Project Summary

Batch details	Hyderabad, Jan 22
Team members	1.Akshaya E S 2.Nikita Ann George 3.Pala Sheshu 4.Ujjwal Kumar Sharma 5.Venkat Paritala 6.Viresh Raj Sah
Domain of Project	Transportation Industry
Proposed project title	Predicting customer satisfaction for airlines
Group Number	3
Team Leader	
Mentor Name	Srikar Muppidi

Dataset name: ***Airline Passenger Satisfaction***

Introduction to the problem/domain/background details:

The domain chosen for the capstone project is from the transportation industry. We hope to gain insights on the various factors affecting a customer's in-flight experience

Airline businesses around the world are decimated by Covid-19 as most international air travel has been grounded. In fact, some airlines such as Thai Airways have already filed for bankruptcy. Nonetheless, once the storm is over, demand for air travel is expected to surge as people rush back for overseas holidays. What can airlines prepare to give themselves a competitive edge when the crowd finally arrives?

To answer this we intend on building a classification problem to predict the customer satisfaction

Problem Statement:

With this project we hope to gain the following insights:

- **To predict customer satisfaction**

- **To create a highly precise classification model for airlines to identify critical bottleneck to raise passenger satisfaction**
- **To understand how to enhance customer experience and build on customer loyalty**

Business problem/ Impact in business of your problem/Need for this study/Abstract :

Why predict customer satisfaction?

As the old adage goes: ”Customers are king.”
This is an age old mantra reflection the importance of customers in every business. With the recent pandemic completely disrupting the way we travel and having a huge impact on the aviation industry, one must now ponder on how we can get back to the norm.

A huge factor effecting the airline business is the customer experience. To recover from the disastrous impact the pandemic had on the industry, we need to find a way to engage passengers in a better manner, by enhancing their experience and ultimately gaining their loyalty .

The research is aimed at analyzing the aviation industry and what factors are the keys to its success. This study uses several classification models such as KNN, Logistic Regression, Gaussian NB, Decision Trees and Random Forest which will later be compared.

Variable identification:

Independent Variables: *There are 23 independent variables including id.22 variables are listed below.*

1. Gender	2. Customer Type
3. Age	4. Type of Travel
5. Class	6. Flight Distance
7. Inflight wifi Service	8. Departure/Arrival time convenient
9. *Ease of Online booking*	10. *Gate location*
11. *Food and drink*	12. *Online boarding*
13. Seat Comfort	14. *Inflight entertainment*
15. *On-board service*	16. *Leg room service*
17. *Baggage handling*	18. *Check-in service*
19. *Inflight service*	20. *Cleanliness*
21. *Departure Delay in Minutes*	22. *Arrival Delay in Minutes*

Target Variables:

1. Satisfaction

Variable information/Data description: *The description of each variable along with their datatype as given in the dataset*

VARIABLE	DATATYPE	DESCRIPTION
1. Id	numeric	Unique identifier
2. Gender	object	Gender of the passengers (Female, Male)
3. Customer Type	object	The customer type (Loyal customer, disloyal customer)
4. Age	numerical	The actual age of the passengers
5. Type of Travel	object	Purpose of the flight of the passengers (Personal Travel, Business Travel)
6. Class	object	Travel class in the plane of the passengers (Business, Eco, Eco Plus)
7. Flight distance	numeric	The flight distance of this journey
8. Inflight wifi service	numeric	Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
9.Departure/Arrival time convenient	numeric	Satisfaction level of Departure/Arrival time convenient
10. Ease of Online booking	numeric	Satisfaction level of online booking
11. Gate location	numeric	Satisfaction level of Gate location
12. Food and drink	numeric	Satisfaction level of Food and drink
14. Online boarding	numeric	Satisfaction level of online boarding
15. Seat comfort	numeric	Satisfaction level of Seat comfort
16. Inflight entertainment	numeric	Satisfaction level of inflight entertainment
17. On-board service	numeric	Satisfaction level of On-board service

18. Leg room service	numeric	Satisfaction level of Leg room service
19. Baggage handling	numeric	Satisfaction level of baggage handling
20. Check-in service	numeric	Satisfaction level of Check-in service
21. Inflight service	numeric	Satisfaction level of inflight service
22. Cleanliness	numeric	Satisfaction level of Cleanliness
23. Departure Delay in Minutes	numeric	Minutes delayed when departure
24. Arrival Delay in Minutes	numeric	Minutes delayed when Arrival
25. Satisfaction	object	Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Business Understanding - It focuses on determining the business requirements/objectives and understanding what outcome to achieve. Also, determine the business units being affected. Convert this business problem into a data mining problem and carve out an initial plan.

- Determine the business objectives: Understand what is needed to be accomplished for the customer.

- Assess the situation: Determine resources availability, and project requirements, assess risks and contingencies, and conduct a cost-benefit analysis.
- Determine data mining goals: Convert a business problem to a data mining problem and recognize the data mining problem type such as classification, regression or clustering, etc.
- Produce a project plan: Devise a step-to-step plan for executing the project.

Data understanding -The data is collected from Kaggle.The dataset contains 129880 rows and 25 variables. Univariate analysis is done to understand the features of the different variables. The quality of data is checked using the missing values treatment, outliers treatment, checking the presence of duplicates, etc.

- Collect initial data: The initial data is loaded from Kaggle and we are using the Jupyter Notebook as our primary tool.
- Describe data: The data type of each variable is checked. The summary statistics of each variable are studied separately.
- Explore data: We have done a basic univariate analysis to understand the variables.
- Verify data quality: The dataset does not contain many null values except for in the variable Arrival Delay in Minutes. Also, we understood that 14 of the variables are in the form of ratings by the passengers on a scale of 1-5.

Data preparation - This stage, which is often referred to as data wrangling, has the objective to develop the final data set for EDA and modeling. Covers all activities to construct the final dataset from the initial raw data. Some of the tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

- Select data: The variables id and Unnamed:0 are dropped as they are irrelevant.
- Clean data: The null values were dropped as they are present in a very small percentage.
- Extract data and Integrate data: The feature engineering aspects of the dataset is yet to be explored.

Future Work/Methodology (Details of algorithms):

1. **Modeling -** In this stage, we build and assess different models built using various techniques from the training dataset. The modeling techniques are yet to be done. Although in this study we are planning to use the classification models such as KNN, Logistic Regression, Gaussian NB, Decision Tree, and Random Forest.

2. **Evaluation -** Evaluate the models and review the steps executed to construct the model to be certain it properly achieves the business objectives.
   Evaluation of the model will be done after the model building.