What Makes a Happy Airline Passenger?

BayTech

Warren Ngoun
Jennah Yasin
Yukio Rivera
Luis Jimenez Barrios

Intro

Research Question: What are the key features that determine a flight passengers overall satisfaction?

Importance:

- Advising airlines with passenger's reviews in order to maximize satisfaction

Project Plan:

- Predicting which factors are most relevant to passenger's satisfaction.
- Dropping features that aren't significant
- Using "satisfaction" feature as target variable
- Finding top features using forward feature search

Hypothesis:

- Based on our personal preferences and past experience as passengers, we believe that the most important features that affect satisfaction would be: Seat comfort, in-flight entertainment, cleanliness, and food & drinks.

Selection of Data

Source of dataset:
https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

Some predictors from the dataset:

- Satisfaction: Airline satisfaction level(Satisfaction or `0`, neutral/dissatisfied or `1`)
- Customer Type: The customer type (Loyal customer, disloyal customer)
- Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)
- Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
- Flight distance: The flight distance of this journey
- Gate location: Satisfaction level of Gate location
- Food and drink: Satisfaction level of Food and drink
- Online boarding: Satisfaction level of online boarding
- Seat comfort: Satisfaction level of Seat comfort
- Inflight service: Satisfaction level of inflight service

Feature Engineering:

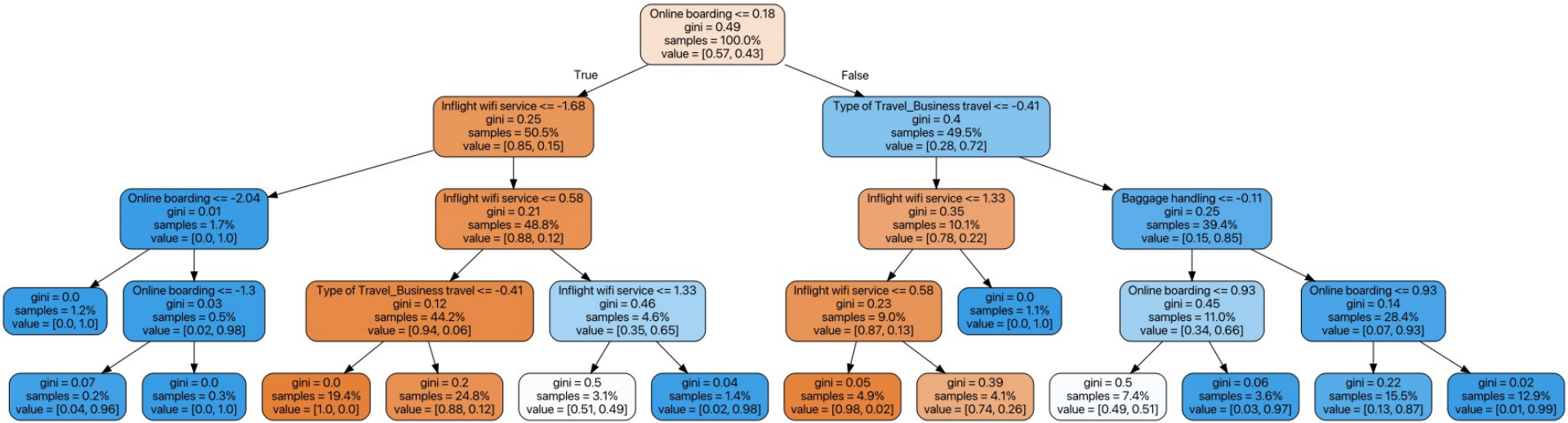
- **One-Hot Encoding for Categorical Columns (like Gender)**

Methods

Materials and Tools:

- Kaggle to find the Airline Passenger Satisfaction data set
- VS Code & GitHub
- Pandas, NumPy, SciKit-Learn for implementation of project
- Course materials: Previous labs & assignments
- Machine Learning
    - Decision Trees
    - Forward Feature Selection

Decision Tree Representing Final Model



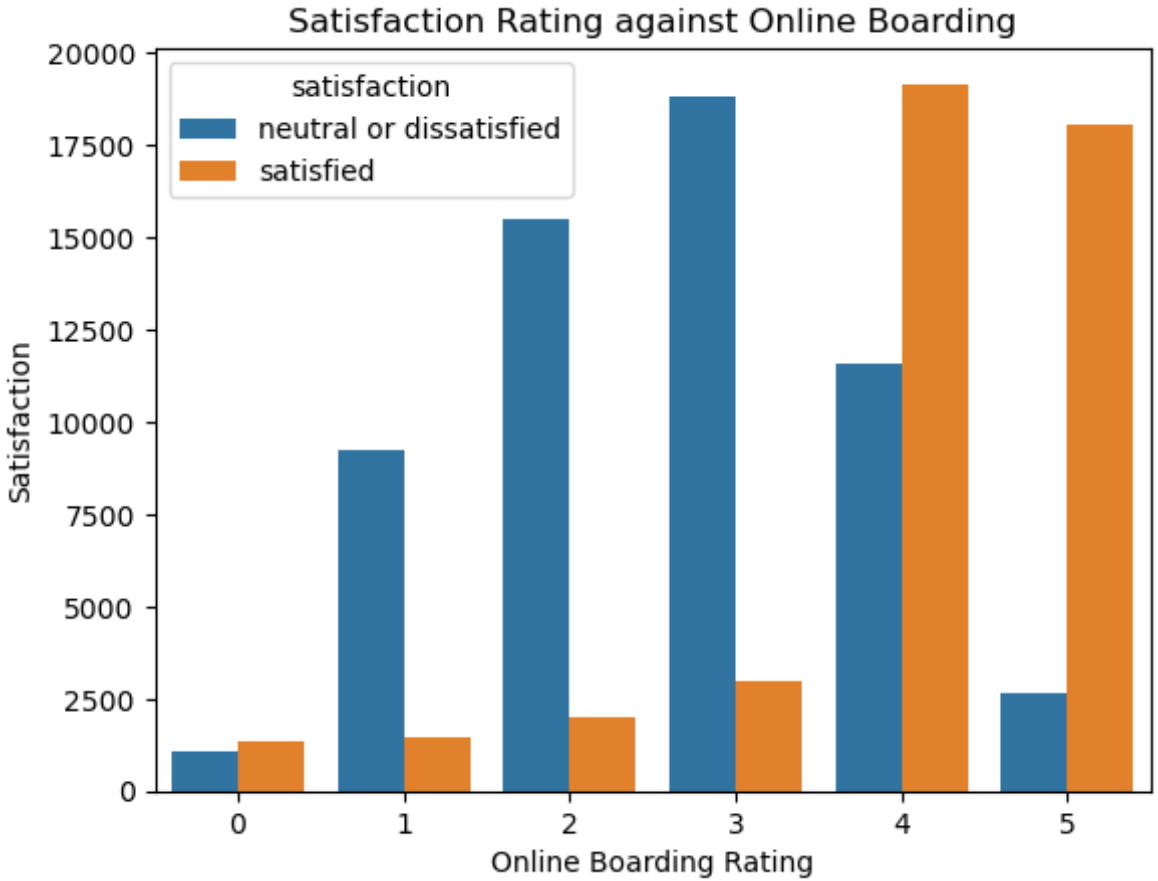
Results

Based on our data analysis and manipulation, we were able to find the top 10 most effective features of our data set. The top 10 features were found to be:

- Online boarding
- Type of Travel_Business travel
- Inflight wifi service
- Gate location
- Baggage handling
- Customer Type_disloyal Customer
- Class_Business
- Inflight Service
- Seat comfort
- Customer Type_Loyal Customer

However, from these top 10 features, we decided to use the first top 5 as our predictors since the overall accuracy didn't significantly increase when we added more of them.

Top feature - Online Boarding:



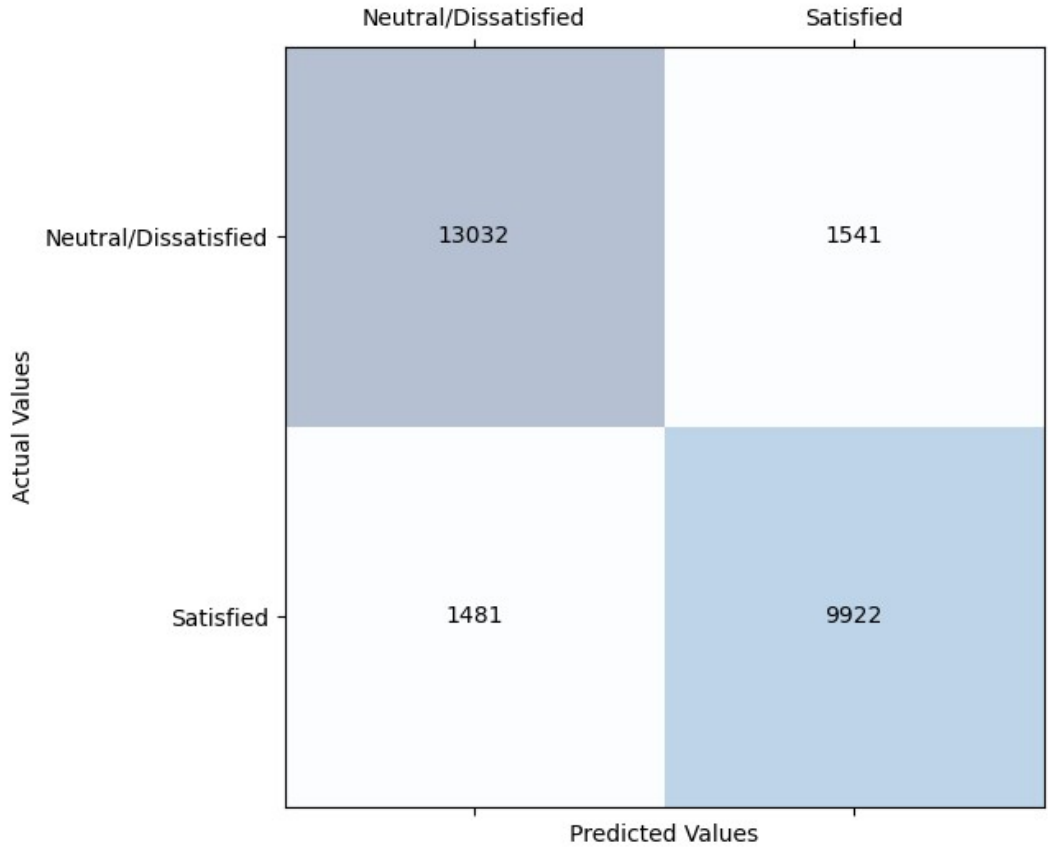
Discussion

Significance of findings:

- The answer from the project implies that Decision Trees can provide a quick and accurate prediction for customer satisfaction with air travel based on a set of predictors.
- For future researchers, our findings of top predictors, such as Online boarding and Inflight Wifi, are strong indicators of customer satisfaction. The identification of the best 5 features from a larger list of 26 is a unique contribution to the field.

In terms of tools, the project investigated the use of Python libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib, which are widely used in data analysis and machine learning.

Confusion Matrix



Summary

Through analyzing the Airline Passenger Satisfaction dataset, we were able to get a better idea of:

- Cleaning up data
- The importance of converting categorical data into numerical values for testing
- Experimenting with kNN and Decision Trees
- Determining that Decision Trees were more significant for our project
- Realizing how effective predictors really are in terms of calculating accuracy