Index

1.Introduction:

Project Overview:

Flight delays are inevitable, and it plays an important role in every passenger's life. It plays a vital role in airline's growth also. For airlines, estimating flight delays correctly is essential since the data may be used to boost client happiness and revenue for airline agencies. There have been many Machine Learning models which are trained for predicting flight delays, where most of them have been trying to predict the delay through extracting important characteristics and most related features. These character extraction is done with different ML algorithms. One such algorithm is Decision Tree, which is used in our project. A Webpage is created for the user interaction to feed in the flight details. These data are then send to pretrained ML model in the cloud, where the model predicts the delay time of the flight and then send back the output to the user machine. Now the user will be provided with flight delay details.

Purpose:

Our project is all about developing a user-friendly webpage which allows user to interact and know about the delay time of the flight they are about to travel.

This allows the users not to lose their precious time and spend it efficiently on other stuffs.

2. LITERATURE SURVEY:

Existing Problem:

Due to its quickness and, in some circumstances like comfort, air travel has become more and more popular with tourists and travelers around the world. The result has been a spectacular increase in land traffic and air traffic. Massive levels of aircraft delays on the ground and in the air have also been brought on by an increase in air traffic. These delays are responsible for large economic and environmental losses to the globe. The passengers are sick of waiting for flight to arrive or for the departure.

Problem statement:

The main objective is to develop a Machine learning model to optimize flight operations and reduce delays. The model's primary goal is to estimate flight delays accurately. This allows the passengers to gain knowledge about the flight delay time. This will help the passengers to spend their time on some other stuffs.

SNO	LITERATURE PAPER	AUTHOR	PROPOSED METHOD	ACCURACY	YEAR
1	A Review on Flight Delay Prediction	Alice Sternberg Jorge Soares Eduardo Ogasawara Diego Carvalho	The methods commonly used include k-Nearest Neighbor, neural networks, SVM, fuzzy logic, and random forests. They were mainly used for classification and prediction.	87.5%	2021
2	Predicting flight delay based on multiple linear regression	Yi Ding	A prediction model based on the multiple linear regression model, was used to predict the delay	80%	2017
3	A Multilayer Perceptron Neural Network with Selective-Data Training for Flight Arrival Delay Prediction	Hajar Alla Lahcen Moumoun Youssef Balouki	Multilayer Perceptron Neural Network	MLP-95.8% ANN-91%	2021

4	Flight delay prediction model based on dual-channel convolutional neural network	WU Renbiao, LI Jiayi, QU Jingyi	Flight delay prediction model based on Dual-Channel Convolutional Neural Network.	92.1%	2018
5	Airport flight delay prediction based on SVM regression	HE Yang, ZHU Jinfu, ZHOU Qinyan	Multivariate linear regression model and SVM regression model are applied to test the current model.	93.2%	2017

3. IDEATION & PROPOSED SOLUTION:

Empathy Map Canvas:

Ideation & Brainstorming:

S. No.	Ideas
1	We can have a decision tree based on the dataset and infer on what will happen for the current day
2	This application can be a web-app so that the users can access the application without any downloading and can use in all devices
3	If the dataset has information about the delay, we can have an advanced application where current scenario can influence the result.

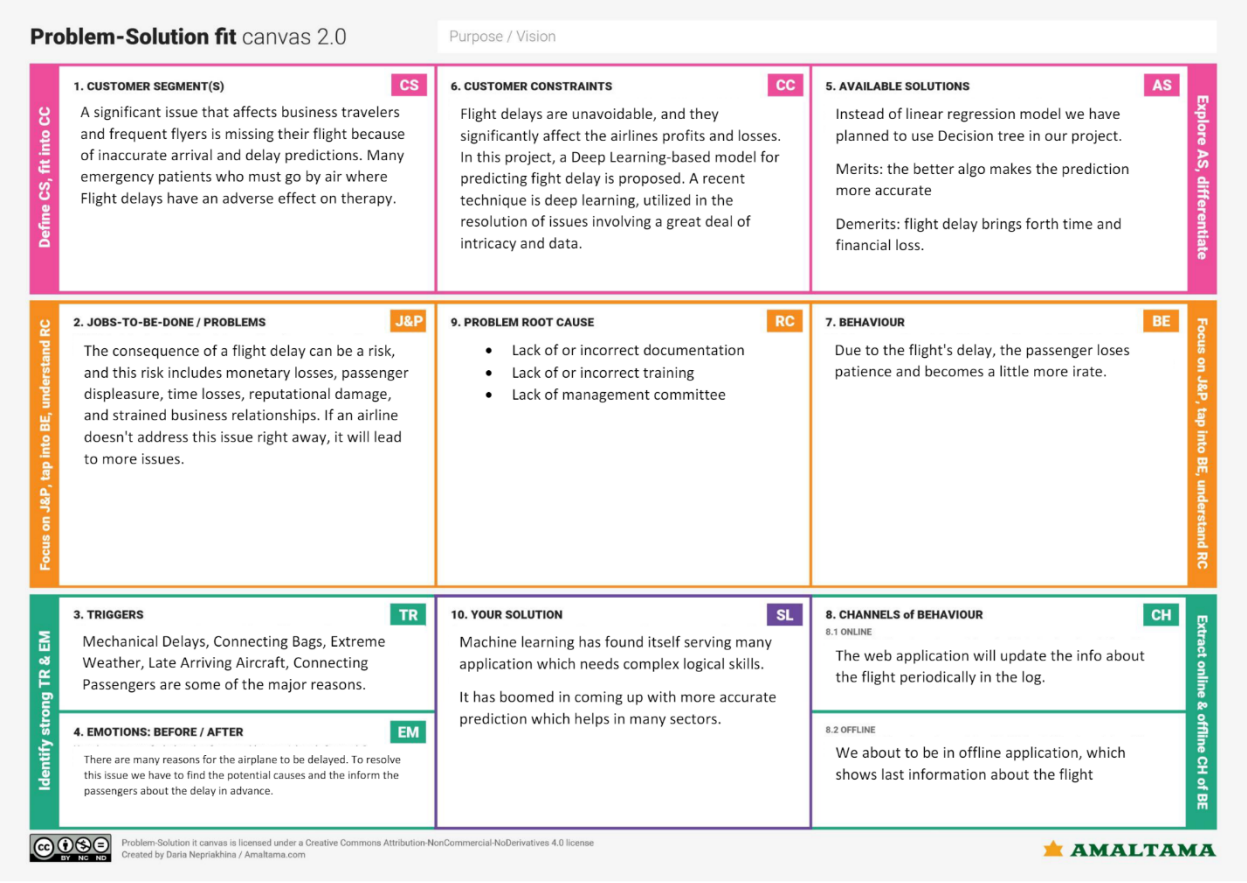
Proposed Solution:

S.No.	Parameter	Description

1	Problem Statement (Problem to be solved)	Developing A Flight Delay Prediction Model Using Machine Learning.
2	Idea / Solution description	The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize the aftereffects that caused by the delay. Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when difference between scheduled and actual arrival times is greater than 10 minutes
3	Novelty /Uniqueness	Delay Detection using simple decision tree algorithm
4	Social Impact / Customer Satisfaction	By predicting the flight delay with more accuracy, the optimized results will help the passengers by alerting them, which will not lead them to miss the flight or helps them to prepare for the worst-case scenario. In the case of the medical field, if a doctor misses a flight, it can result in a life-or-death scenario. Our project helps them to stay aware of their flights.

5	Business Model (Revenue Model)	Business to Consumer model The solution is a low-cost airline model planned to be created as an application with which the consumers can interact directly to know the details of their flight. It follows a non-monetary revenue model where the consumers aren't charged for what they get but are asked to provide their flight details and ratings which can be used to improve the model and shared with the airline in return for airline's flight data.
6	Scalability of the Solution	The present solution is drafted with the aim of experimenting with airlines based out of the United States of America. If there is a possibility to acquire data of a broader region (say North America, other continents), then the solution can be developed to benefit a wider range of people.

Problem Solution fit:



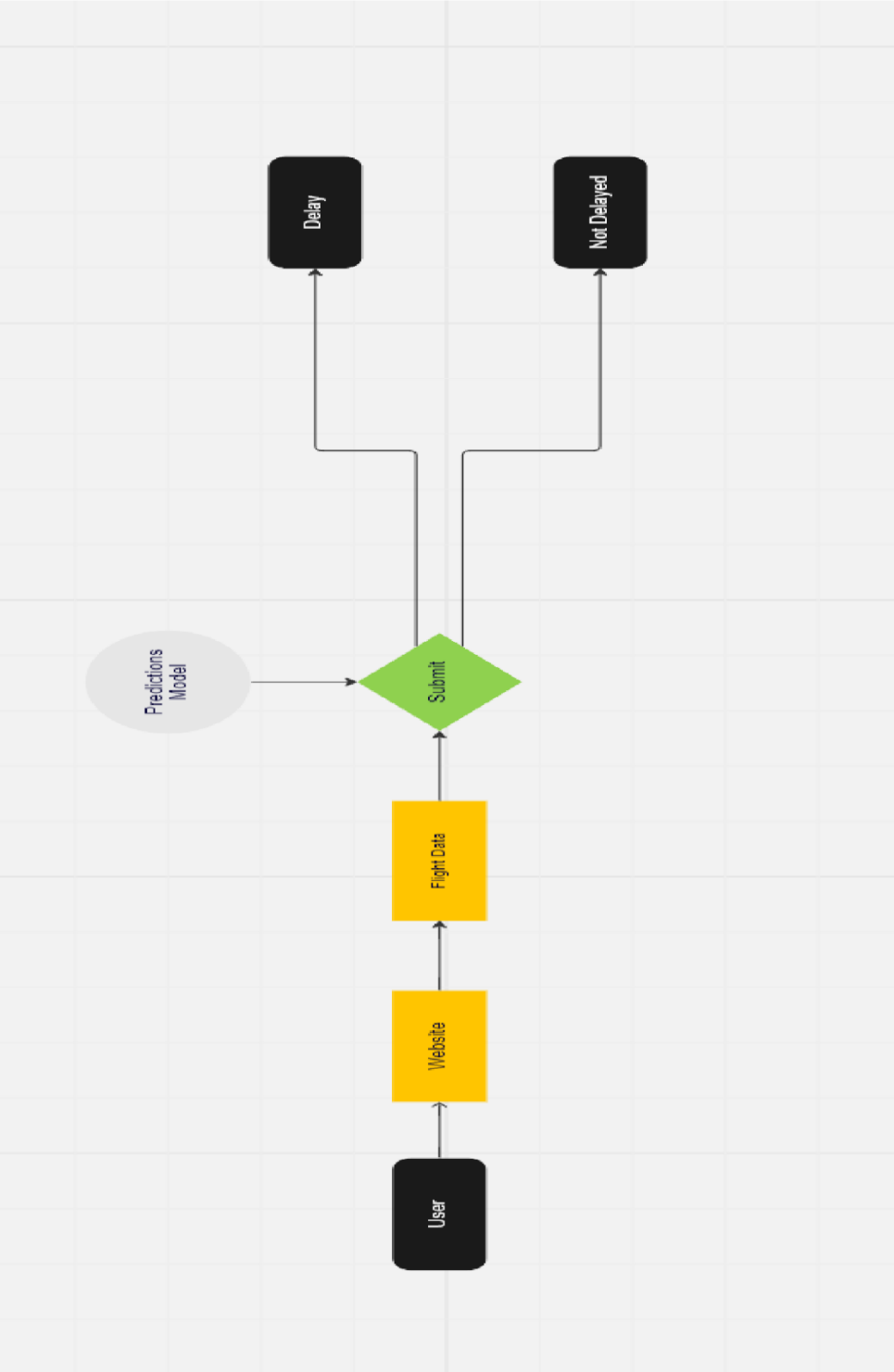
4. EQUIREMENT ANALYSIS:

Functional Requirements	Actions
Flight Search Validation	Check whether the given flight id is valid flight id present in the data
Prediction Algorithm	Train a model using a given dataset for predicting whether the flight will be delayed or not
Flight Data	Getting Flights data from the user and from the airlines
Data Verification	Verifying whether the given data is valid
Algorithm Used	Decision tree algorithm is used to train the model to predict whether the flight will be delayed or not

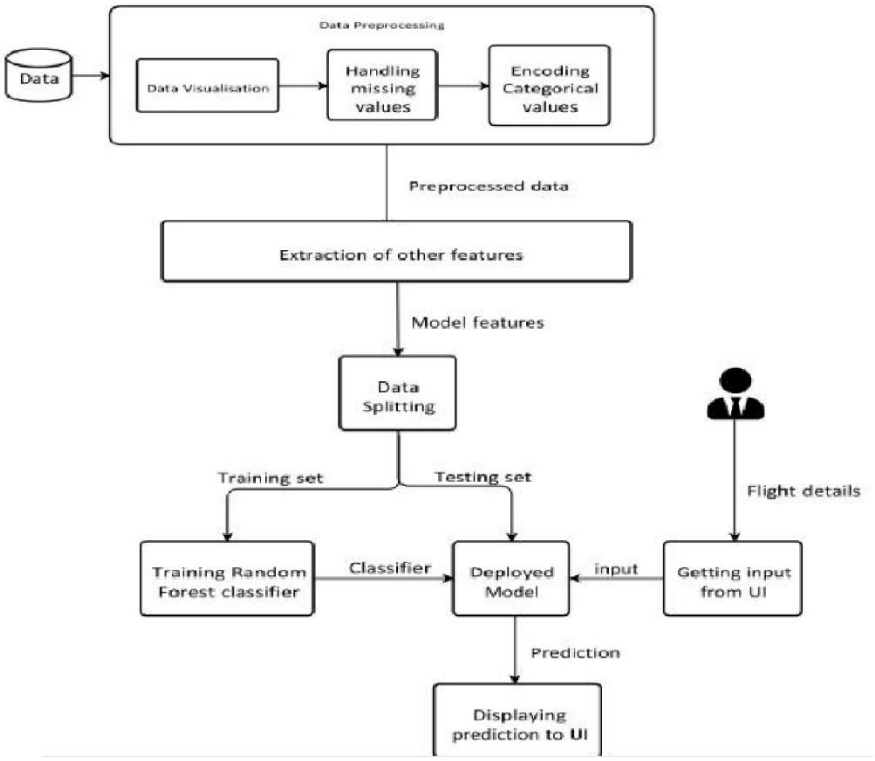
Technical Requirement:

S.No	Component	Description	Technology
1.	User Interface	Ways through which the user interacts with the application.	HTML, CSS, JavaScript
2.	Application Logic	Prediction model to predict the delay	Python
3.	Database	Stores the flight	MySQL

5. Project Design Phase:
Data Flow Diagram:



Solution Architecture:



User Stories:

User Type	Functional Requirements	User Story / Task	Priority
Customer	Provide Flight Data	As a Customer, I can give flight data to check whether the flight will be delayed or not	High
Admin	Model Creations	As an Admin, I can create a model to predict whether the flight will be delayed or not.	High
Admin	Model Updation	As an Admin, I can update the model with newer dataset to improve the performance of the model.	Medium

Customer	Website Access	As a customer, I can access the website that is used to check whether the flight will be delayed or not.	High

6.Project Planning and Scheduling:

Scheduling:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I must be able to enter the Flight details and check delay time.	2	High	DHANUSH LOGESH
Sprint-1	Interactive	USN-2	The webpage is build with CSS and HTML in a interactive way.	1	High	NISHOK KRISH
Sprint-2	Model Creation	USN-3	A Delay prediction model is created and the application is connected to IBM Cloud network.	2	Low	DHANUSH LOGESH

Planning

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)

Sprint-1	20	6 Days	24 Oct 2022	Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	7 Nov 2022	Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	Nov 2022	20	19 Nov 2022

Scheduling:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I must be able to enter the Flight details and check delay time.	2	High	DHANUSH LOGESH
Sprint-1	Interactive	USN-2	The webpage is build with CSS and HTML in a interactive way.	1	High	NISHOK KRISH
Sprint-2	Model Creation	USN-3	A Delay prediction model is created and the application is connected to IBM Cloud network.	2	Low	DHANUSH LOGESH
Sprint-3	Cloud	USN-4	Flask is built in order to integrate the model with the Webpage.	1	High	NISHOK KRISH
Sprint -4	Dashboard	USN-4	Linking the model and HTML webpage with the built Flask library.	2	High	DHANUSH KRISH

Velocity:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

7.Coding and Solution:

Code:

Sprint 3:

```
from flask import Flask , request , render_template , flash
import numpy as np
import os
import pandas as pd
import utils
app= Flask(__name__, template_folder='templates',
static_folder='static/css')
app.config.update(SECRET_KEY=os.urandom(24))
@app.route('/predict' , methods=['GET', 'POST'])
def predict():
    result = ""
    err = 0
    flightNumber = request.form.get("flightNumber")
    availableFlightNumbers = utils.getFlightNumbers()
    if int(flightNumber) not in availableFlightNumbers:
        err = 1
        return render_template('home.html', result = "Invalid Flight Number")
    date = request.form.get("date")
    formattedArray = utils.convertDateToFormat(date)
    month = formattedArray[0]
```

```
    dayOfMonth = formattedArray[1]
    dayOfWeek = formattedArray[2]
    origin = request.form.get("origin")
    if (origin == "ATL"):
        origin = 0
    elif (origin == "DTW"):
        origin = 1
    elif (origin == "JFK"):
        origin = 2
    elif (origin == "MSP"):
        origin = 3
    elif (origin == "SEA"):
        origin = 4
    else:
        origin = 5
    if origin > 4 :
        err = 1
        return render_template('home.html', result = "Please enter a valid
origin airport")
    dest = request.form.get("destination")
    if (dest == "ATL"):
        dest = 0
    elif (dest == "DTW"):
```

```
        dest = 1
    elif (dest == "JFK"):
        dest = 2
    elif (dest == "MSP"):
        dest = 3
    elif (dest == "SEA"):
        dest = 4
    else :
        dest = 5


    if dest > 4 :
        err = 1
        return render_template("home.html", result = "Please select a valid
destination airport")
    depDelay = request.form.get("depDelay")
    if (depDelay == "yes"):
        depDelay = 1
    else:
        depDelay = 0
    scheduledArrivalTime = request.form.get("scheduledArrivalTime")
    scheduledArrivalTime = int(str(scheduledArrivalTime).split(":")[0])


    #Load the model
```

```
    print(flightNumber , month , dayOfMonth , dayOfWeek , origin , dest,
scheduledArrivalTime, depDelay)

    if err == 0:

        result = utils.get_prediction(flightNumber , month , dayOfMonth ,
dayOfWeek , origin , dest, scheduledArrivalTime, depDelay)

        if (int(result) == 0):

            print("The Flight will be on time")

            msg = "The Flight will be on time"

        elif (int(result) == 1):

            print("The Flight will be delayed")

            msg = "The Flight will be delayed"

        return render_template('home.html' , result = msg)


@app.route('/', methods=['POST' , 'GET'])

def home():

    return render_template('home.html')


# predict()

if (__name__ == '__main__'):

    app.run(debug=True)
```

Sprint 4:

```
import pandas as pd

def convertDateToFormat(date):
    date = datetime.strptime(date, '%Y-%m-%d')
    dayofWeek = date.weekday()
```

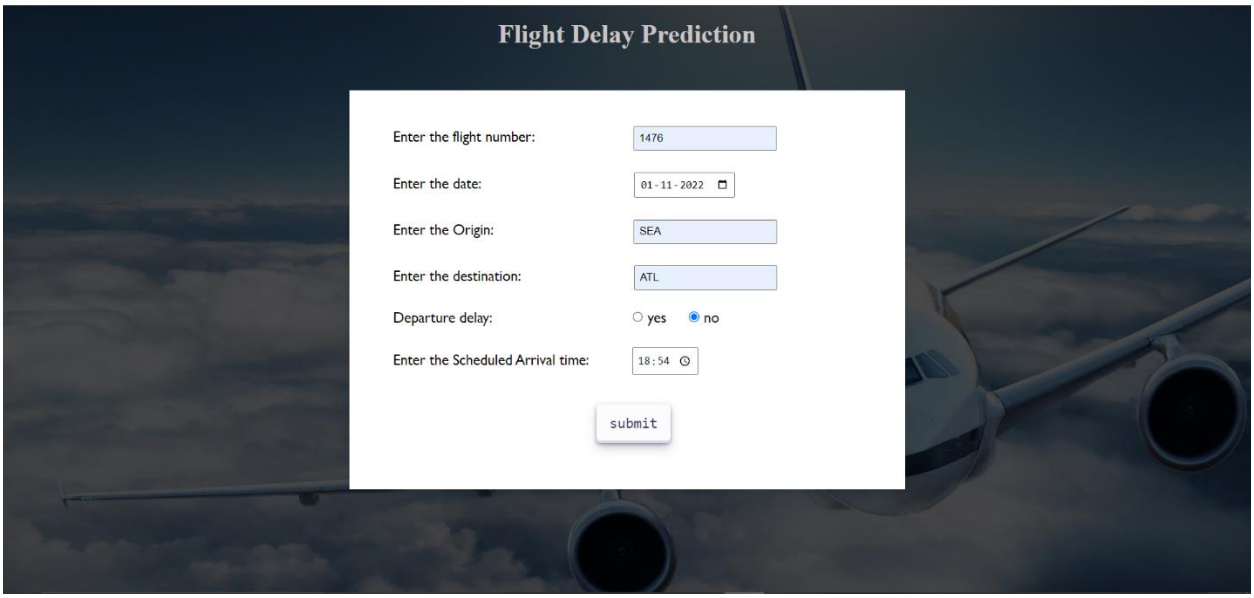
```
    month=date.month
    dayOfMonth=date.day

    return [month, dayOfMonth, dayofWeek]


def getFlightNumbers():
    df = pd.read_csv("flightData.csv")
    flightNumbers = df['FL_NUM'].unique()
    return flightNumbers

def get_prediction(flightNumber = 39, month =12 , dayofMonth = 9,
dayofWeek = 5 , origin = 1 , dest = 3, scheduledArrivalTime = 12, depDelay
= 0 ):
    values = [[flightNumber, month, dayofMonth, dayofWeek, origin, dest,
scheduledArrivalTime, depDelay]]


    # NOTE: you must manually set API_KEY below using information
retrieved from your IBM Cloud account.
    API_KEY = "kr_VNpRFqz0Mte9mcgQLk25-SFN2SevdqscISORzvMzj"
    token_response =
requests.post('https://iam.cloud.ibm.com/identity/token',
data={"apikey":API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-
type:apikey'})
    mltoken = token_response.json()["access_token"]

    header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' +
mltoken}

    # NOTE: manually define and pass the array(s) of values to be scored in
the next line
    payload_scoring = {"input_data": [{"field":
["FL_NUM","MONTH","DAY_OF_MONTH","DAY_OF_WEEK","ORIGIN","D
EST","CRS_ARR_TIME","DEP_DEL15","ARR_DEL15"], "values": values}]}

    response_scoring = requests.post('https://us-
south.ml.cloud.ibm.com/ml/v4/deployments/d68c1881-d3af-408c-a068-
e97db487099a/predictions?version=2022-11-04', json=payload_scoring,
    headers={'Authorization': 'Bearer ' + mltoken})
```

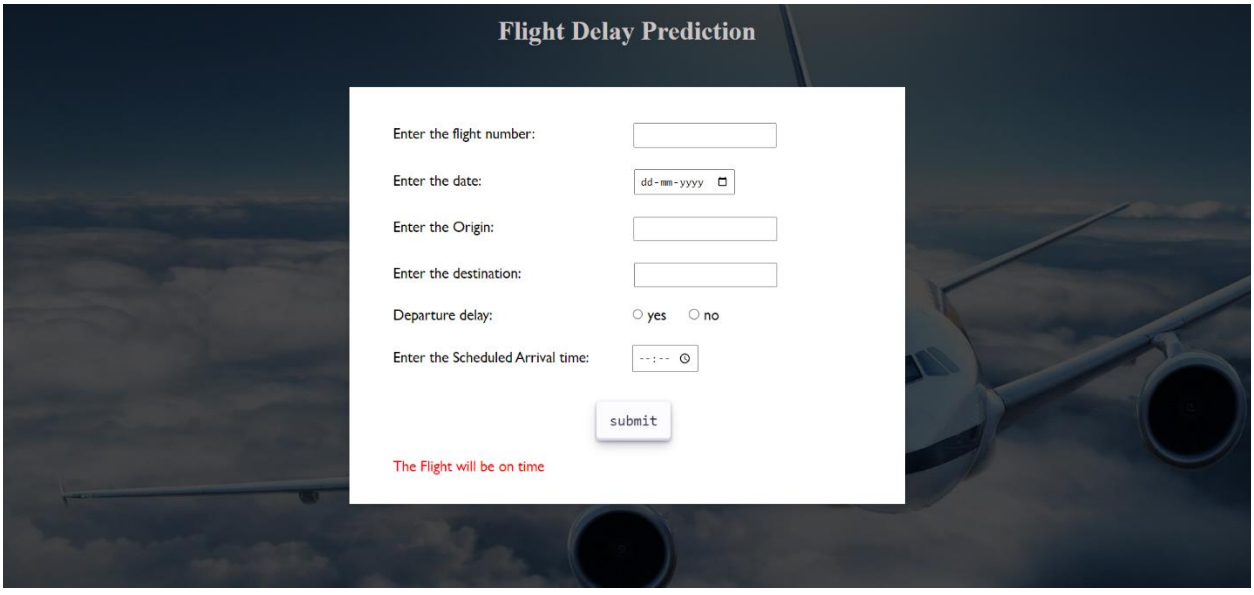
```
predictionResult = response_scoring.json()
print(predictionResult)
result = predictionResult["predictions"][0]["values"][0][0]

return result
```

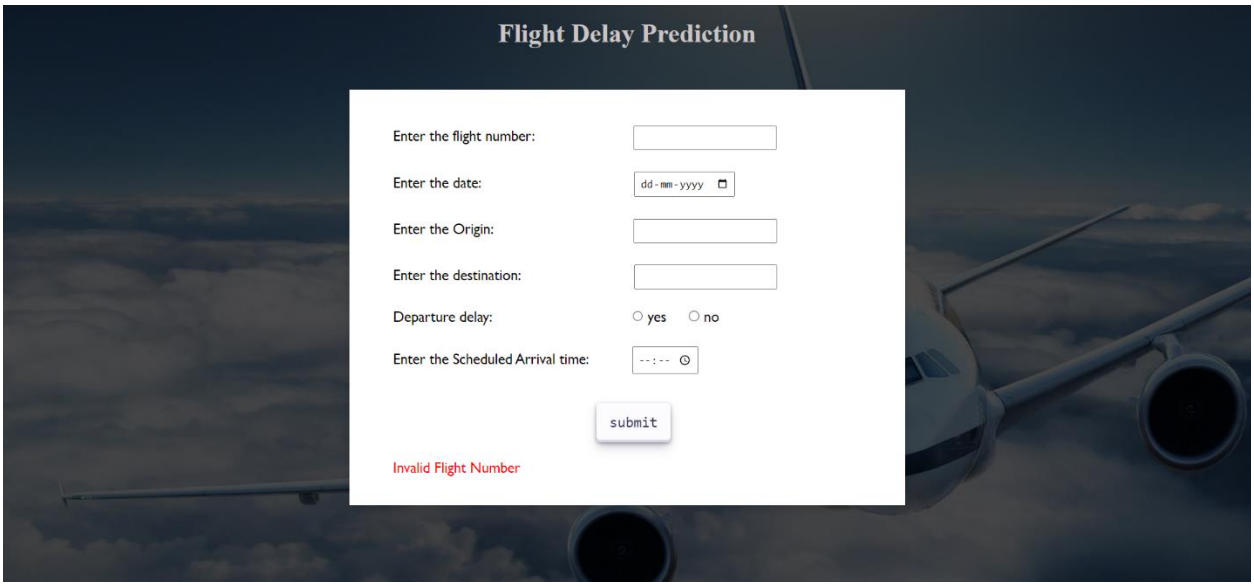
8.Results:

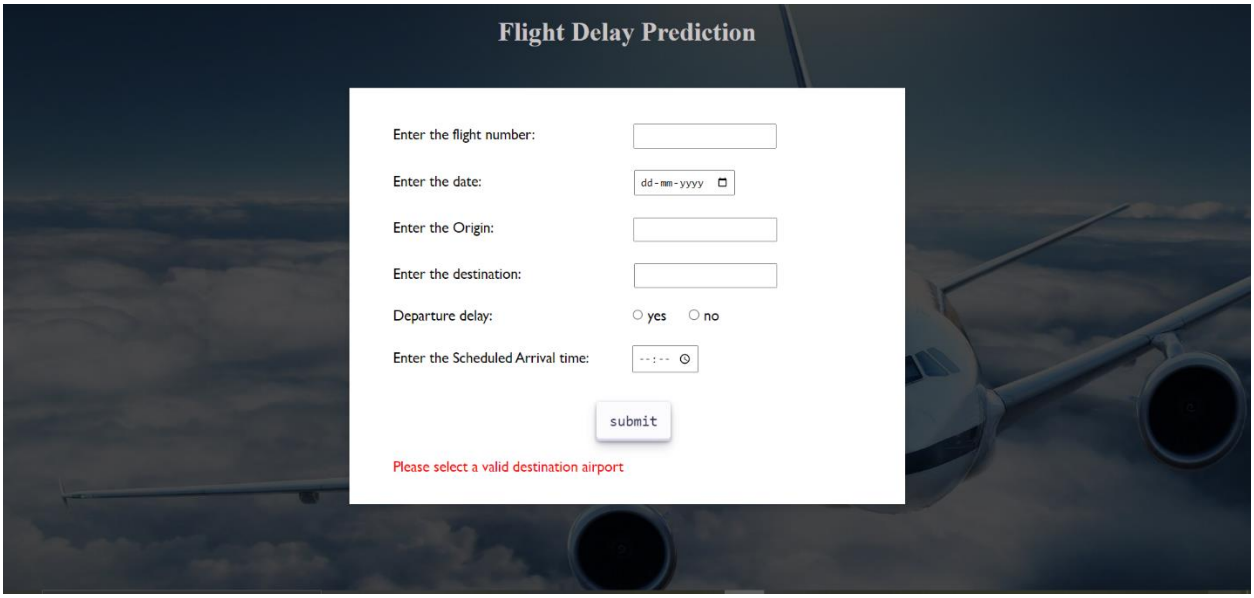
The model is created in such a way that the user finds it easy to interact with and gain the required output for the imparted input.

Output:

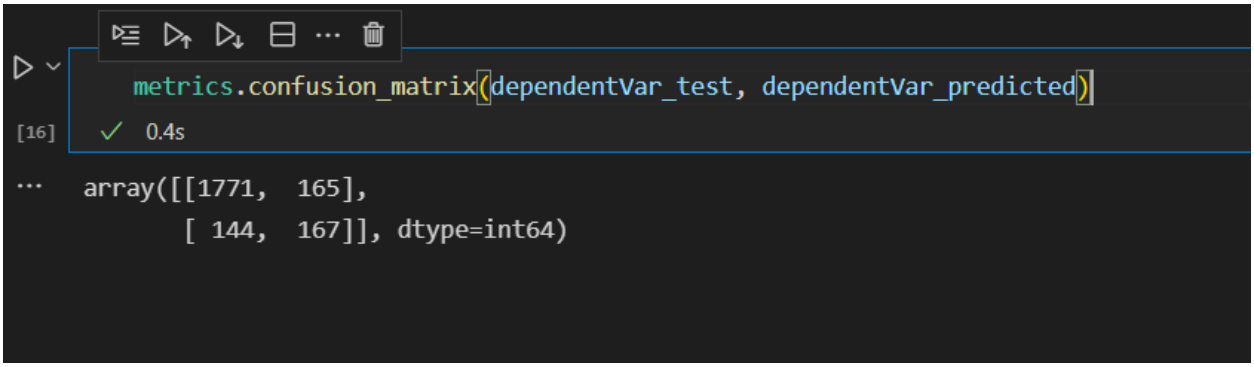





Flight Delay Prediction
Enter the flight number:
Enter the date:
dd-mm-yyyy
Enter the Origin:
Enter the destination:
Departure delay:
yes
no
Enter the Scheduled Arrival time:
--:--
submit
Invalid Flight Number


Flight Delay Prediction
Enter the flight number:
Enter the date:
dd-mm-yyyy
Enter the Origin:
Enter the destination:
Departure delay:
yes
no
Enter the Scheduled Arrival time:
--:--
submit
Please select a valid destination airport

Performance Metrics:



	Positive (1)	Negative (0)
Positive (1)	1771(True Positive)	165(False Positive)
Negative (0)	144(False Negative)	167(True Negative)

9. Advantages and disadvantages:

Advantages:

Accuracy:

This Machine Learning works at its best to predict delay time of flights with an impulsive accuracy rate of 89.4%.

Quick:

Since this model is already trained and running on a cloud platform, the model can quickly process the data and come up with an answer.

Less resource:

Since the computation is done on the cloud, less user computation power is required.

Disadvantages:

Internet:

The model is running on a cloud platform which means the user needs to connect to internet to interact with the application.

10. Conclusion:

Although the flight delay are predicted in advance based on record, some natural causes such as climate change, machine malfunction, etc. These causes will stand in our way and are quite problematic.

The developed ML model will predicts the flight delay with at most accuracy. The Decision Tree algorithm is used to boost up the performance of the model. This algorithm will enhance the prediction and give a accurate results to the users.

11. Future scope:

Flight delay prediction application can be integrated with some other applications like ticket booking system, travel map as an API. We can also have a dashboard for the users and track there frequent traveling path, flight and predict the delay time. We can also gain data from users to update our model periodically.

User Interface can be improved in such a way that it is more interactive. Many addition attributes can be added along with the model to enhance user experience.

12.Appendix:

GitHub and Source code link: https://github.com/IBM-EPBL/IBM-Project-25791-1659973269/tree/main/Project%20Development%20Phase