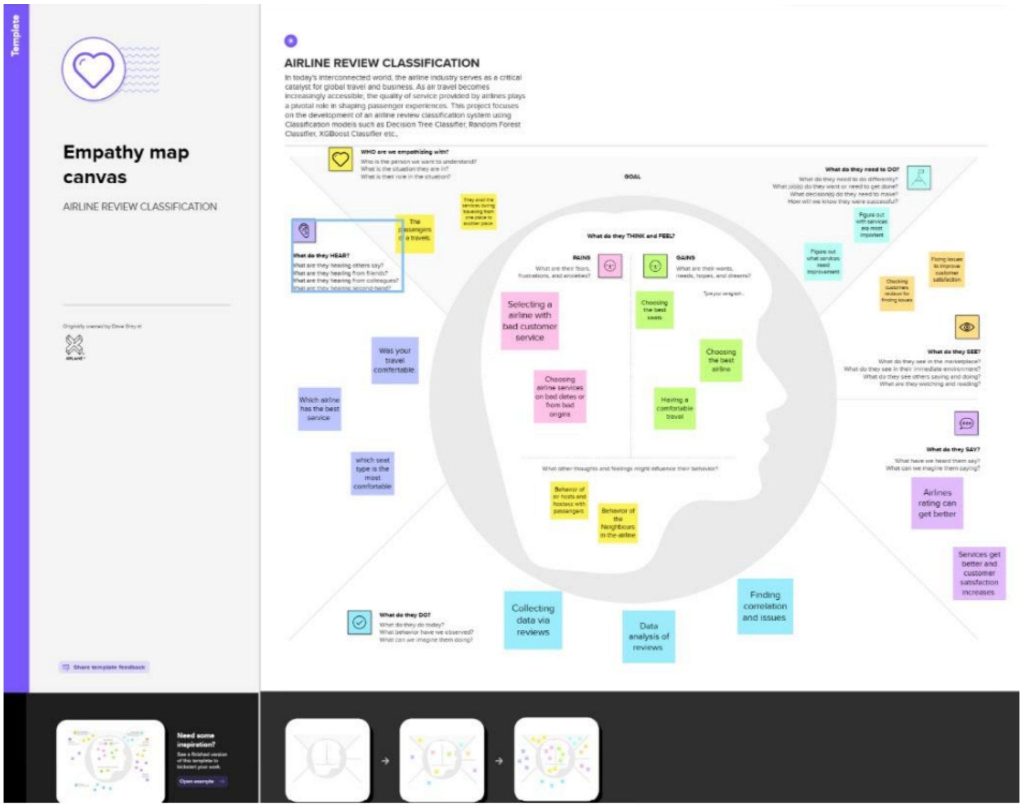
Introduction

In today's interconnected world, the airline industry serves as a critical catalyst for global travel and business. As air travel becomes increasingly accessible, the quality of service provided by airlines plays a pivotal role in shaping passenger experiences.
This project focuses on the development of an airline review classification system using Classification models such as Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier etc.
The proliferation of social media platforms, travel websites, and online forums has given rise to a wealth of usergenerated content, including airline reviews. Extracting actionable insights from this vast pool of unstructured text data has the potential to provide airlines with valuable information for refining their services and elevating passenger satisfaction.
Throughout this report, we will delve into the methodology employed to pre process the raw text data, the process of selecting pertinent features, the training and evaluation of the classification model, and the subsequent interpretation of the obtained results.

Ideation Phase:

Empathy mapping:

Classifying airline reviews using machine learning and creating an empathy map can be a valuable approach to gain insights into customer sentiment and improve airline services.


Template
Empathy map canvas
AIRLINE REVIEW CLASSIFICATION
Originally created by Dave Gray at
XPLANE
Share template feedback
Need some inspiration?
See a finished version of this template to kickstart your work.
Open example

AIRLINE REVIEW CLASSIFICATION

In today's interconnected world, the airline industry serves as a critical catalyst for global travel and business. As air travel becomes increasingly accessible, the quality of service provided by airlines plays a pivotal role in shaping passenger experiences. This project focuses on the development of an airline review classification system using Classification models such as Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier etc.,

WHO are we empathizing with?
Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?

The passengers o a travels

They avail the services during traveling from one place to another place.

GOAL

What do they need to DO?
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?

Figure out with services are most important

Figure out what services need improvement

What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?

Was your travel comfertable

Which airline has the best service

which seat type is the most comfortable

What do they THINK and FEEL?

PAINS
What are their fears, frustrations, and anxieties?

Selecting a airline with bad customer service

Choosing airline services on bad dates or from bad origins

GAINS
What are their wants, needs, hopes, and dreams?

Type your paragraph...

Choosing the best seats

Choosing the best airline

Having a comfortable travel

What other thoughts and feelings might influence their behavior?

Behavior of air hosts and hostess with passengers

Behavior of the Neighbours in the airline

What do they SEE?
What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?

Fixing issues to improve customer satisfaction

Checking customers reviews for finding issues

What do they SAY?
What have we heard them say?
What can we imagine them saying?

Airlines rating can get better

Services get better and customer satisfaction increases

What do they DO?
What do they do today?
What behavior have we observed?
What can we imagine them doing?

Collecting data via reviews

Data analysis of reviews

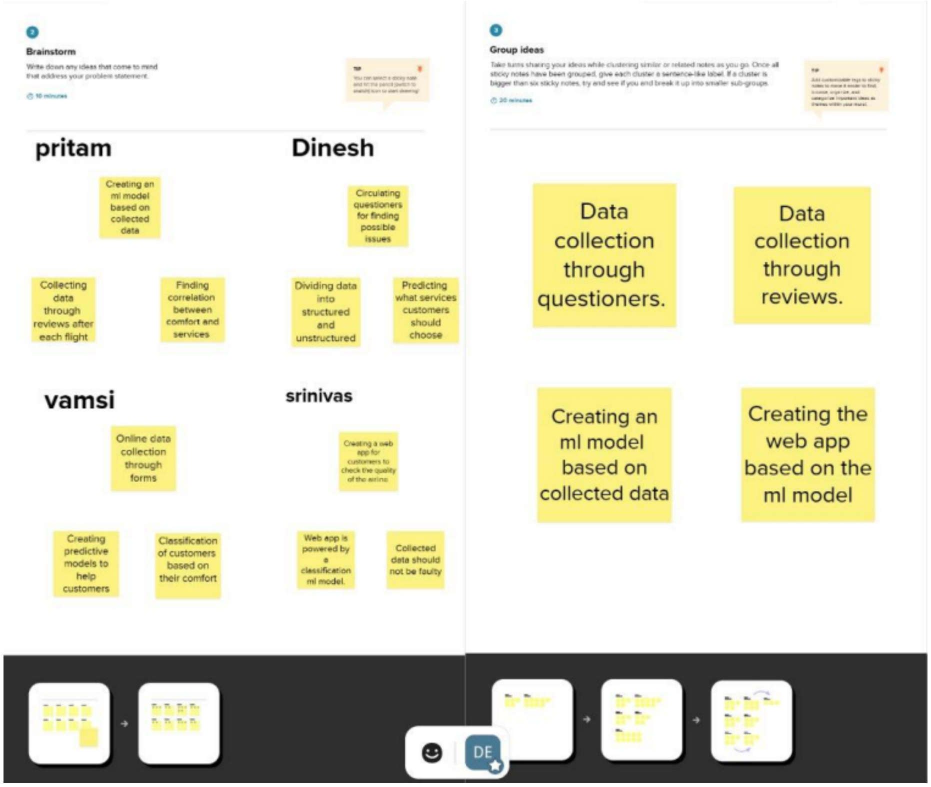
Finding correlation and issues

Data Collection:

Gather a diverse dataset of airline reviews. This data should include text reviews along with associated labels (e.g., positive, negative, or neutral sentiment). You can obtain this data from sources like customer review websites, social media, or surveys

Brainstorming:

Classifying airline reviews using machine learning can involve various ideas and approaches. Here's a brainstorming session with some ideas, followed by a prioritization of these ideas Sentiment Analysis:



2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

pritam

Creating an ml model based on collected data

Collecting data through reviews after each flight

Finding correlation between comfort and services

Dinesh

Circulating questioners for finding possible issues

Dividing data into structured and unstructured

Predicting what services customers should choose

vamsi

Online data collection through forms

Creating predictive models to help customers

Classification of customers based on their comfort

srinivas

Creating a web app for customers to check the quality of the airline

Web app is powered by a classification ml model.

Collected data should not be faulty

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

Data collection through questioners.

Data collection through reviews.

Creating an ml model based on collected data

Creating the web app based on the ml model

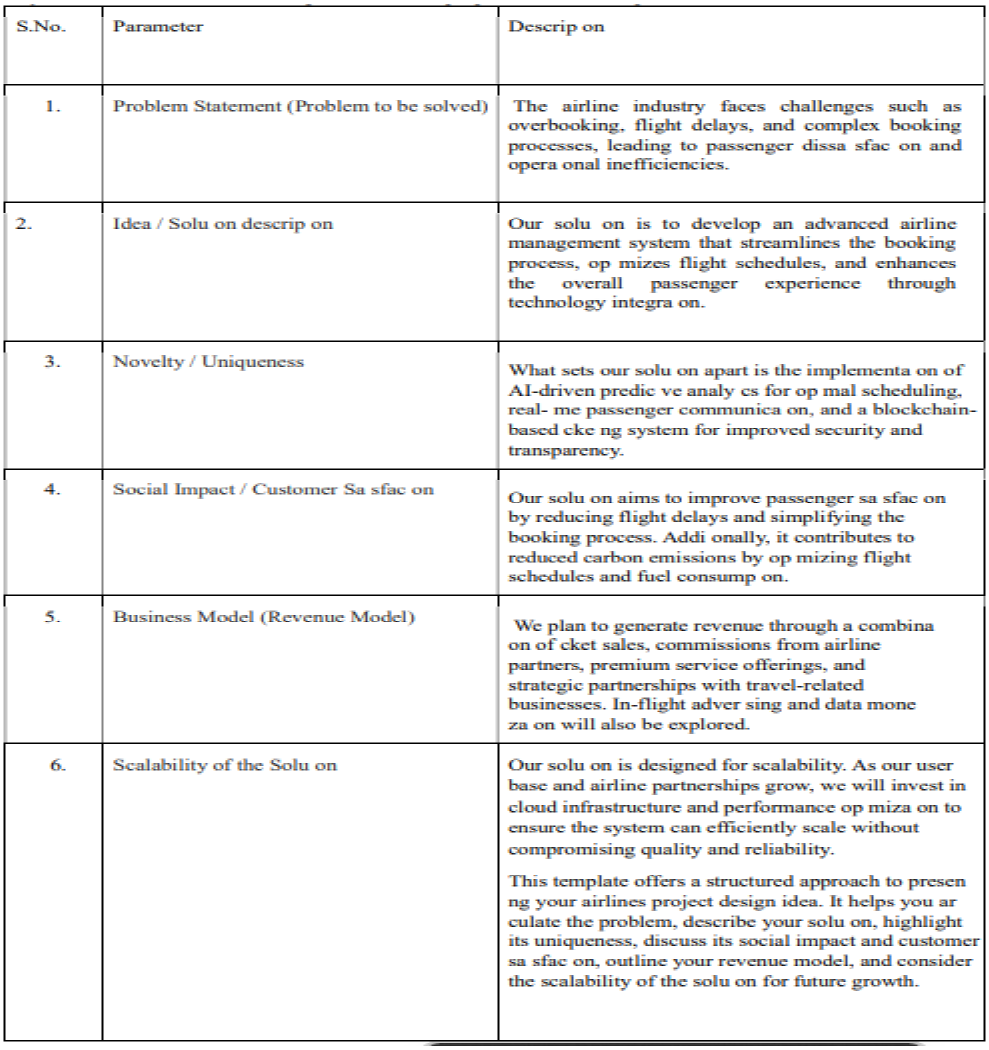
Develop a sentiment analysis model to categorize reviews as positive, negative, or neutral based on the expressed sentiment.
 Feedback Clustering

Use clustering techniques to group similar reviews together. This can help in identifying common issues and solutions.
When considering which ideas to pursue, assess your goals, available resources, and the specific challenges or opportunities your airline faces. Start with high-impact ideas and gradually expand into other areas as resources and capabilities allow.

Proposed solution Template:

creating a well-structured proposal or solution template is essential for effectively presenting ideas,projects,or solutions to stakeholders, clients,or team members.Below is a general template that you can use as a starting point, which you can adapt to your specific needs and requirements
1) Problem statement
2) Idea/solution on description
3)Novelty/ Uniqueness

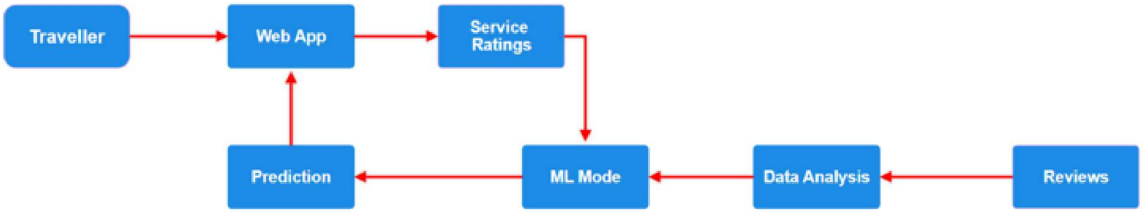
4)Social Impact
5) Business Model
6)Scalability of the Solution



S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The airline industry faces challenges such as overbooking, flight delays, and complex booking processes, leading to passenger dissatisfaction and operational inefficiencies.
2.	Idea / Solution description	Our solution is to develop an advanced airline management system that streamlines the booking process, optimizes flight schedules, and enhances the overall passenger experience through technology integration.
3.	Novelty / Uniqueness	What sets our solution apart is the implementation of AI-driven predictive analytics for optimal scheduling, real-time passenger communication, and a blockchain-based ticketing system for improved security and transparency.
4.	Social Impact / Customer Satisfaction	Our solution aims to improve passenger satisfaction by reducing flight delays and simplifying the booking process. Additionally, it contributes to reduced carbon emissions by optimizing flight schedules and fuel consumption.
5.	Business Model (Revenue Model)	We plan to generate revenue through a combination of ticket sales, commissions from airline partners, premium service offerings, and strategic partnerships with travel-related businesses. In-flight advertising and data monetization will also be explored.
6.	Scalability of the Solution	Our solution is designed for scalability. As our user base and airline partnerships grow, we will invest in cloud infrastructure and performance optimization to ensure the system can efficiently scale without compromising quality and reliability. This template offers a structured approach to presenting your airlines project design idea. It helps you articulate the problem, describe your solution, highlight its uniqueness, discuss its social impact and customer satisfaction, outline your revenue model, and consider the scalability of the solution for future growth.

Solution Architecture:

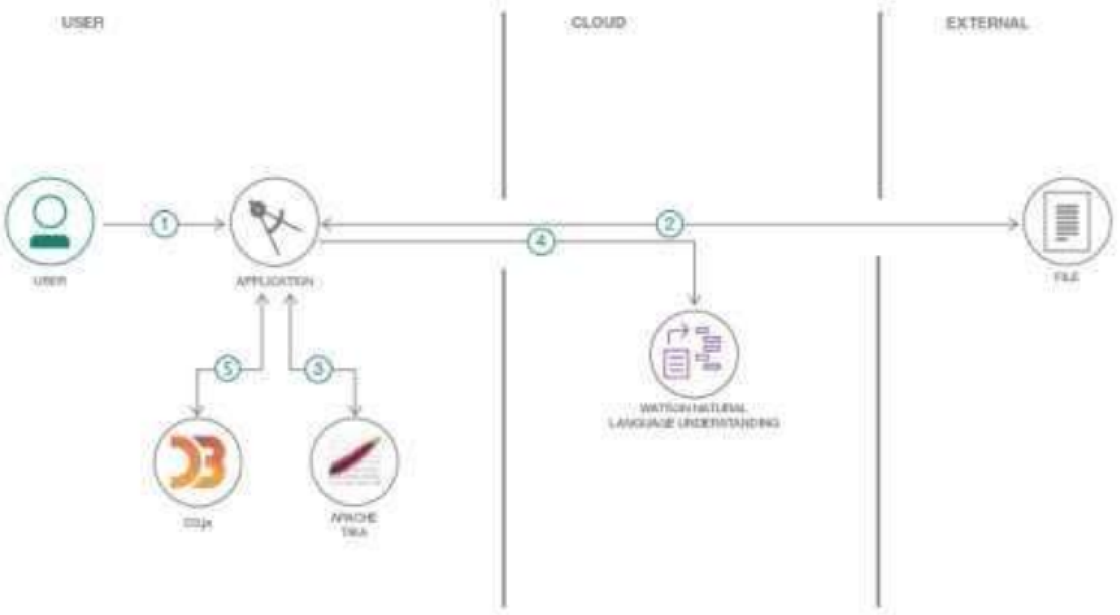
Creating a solution architecture is crucial for designing and implementing complex systems or projects effectively . A solution architecture outlines the high-levels structure, components, and interactions of the solution, ensuring that it aligns with business goals and technical requirements. Here's a template you can as a starting point for a solution architecture

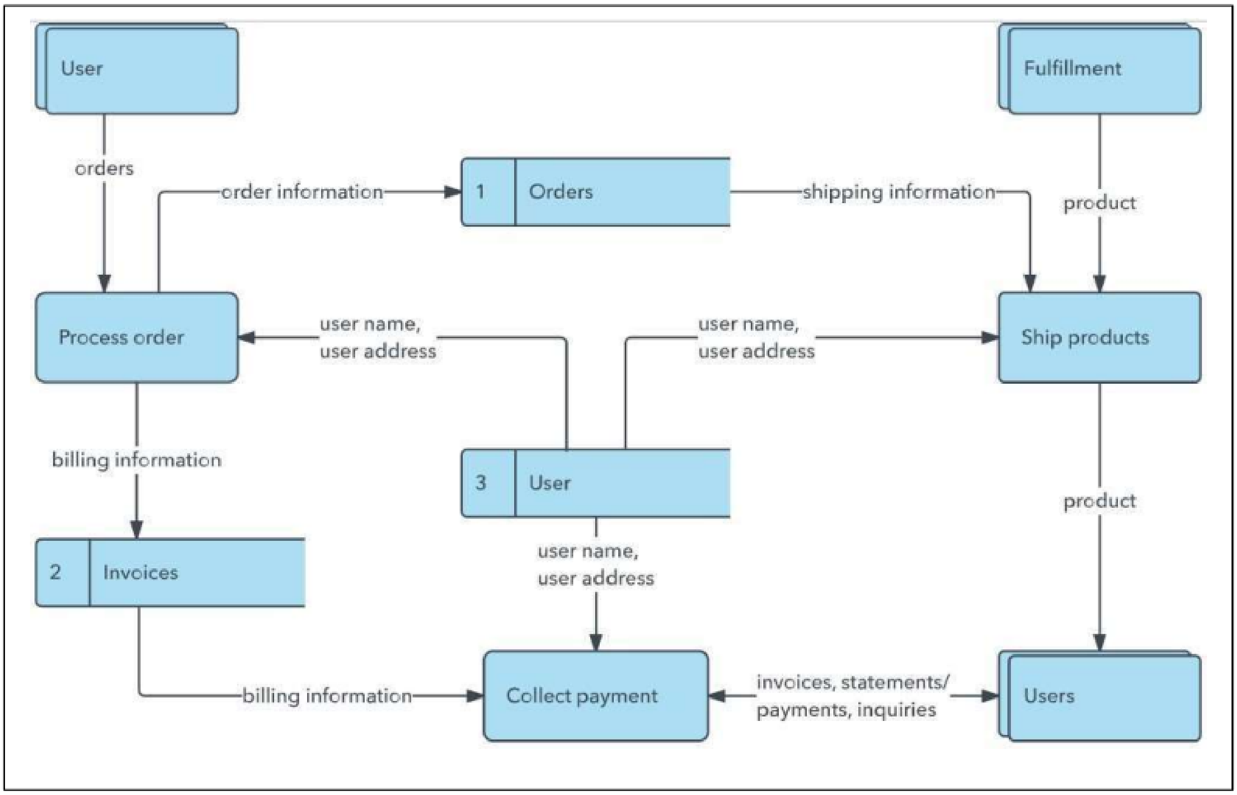


Data Flow :

Data flow refers to the movement of data from one point or process to another within a system or network. It involves the transfer, transformation, and processing of data throughout its lifecycle.In a typical data flow, data originates from a source or input and flows through various stages or components within a system. This flow can be represented as a series of steps or stages, including data collection, storage, processing, analysis, and output.

Flow:





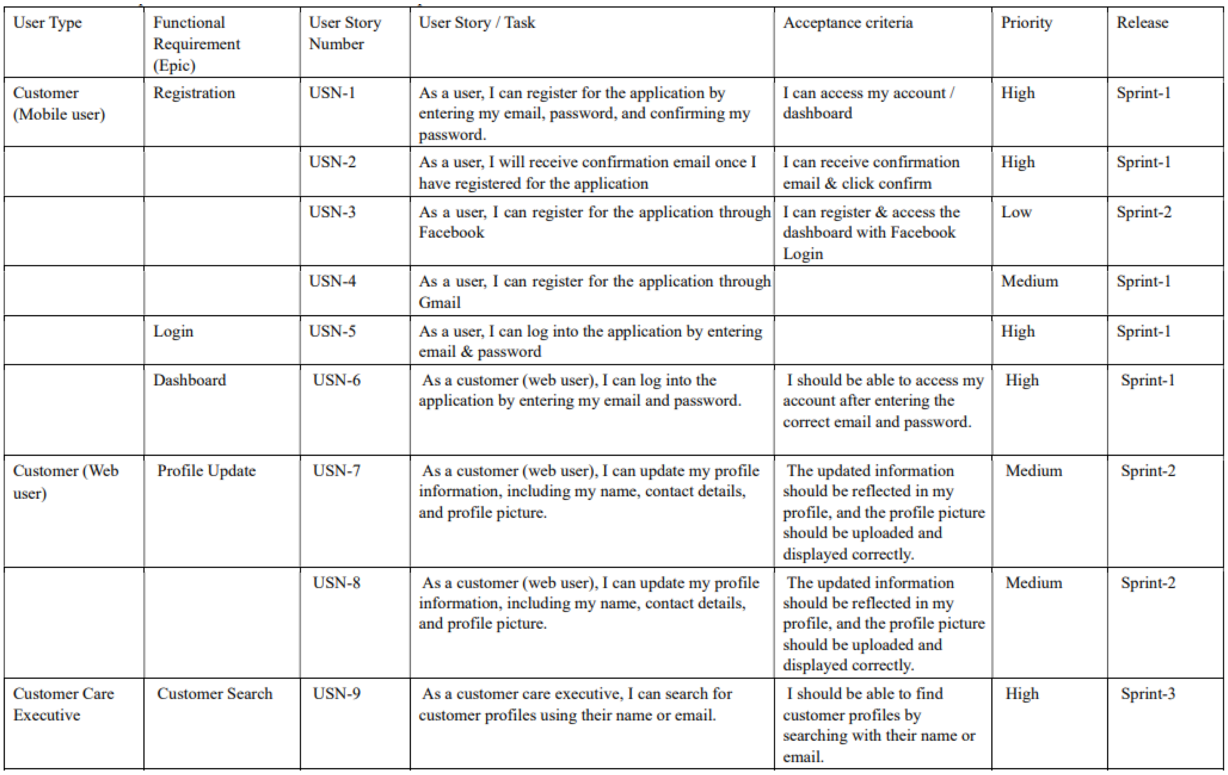
User Stories:

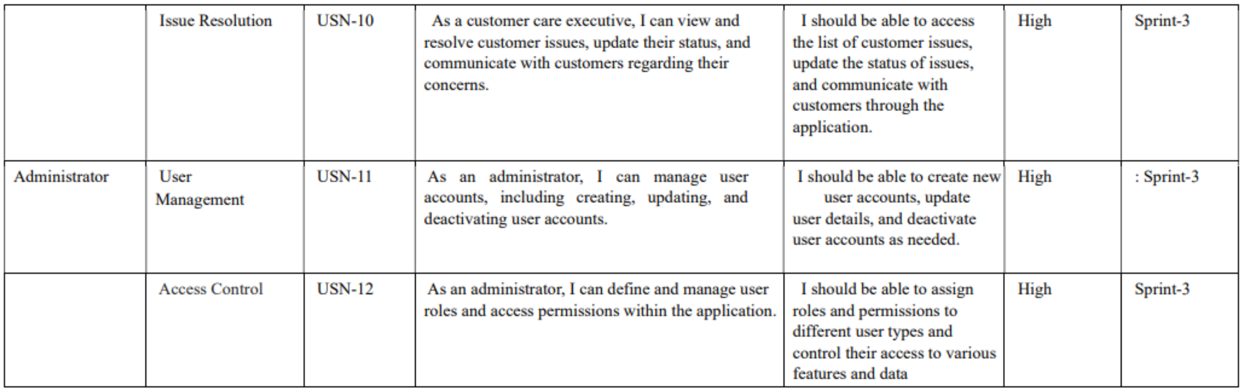
User stories are a common way to document functional requirements in agile software development. They provide a clear, user-centric description of a feature or functionality from the perspective of the end user. User stories are typically written in a simple, structured format and serve as a communication tool between product owners, developers, and other stakeholders. Here's a template for writing user stories:

As a [type of user]": This part identifies who the user is. It can be a specific role, such as "As a customer" or "As an administrator." The goal is to specify the perspective from which the user is making the request.

"I want [an action]": This is the action or functionality the user is requesting. It should be specific, clear, and focused on what the user needs to do or achieve. For example, "I want to add items to my shopping cart."

"So that [benefit/value]": This part explains the value or benefit that the user expects from the requested action. It helps in understanding the underlying motivation behind the user story. For example, "So that I can review and purchase them later."





User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	As a customer (web user), I can log into the application by entering my email and password.	I should be able to access my account after entering the correct email and password.	High	Sprint-1
Customer (Web user)	Profile Update	USN-7	As a customer (web user), I can update my profile information, including my name, contact details, and profile picture.	The updated information should be reflected in my profile, and the profile picture should be uploaded and displayed correctly.	Medium	Sprint-2
		USN-8	As a customer (web user), I can update my profile information, including my name, contact details, and profile picture.	The updated information should be reflected in my profile, and the profile picture should be uploaded and displayed correctly.	Medium	Sprint-2
Customer Care Executive	Customer Search	USN-9	As a customer care executive, I can search for customer profiles using their name or email.	I should be able to find customer profiles by searching with their name or email.	High	Sprint-3
	Issue Resolution	USN-10	As a customer care executive, I can view and resolve customer issues, update their status, and communicate with customers regarding their concerns.	I should be able to access the list of customer issues, update the status of issues, and communicate with customers through the application.	High	Sprint-3
Administrator	User Management	USN-11	As an administrator, I can manage user accounts, including creating, updating, and deactivating user accounts.	I should be able to create new user accounts, update user details, and deactivate user accounts as needed.	High	: Sprint-3
	Access Control	USN-12	As an administrator, I can define and manage user roles and access permissions within the application.	I should be able to assign roles and permissions to different user types and control their access to various features and data	High	Sprint-3

As a researcher in material science, I want to reuse a sample after performing Energy Dispersion X-Ray Spectroscopy (EDS) to reduce waste and optimize resources.

As a researcher preparing samples for Transmission Electron Microscopy (TEM), I want to understand the various techniques used to prepare samples, such as ultramicrotomy, ion milling, and mechanical grinding/polishing, to ensure high-quality TEM analysis.

As a researcher utilizing X-ray Photo electron Spectroscopy (XPS), I want to know the type of information it provides about the surface of a sample, such as chemistry, to accurately analyze the surface composition and understand its chemical properties.

Project Planning

Technical Architecture:
The technical architecture of a software system refers to the overall structure and organization of its components, including the hardware, software, and network elements that make up the system.

Open Source Frameworks:

Open source frameworks are pre-built software components that provide a set of common functionality for specific tasks or domains.

Third-party APIs:

Third-party APIs (Application Programming Interfaces) are software intermediaries that allow applications to communicate and exchange data with each other.

Cloud Deployment:

Cloud deployment refers to the process of hosting and running software applications on cloud computing platforms, such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP).

PROJECT OVERVIEW :

The Airline Reservation System project is an implementation of a general Airline Ticketing website like Orbitz, which helps the customers to search the availability and prices of various airline tickets, along with the different packages available with the Reservations.

Project Planning:

1.Create a detailed project plan outlining the tasks, timelines, and resources required.

2.Develop a budget that includes all projected costs and potential risks.

Infrastructure Development:

1.Evaluate the need for airport infrastructure development and expansion.

2.Coordinate with airport authorities for facilities and service requirements.

Evaluation and Monitoring:

1.Establish key performance indicators (KPIs) to measure the success of the project.

2.Conduct regular evaluations to assess the project's progress and make necessary adjustments.

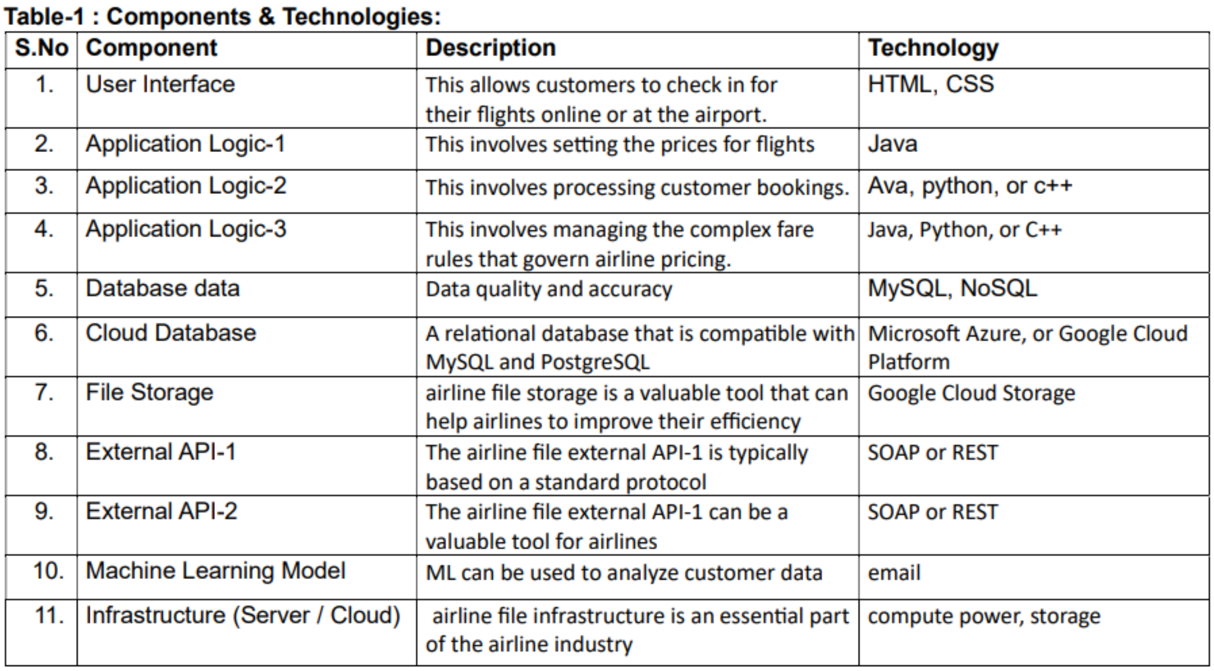
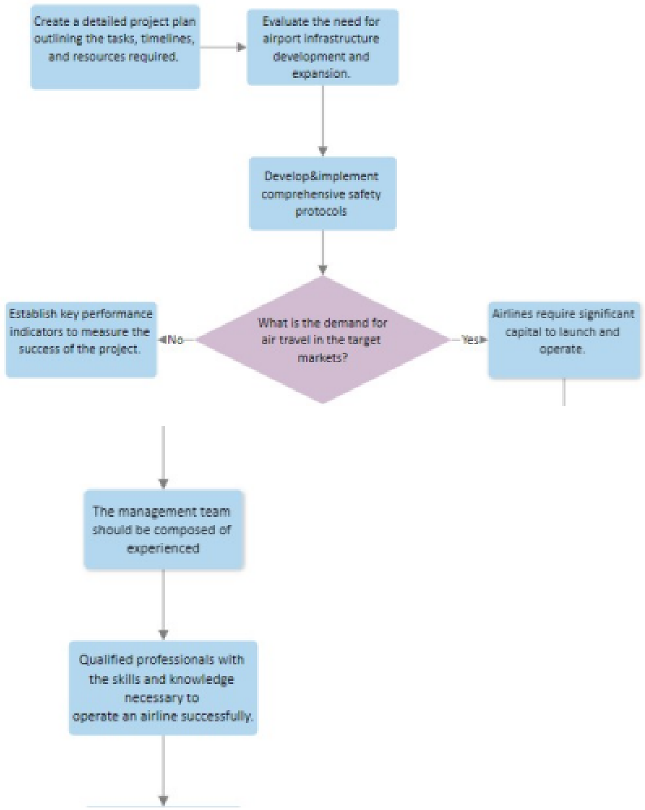
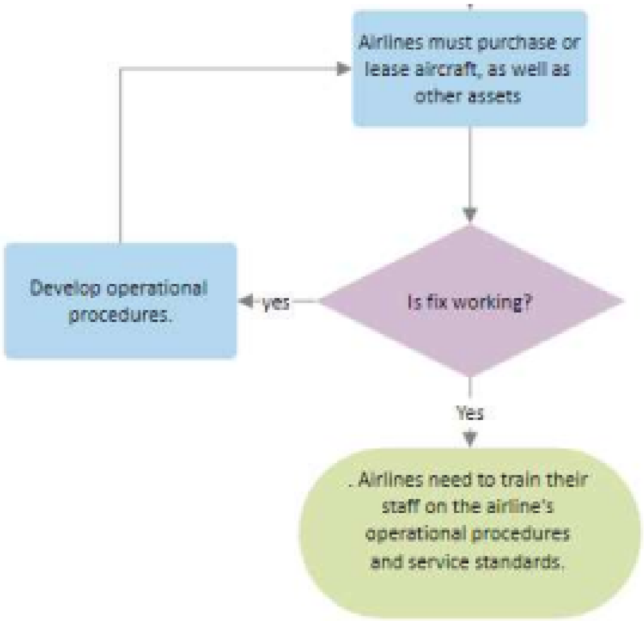


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	This allows customers to check in for their flights online or at the airport.	HTML, CSS
2.	Application Logic-1	This involves setting the prices for flights	Java
3.	Application Logic-2	This involves processing customer bookings.	Ava, python, or c++
4.	Application Logic-3	This involves managing the complex fare rules that govern airline pricing.	Java, Python, or C++
5.	Database data	Data quality and accuracy	MySQL, NoSQL
6.	Cloud Database	A relational database that is compatible with MySQL and PostgreSQL	Microsoft Azure, or Google Cloud Platform
7.	File Storage	airline file storage is a valuable tool that can help airlines to improve their efficiency	Google Cloud Storage
8.	External API-1	The airline file external API-1 is typically based on a standard protocol	SOAP or REST
9.	External API-2	The airline file external API-1 can be a valuable tool for airlines	SOAP or REST
10.	Machine Learning Model	ML can be used to analyze customer data	email
11.	Infrastructure (Server / Cloud)	airline file infrastructure is an essential part of the airline industry	compute power, storage

Data Flow:


Create a detailed project plan outlining the tasks, timelines, and resources required.
Evaluate the need for airport infrastructure development and expansion.
Develop&implement comprehensive safety protocols
What is the demand for air travel in the target markets?
No
Establish key performance indicators to measure the success of the project.
Yes
Airlines require significant capital to launch and operate.
The management team should be composed of experienced
Qualified professionals with the skills and knowledge necessary to operate an airline successfully.

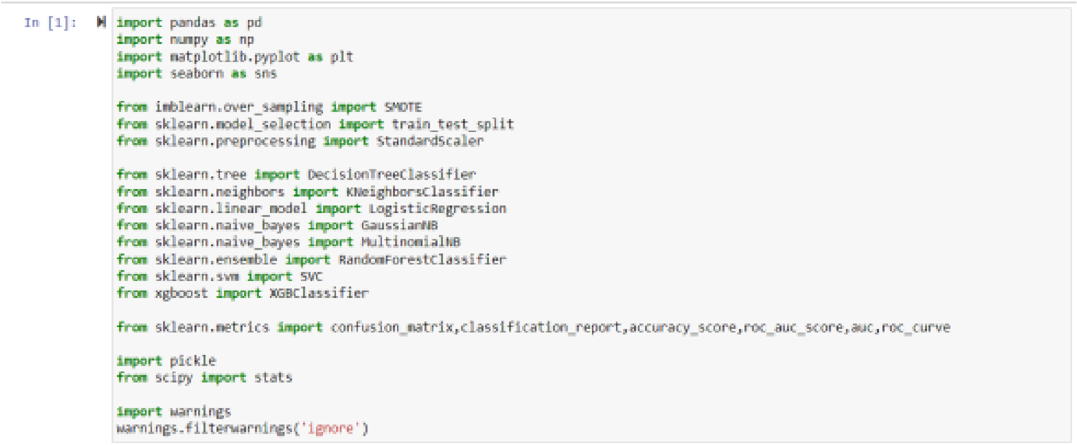

Airlines must purchase or lease aircraft, as well as other assets
Is fix working?
yes
Develop operational procedures.
Yes
. Airlines need to train their staff on the airline's operational procedures and service standards.

Project Development

The stages of project development are:

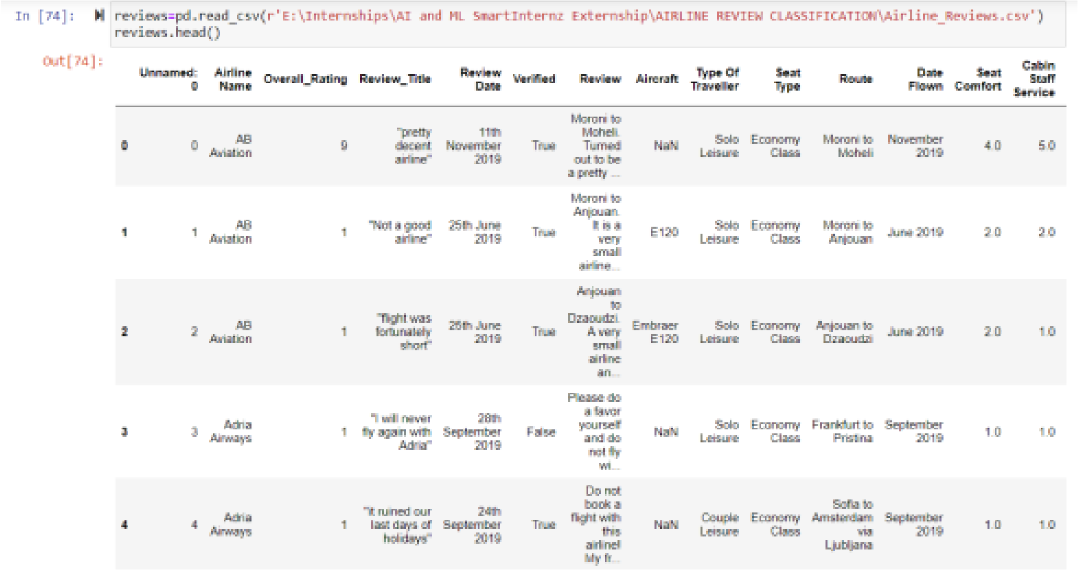
1. Importing Datasets and Libraries

a. Importing Libraries



```
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns

         from imblearn.over_sampling import SMOTE
         from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler

         from sklearn.tree import DecisionTreeClassifier
         from sklearn.neighbors import KNeighborsClassifier
         from sklearn.linear_model import LogisticRegression
         from sklearn.naive_bayes import GaussianNB
         from sklearn.naive_bayes import MultinomialNB
         from sklearn.ensemble import RandomForestClassifier
         from sklearn.svm import SVC
         from xgboost import XGBClassifier

         from sklearn.metrics import confusion_matrix,classification_report,accuracy_score,roc_auc_score,auc,roc_curve

         import pickle
         from scipy import stats

         import warnings
         warnings.filterwarnings('ignore')
```

b. Importing Dataset



```
In [74]:  reviews=pd.read_csv(r'E:\Internships\AI and ML SmartInternz Externship\AIRLINE REVIEW CLASSIFICATION\Airline_Reviews.csv')
          reviews.head()
```

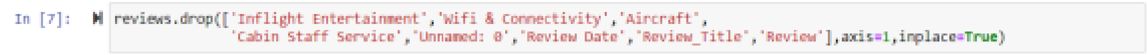
Out[74]:

	Unnamed: 0	Airline Name	Overall_Rating	Review_Title	Review Date	Verified	Review	Aircraft	Type Of Traveller	Seat Type	Route	Date Flown	Seat Comfort	Cabin Staff Service
0	0	AB Aviation	9	"pretty decent airline"	11th November 2019	True	Moroni to Moheli. Turned out to be a pretty ...	NaN	Solo Leisure	Economy Class	Moroni to Moheli	November 2019	4.0	5.0
1	1	AB Aviation	1	"Not a good airline"	25th June 2019	True	Moroni to Anjouan. It is a very small airline...	E120	Solo Leisure	Economy Class	Moroni to Anjouan	June 2019	2.0	2.0
2	2	AB Aviation	1	"flight was fortunately short"	25th June 2019	True	Anjouan to Dzaoudzi. A very small airline an...	Embraer E120	Solo Leisure	Economy Class	Anjouan to Dzaoudzi	June 2019	2.0	1.0
3	3	Adria Airways	1	"I will never fly again with Adria"	28th September 2019	False	Please do a favor yourself and do not fly wi...	NaN	Solo Leisure	Economy Class	Frankfurt to Pristina	September 2019	1.0	1.0
4	4	Adria Airways	1	"it ruined our last days of holidays"	24th September 2019	True	Do not book a flight with this airline! My fr...	NaN	Couple Leisure	Economy Class	Sofia to Amsterdam via Ljubljana	September 2019	1.0	1.0

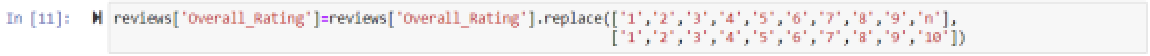
2. Data Preprocessing: Removal of Null values, removing useless columns, converting categorical to numeric values are done in this step.

a. Handling NULL Values

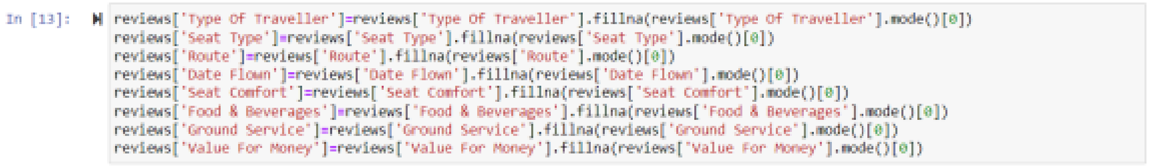
Dropping the Unnecessary Columns



```
In [7]:  reviews.drop(['Inflight Entertainment','Wifi & Connectivity','Aircraft',
                        'Cabin Staff Service','Unnamed: 0','Review Date','Review_Title','Review'],axis=1,inplace=True)
```

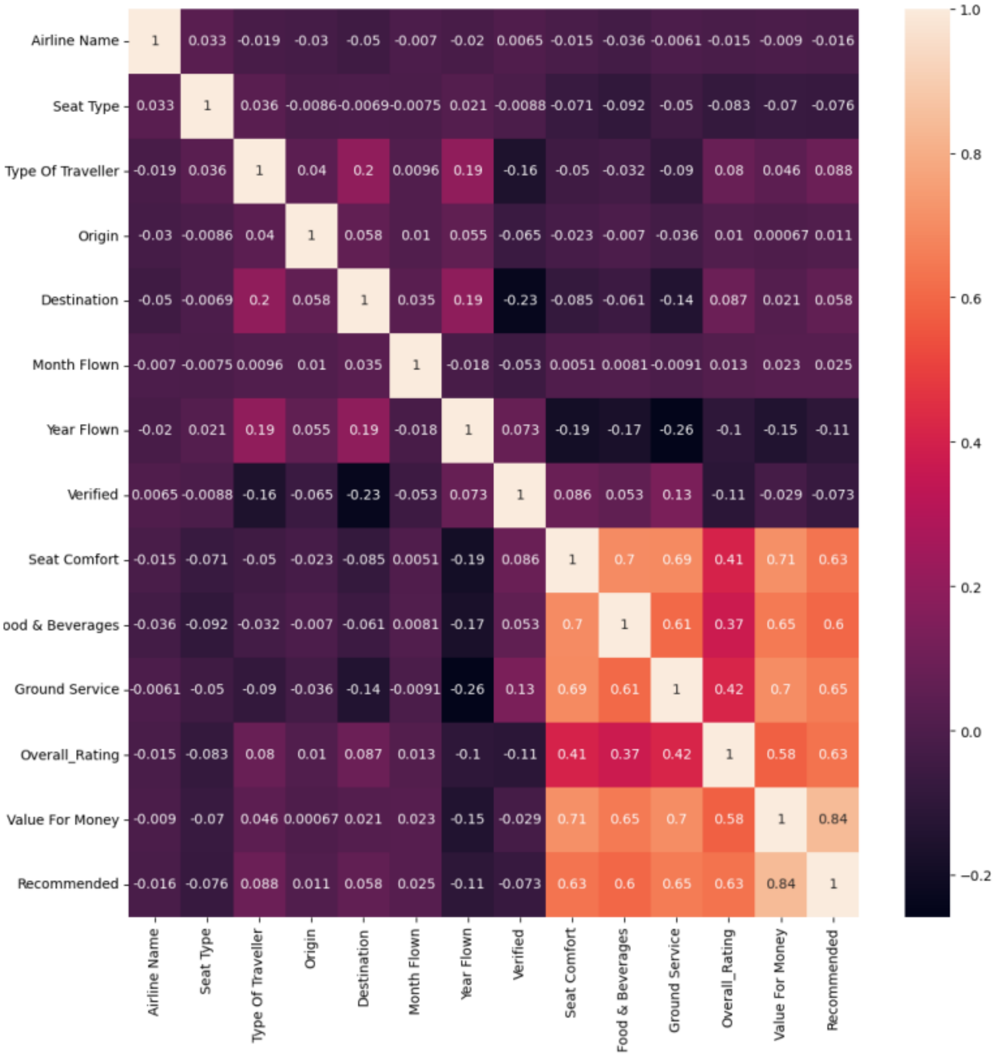


```
In [11]:  reviews['Overall_Rating']=reviews['Overall_Rating'].replace(['1','2','3','4','5','6','7','8','9','n'],
                                                                     ['1','2','3','4','5','6','7','8','9','10'])
```



```
In [13]:  reviews['Type Of Traveller']=reviews['Type Of Traveller'].fillna(reviews['Type Of Traveller'].mode()[0])
          reviews['Seat Type']=reviews['Seat Type'].fillna(reviews['Seat Type'].mode()[0])
          reviews['Route']=reviews['Route'].fillna(reviews['Route'].mode()[0])
          reviews['Date Flown']=reviews['Date Flown'].fillna(reviews['Date Flown'].mode()[0])
          reviews['Seat Comfort']=reviews['Seat Comfort'].fillna(reviews['Seat Comfort'].mode()[0])
          reviews['Food & Beverages']=reviews['Food & Beverages'].fillna(reviews['Food & Beverages'].mode()[0])
          reviews['Ground Service']=reviews['Ground Service'].fillna(reviews['Ground Service'].mode()[0])
          reviews['Value For Money']=reviews['Value For Money'].fillna(reviews['Value For Money'].mode()[0])
```

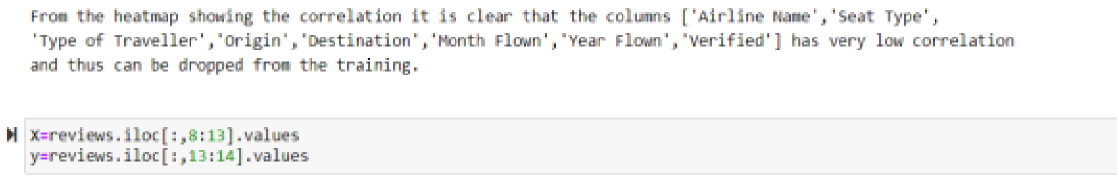
3. Exploratory Data Analysis: 1-Dimentional and 2-Dimentional data analysis is done to find important relations and insights.



	Airline Name	Seat Type	Type Of Traveller	Origin	Destination	Month Flown	Year Flown	Verified	Seat Comfort	Food & Beverages	Ground Service	Overall_Rating	Value For Money	Recommended
Airline Name	1	0.033	-0.019	-0.03	-0.05	-0.007	-0.02	0.0065	-0.015	-0.036	-0.0061	-0.015	-0.009	-0.016
Seat Type	0.033	1	0.036	-0.0086	-0.0069	-0.0075	0.021	-0.0088	-0.071	-0.092	-0.05	-0.083	-0.07	-0.076
Type Of Traveller	-0.019	0.036	1	0.04	0.2	0.0096	0.19	-0.16	-0.05	-0.032	-0.09	0.08	0.046	0.088
Origin	-0.03	-0.0086	0.04	1	0.058	0.01	0.055	-0.065	-0.023	-0.007	-0.036	0.01	0.00067	0.011
Destination	-0.05	-0.0069	0.2	0.058	1	0.035	0.19	-0.23	-0.085	-0.061	-0.14	0.087	0.021	0.058
Month Flown	-0.007	-0.0075	0.0096	0.01	0.035	1	-0.018	-0.053	0.0051	0.0081	-0.0091	0.013	0.023	0.025
Year Flown	-0.02	0.021	0.19	0.055	0.19	-0.018	1	0.073	-0.19	-0.17	-0.26	-0.1	-0.15	-0.11
Verified	0.0065	-0.0088	-0.16	-0.065	-0.23	-0.053	0.073	1	0.086	0.053	0.13	-0.11	-0.029	-0.073
Seat Comfort	-0.015	-0.071	-0.05	-0.023	-0.085	0.0051	-0.19	0.086	1	0.7	0.69	0.41	0.71	0.63
ood & Beverages	-0.036	-0.092	-0.032	-0.007	-0.061	0.0081	-0.17	0.053	0.7	1	0.61	0.37	0.65	0.6
Ground Service	-0.0061	-0.05	-0.09	-0.036	-0.14	-0.0091	-0.26	0.13	0.69	0.61	1	0.42	0.7	0.65
Overall_Rating	-0.015	-0.083	0.08	0.01	0.087	0.013	-0.1	-0.11	0.41	0.37	0.42	1	0.58	0.63
Value For Money	-0.009	-0.07	0.046	0.00067	0.021	0.023	-0.15	-0.029	0.71	0.65	0.7	0.58	1	0.84
Recommended	-0.016	-0.076	0.088	0.011	0.058	0.025	-0.11	-0.073	0.63	0.6	0.65	0.63	0.84	1

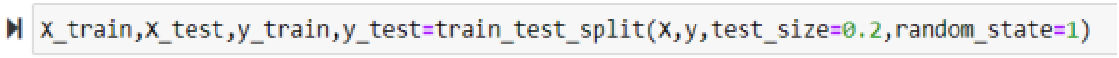
4. Model Preparation: The columns important for model creation are kept while the other columns are removed. The data is also divided into training and testing data. Scaling of the data is also done.

a. Splitting Data into Training and Testing Sets



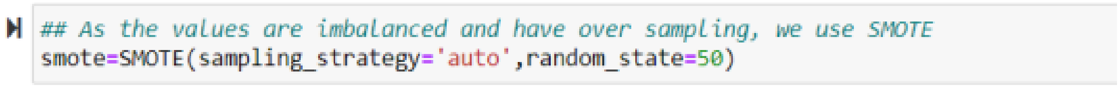
From the heatmap showing the correlation it is clear that the columns ['Airline Name','Seat Type', 'Type of Traveller','Origin','Destination','Month Flown','Year Flown','Verified'] has very low correlation and thus can be dropped from the training.

```
X=reviews.iloc[:,8:13].values
y=reviews.iloc[:,13:14].values
```

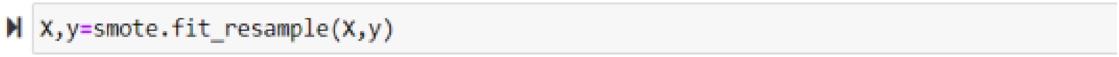


```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=1)
```

b. Removing Class Imbalance

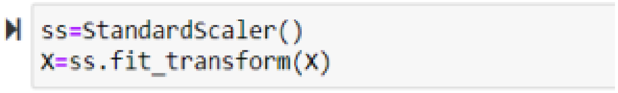


```
## As the values are imbalanced and have over sampling, we use SMOTE
smote=SMOTE(sampling_strategy='auto',random_state=50)
```



```
X,y=smote.fit_resample(X,y)
```

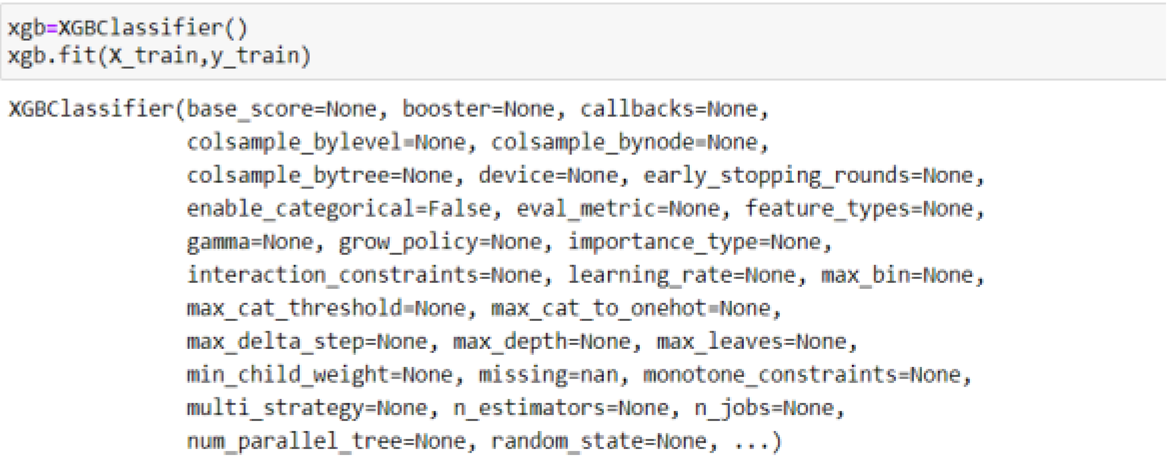
c. Scaling the column values



```
ss=StandardScaler()
X=ss.fit_transform(X)
```

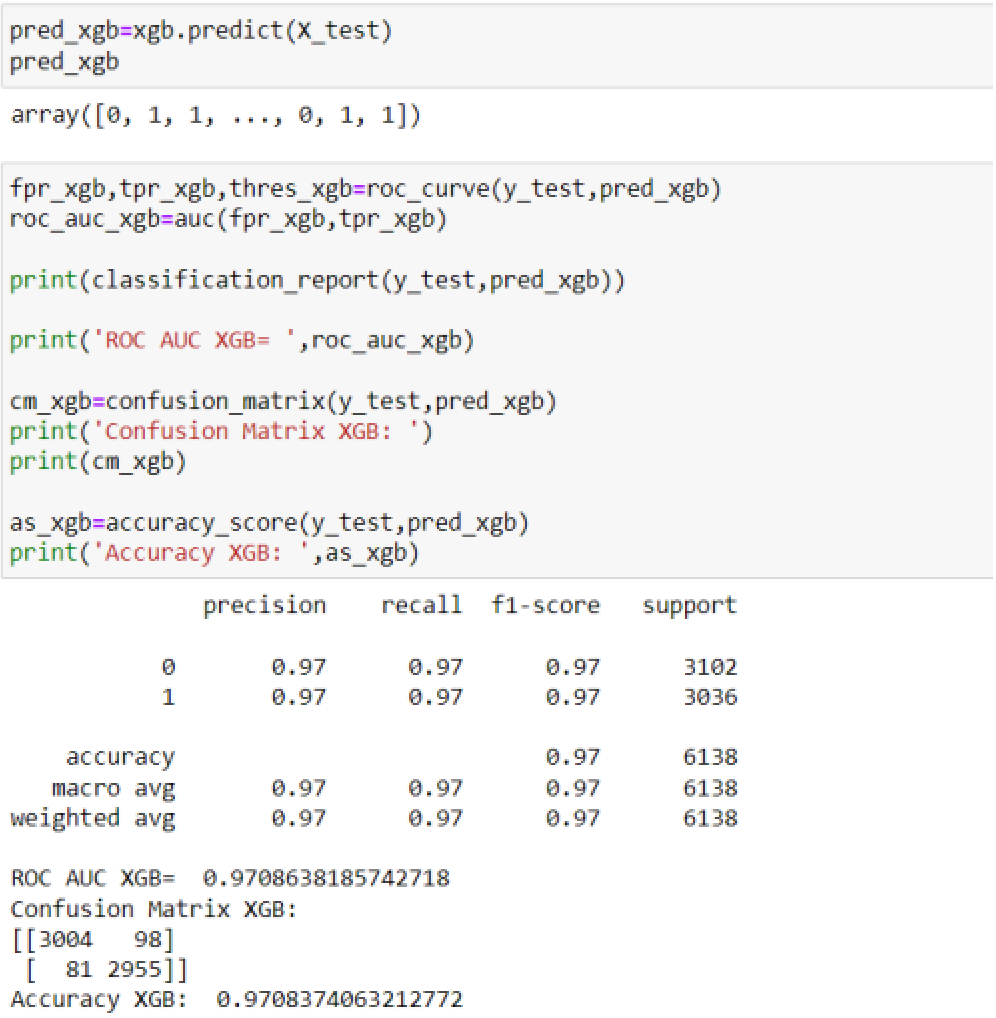
5. Model Training: Model based several classification types are trained to find the model with the best performance. The best model is saved for further predictions.

g. XG Boost



```
xgb=XGBClassifier()
xgb.fit(X_train,y_train)
```

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None,
              num_parallel_tree=None, random_state=None, ...)
```



```
pred_xgb=xgb.predict(X_test)
pred_xgb
```

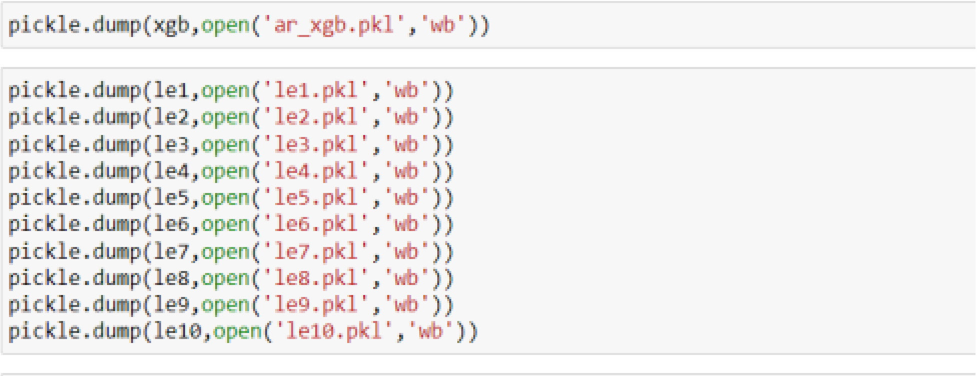
```
array([0, 1, 1, ..., 0, 1, 1])
```

```
fpr_xgb,tpr_xgb,thres_xgb=roc_curve(y_test,pred_xgb)
roc_auc_xgb=auc(fpr_xgb,tpr_xgb)

print(classification_report(y_test,pred_xgb))

print('ROC AUC XGB= ',roc_auc_xgb)

cm_xgb=confusion_matrix(y_test,pred_xgb)
print('Confusion Matrix XGB: ')
print(cm_xgb)

as_xgb=accuracy_score(y_test,pred_xgb)
print('Accuracy XGB: ',as_xgb)
```

```
              precision    recall  f1-score   support

           0       0.97      0.97      0.97      3102
           1       0.97      0.97      0.97      3036

    accuracy                           0.97      6138
   macro avg       0.97      0.97      0.97      6138
weighted avg       0.97      0.97      0.97      6138

ROC AUC XGB=  0.9708638185742718
Confusion Matrix XGB: 
[[3004   98]
 [  81 2955]]
Accuracy XGB:  0.9708374063212772
```

6. Saving the model: It uses pickle or joblib libraries to save the model

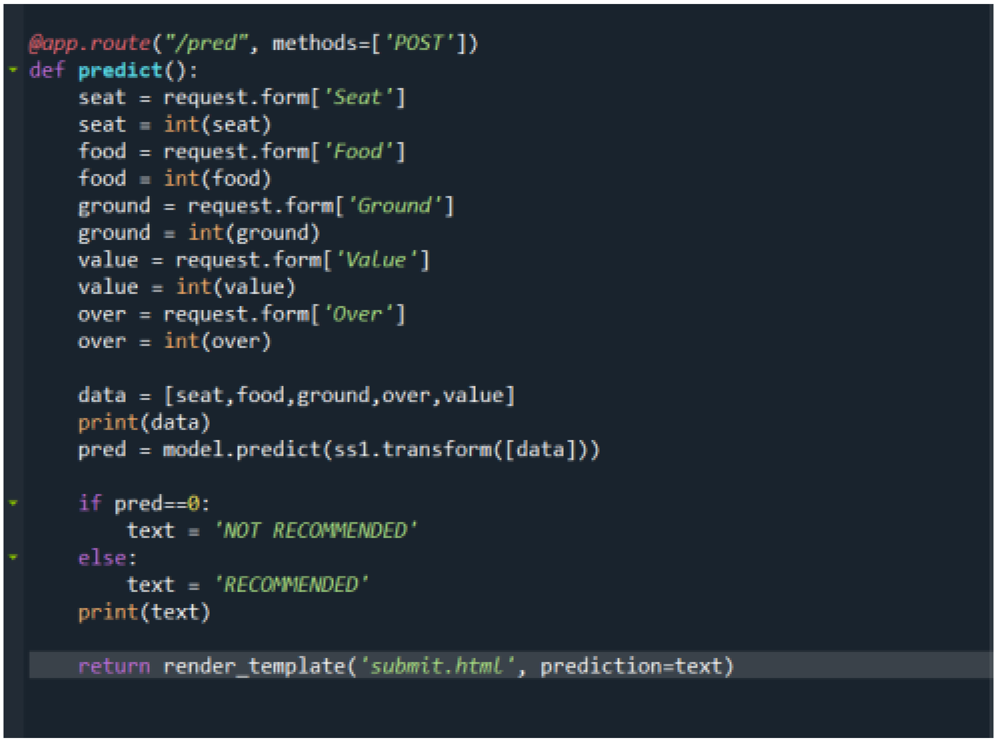
so that is can be used for further prediction.



```
pickle.dump(xgb,open('ar_xgb.pkl','wb'))
```

```
pickle.dump(le1,open('le1.pkl','wb'))
pickle.dump(le2,open('le2.pkl','wb'))
pickle.dump(le3,open('le3.pkl','wb'))
pickle.dump(le4,open('le4.pkl','wb'))
pickle.dump(le5,open('le5.pkl','wb'))
pickle.dump(le6,open('le6.pkl','wb'))
pickle.dump(le7,open('le7.pkl','wb'))
pickle.dump(le8,open('le8.pkl','wb'))
pickle.dump(le9,open('le9.pkl','wb'))
pickle.dump(le10,open('le10.pkl','wb'))
```

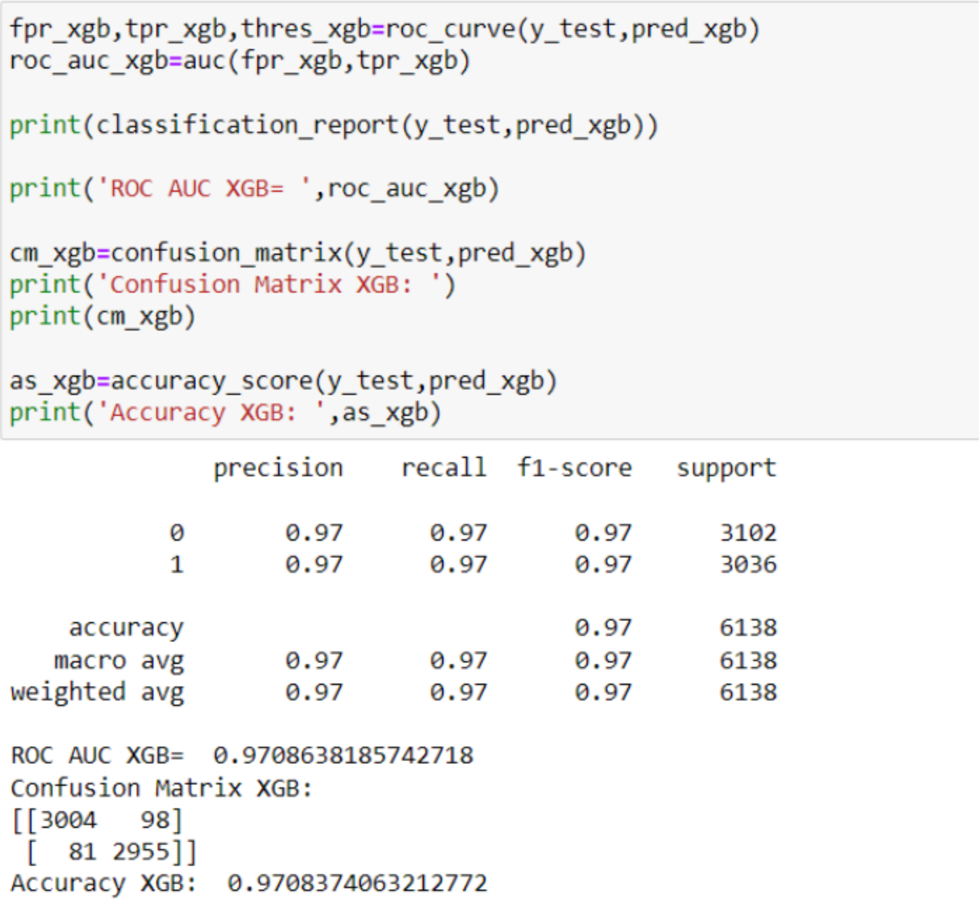
7. FLASK Application Development: This step includes connecting the ML model in the backend to web pages to create an user interactive portal which can show the results of the model, and also collect additional data.



```
@app.route("/pred", methods=['POST'])
def predict():
    seat = request.form['Seat']
    seat = int(seat)
    food = request.form['Food']
    food = int(food)
    ground = request.form['Ground']
    ground = int(ground)
    value = request.form['Value']
    value = int(value)
    over = request.form['Over']
    over = int(over)

    data = [seat,food,ground,over,value]
    print(data)
    pred = model.predict(ss1.transform([data]))

    if pred==0:
        text = 'NOT RECOMMENDED'
    else:
        text = 'RECOMMENDED'
    print(text)

    return render_template('submit.html', prediction=text)
```

PERFORMANCE TESTING:

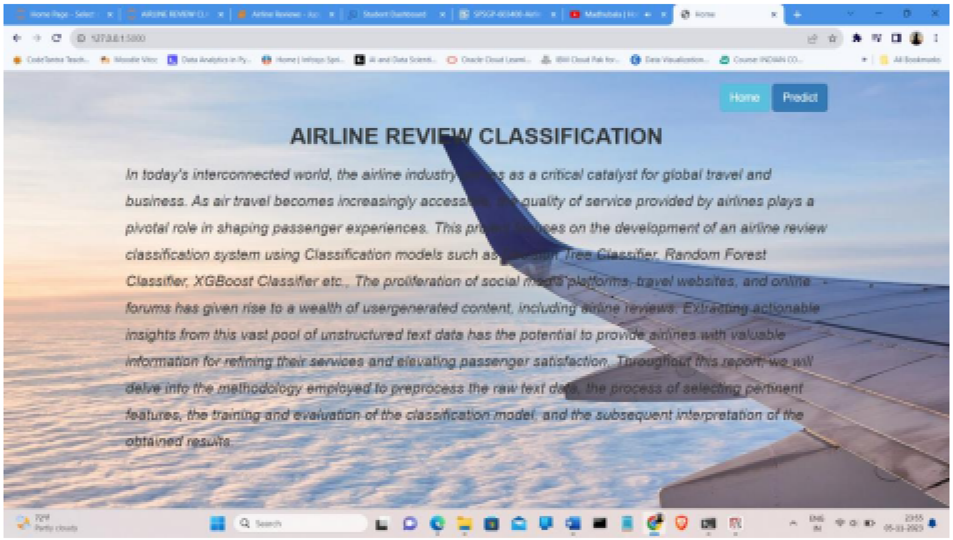
Metrics: Confusion Matrix - , Accuracy Score- & Classification Report -



```
fpr_xgb,tpr_xgb,thres_xgb=roc_curve(y_test,pred_xgb)
roc_auc_xgb=auc(fpr_xgb,tpr_xgb)

print(classification_report(y_test,pred_xgb))

print('ROC AUC XGB= ',roc_auc_xgb)

cm_xgb=confusion_matrix(y_test,pred_xgb)
print('Confusion Matrix XGB: ')
print(cm_xgb)

as_xgb=accuracy_score(y_test,pred_xgb)
print('Accuracy XGB: ',as_xgb)
```

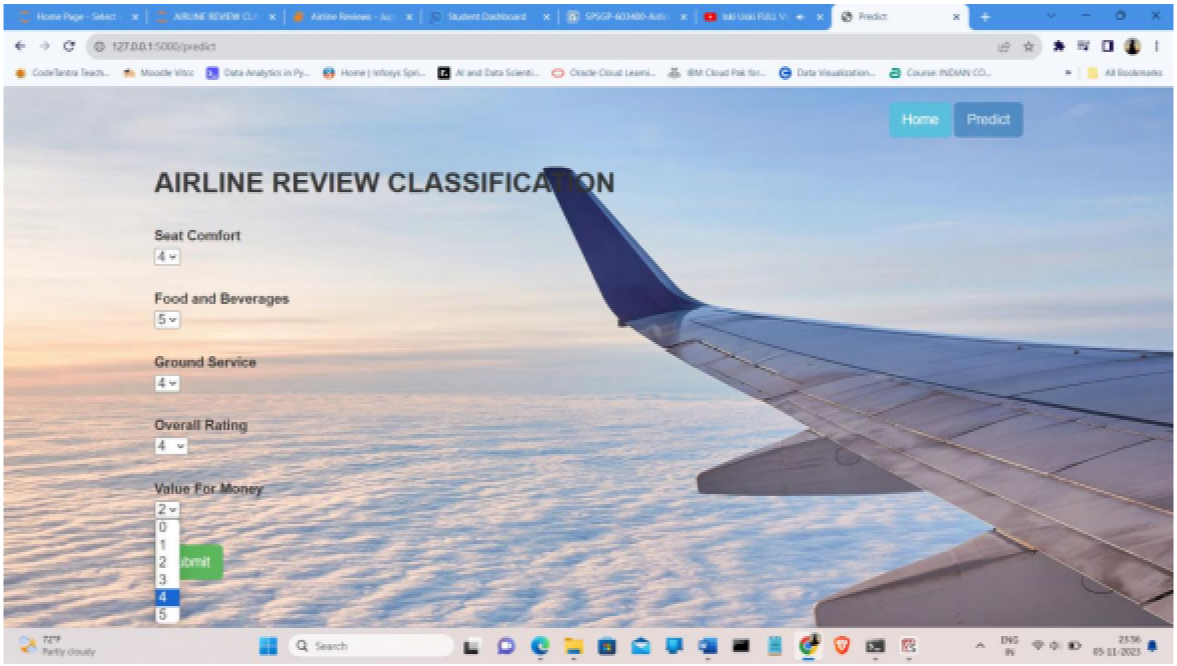
```
              precision    recall  f1-score   support

           0       0.97      0.97      0.97      3102
           1       0.97      0.97      0.97      3036

    accuracy                           0.97      6138
   macro avg       0.97      0.97      0.97      6138
weighted avg       0.97      0.97      0.97      6138

ROC AUC XGB=  0.9708638185742718
Confusion Matrix XGB: 
[[3004   98]
 [  81 2955]]
Accuracy XGB:  0.9708374063212772
```

Results



AIRLINE REVIEW CLASSIFICATION

In today's interconnected world, the airline industry serves as a critical catalyst for global travel and business. As air travel becomes increasingly accessible, the quality of service provided by airlines plays a pivotal role in shaping passenger experiences. This project focuses on the development of an airline review classification system using Classification models such as Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier etc.. The proliferation of social media platforms, travel websites, and online forums has given rise to a wealth of usergenerated content, including airline reviews. Extracting actionable insights from this vast pool of unstructured text data has the potential to provide airlines with valuable information for refining their services and elevating passenger satisfaction. Throughout this report, we will delve into the methodology employed to preprocess the raw text data, the process of selecting pertinent features, the training and evaluation of the classification model, and the subsequent interpretation of the obtained results.



AIRLINE REVIEW CLASSIFICATION

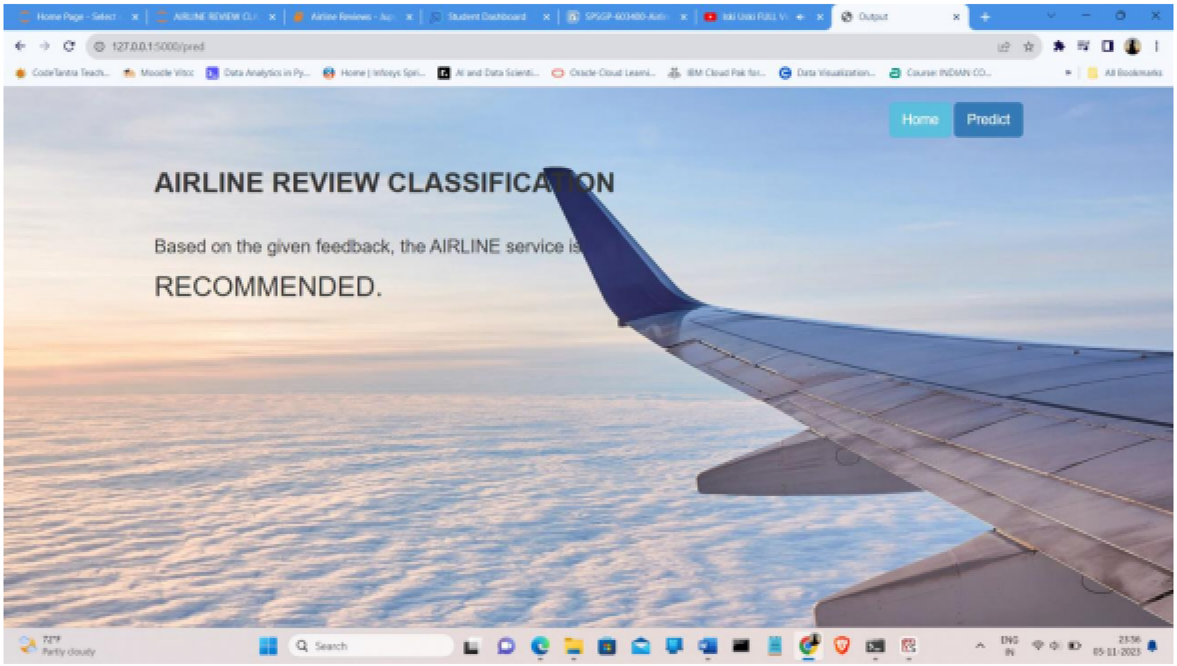
Seat Comfort
4

Food and Beverages
5

Ground Service
4

Overall Rating
4

Value For Money
2



AIRLINE REVIEW CLASSIFICATION

Based on the given feedback, the AIRLINE service is

RECOMMENDED.

GITHUB LINK

https://github.com/smartinternz02/SI-GuidedProject-603400-1699031496

DEMO Link

https://drive.google.com/file/d/1HXf0kMcoRRbYShC-QGNkszpjpL2pVPDh/view?usp=sharing