

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	Akshaya E S
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. <i>Ease of Online booking</i>	10. <i>Gate location</i>
11. <i>Food and drink</i>	12. <i>Online boarding</i>
13. Seat Comfort	14. <i>Inflight entertainment</i>
15. <i>On-board service</i>	16. <i>Leg room service</i>
17. <i>Baggage handling</i>	18. <i>Check-in service</i>
19. <i>Inflight service</i>	20. <i>Cleanliness</i>
21. <i>Departure Delay in Minutes</i>	22. <i>Arrival Delay in Minutes</i>

Target Variables:

1. Satisfaction

Variable information/Data description:

The description of each variable along with their datatype as given in the dataset

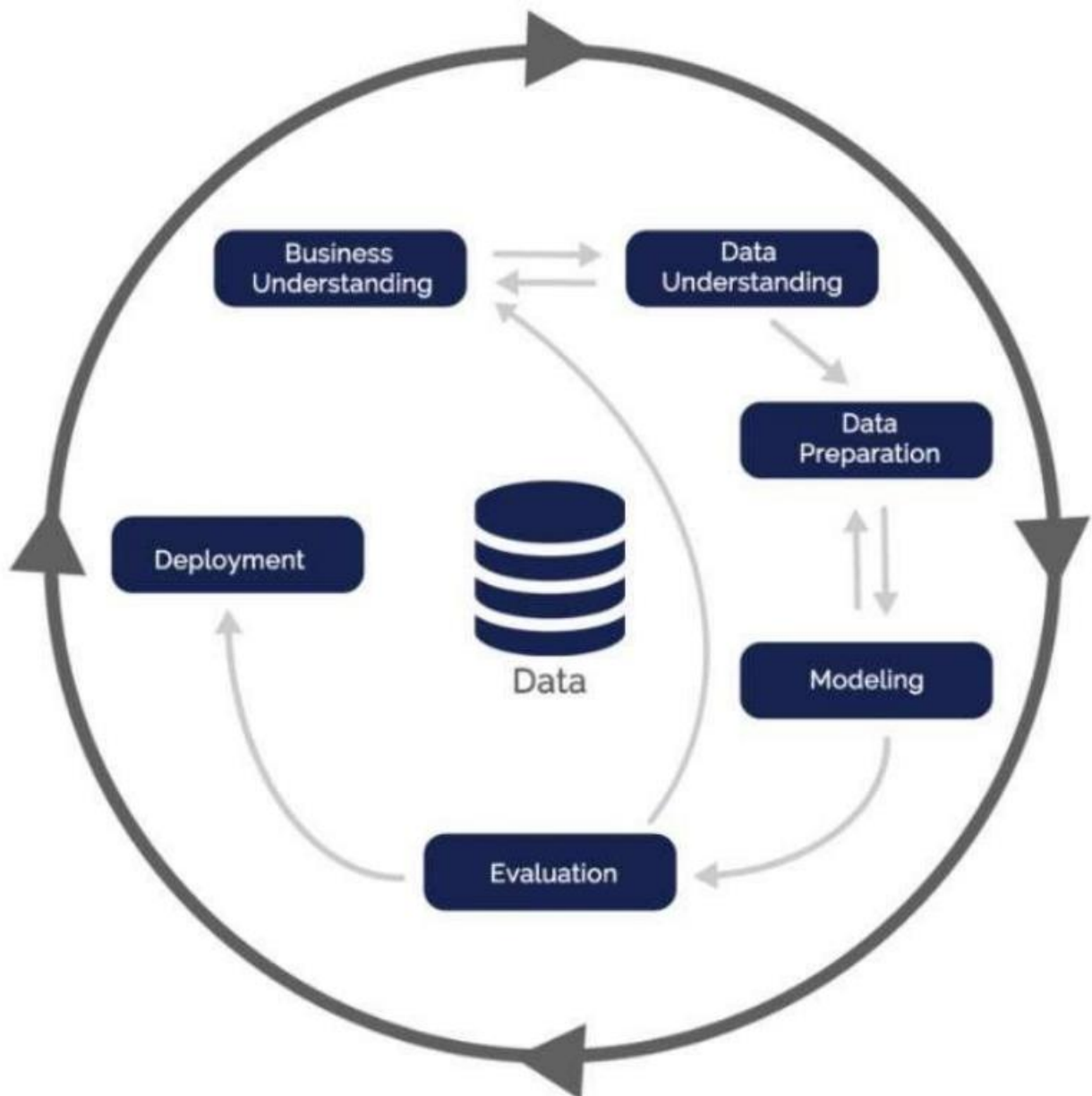
VARIABLE	DATATYPE	DESCRIPTION
1. Id	numeric	Unique identifier
2. Gender	object	Gender of the passengers (Female, Male)
3. Customer Type	object	The customer type (Loyal customer, disloyal customer)
4. Age	numerical	The actual age of the passengers
5. Type of Travel	object	Purpose of the flight of the 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
	numeric	

15. Seat comfort		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)

Future Work/Methodology (Details of algorithms):

Methodology to be followed:

CRISP-DM which stands for Cross Industry Standard Process for Data Mining is a methodology created to help shape data mining projects. It describes the different phases/tasks involved in the project and provides an overview of data mining life cycle.



1. 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 situation: Determine resources availability, project requirements, assess risks and contingencies, and conduct a cost-benefit analysis.
- Determine data mining goals: Convert 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.

2. Data understanding - This phase starts with collecting the data and then examining the data for its surface properties like data format, number of records, etc. The next step is to better understand the data by understanding each attribute and perform basic statistics on them. Understand the relationship between different attributes. Determine the quality of data by checking the missing values, outliers, duplicates, etc.

- Collect initial data: Acquire the data and load it into the analysis tool to be used.
- Describe data: Examine the data and document its surface properties like data format, number of records, or field identities. Understand the meaning of each attribute and attribute value in business terms. For each attribute, compute basic statistics so as to get a higher-level understanding.
- Explore data: Find insights from the data. Query it, visualize it, and identify relationships among the data.
- Verify data quality: Identify special values, missing attributes and null data. Determine how clean/dirty is the data.

3. Data preparation - This stage, which is often referred to as data wrangling, has the objective to develop the final data set for EDA and modelling. 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 modelling tools.

- Select data: Determine which attributes/features will be used and document reasons for inclusion/exclusion.
- Clean data: Correct, impute and remove the improper data.
- Extract data: Derive new attributes from the existing ones
- Integrate data: Create features by combining data from multiple sources.
- Format data: Re-format data as necessary. For example, convert string values to numeric values so as to perform mathematical operations.

4. Modeling - In this stage we build and assess different models built using various techniques from the training dataset.

- Select modelling technique: Determine the algorithms to be used to model the data based on the business requirement.
- Generate test design: In order to build and test the model, we need to divide the dataset into training and testing data set. In this step we divide the data into train and test data set.
- Build model: Based on the modelling technique selected, build the model on the input data set.
- Assess model: Compare the results of different models based on confusion matrix. The outcome of this step frequently leads to model tuning iterations until the best model is found.

5. Evaluation - Evaluate the models and review the steps executed to construct the model to be certain it properly achieves the business objectives.

- Evaluate results: Understand the data mining results and check how impactful they are in achieving the data mining goal. Select appropriate model based on confusion matrix.
- Review process: Review the work accomplished and make sure that nothing was overlooked and all steps were properly executed. Summarize the findings and correct anything if needed.
- Determine next steps: Based on the previous three tasks, determine whether to proceed to deployment, iterate further, or initiate new projects.

Timeline Chart (Weekly plan):

Tentative weekly plan that will be followed:

<u>WEEK</u>	<u>ACTIVITIES</u>
Week 1	Data processing and understanding
Week 2-3	EDA and Statistics
Week 4	Interim Presentation
Week 5-7	Model Building, Feature Engineering
Week 8	Model Validation & Presentation

References (Data set source/Journals/articles)

1. <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>
2. <https://towardsdatascience.com/predicting-satisfaction-of-airline-passengers-with-classification-76f1516e1d16>
3. https://www.researchgate.net/publication/350552031_Predicting_Airline_Passenger_Satisfaction_with_Classification_Algorithms

Declaration: This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated to it and can be used to showcase the work we do on it as a presentation in Great Learning.