Group 2 Project Proposal

Anqi Xue (ax2170), Tanisha Aggrawal (ta2709), Vishal Bhardwaj (vb2573)

1. Background and context to the problem statement:

Air travel is an essential mode of transportation for many people around the world, whether for business or leisure. However, the quality of the airline experience can vary greatly from one airline to another. Airline passengers have different expectations when it comes to their travel experience, ranging from comfort and convenience to safety and on-time performance.

Airlines often conduct passenger satisfaction surveys to better understand passenger satisfaction and identify key factors that influence it. These surveys can cover a range of topics, including the booking process, check-in, boarding, in-flight experience, and arrival.

The Airline Passenger Satisfaction classification problem seeks to analyze data from one such survey. The dataset contains a variety of factors related to the passenger experience, such as the airline, travel class, age, gender, inflight Wi-Fi service, cleanliness, departure/arrival time, etc.

The goal of the problem is to build a classification model that can accurately predict whether a passenger is satisfied or dissatisfied based on these factors. By analyzing the data and identifying which factors are highly correlated with passenger satisfaction, airlines can make informed decisions about improving the passenger experience and, ultimately, improving overall performance and profitability.

2. Identification and description of the dataset we are planning on using along with their source:

Airline Passenger Satisfaction (https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction)

Content: Gender, Customer Type, Age, Type of Trave, Class, Flight distance, Inflight Wi-Fi service, Departure/Arrival time convenience, Ease of Online booking, Gate location, Food and drink, Online boarding, Seat comfort, Inflight entertainment, On-board service, Leg room service, Baggage handling, Check-in service, Inflight service, Cleanliness, Departure Delay in Minutes, Arrival Delay in Minutes, Satisfaction

3. Proposed ML techniques we are proposing on applying to solve the problem:

(1) Logistic Regression (2) XGBoost (3) Random Forest