainment" contribute modestly to satisfaction, maintaining a baseline standard in these areas is advisable.
4. **Strategic Resource Allocation:** Features like "Seat Type" and "Wifi & Connectivity" have minimal impact on recommendations. Airlines should avoid overinvesting here unless targeting specific customer segments or premium offerings.

Conclusion

This analysis demonstrates the importance of focusing on factors that significantly influence customer recommendations, such as perceived value and service quality. By leveraging the insights from this project, airlines can strategically allocate resources to optimize customer satisfaction and loyalty.

Introduction

Business Idea
The airline industry is one of the most competitive and customer-centric markets in the world. With countless options available to travelers, understanding what truly drives customer satisfaction and loyalty is not just an advantage—it's a necessity for standing out as a competitor in the market.
In this project, we set out to answer a critical question for the industry:

> **“How do the various aspects of a flight experience influence a customer’s overall satisfaction and their likelihood to recommend the airline?”**

This question is not just relevant for airlines seeking to optimize their services and maximize customer loyalty, but also for travelers who increasingly rely on reviews and recommendations to guide their decisions. By applying advanced analytical techniques and leveraging real-world customer review data, we aimed to uncover actionable insights that could help airlines enhance the customer journey and gain a competitive edge.

Why It’s Important
Customer service and satisfaction is at the core of success in the airline industry, where loyalty and customer recommendations can heavily influence an airline’s reputation, and thus their profitability. In an era where online reviews are increasingly accessible and can significantly influence consumer behavior and outlooks, understanding the factors that drive customer satisfaction has never been more critical.

This project aims to bridge the gap between raw customer feedback and actionable business insights by harnessing machine learning models and techniques. By identifying the key aspects of the flight experience that most impact customer satisfaction and recommendation likelihood, airlines can:

- Strategically allocate resources to the areas that matter most to customers.
- Enhance the overall customer experience, fostering loyalty in a highly competitive market.
- Make data-driven decisions to design services and packages that align with customer priorities, improving operational efficiency and profitability.

Additionally, this analysis contributes to a broader understanding of consumer behavior in the airline industry, offering insights that can benefit not only airlines but also passengers and travel agencies.

Airline Reviews

Data

The dataset we are using for this project was retrieved from Kaggle, posted by user Juhi Bhojani. The data was originally scraped from the website airlineequality.com, which serves as an incredible online resource for exploring data and gathering insights related to airlines, airports, and customer flight satisfaction.
This dataset provides in-depth information about the various details associated with a flight, with the focus being on whether or not a particular reviewer recommended their flight or not.

At A Glance

1. *23,171 reviews (rows)*
2. *19 flight details (columns/features)*

Attribute Descriptions

Categorical Attributes

Airline Name Name of the airline	Overall Rating General rating given to the airline.
Type of Traveller Solo Leisure, Couple Leisure, Business, Family Leisure	Seat Type Economy Class, Business Class, Premium Economy, First Class
Verified Whether the review is verified or not.	**Recommended (Target Variable)** Whether or not the reviewer ultimately recommended their airlines (Yes or No)

Numerical Attributes
(each rating on a scale of 1-5)

Seat Comfort Rating of seat comfort during the flight	Cabin Staff Service Rating of the cabin crew's service quality
Food & Beverages Rating of the quality and variety of food and drinks	Ground Service Rating of services at the airport
Inflight Entertainment Rating of available entertainment options	Wifi & Connectivity Rating of in-flight Wi-Fi quality and speed
Value for Money Rating of how passengers perceive the value of the flight for the price paid	

Textual/Date Attributes

Review_Title Title of the review	Review Date Date the review was posted
Review Text of the review	Route Flight route (origin-destination pair)
Date Flown When the flight occurred	

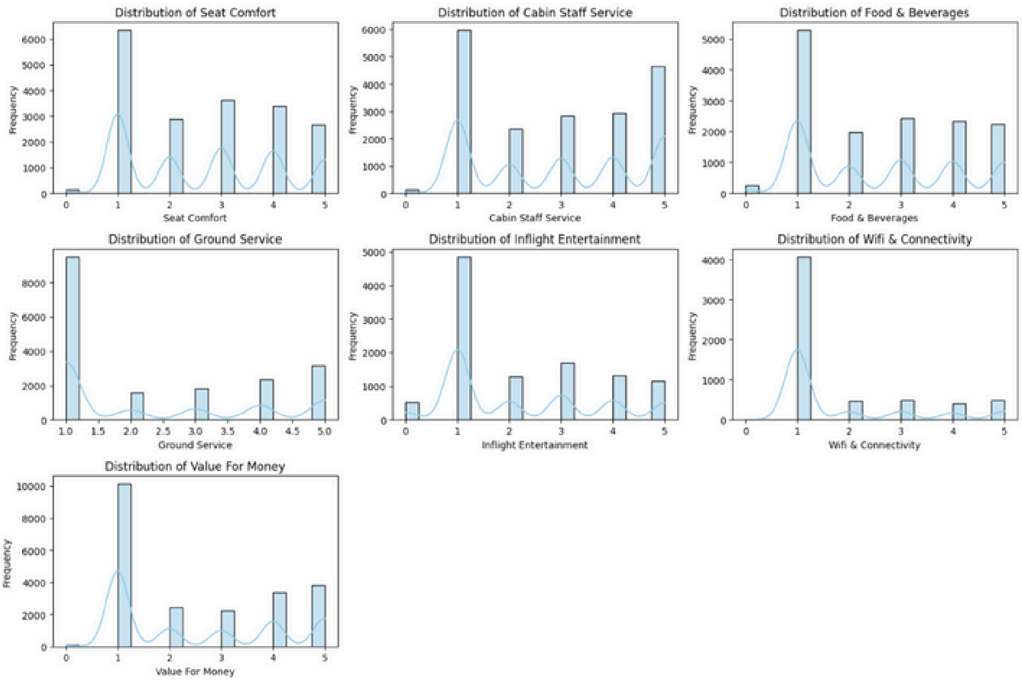
Note for textual/date attributes: We dropped these attributes in our final models since we aren't using NLP/text analysis or Time-Series Models.

Airline Reviews

Data

Exploratory Data Analysis Visualizations

Distribution of Numerical Ratings



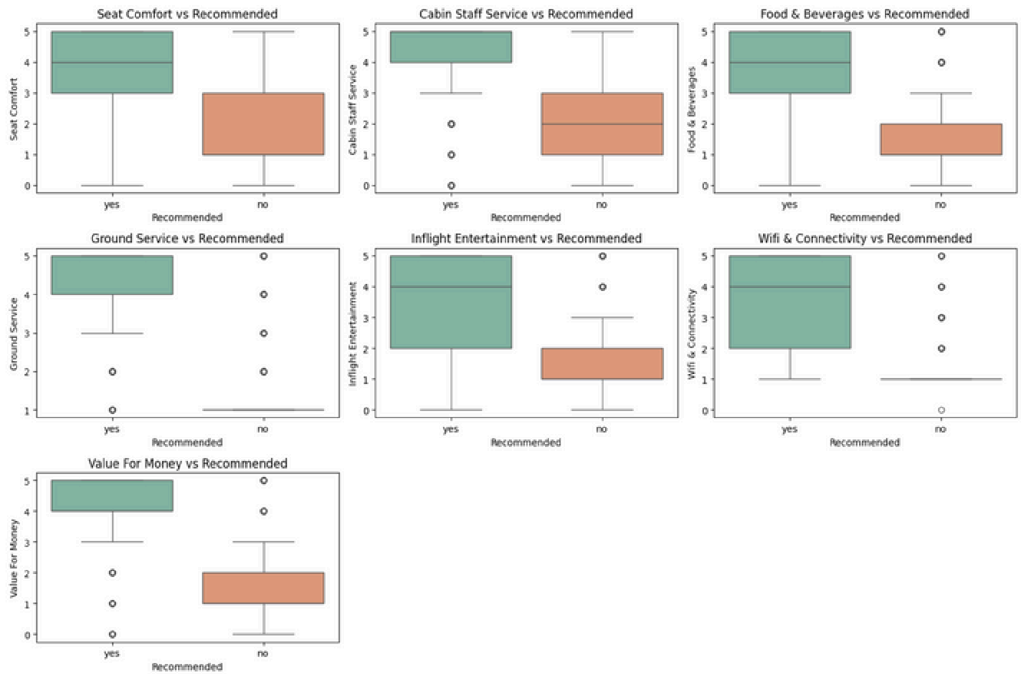
During the EDA phase, we generated various visualizations to help us develop a better understanding of our raw data and how we might need to preprocess and clean our data for our final models.

Below are some of the most insightful visualizations produced during this phase of EDA:

These histograms provided an overview of the frequency distribution of each numerical column, showing how ratings are spread.

Additionally, it also helped us see which rating attributes had “error” entries that came as 0’s. With this, we knew to remove rows that contained 0 values in any of their rating features.

Relationship Between Ratings and Recommendation



These boxplots visualized how ratings differed for reviews where customers recommended the airline vs where they did not.

Airline Reviews

Pre-Processing

This section explores the preprocessing steps taken for all of our models:

Model I

Decision Tree
a. *Benchmark Model*
b. *Full Model*

Model II

Random Forest
a. Benchmark Model
b. Full Model

Minimal Preprocessing (for Benchmark Models)

Although a major project requirement was that benchmark models were supposed to run off the raw dataset, doing so caused many issues with our models, preventing them from running at all.

To go around this, we applied very minimal preprocessing measures for our benchmark models to ensure that our dataset was compatible with the tree models. These were the bare minimum preprocessing steps that ensured that our benchmark models weren't as processed as our full models, but still allowed the benchmark models to run.

Step 1: Dropping attributes that caused errors
Since tree models in scikit-learn use only numerical features, many string-based attributes caused errors that prevented our benchmark models from running at all.

It is also important to note that although we could have applied feature engineering (encoding) to some of these variables to make use of them, we also dropped them by intuition as we knew that they would have little to no impact on customer satisfaction based on the situation context.

There were columns that were:

a. *invalid in the problem analysis context*
   i. E*.g. "Overall_Rating" was just another target variable similar to "Recommendation"*

b. contain insights that wouldn't have an influence on the customers' Recommendation
   i. E*.g. "Review Title" wasn't an actual aspect of the flight experience*

c. *contain string/text-based insights (since we aren't using NLP or text analysis)*
   i. *E.g. "Review" (which is just the actual text review that each reviewer included as a part of their overall flight experience review) could have included great insights for us to use, but deriving meaningful insights from the Reviews of such a huge dataset would have required using text analysis methods/NLP*

d. *time-sensitive (since we aren't using a time-series model)*
   i. *E.g. "Review Date" and "Date Flown"*

To solve this, we dropped the following attributes: Airline Name, Review_Title, Review Date, Review, Aircraft, Route, Date Flown

Step 2: Use label encoding for string values in "Type of Traveller" and "Seat Type"
For the string-based attributes "Type of Traveller" and "Seat Type", we decided to use feature engineering (label encoding) to make them compatible with our tree models. In doing this, we wouldn't have to drop these attributes, which have significant influence on the Recommendation variable.

Both features exhibit a natural hierarchy or order that label encoding effectively preserves.

- For instance, in 'Type of Traveller', categories like 'Solo Leisure' (0), 'Couple Leisure' (1), 'Business' (2), and 'Family Leisure' (3) reflect an increasing degree of group size or trip complexity.
- Similarly, in 'Seat Type', categories such as 'Economy Class' (0), 'Business Class' (1), 'Premium Economy' (2), and 'First Class' (3) represent progressively higher levels of luxury or service.
- Label encoding helps maintain these inherent relationships between the categories.

Label encoding is computationally simple and provides a numeric representation that facilitates faster model training while ensuring compatibility with our chosen algorithms. It strikes a balance between preserving the structure of the data and keeping the feature space manageable.

Step 3: Transform values in Recommended column as (Yes = 1) and (No = 0).
We transformed the 'Recommended' column values from 'Yes' and 'No' to **1** and **0** because tree-based models, like Decision Trees and Random Forests, require numerical inputs. These models cannot process categorical values like 'Yes' and 'No' directly for splitting and decision-making. By encoding them as 1 (for "Yes") and 0 (for "No"), we enable the model to perform efficient calculations and make binary classifications, improving model performance.

Airline Reviews

Pre-Processing

Full Preprocessing (for Full Models)

In addition to the minimal preprocessing steps used for benchmark models, our full preprocessing plan for our full models included feature engineering, feature selection, SMOTE resampling, and hyperparameter tuning.

Step 0: Exploratory Data Analysis (EDA)
Before we began full preprocessing, we utilized various pandas functions to further explore our raw dataset, develop a deeper understanding of the data available to us, and get a better idea of what specific areas needed attention during the full preprocessing stages.

- Some columns, like **Aircraft** and **Wifi & Connectivity**, had a high percentage of missing values.
- Rating columns (e.g. seat comfort, cabin staff service, etc) are numeric, with some missing data.
- The dataset has a few redundant columns, such as **Unnamed: 0.**
- There is a vast imbalance in class distribution:

```
Class Distribution: Recommended
no     15364
yes     7807
```

Step 4: Transform null values if applicable
Many reviews had null values in rating columns where the reviewer may have just felt neutral

- (e.g. Wifi & Connectivity, Inflight Entertainment, etc)
- If there was a null rating, just set it to "3.0"
  - Rationale: Assuming they just felt neutral, 3.0 was the neutral value on the 1-5 scale

Step 5: Drop all rows with null values in non-transformed attributes
Many rows still had null values in various columns outside of the transformed values mentioned in Step 3.

Step 6: Drop all rows that are not Verified
To get more reliable insights, we dropped reviewers where their “Verified” column value was “False”.

According to airlinequality.com, the “editorial staff have inspected a copy of an e-ticket, booking details or a boarding pass, with the customer name confirming the trip written about in the review”.

Note: With the inclusion of this preprocessing step, it is important to note that our full models are trained and tested to predict the Recommendation values for “Verified” reviews.

Step 7: Resampling with SMOTE
During our EDA, we noticed a vast imbalance in class distribution for the target variable.
To improve the performance and accuracy of our model, we decided to target this issue by utilizing SMOTE resampling.

- To Balance the Target Variable Distribution
- To Improve Model Performance on the Minority Class

Step 8: Employing Hyperparameter Tuning
This ensured that our full Decision Tree and Random Forest models utilized the best possible parameters to achieve best performance.

```
Best Parameters for Decision Tree:
{'criterion': 'gini', 'max_depth': 5, 'min_samples_leaf': 5, 'min_samples_split': 2}

Decision Tree Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       803
           1       0.92      0.86      0.89       241

    accuracy                           0.95      1044
   macro avg       0.94      0.92      0.93      1044
weighted avg       0.95      0.95      0.95      1044
```

```
Best Parameters for Random Forest:
{'max_depth': 6, 'min_samples_leaf': 2, 'min_samples_split': 6}

Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.96      0.96      2604
           1       0.90      0.93      0.92      1092

    accuracy                           0.95      3696
   macro avg       0.94      0.95      0.94      3696
weighted avg       0.95      0.95      0.95      3696
```

Airline Reviews

Analysis

Model I : Decision Trees

Why This Model

Decision Tree models are simple to understand, interpret, and visualize, making them a great choice for datasets where explanation is important. They can handle both categorical and numerical data effectively and are robust to outliers. Additionally, Decision Trees naturally manage non-linear relationships in the data, making them versatile for various types of problems. The model's ability to handle missing values ensures that we can achieve reliable results without extensive preprocessing.

Benchmark Model

Accuracy

```
Train Accuracy: 0.941858314322708
Test Accuracy: 0.9344073647871116
F1 Score: 0.9342660371181055
Cross-validated Stratified Score (Accuracy): 0.9408717351104606
```

Proportion

```
dtype: object
Class Distribution: Recommended
0    0.777394
1    0.222606
Name: count, dtype: float64
```

Pre-processing

After the preprocessing and hyperparameter tuning that mention in the previous part, we get the best parameters for decision tree show below:

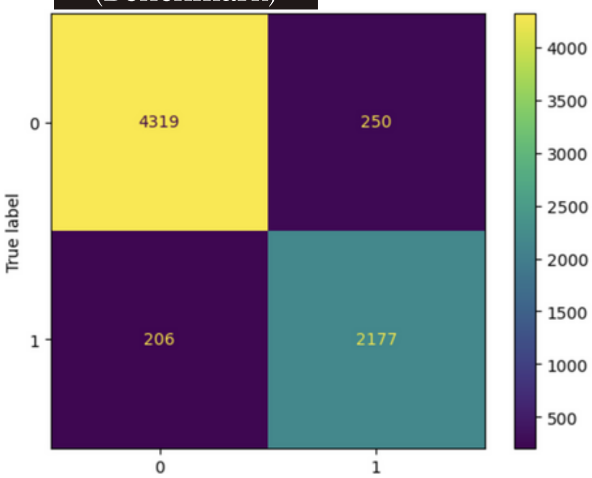
```
Best Parameters for Decision Tree:
{'criterion': 'gini', 'max_depth': 5, 'min_samples_leaf': 5, 'min_samples_split': 2}

Decision Tree Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       803
           1       0.92      0.86      0.89       241

    accuracy                           0.95      1044
   macro avg       0.94      0.92      0.93      1044
weighted avg       0.95      0.95      0.95      1044
```

Predicted Result

Confusion Matrix

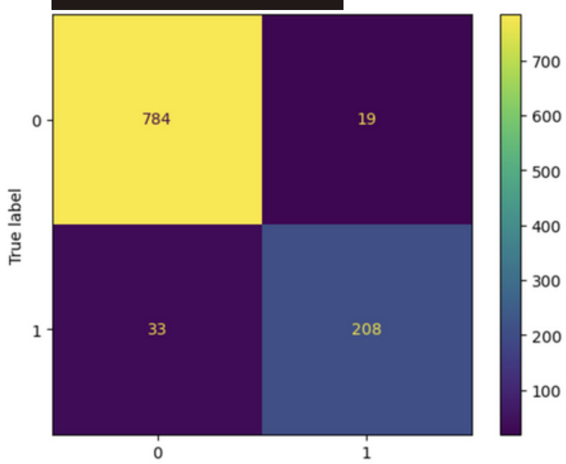
Based on the confusion matrix results, the benchmark model achieves an accuracy of 93%, while the final model reaches 95%. The model with full preprocessing performs significantly better compared to minimal preprocessing. Additionally, the error rate decreases notably, indicating that the hyperparameter adjustments were effective.

Confusion Matrix (Benchmark)



The overall accuracy is 93%; There are false positives (250) and false negatives (206) that may impact decision making.

Confusion Matrix (Decision Trees)



The overall accuracy is 95%; There are only false positives (19) and false negatives (33) detected, which decreased a lot.

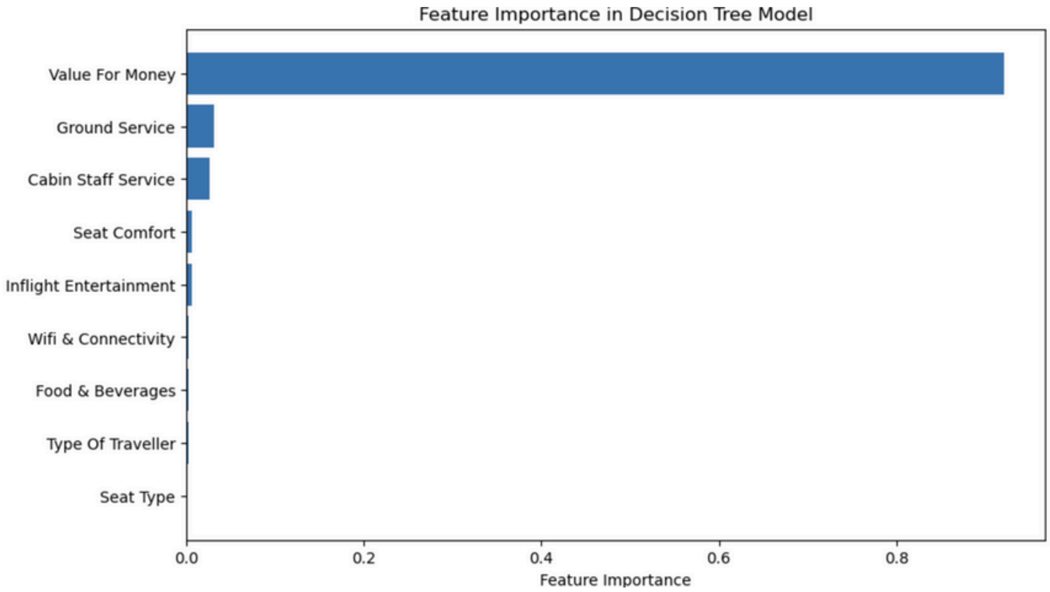
Airline Reviews

Analysis

Accuracy Comparison

Accuracy Comparison	Accuracy	F1 Score	Cross-Validated Stratified Score
Benchmark	0.9344	0.9343	0.9409
With Preprocessing	0.9502	0.9507	0.9613

Feature Importance



Ranking by feature Importance (Highest to lowest)

	Feature	Importance
8	Value For Money	0.921349
5	Ground Service	0.031378
3	Cabin Staff Service	0.026667
2	Seat Comfort	0.006424
6	Inflight Entertainment	0.006135
7	Wifi & Connectivity	0.003084
4	Food & Beverages	0.002199
0	Type Of Traveller	0.001997
1	Seat Type	0.000767

The top three contributors to customer satisfaction were identified as "Value for Money", "Cabin Staff Service", and "Ground Service". These findings are consistent with general expectations in the airline industry, highlighting the critical roles of cost-effectiveness and service quality in shaping customer experiences.
Overall, the decision tree proved to be an effective baseline model. It delivered reasonable accuracy while providing valuable insights into the key factors driving customer recommendations.

Analysis

Model II : Random Forest

Why This Model

Random Forest has the advantage of avoiding overfitting, flexibility to handle both regression and classification with high accuracy, and easy to determine feature importance. Our dataset contains both categorical and numerical data, and many missing values. Random Forest also can effectively manage missing values.

Benchmark Model

Accuracy

```
Train Accuracy: 0.9818114557001049
Test Accuracy: 0.9401611047180668
F1 Score: 0.9402478511859288
Cross-validated Stratified Score (Accuracy): 0.9429681789034337
```

- Both our train accuracy and test accuracy is high, which means our model captures the patterns in the sets well. The difference between train accuracy and test accuracy is only 4%, suggesting that our model is not overfitting and generalizes well to the unseen data. While the minor tuning might further optimize the test performance, there's no major concerns of overfitting or underfitting based on the accuracy.

Proportion

```
dtype: object
Class Distribution: Recommended
0    0.66307
1    0.33693
Name: count, dtype: float64
```

- The result shows that the dataset is imbalanced significantly, biased toward the majority class (Class 0) and it may lead to predicting the majority class (Class 0) more often and poor performance for the minority class (Class 1).

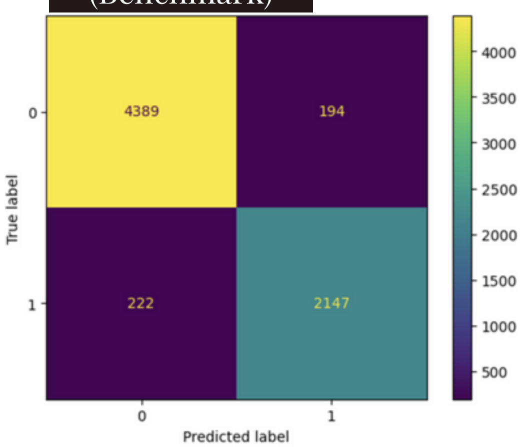
Pre-processing

After the preprocessing and hyperparameter tuning that mention in the previous part, we get the best parameters for random forests show below:

```
Best Parameters for Random Forest:
{'max_depth': 6, 'min_samples_leaf': 2, 'min_samples_split': 6}

Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.96      0.96      2604
           1       0.90      0.93      0.92      1092

    accuracy                           0.95      3696
   macro avg       0.94      0.95      0.94      3696
weighted avg       0.95      0.95      0.95      3696
```

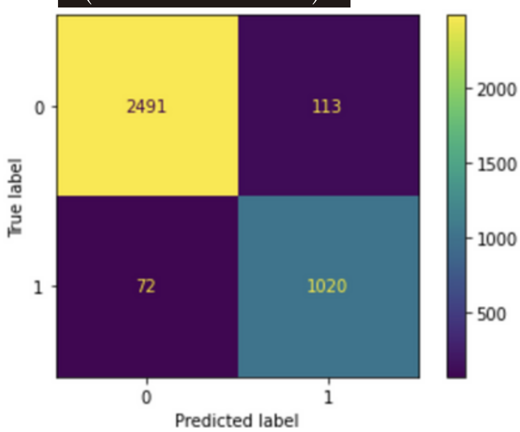
Predicted Result

Confusion Matrix (Benchmark)



According to the confusion matrix of benchmark, the precision of Class 1 is 91% and the overall accuracy is 94%, which performs well in identifying both classes and there are false positives (194) and false negatives (222) that may impact decision making. In order to improve performance, we should do some adjustments.

Confusion Matrix (Random Forest)



According to the confusion matrix of the final model, the overall accuracy is 95%, which is much better than the one of the decision tree. Also, the error rate decreases significantly, meaning that the adjustments of hyperparameters are useful.

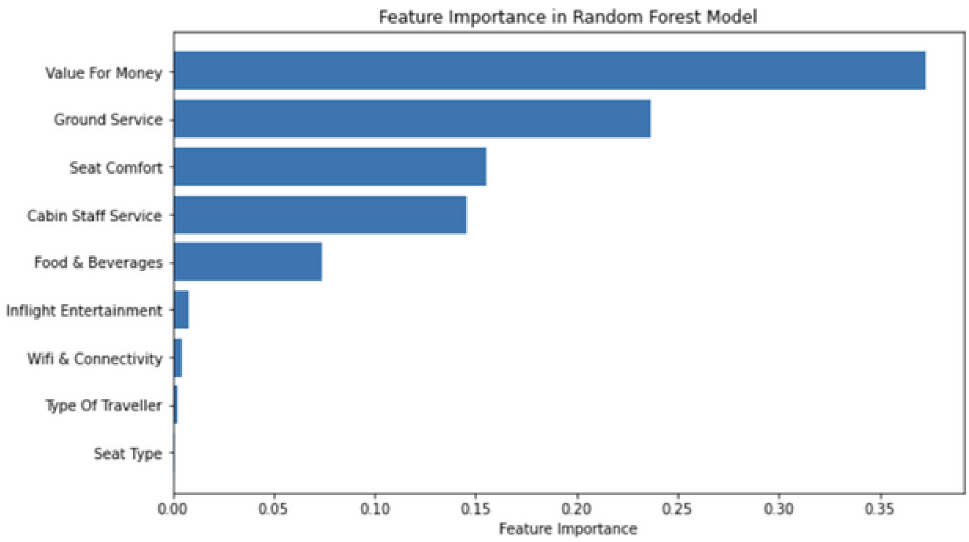
Airline Reviews

Analysis

Feature Importance

In the end, we can conclude how much each feature improves the accuracy. According to the graph above, "Value For Money" plays an important role in predicting our target variable, contributing 37% to the customer decisions.

The feature importance of random forest's top 2 ranking and the last one is the same as the one of the decision tree, which is "Value For Money", "Ground Service" and the last one is "Seat Type". However, the ranking from 3-8 is different. In the model of random forest, "Cabin Staff Service", "Food & Beverages", "Inflight Entertainment", "Wifi & Connectivity" and "Type of Traveller" are ranked from3 to 8. But the model of decision tree starts from "Cabin Staff Service", "Seat Comfort", "Food & Beverages", "Type of Traveller", "Wifi & Connectivity", and "Inflight Entertainment".



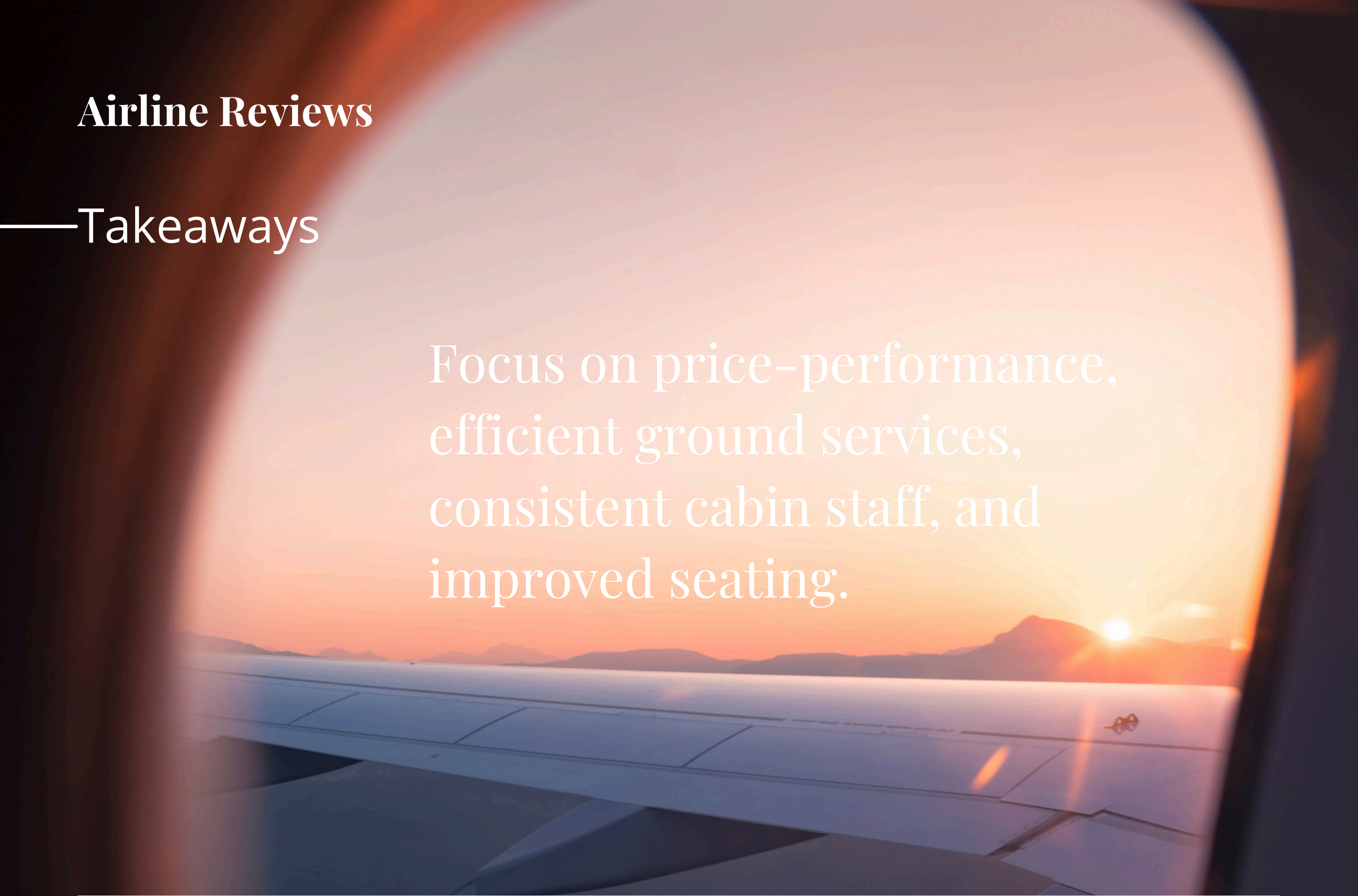
The table below shows much clearer of the difference between random forest and decision tree:

Random Forest

	Feature	Importance
8	Value For Money	0.372761
5	Ground Service	0.236362
2	Seat Comfort	0.155358
3	Cabin Staff Service	0.145830
4	Food & Beverages	0.073955
6	Inflight Entertainment	0.008240
7	Wifi & Connectivity	0.004162
0	Type Of Traveller	0.002249
1	Seat Type	0.001083

Decision Trees

	Feature	Importance
8	Value For Money	0.910149
5	Ground Service	0.036319
3	Cabin Staff Service	0.031605
2	Seat Comfort	0.008192
4	Food & Beverages	0.004919
0	Type Of Traveller	0.004294
7	Wifi & Connectivity	0.001924
6	Inflight Entertainment	0.001704
1	Seat Type	0.000895



Takeaways

Focus on price-performance, efficient ground services, consistent cabin staff, and improved seating.

Important Attributes

According to our result, “Value For Money” are the most important attributes for airline customers’ decisions. To address this, airlines should focus on strategies that optimize the price-performance ratio. Example of implementing promotional pricing, offering bundle service packages and introducing loyalty rewards programs.

Second one is “Ground Service”, improving its quality is essential for creating a better end-to-end customer experience. Examples include streamlining the check-in and boarding processes, enhancing baggage handling efficiency, and providing more responsive and courteous customer support at the airport.

The third is “Cabin Staff Service” in the decision tree and “Seat Comfort” in the random forest. For “Cabin Staff Service”, since it doesn’t affect too much for the customer satisfaction, we would not suggest the airlines reinvest too much on it, maintaining consistent service standards but provide comprehensive training, offer personalized assistance, and use passenger feedback to drive improvements.

For “Seat Comfort”, we suggest that airlines can offer more legroom options, improving seat design and cushioning, and providing adjustable features.