Developing A Flight Delay Prediction Model Using Machine Learning

Project ID:PNT2022TMID24546

PRIYANKA S(TEAM LEADER)

PRIYADHARSHNINI U(MEMBER 1)

KOHILA K(MEMBER 2)

HARITHA K(MEMEBER 3)

INTRODUCTION

Project Overview

Over the last twenty years, air travel has been increasingly preferred among travellers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. An increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. These delays are responsible for large economic and environmental losses. The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays.

Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vectors like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use a decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when the difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifiers with logistic regression and a simple neural network for various figures of merit.

Purpose

By the end of this project:

- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process/clean the data using different data preprocessing techniques.
- You will be able to analyze or get insights into data through visualization.
- Applying different algorithms according to the dataset and based on visualization.
- You will be able to know how to build a web application using the Flask framework.

ADVANTAGES AND DISADVANTAGES:

- Flight delay is inevitable and it plays an important role in both profit and loss of the airlines. An accurate estimation of flight delay is critical for airlines because the results can be applied to increase customer satisfaction and incomes of airlines agencies.
- Sometimes the prediction may not be accurate.
- Wrong prediction leads to may unwanted problems.

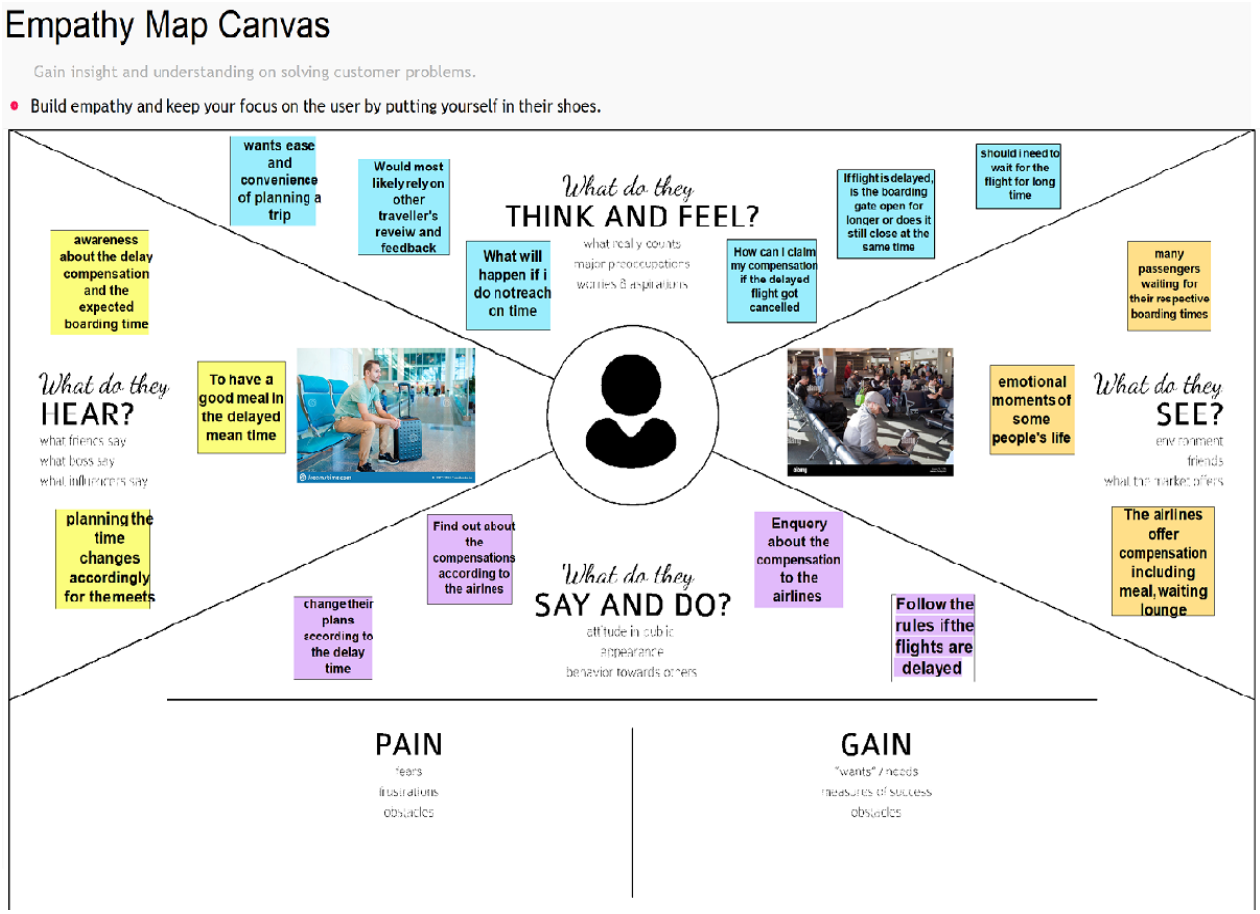
LITERATURE SURVEY

LITERATURE SURVEY				
TITLE	AUTHOR	ALGORITHM	ADVANTAGES	DISADVANTAGES
FLIGHT DELAY PREDICTION BASED ON DEEP LEARNING AND LEVENBERG-MARQUART ALGORITHM	Maryam Farshchian Yazdi, Seyed Reza Kamel , Seyyed Javad Mahdavi Chabok and Maryam Kheirabadi	DEEP LEARNING	Accuracy of the proposed prediction model analysed and compared to previous prediction method. results of three models on both imbalanced and balanced datasets shows that precision, accuracy, sensitivity, recall and F-measure of SDA-LM model with imbalanced and balanced dataset is improvement than SAE-LM and SDA models. The results also show that accuracy of the proposed model in forecasting flight delay on imbalanced and balanced dataset respectively has greater than previous model called RNN	However, this model has drawbacks of over fitting, that researchers have solved that through typical data dropout technique for each step of repeated training procedure. Moreover, application of this method decreases the computation time and memory space during the training. The next drawback is the noise of input data. However, the researcher neglects the noise during prediction.
FLIGHT DELAY PREDICTION USING MACHINE LEARNING ALGORITHM XGBOOST	Nathalie Kuhn and Navaneeth Jamadagni	MACHINE LEARNING	Different ML algorithms to predict if a flight will be delayed or not before it is even announced on the departure boards. So it will not be aiming to get the highest accuracy possible, because it would be quite easy by adding a series of features that will biased the model in terms of predictive power. So this information was looked at as part of the Exploratory Data Analysis (EDA).	The authors compare various machine learning algorithms to predict flight delays, but Failed to consider simple neural networks and decision tree classifiers. So simple machine learning algorithms like decision tree and simple neural networks to be implemented to predict flight delays, and investigate if we can predict flight delay with fewer feature-set accurately.
BIG DATA IN FLIGHT DELAY PREDICTION FOR MAINTAINING DATABASE OF THE FLIGHTS DEPARTING THE PORT	Yushan Jiang Yongxin Liu	BIG DATA	The data source and pre-processing steps Including the data merging and cleansing will be introduced. As each dataset is untidy with messy redundant records and missing values .The data used in this paper consists of two parts, Airline On-Time Performance Data (AOTP) and Quality Controlled Local Climatological Data (QCLCD) in the year 2016.It contains basic hourly airport weather data including temperature	Data is thoroughly examined for integrity criteria as well. Since expected model is to work with all the forms like offline, near line and online data, the irrelevant and unnecessary parameters that could overburden the dataset is reduced .Dropped the null values and assigned zero to Not a Number (Na N) values as one of the data cleansing activities. The data types of time factors such as scheduled time, airtime etc., are found to be in float point and proper conversion of

			extremes, visibility, air pressure, humidity and wind	input time to standard date time format. Finally, the data is analysed for distribution, converting and pre-processing. Then different datasets such the airline, flight, airports and weather datasets are integrated and normalized to identify the correlating factors that affect the flight cancellations
PREDICTION OF DELAY OF FLIGHT USING DATA MINING	L.BeLCASTRO Fabrizio marozo Domencia Talia	DATA MINING	Two open datasets of airline flights and weather observations have been collected and exploratory data analysis has been performed to discover initial insights, evaluate the quality of data, and identify potentially interesting subsets The data preparation and mining tasks have been implemented as Map Reduce programs. Other than providing the necessary computing resources for our experiments, the Cloud makes the proposed process more general. If the amount of data increases (e.g., by extending the analysis to many years of flight and weather data), the Cloud can provide the required resources with a high level of elasticity, reliability, and scalability.	The air time and flight distance would also have a greater impact on on-time performance of specific flight; Different carriers and specific aircraft would also have a slight influence of on time performance. Accuracy of this model is low because detailed weather and aircraft data could not be collected. A research analyses flight information operated by American Airlines, predicting possible arrival delay of the flight using Data Mining . Due to the imbalanced data, Over-Sampling technique, Randomized SMOTE was applied for Data Balancing. The Gradient Boosting Classifier Model was deployed by training and then Grid Search on Gradient Boosting Classifier Model on flight data, caused hyper-parameter tuned and achieving a maximum accuracy of 85.73%. Result showed that deleting some features affected the value of accuracy and reduced it..

IDEATION & PROPOSED SOLUTION

Empathy Map Canvas



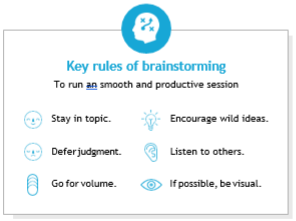
Ideation & Brainstorming

1 Define your problem statement

The passenger needs a best predicting tool to find out the flight status so," A Flight Delay Prediction" can be applied to get notified if the flight is delayed beforehand

PROBLEM

How might we create a user-friendly Prediction Tool



2 Brainstorm

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

10 minutes

Priyanka

- User-friendly
- The prediction should be comfortable for all users
- The prediction should be accurate

Priyadharshini

- Many languages should be added
- All datasets should be collected for prediction
- The Tool should have a tutorial in the beginning

Kohila

- It should be creative
- The tool should be simple to be used by all age people
- We should gather more information about the delayed flights

Haritha

- We should decide about the technology we should use
- The tool should contain the Helpline
- The Tool must be able to operate in all areas including remote

3 Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, 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

DATA

- The flight delay details inputed should be accurate
- The customer details should be verified properly

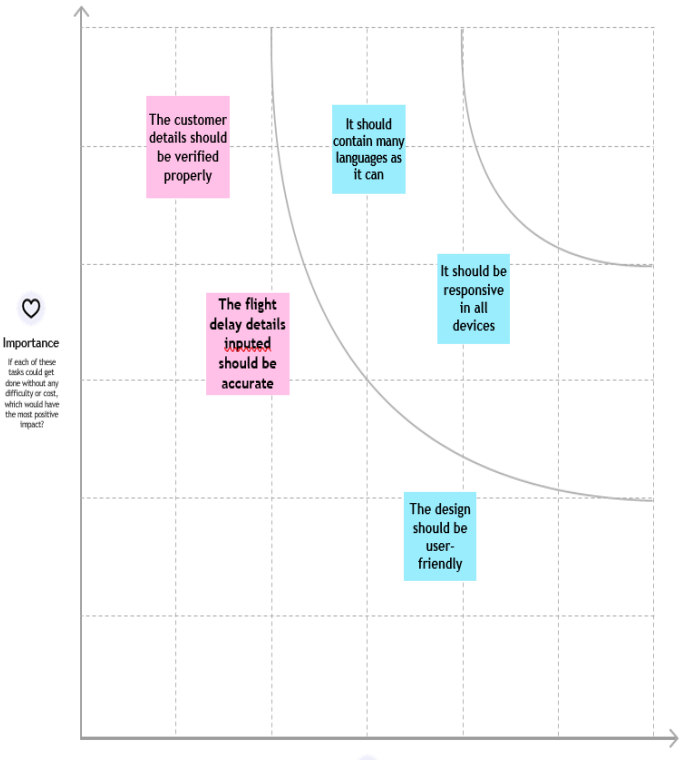
APP DESIGN

- The design should be user-friendly
- It should be responsive in all devices
- It should contain many languages as it can

4 Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

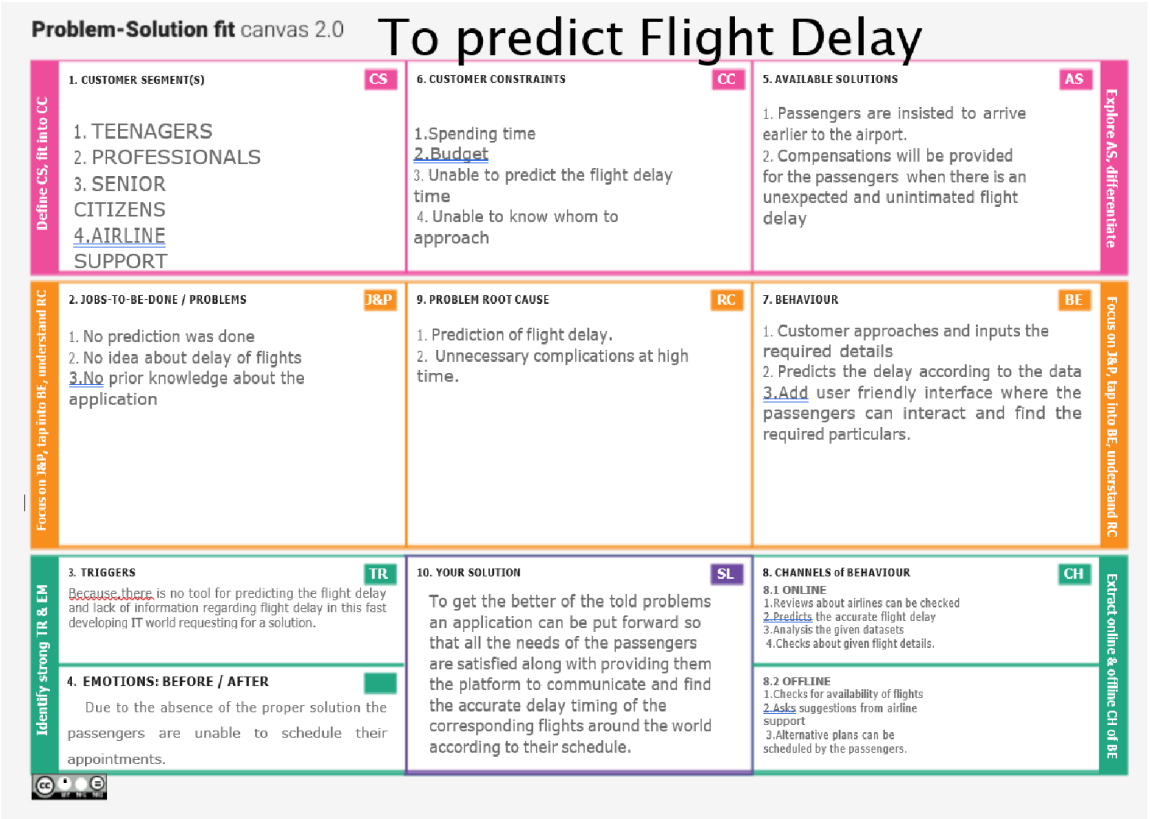


Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Flight delays cause significant financial and other losses to airlines, airports, and passengers. Their prediction is crucial during the decision-making process for all players of Aviation industry. Therefore, predicting the likelihood of delay based on flights features bridges an Important information between airlines and passengers.
2.	Idea description	The dataset is collected and identify the main Factors influencing the delays. These factors will then be combined in different ways and with different methods and will describe how to implement these data efficiently for Prediction. Finally, different methods are introduced to measure the performances of the models. These methods will be first used to optimize the models, in order to get the best possible predictions, and then will be evaluated using different use cases.
3.	Uniqueness	To solve the flight delay problems, a long short-term memory network of delay prediction with an attention mechanism is established. At present, some airport only stores their detailed indirect factors, so it is difficult to obtain information of all airports in network. At the same time, to ensure a good prediction accuracy, this model needs to be updated according to the latest data. In order to facilitate the acquisition of data and ensure the speed of training, this model predicts the flight delay by the historical information of a single airport. The direct and indirect factors are considered in this prediction software.
4.	Social Impact	Flight delays are economic, social, and environmental problems that cause inconvenience for both airline companies and passengers. Flight delays not only irritate air passengers and disrupt their schedules but also cause a decrease in efficiency, an increase in capital costs, reallocation of flight crews and aircraft, and additional crew expenses. As flight delays require the consumption of extra labour, capital, and other inputs necessary in the process, higher operating costs are inevitable for airline companies. Consumers may consider the potential for the delay before choosing to make a booking. As a result,an airline's record of flight delays may have a negative impact on passenger demand.

5.	Business Model (Revenue Model)	As far as delay is concerned there is no application that specifies the flight delay. The one that we going to develop will have a features that analyse the data given by the passengers (flight details) to predict the delay. The delay can be predicted with respect to the climatic condition, take off time and many more inputted data sets. By using this application, the passenger can able to known the delay before he/she arrives at the airport. So this may help in various situation like avoiding last minute surprises, waiting for a long time etc.
6.	Scalability of the Solution	The algorithms used here is efficient enough in terms of speed, this improves the data for further analysis which leads to high accurate Results

Problem Solution Fit



REQUIREMENT ANALYSIS

Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Flight code validation	Validation through airline database
FR-4	Delay prediction	Prediction result through Gmail Prediction result through Message
FR-5	Airline Support	Collaboration with the airlines customer support
FR-6	Algorithm for prediction	Train the data Predict the data

Non-Functional requirement

Following are the non-functional requirements of the proposed solution.

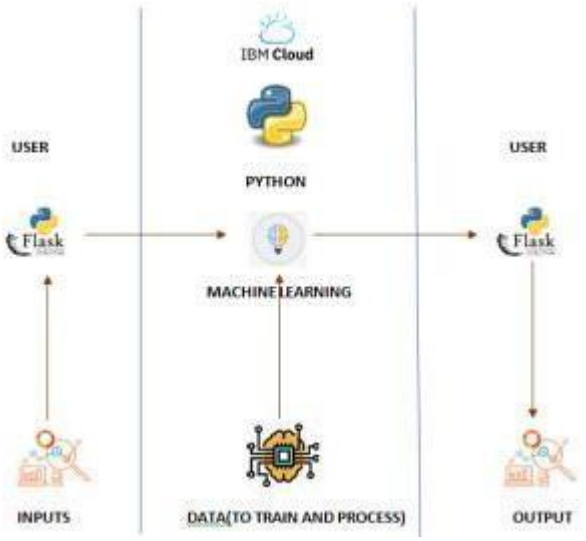
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The UX/UI will be user-friendly and will be highly Responsive
NFR-2	**Security**	The data security will be ensured by using appropriate security measures
NFR-3	**Reliability**	The application performs for all proper data Efficiently
NFR-4	**Performance**	The appliation response time will be faster and the data will be processed at a higher speed
NFR-5	**Compatibility**	The application is highly compatabile in any devices.
NFR-6	**Scalability**	The application is designed in such a way that it can handle ,if any high traffic occurs

PROJECT DESIGN

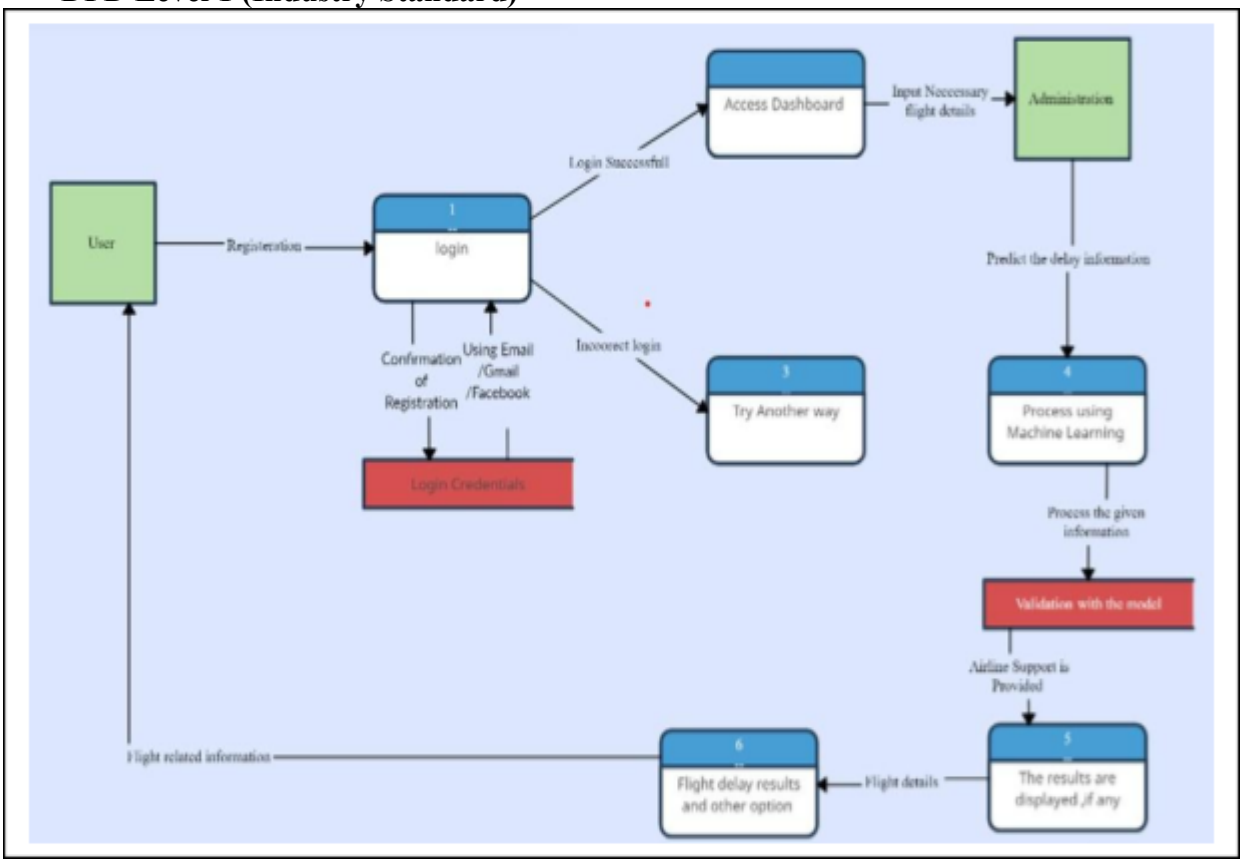
Data Flow Diagrams

Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

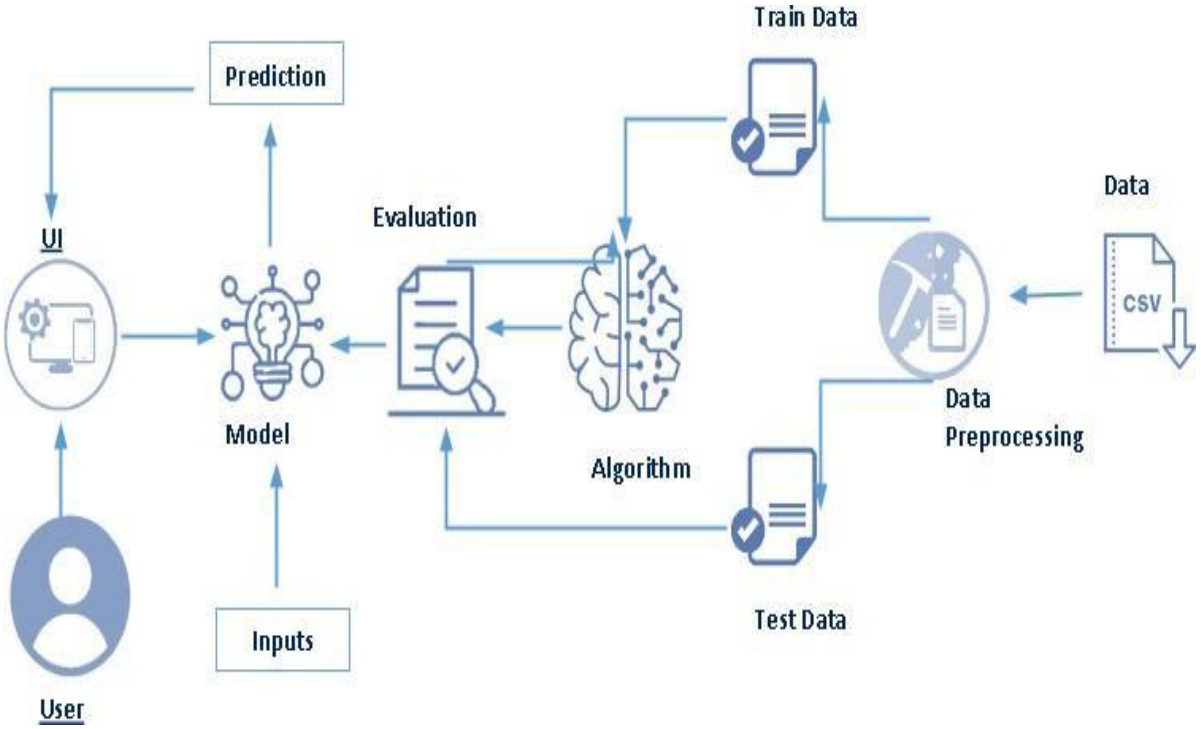
Simplified

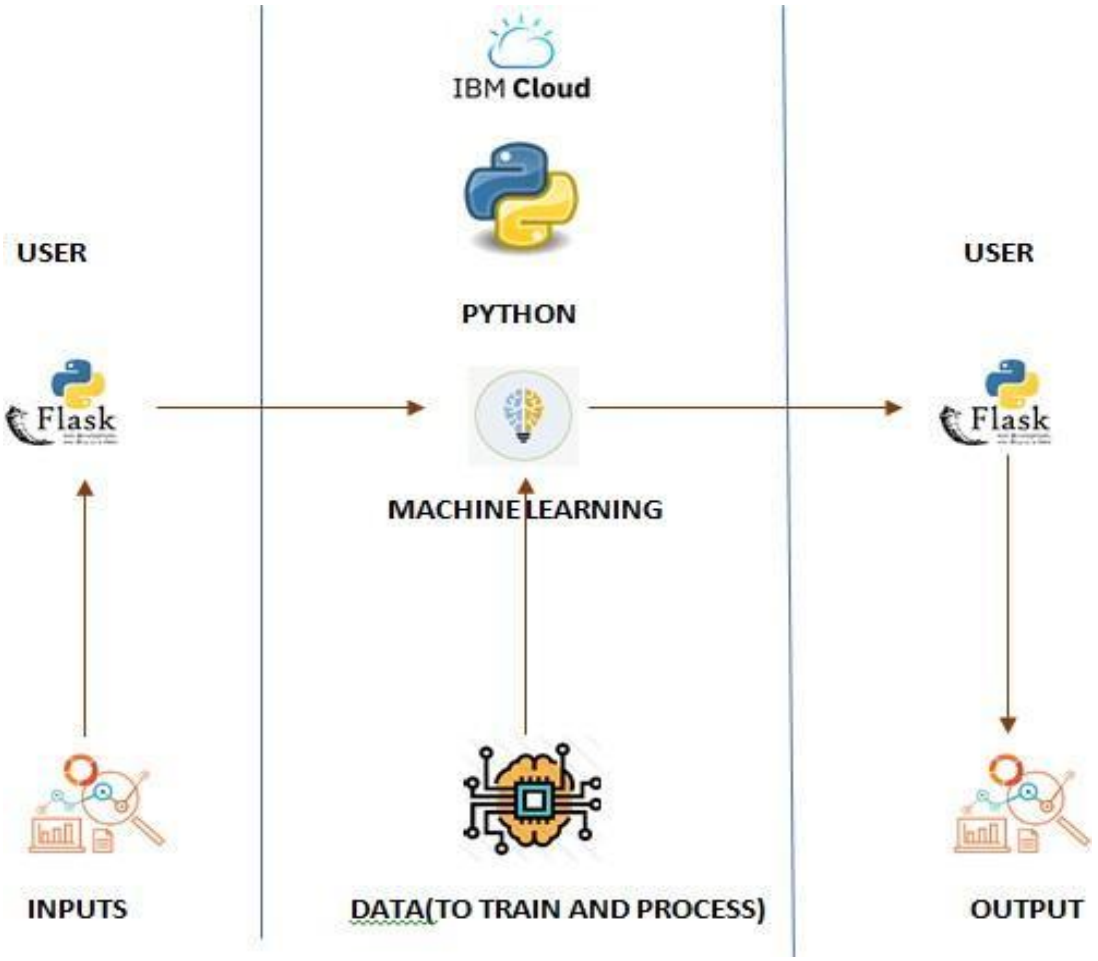


DFD Level 1 (Industry Standard)



Solution & Technical Architecture





User Stories

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	I can register through Gmail	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can login into the application by using email and password	High	Sprint-1
	Dashboard	USN-6	As a user, I can able to access the dashboard after successfully login	I can access the dashboard after login	High	Sprint-2
Customer (Web 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	I can register through Gmail	Medium	Sprint-1
		USN-5	As a user, I can log into the application by entering email & password	I can login into the application by using email and password	High	Sprint-1
Customer Care Executive	Airline Support	USN-1	As a user, I want to contact the customer care whenever I find a need	I can contact the customer care easily and services are also provided quickly	High	Sprint-3
Administrator	Data processing and prediction	USN-1	As a user, I want to predict and analysis the data correctly to provide the results	I can receive the results accurately and precisely	High	Sprint-4

PROJECT PLANNING & SCHEDULING

Sprint Planning & Execution:

Sprint	Functional Requirement (EPIC)	User Story Number	User Story/Task	Story Points	Priority	Team Members
Sprint -1	REGISTERATION	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	high	3
Sprint -1		USN-2	As a user, I will receive a confirmation	1	high	1

			email once I have registered for the application			
Sprint -2		USN-3	As a user, I can register for the application through Facebook	2	low	1
Sprint -1		USN-4	As a user, I can register for the application through google account	2	high	1
Sprint -1	LOGIN	USN-5	As a user, I can log into the application by entering email & password	1	high	1
Sprint -1	DASHBOARD	USN-6	As a customer I can check with all the Flight details available on the website and choose the correct prediction	3	high	3
Sprint -4	Customer support	USN-7	As a user I want to contact with the customer support when there is any query with the the application	2	low	2
Sprint -1	User details display	USN-8	As a customer I should be able to see all my given details filled with the registration process	2	high	2
Sprint -2	algorithm	USN-9	As a customer, I should be able to get the accurate prediction of the flight details available on the website	5	high	4
Sprint -2		USN-10	As a customer I should be updated with various best available.	2	medium	2

Sprint -3	IBM watson for storage and organization	USN-11	As a customer I should be able to give all the necessary details and predict the flight arrival and provide best user experience.	3	high	2
Sprint -4	Flight Management	USN-12	As a customer I can manage the details of the flight arrival timings and schedule changes.	5	high	4

Sprint Delivery Schedule:

Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date**	**Story Points Completed (as on Planned End Date)**	**Sprint Release Date (Actual)**
Sprint -1	20	6 Days	24 Oct 2022	29 Oct 2022		
Sprint -2	20	6 Days	31 Oct 2022	05 Nov 2022		
Sprint -3	20	6 Days	07 Nov 2022	12 Nov 2022		
Sprint -4	20	6 Days	14 Nov 2022	19 Nov 2022		

CODING AND SOLUTIONING:

Feature 1:

The application for the project

index.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta http-equiv="X-UA-Compatible" content="IE=edge">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
        <link  href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-Zenh87qX5JnK2Jl0vWa8Ck2rdkQ2Bzep5IDxbcnCeuOxjzrPF/et3URy9Bv1WTRi" crossorigin="anonymous">
   <link rel="stylesheet" href="style.css">
   <title>Project</title>
</head>
<body>
   <section class="image ">
      <h1 class="text text-center">Prediction of Flight Delay</h1>
      <form class="container form p-4 mt-5">
         <div class="container text-center">
            <div class="row">
               <div class="col">
                        <label class="form-label me-4" for="input1">Enter the Flight Number</label>
               </div>
               <div class="col">
                  <input type="text" class="input" id="input1" placeholder="Flight Number">
               </div>
            </div>
            <div class="row mt-3">
```

```
    <div class="col">
        <label class="form-label me-5" for="input2">Month</label>
    </div>
    <div class="col">
        <input type="text" class="input" id="input2" placeholder="Month">
    </div>
</div>

<div class="row mt-3">
    <div class="col">
        <label class="form-label me-4" for="input3">Day of Month</label>
    </div>
    <div class="col">
        <input type="text" class="input" id="input3" placeholder="Day of Month">
    </div>
</div>

<div class="row mt-3">
    <div class="col">
        <label class="form-label me-4" for="input4">Day of Week</label>
    </div>
    <div class="col">
        <input type="text" class="input" id="input4" placeholder="Day of Week">
    </div>
</div>

<div class="row mt-3 ms-1">
    <div class="col">
        <label class="form-label me-5" for="drop1">Origin</label>
    </div>
```

```
            <div class="col ms-4">
                    <select class="form-select select text-center" aria-label="Default select example">
                <option selected>MSP</option>
                <option value="1">One</option>
                <option value="2">Two</option>
                <option value="3">Three</option>
              </select>
            </div>
          </div>

          <div class="row mt-3 ms-1">
            <div class="col">
              <label class="form-label me-5" for="drop2">Destination</label>
            </div>
            <div class="col ms-4">
              <div class="dropdown">
                    <select class="form-select select text-center" aria-label="Default select example">
                <option selected>MSP</option>
                <option value="1">One</option>
                <option value="2">Two</option>
                <option value="3">Three</option>
              </select>
            </div>
            </div>
          </div>

          <div class="row mt-3">
            <div class="col">
```

```
                        <label class="form-label me-4" for="input5">Scheduled Departure Time</label>
                    </div>
                    <div class="col">
                        <input type="text" class="input" id="input5" placeholder="Scheduled Departure Time">
                    </div>
                </div>

                <div class="row mt-3">
                    <div class="col">
                        <label class="form-label me-4" for="input6">Scheduled Arrival Time</label>
                    </div>
                    <div class="col">
                        <input type="text" class="input" id="input6" placeholder="Scheduled Arrival Time">
                    </div>
                </div>

                <div class="row mt-3">
                    <div class="col">
                        <label class="form-label me-4" for="input7">Actual Departure Time</label>
                    </div>
                    <div class="col">
                        <input type="text" class="input" id="input7" placeholder="Actual Departure Time">
                    </div>
                </div>
                <button class="btn btn-warning mt-3">Submit</button>

            </div>
        </form>
```

```
    </section>



     <script src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.11.6/dist/umd/popper.min.js" integrity="sha384-oBqDVmMz9ATKxIep9tiCxS/Z9fNfEXiDAYTujMAeBAsjFuCZSmKbSSUnQlmh/jp3" crossorigin="anonymous"></script>

<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.min.js" integrity="sha384-IDwe1+LCz02ROU9k972gdyvl+AESN10+x7tBKgc9I5HFtuNz0wWnPclzo6p9vxnk" crossorigin="anonymous"></script>

</body>

</html>
```

style.css:

```
.image{
                                                                    background-image: url("https://flightsimulator-storage-cdn.azureedge.net/videos/hero-21.9.jpg");
    background-repeat: no-repeat;
    background-size: cover;
    width:100% ;
    color: azure;
    padding-bottom: 4%;


}
.text{
    font-family: Georgia, 'Times New Roman', Times, serif;
    font-weight: 800;
    color: azure;
}


.form{
    margin: 0 auto;
    border: 1px solid black;
    box-shadow: white 2px 2px 7px 4px;
```

```
    width: 50%;
}

.input{
    text-align: center;
    outline: none;
}
.input:hover{
    box-shadow: black 4px 2px 1px 1px;
}
.select{
    width: 40%;
}
```

app.py:

```
from flask import Flask,request,render_template
import numpy as np
import pandas as pd
import pickle
import os
model = pickle.load(open('flight.pkl','rb'))
app = Flask(__name__)
@app.route('/')
def home():
    return render_template("index.html")

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

def predict():
    name = request.form['name']
    month = request.form['month']
```

```
    dayofmonth = request.form['dayofmonth']
    dayofweek = request.form['dayofweek']
    origin = request.form['origin']
    if origin=="msg":
        origin1,origin2,origin3,origin4,origin5 = 0,0,0,0,1
    if origin =="dtw":
        origin1,origin2,origin3,origin4,origin5 = 1,0,0,0,0
    if origin =="jfk":
        origin1,origin2,origin3,origin4,origin5 = 0,0,1,0,0
    if origin =="sea":
        origin1,origin2,origin3,origin4,origin5 = 0,1,0,0,0
    if origin =="alt":
        origin1,origin2,origin3,origin4,origin5 = 0,0,0,1,0
    destination = request.form['destination']
    if destination =="map":
        destination1,destination2,destination3,destination4,destination5=0,0,0,0,1
    if destination =="dtw":
        destination1,destination2,destination3,destination4,destination5=1,0,0,0,0
    if destination =="jfk":
        destination1,destination2,destination3,destination4,destination5=0,0,1,0,0
    if destination =="sea":
        destination1,destination2,destination3,destination4,destination5=0,1,0,0,0
    if destination =="alt":
        destination1,destination2,destination3,destination4,destination5=0,0,0,1,0
    dept = request.form['dept']
    arrtime = request.form['arrtime']
    actdept = request.form['actdept']
    dept15=int(dept)-int(actdept)
                                                                                total
=[[name,month,dayofmonth,dayofweek,origin1,origin2,origin3,origin4,origin5,destination1,d
estination2,destination3,destination4,destination5,dept,arrtime,actdept,dept15]]
```

```
    y_pred = model.predict(total)
    print(y_pred)
    if y_pred==[0.] :
        ans="The Flight Will Be On Time"
    else:
        ans="The Flight Will Be Delayed"
    return render_template("index.html",showcase = ans)
```

flight prediction.py:

```
#!/usr/bin/env python
# coding: utf-8

# In[39]:



import sys
import numpy as np
import pandas as pd
import seaborn as sns
import pickle
get_ipython().run_line_magic('matplotlib', 'inline')
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
import sklearn.metrics as metrics
```

```
# In[40]:

pwd

# In[41]:

import os, types

import pandas as pd

from botocore.client import Config

import ibm_boto3


def __iter__(self): return 0


# @hidden_cell

# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.

# You might want to remove those credentials before you share the notebook.

cos_client = ibm_boto3.client(service_name='s3',

    ibm_api_key_id='YrQegtEAFTtgFLwpCLbzHKmafzrTqvXux9ly58CLUtL4',

    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",

    config=Config(signature_version='oauth'),

    endpoint_url='https://s3.private.us.cloud-object-storage.appdomain.cloud')


bucket = 'flightdelayprediction-donotdelete-pr-6pklnqsuhe0vn6'

object_key = 'flightdata.csv'


body = cos_client.get_object(Bucket=bucket,Key=object_key)['Body']

# add missing __iter__ method, so pandas accepts body as file-like object

if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )


dataset = pd.read_csv(body)
```

```
# In[42]:

dataset.head()

# In[43]:

dataset.info()

# In[44]:

dataset.describe()

# In[45]:

dataset.isnull().sum()

# In[46]

dataset["DEST"].unique()

# In[47]:

sns.scatterplot(x='ARR_DELAY',y='ARR_DEL15',data=dataset)

# In[48]:

sns.catplot(x="ARR_DEL15",y="ARR_DELAY",kind='bar',data=dataset)

# In[49]:

sns.heatmap(dataset.corr())

# In[50]:

dataset=dataset.drop('Unnamed: 25',axis=1)

dataset.isnull().sum()

# In[51]:

dataset
=dataset[["FL_NUM","MONTH","DAY_OF_MONTH","DAY_OF_WEEK","ORIGIN","DE
ST","CRS_ARR_TIME","DEP_DEL15","ARR_DEL15"]]

dataset.isnull().sum()

# In[52]:

dataset=dataset.fillna({'ARR_DEL15':1})

dataset=dataset.fillna({'DEP_DEL15':0})

dataset.iloc[117:185]

# In[53]:
```

```
import math


for index,row in dataset.iterrows():
    dataset.loc[index,'CRS_ARR_TIME'] = math.floor(row['CRS_ARR_TIME']/100)
dataset.head()
# In[54]:
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset['DEST'] = le.fit_transform(dataset['DEST'])
dataset['ORIGIN'] = le.fit_transform(dataset['ORIGIN'])
# In[55]:
dataset.head(5)
# In[56]:
from sklearn.preprocessing import OneHotEncoder
x=dataset.values
oh = OneHotEncoder()
z=oh.fit_transform(x[:,4:5]).toarray()
t=oh.fit_transform(x[:,5:6]).toarray()
# In[57]:
z
# In[58]:
t
# In[59]:
dataset = pd.get_dummies(dataset,columns=['ORIGIN','DEST'])
dataset.head()
# In[60]:
x=dataset.iloc[:,0:8].values
y=dataset.iloc[:,8:9].values
```

```
# In[61]:
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
# In[62]:
x_test.shape
# In[63]:
y_train.shape
# In[64]:
x_train.shape
# In[65]:
y_test.shape
# In[66]:
from sklearn.preprocessing import StandardScaler
sc= StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)



# In[67]
from sklearn.tree import DecisionTreeClassifier
classifier =DecisionTreeClassifier(random_state = 0)
classifier.fit(x_train,y_train)
# In[68]:
decisiontree = classifier.predict(x_test)
# In[69]:
decisiontree
# In[70]:
from sklearn.metrics import accuracy_score
```

```
desacc = accuracy_score (y_test,decisiontree)
# In[71]:
desacc

# In[72]:
import pickle
# In[73]:
pickle.dump(classifier,open('flight.pkl','wb'))
# In[74]:
pip install ibm_watson_machine_learning
# In[75]:
from ibm_watson_machine_learning import APIClient
wml_credentials={
   "url":"https://us-south.ml.cloud.ibm.com",
   "apikey":"SIPG1xOrBleCVP2trdmcnOuwL2SD_fe9i5doB2akj1gb"
}
client = APIClient(wml_credentials)



# In[76]:
def guid_from_space_name(client,space_name):
   space=client.spaces.get_details()
                        return (next(item for item in space['resources'] if
item['entity']["name"]==spacename)['metadata']['id'])
# In[ ]:
space_uid = guid_from_space_name(client,'model')
print("Space UID = "+space_uid)
# In[ ]:
client.set.default_space(space_uid)
```

TESTING:

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

CONCLUSION:

- As flight delay cost a lot to the airlines as well as passenger in financial and environmental terms, flight delay is the talk of the hour.
- Flight delay causes surging of prices by costing a lot on operational purposes. They may increase to customers too.
- As the outcome is directly associated with the passengers and airlines which in turn it is liked to another set of passengers and airlines it is very crucial to get real time delay for each player within the air transport system, hence there is a requirement to develop a system to predict the delay in flights to scale back monetary loss and for the higher and smooth operation.

FUTURE SCOPE:

Further supportive study is required to correlate all the problem, scope, and method for getting most accurate result. Although weather conditions are the major reasons for flight delay, other unprecedented events such as major calamities, natural or manmade can cause major delay in flight.