# GOVERNMENT COLLEGE OF ENGINEERING
# (Formerly IRTT)
# ERODE-638 316

## BONAFIDE CERTIFICATE

Certified that this project titled “ **Developing a Flight Delay Prediction Model using Machine Learning”** is the bonafide work of

| | |
|---|---|
| **HARI PRASATH N** | **(731119106008),** |
| **SARATHI S S** | **(731119106029),** |
| **SHAMRAJBABU T** | **(731119106032),** |
| **VEERASURIYAN K B** | **(731119106038),** |
| **VIGNESH B S** | **(731119106039)** |

who carried out the project work under my supervision.

**SIGNATURE OF HOD**
Dr,R.VALARMATHI,M.E.,Ph.D.,
HEAD OF THE DEPARTMENT
DEPARTMENT OF ECE,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE – 638316

**SIGNATURE OF SPOC**
Dr,G.GOWRISON,M.E.,Ph.D.,
ASSISTANT PROFESSOR(SR)
DEPARTMENT OF ECE,
GOVERNMENT COLLEGE OF
ENGINEERING,ERODE - 638316

**SIGNATURE OF FACULTY MENTOR**
Dr.R.SENTHIL KUMAR.,Phd.,
ASSISTANT PROFESSOR(SR)
DEPARTMENT OF ECE,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE – 638316

**SIGNATURE OF FACULTY EVALUTOR**
Dr.S.BHARATHI.M.Tech.,Phd.,
ASSISTANT PROFESSOR(SR)
DEPARTMENT OF ECE,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE - 638316

# ABSTRACT

Flight Delay Prediction aims to predict the delay in the aircrafts due to the increasing number of travellers in recent times. An aircraft arrival is considered to be delayed if the aircraft is late over 15 minutes between the scheduled time and the arrival time. Flight Delay Prediction takes into consideration various attributes of the delay which includes scheduled time, source and destination of the flight, arrival time of the flight and departure time of the flight and many more attributes to predict the delay in the flight arrivals. These flight delays help the user massively to select the airlines, to select the source station and other economic aspects of the travellers. At the same time, Flight Delay prediction also helps the airlines to focus on the major reasons of the flight delay and minimise delay time on future occasions. The aviation industry is also benefited with the help of the Flight Delay Prediction.

TABLE OF CONTENTS

# CHAPTER 1

# INTRODUCTION

## 1. INTRODUCTION

Flight Delay Prediction aims to predict the delay in the aircrafts due to the increasing number of travellers in recent times. An aircraft arrival is considered to be delayed if the aircraft is late over 15 minutes between the scheduled time and the arrival time. Flight Delay Prediction takes into consideration various attributes of the delay which includes scheduled time, source and destination of the flight, arrival time of the flight and departure time of the flight and many more attributes to predict the delay in the flight arrivals. These flight delays help the user massively to select the airlines, to select the source station and other economic aspects of the travellers. At the same time, Flight Delay prediction also helps the airlines to focus on the major reasons of the flight delay and minimise delay time on future occasions. The aviation industry is also benefited with the help of the Flight Delay Prediction.

### 1.1 Project Overview

The flight delay prediction mostly depends on the initial steps that are being carried out. Firstly, the dataset collection. The data that is collected for the prediction must be accurate and concise in nature. Any discrepancies in the dataset will cost the accuracy of the flight delay prediction directly. The second step is Data pre-processing. The collected data is improper i.e., those data will have outliers, missing values and the number of attributes may also be huge. At times the data can also be unstructured. In order to solve this issue, the data must be cleaned and pre-processed in a proper manner. The next important issues arise with the data consistency, the flight delay data must be consistent.The time format must be the same across all the dataset. Similarly, the time zone varies from location to location. These inconsistencies must be solved before training the model with the data. The issue also occurs due to abnormalities. For example, the flight delay can be caused due to bad weather or gets cancelled due to any natural calamities. The model cannot predict the flight delays in these abnormalities.

### 1.2 Purpose

The problem majorly affects the travellers, airline and the aviation industries. The travellers have to look for alternatives in the case of delay of the expected flight arrivals. The airline agency will lose customer trust which is the most important factor. At the same time, Flights are also used to transport goods, the people as well as the organisation who are dependent upon the flight arrival will also be affected due to the flight delay. The aviation industries are also responsible for delivering good products through which the delay of the flights can be reduced. The flight delay prediction may help the aviation industry hugely to protect them from their economic and financial losses. This delay prediction can help the travellers hugely to plan ahead and save their valuable time. The cost associated with the flights can also be majorly decreased when these delays can be predicted correctly. The reputation of the airlines can be majorly dependent on these delay predictions because the delay prediction has a direct hand in determining the customer trust on the airline agency.

# CHAPTER 2

# 2.LITERATURE SURVEY

## 2.1 Existing Problem

### 2.1.1.Study of Flight Departure Delay and Analysis

**ABSTRACT**: Analysis of flight delay and causal factors is crucial in maintaining airspace efficiency and safety. However, delay samples are not independent since they always show a certain aggregation pattern. Therefore, this study develops a novel spatial analysis approach to explore the delay and causal factors which is able to take dependence and the possible problem involved including error correlation and variable lag effect of causal factors on delayinto account. The study first explores the delay aggregation pattern by measuring and quantifying the spatial dependence of delay. The spatial error model (SEM) and spatial lag model (SLM) are then established to solve the error correlation and the variable lag effect, respectively. Results show that the SEM and SLM achieve better fit than ordinary least square (OLS) regression, which indicates the effectiveness of considering dependence by employing spatial analysis. Moreover, the outcomes suggest that, aside from the well-known weather and flow control factors, delay-reduction strategies also need to pay more attention toreducing the impact of delay at the previous airport.

**REFERENCE:** https://go.gale.com/ps/i.do?p=AONE&u=googlescholar&id=GALE|A610927245&v=2.1&it=r&sid=AONE&asid=8b2120b0

**2. A Machine Learning approach for prediction of on time performance of flights. ABSTRACT:** A two-stage predictive model was developed to efficiently predict the departure and arrival delays of flights using flight schedule and weather features. Various supervised machine learning algorithms were implemented. It was found that the departure delay prediction had comparatively higher error rates due to a weak feature set. Furthermore, a Decision Support Tool was developed using the model to predict real-time flight delays. In the future, more data can be extracted by considering a larger number of airports over a longer time frame to improve the model and other deep architectures can also be implemented.

**REFERENCE:**https://ieeexplore.ieee.org/document/8102138

**3.Competing Networks, Spatial and Industrial Concentration in the US Airline Industry**

**ABSTRACT:** The paper uses Gini decomposition analysis to evaluate changes in the spatial distribution and industry shares of total US air traffic, as well as analysing the decomposition components for individual airlines and airports for the period 1990–2002. The paper develops explicit relationships between two of the main decomposition schemes used in the income inequality literature and shows the insights that such analysis may provide for evaluation and examination of air transport networks and traffic distributions. A multidimensional Gini and itsdecomposition are derived using an adjustment method derived from the relationship between the two Gini decomposition schemes.

**REFERENCE:**https://www.researchgate.net/publication/24089448_

**4.Identification, Characterization, prediction of Traffic flow patterns in multi airport systems.**

**ABSTRACT:** The data-driven framework is based on a sequential application of machine learning methods on historical flight tracks, weather forecasts and airport operational data. A multi-layer clustering analysis is performed to mine spatial and temporal trends in flight trajectory data for traffic flow pattern identification. The results revealed significant variability in throughput and delay performance for different metroplex configurations, emphasizing the importance of anticipating the behavior of the metroplex as a system when forecasting individual airport capacity. Future research goes along this direction by exploring the development of higher-fidelity models for airport capacity prediction that take as input detailed weather information and metroplex configuration forecasts in order to deliver probabilistic capacity forecasts for strategic TMI planning.

**REFERENCE:**https://ieeexplore.ieee.org/document/8373742

**5.Comparative Analysis on Propagation Effects of Flight Delays**

**ABSTRACT:** This paper aims to capture the interdependency among the sequence of flight delays due to airline operations in airports, weather, and air traffic control conditions. A copula function is used to determine the distribution of delay sequence and examine the propagation effects. Using the actual data sourced from an airline in Asia Pacific region, it is found that flight delays could propagate to downstream airports/airlines, where the strength of delays was decreased, passed on, or increased. Considering the possible effects of increased delays under air traffic control or airline factors, scenarios that adjust flight schedules with additional buffer time were created and analyzed. Results show that, by adding buffer time efficiently, flight schedules can become more reliable.

**REFERENCE:** https://www.researchgate.net/publication/322796669_

**6.Flight Delay Prediction: Data Analysis and Model Development**

**ABSTRACT**: The proposed model gains insight into factors causing flight delays, cancellations and the relationship between departure and arrival delay using exploratory data analysis. In addition, Random Forest (RF) algorithm is used to train and test the big dataset to help the model development. A web application has also been developed to implement the model and the testing results are presented with the limitation discussed.

**REFERENCE:**https://ieeexplore.ieee.org/document/9594260

**7.Modeling flight delay propagation in airport and airspace network.**

**ABSTRACT:** An Airport-Sector Network Delays model is developed in this paper for flight delay estimation within air transport network. This model takes both airports and airspace capacities into account by iterating among its three main components: a queuing engine, whichtreats each airport in the network as a queuing system and is used to compute delays at individual airport, a Link Transmission Model, which computes delays at individual sector andtransmits all air delays into ground delays, and a delay propagation algorithm that updates flight itineraries and demand rates at each airport on the basis of the local delays computed by the queuing engine and flow control delays computed by the Link Transmission Model. The model has been implemented to a network consisting of the 21 busiest airports in China and 2962 links that represent to 151 enroute control sectors in mainland China, and its performance is evaluated by comparing with the actual delay data and results of Airport Network Delays model. It is found that the proposed model is wellsuited for simulating delays in air transport system where either airports or airspace could be the bottleneck of the system.

**REFERENCE:**https://hal-enac.archives-ouvertes.fr/hal-01897108/document

**8. Flight Delay Prediction Using Machine Learning Algorithm XGBoost.**

**ABSTRACT:** We are proposing machine learning algorithms like XGBoost regressed, Linear regression Techniques. The aim of this research work is to predict Flight Delay, which is highest economy producing field for many countries and among many transportation this one is fastest and comfort, so to identify and reduce flight delay.

**REFERENCE:** https://www.researchgate.net/publication/344227817_Flight_Delay_Prediction_Using_Machine_Learning_Algorithm_XGBoost

**9.Propagation Index on Airport Delays**

**ABSTRACT:** This paper explores the propagation effect of flight delays among airports in theaviation system and proposes a new measure, the propagation index, to effectively analyze the interrelationship among airports in relation to flight delays. This index quantifies the effect of delay propagation by measuring the causality among delay time series. To assess the effectiveness of the proposed index on airport delays, three neural network-based regression models are built. The comparative experiments demonstrate that the propagation index proposed is highly correlated with observed airport delays.

**REFERENCE:** https://journals.sagepub.com/doi/abs/10.1177/0361198119844240

**10: Flight delay prediction for commercial air transport:** A deep learning approach ABSTRACT: The proposed method has proven to be highly capable of handling the challenges of large datasets and capturing the key factors influencing delays. This ultimately enables connected airports to collectively alleviate delay propagation within their network through collaborative efforts

**REFERENCE:** https://www.sciencedirect.com/science/article/abs/pii/S136655451831197

## 2.3 Problem Statement Definition

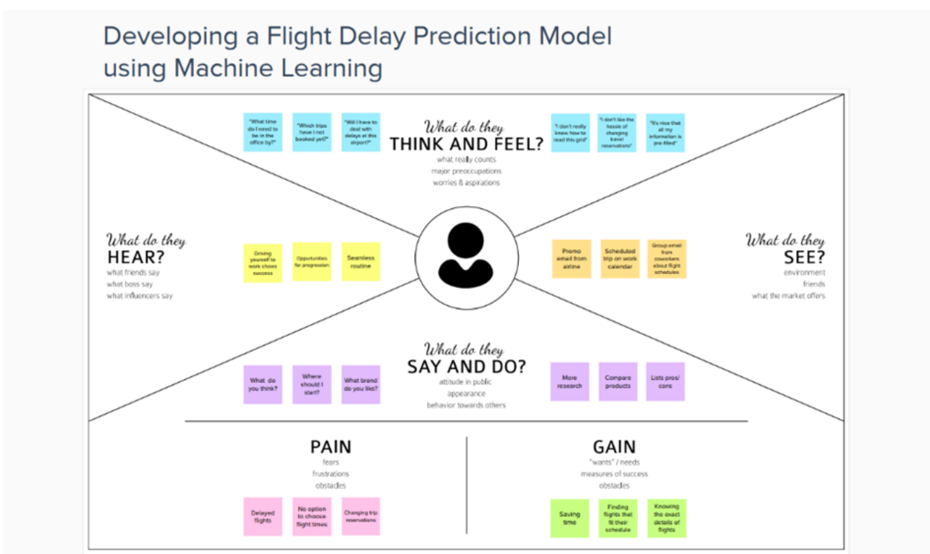
Flight Delay Prediction aims to predict the delay in the aircrafts due to increasing number of travellers in the recent times. An aircraft arrival is considered to be delay if the aircraft is late by over 15 minutes between the scheduled time and the arrival time. Flight Delay Prediction takes into consideration various attributes of the delay which includes scheduled time, source and destination of the flight, arrival time of the flight and departure time of the flight and many more attributes to predict the delay in the flight arrivals. These flight delays help the user massively to select the airlines, to select the source station and other economic aspects of the travellers. At the same time, Flight Delay prediction also helps the airlines to focus on the major reasons of the flight delay and minimize delay time on future occasions. Aviation industry are also benefitted with the help of the Flight Delay Prediction.

# CHAPTER 3

# IDEATION AND PROPOSED SYSTEM

## 3. IDEATION & PROPOSED SOLUTION

### 3.1 Empathy Map Canvas



## 3.2 Ideation & Brainstorming

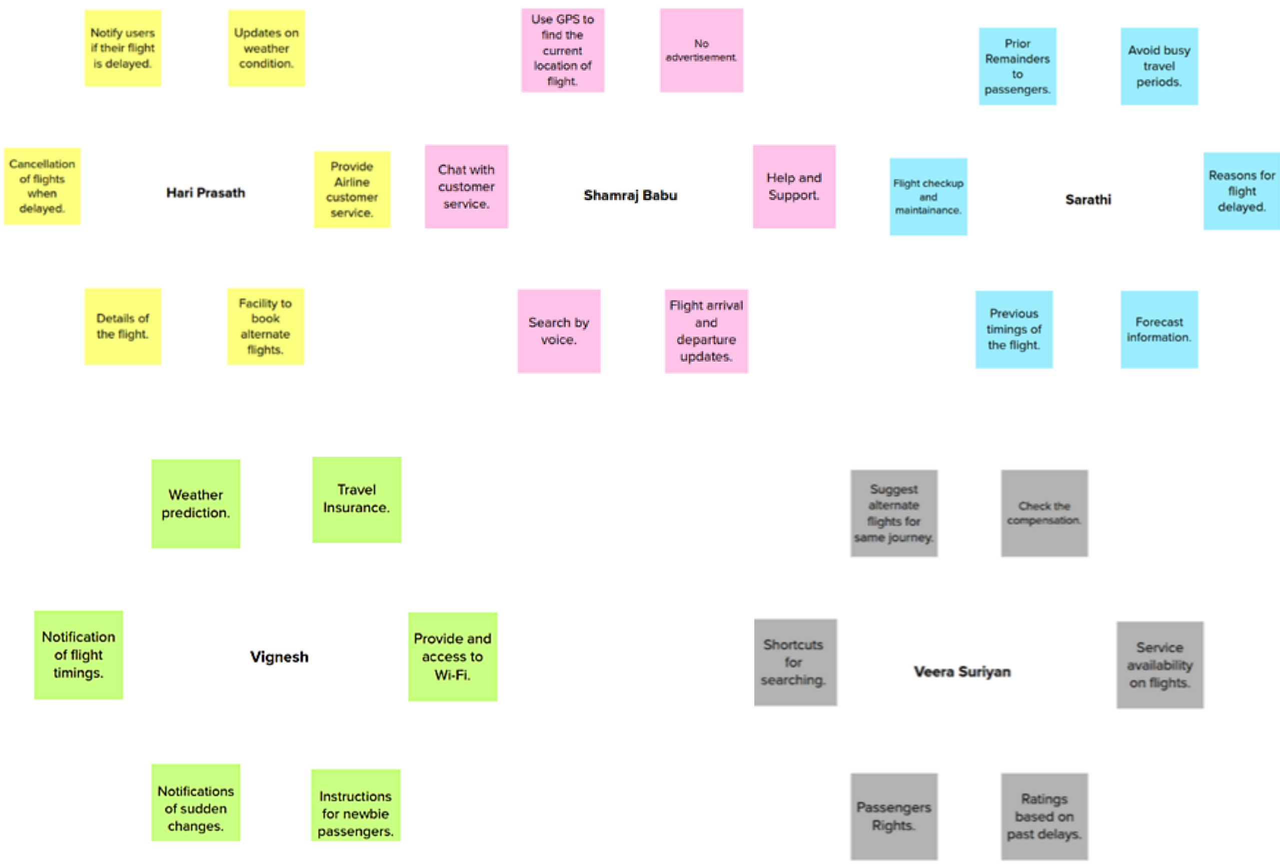



Figure 3.2 Brainstorming

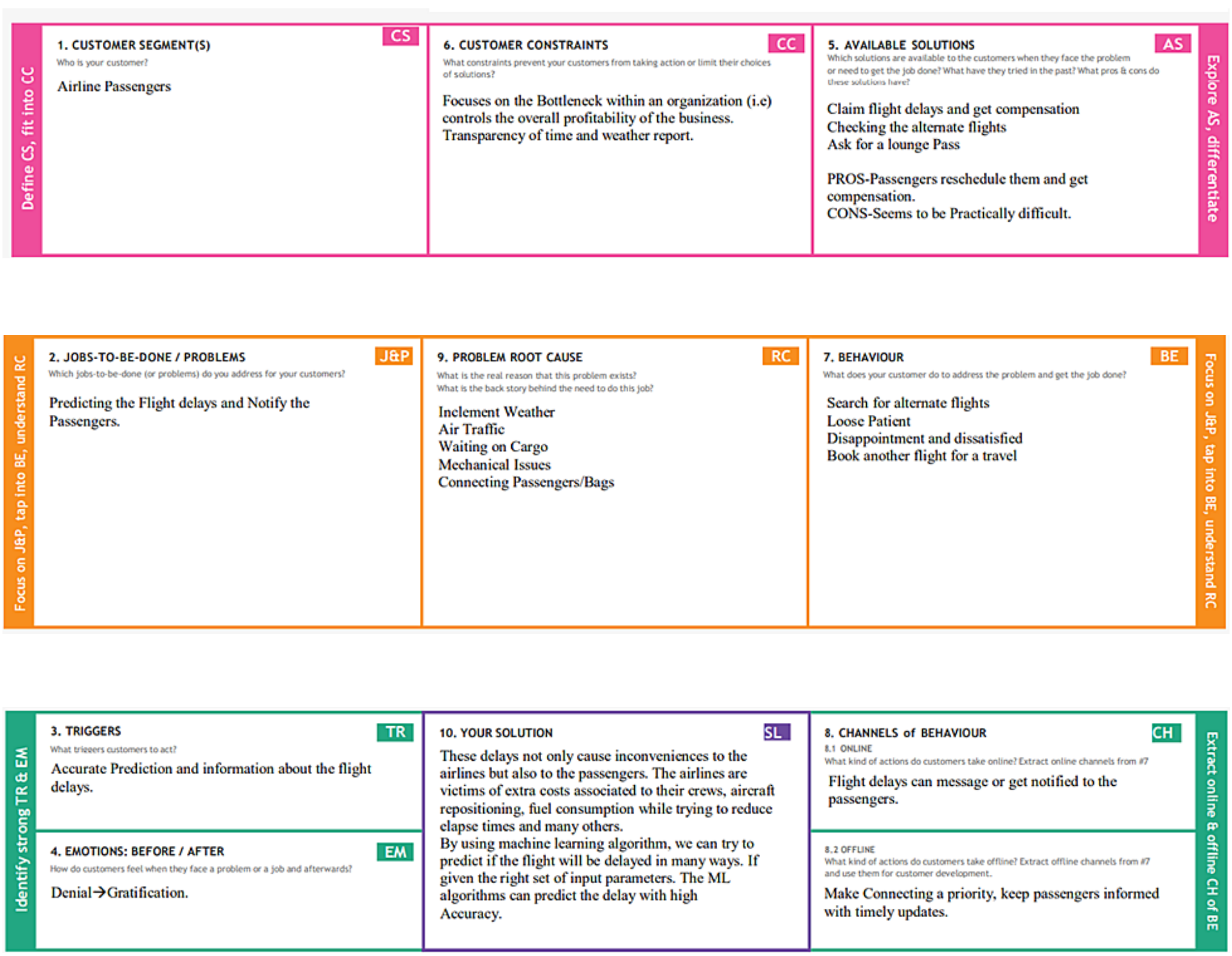
## Group Ideas:


## Prioritise:

### 3.3 Proposed Solution

The main objective of the model is to predict flight delays accurately in orderto optimise flight operations and minimise delays. Using a machine learning model, prediction of flight arrival delays can be done. The input to the algorithm is rows of feature vectors like departure date, departure delay, distance between the two airports, scheduled arrival time etc. Then a decision tree classifier is used to predict if the flight arrival will be delayed or not. Unlike other models here, comparison of decision tree classifiers with logistic regression and a simple neural network for various figures of merit is performed for better efficiency of predict

### 3.4 Problem Solution fit



# CHAPTER 4

# REQUIREMENT ANALYSIS

## 4. REQUIREMENT ANALYSIS

### 4.1 Functional requirement

Following are the functional requirement of proposed system.

4.1 Functional requirements

Table 4.1 Functional Survey

| FR No. | Functional Requirement | Sub Requirement |
|---|---|---|
| FR-1 | User Registration | Registration through g-mail, Websites, google accounts |
| FR-2 | User Verification | Verify where the user is authorized |
| FR-3 | Flight Details | Details of Flights Route and timing of flights |
| FR-4 | Flight Technical Issues | Identify, solve and rectify the issues that occur in flights. |
| FR-1 | Search Flights | The system should allow users to search for their flight details. |
| FR-1 | Flights Status Notification | Notify the passengers about the flight delay and Current status. |

Table 4.2 Non-Functional requirements

| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | Usability | Compact<br>Easy to access |
| NFR-2 | Reliability | Should provide accurate predictions |
| NFR-3 | Performance | High accuracy. High Uninterrupted service. |
| NFR-4 | Availability | The system should be available at all times. |
| NFR-5 | Scalability | Can handle multiple users at the same time Accessible even in remote areas. |
| NFR-6 | Security | The User data will not be used and more securely and safe. User's information can be secured. |

# CHAPTER 5

# PROJECT DESIGN

## 5. PROJECT DESIGN

### 5.1 Data Flow Diagrams

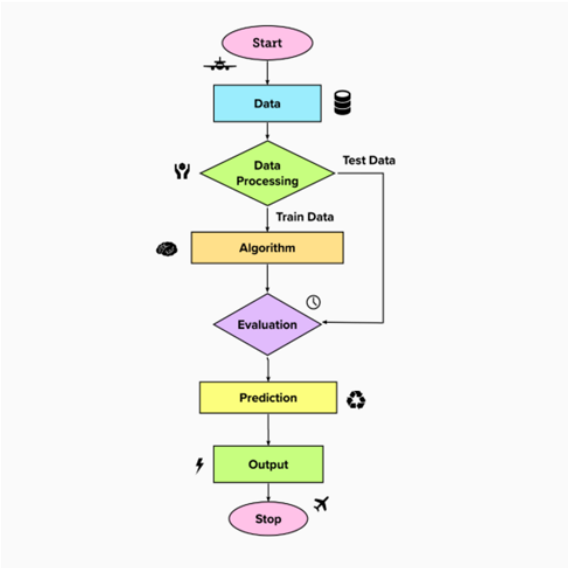
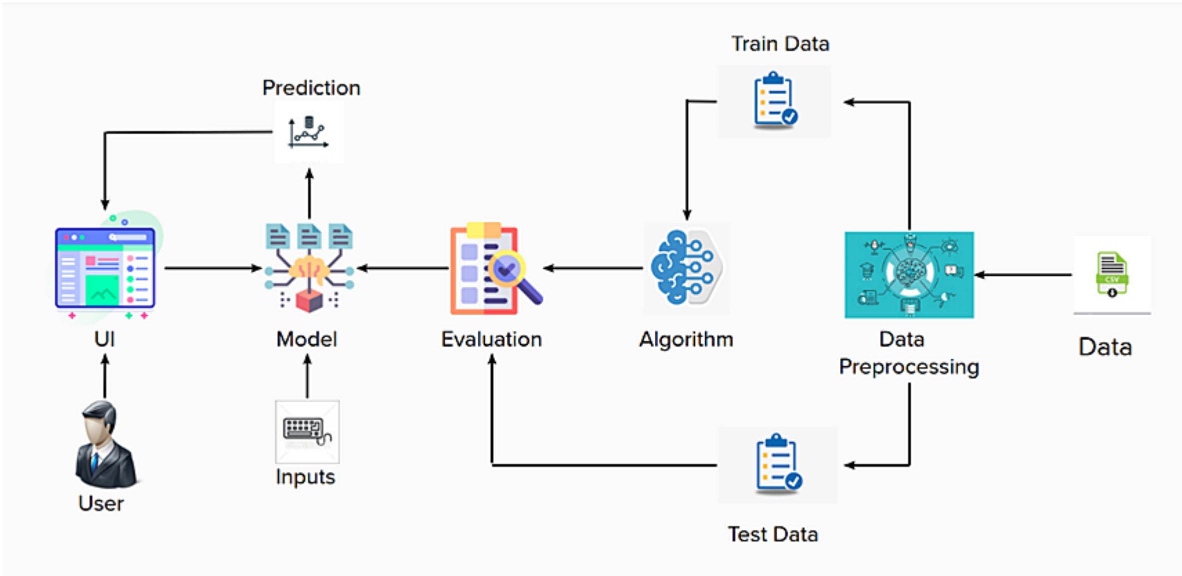


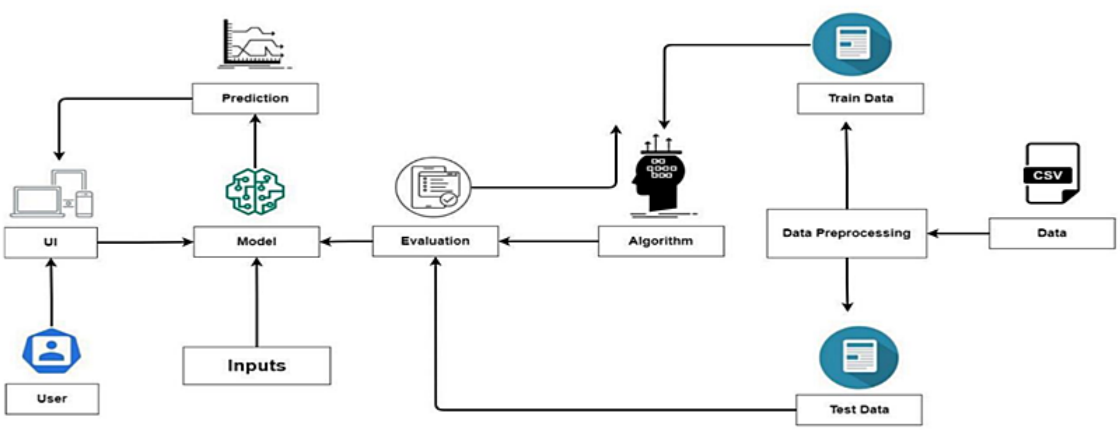

Figure 5.1 Data Flow Diagrams

### 5.2 Solution & Technical Architecture

**Solution architecture:**




**Technical Architecture:**



### 5.3 User

| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criteria | Priority | Release |
|---|---|---|---|---|---|---|
| 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 a confirmation email once I have registered for the application. | I can receive confirmation email & click confirm | High | Sprint-1 |
| | | 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 user, I can check the status of the flight by entering the details of the flight | | High | Sprint-1 |
| | Screen Space | USN-7 | As a web user, I have a better vision of the website | I can engage well with the website | High | Sprint-1 |
| Customer Care Executive | Communication | USN-1 | As a customer care executive, I need to be clear in my responses. | User will have a good understanding of the website | High | Sprint-2 |
| | Knowledge | USN-2 | As a customer care executive, I need to have extensive knowledge about the website | User will comprehend the website better | High | Sprint-2 |
| Administrator | Overall Management | USN-1 | As an administrator, I have to look after the working of the website. | Allows use and growth of the website. | High | Sprint-3 |
| | Security | USN-2 | As an administrator, I would identify if the given user is valid. | The website is protected from unauthorized entry | High | Sprint-3 |

# CHAPTER 6

# PROJECT PLANING

## 6. PROJECT PLANNING & SCHEDULING

### 6.1 Sprint Planning & Estimation

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Point | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Data Collection | USN-1 | As a user, I can't interact anything. Waiting is user's task. User can listen the relationship existbetween the various attributes of data by presentation of developer. | 2 | Low | Veerasuriyan K |
| Sprint-1 | Data Pre-processing | USN-1 | As a user, provide the clean dataset that further Allows the model. | 2 | High | Hari Prasath N<br>Shamrajbabu T<br>Vignesh B S<br>Sarathi S S<br>Veerasuriyan K |
| Sprint-2 | Building the Model | USN-2 | As a user, I can predict flight delay by variousdeveloped ML models by console | 1 | High | Shamrajbabu T<br>Vignesh B S |
| Sprint-2 | Evaluation of Model | USN-3 | As a user, I can predict flight delay by best Modelin various developed ML model by console | 2 | High | Hari Prasath N<br>Vignesh B S |
| Sprint-2 | Model Deployment on IBM Cloud. | USN-4 | As a user, I can use the model by requesting thedeployed model on Cloud. | 1 | Medium | Hari Prasath N |
| Sprint-3 | User interaction Dashboard | USN-5 | As a user, I can use the model or prediction from model by interacting with dashboard | 2 | High | Sarathi S S<br>Veerasuriyan K |
| Sprint-3 | Integrated and Creation of Flask. | USN-6 | As a user, I can use the model or prediction from model by interacting with the Website. | 1 | Medium | Shamrajbabu T<br>Sarathi S S |
| Sprint-4 | Registration | USN-7 | As a user, I can register for the application by entering my email, password, and confirming mypassword. | 2 | High | Hari Prasath N |

Figure 6.1 Sprint planning

### 6.2 Sprint Delivery Schedule

| 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 | 29 Oct 2022 | 20 | 29 Oct 2022 |
| Sprint-2 | 20 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 20 | 05 Nov 2022 |
| Sprint-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 20 | 12 Nov 2022 |
| Sprint-4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 20 | 19 Nov 2022 |

Figure 6.2 Sprint Delivery Schedule

**Velocity:**

We have a 24-day sprint duration, and the velocity of the team is 20 (points per sprint).Thus he team's average velocity (AV) per iteration unit (story points per day) is as follows

AV = sprint duration / velocity

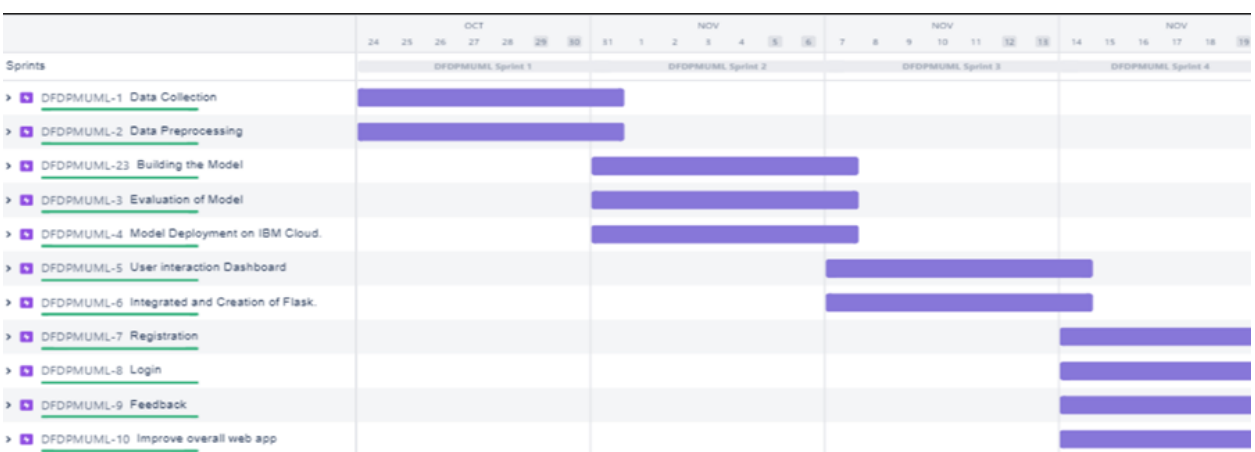
= 24/20

= 1.2

**Burn down chart:**

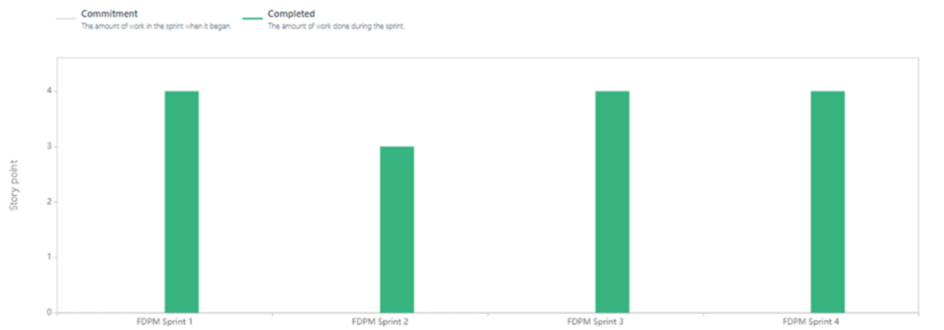

Figure 6.3 Burndown Chart

## 6.3 Reports from JIRA



### 6.3.2 Velocity report

In the velocity report, the team's velocity is calculated by taking the average of the total completed estimates from their last few sprints. The velocity report of the team is shown below:




| Sprint | Commitment | Completed |
|---|---|---|
| FDPM Sprint 1 | 0 | 4 |
| FDPM Sprint 2 | 0 | 3 |
| FDPM Sprint 3 | 0 | 4 |
| FDPM Sprint 4 | 0 | 4 |

Figure 6.12 Velocity report

### 6.3.3 Cumulative flow diagram

Cumulative flow diagram shows the statuses of issues over time. This helps the team identify potential bottlenecks that need to be investigated. The cumulative flow diagram of the team is displayed below

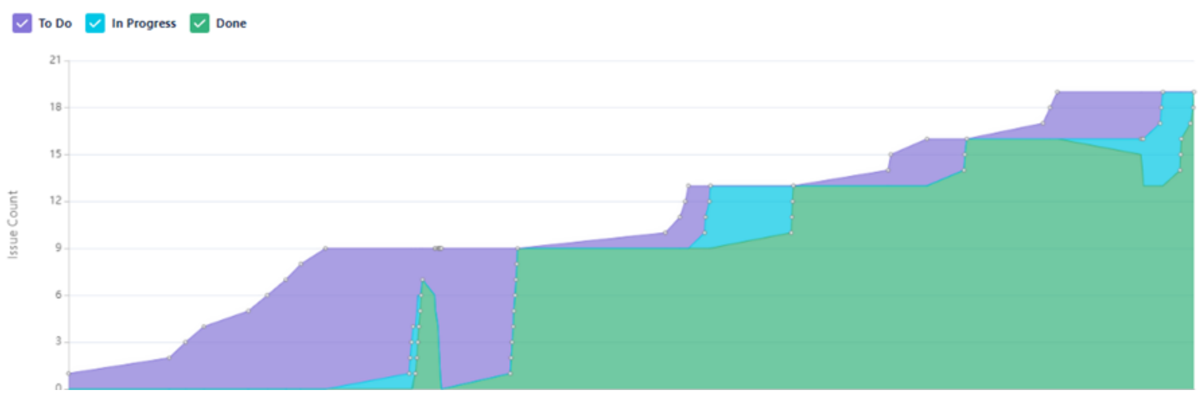



Figure 6.13 Cumulative flow diagram

# CHAPTER 7

# CODING AND SOLUTIONING

**7. CODING & SOLUTIONING**

During the Project Development Phase, we have done four Sprints they are Sprint 1, Sprint 2, Sprint 3 and Sprint 4.In Agile product development, a sprint is a set period of time during which specific work has to be completed and made ready for review.

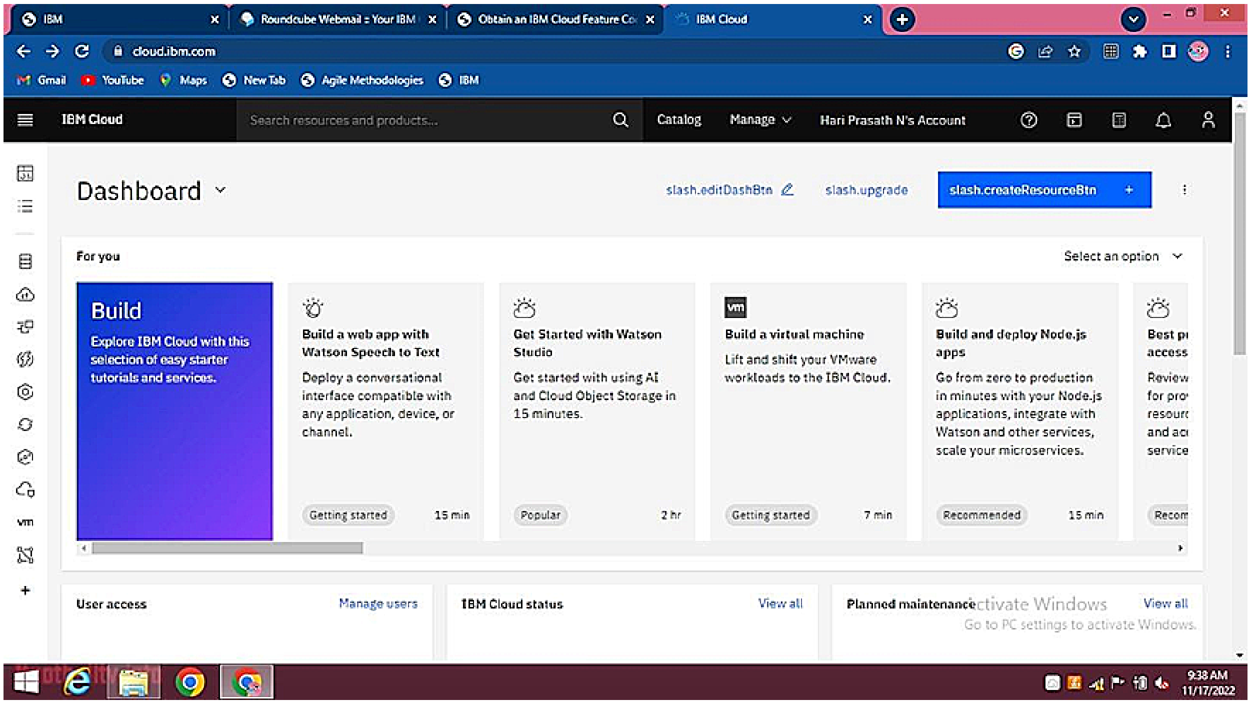
## 7.1 Sprint 1

During Sprint1 we have planned for Downloading the dataset, import the libraries, Read the dataset, understanding data types and summary of features, handling missing values, Replacing the missing values, Label encoding, Split the dataset into dependent and independent variable, split the dataset into train and test set.

## 7.2 Sprint 2

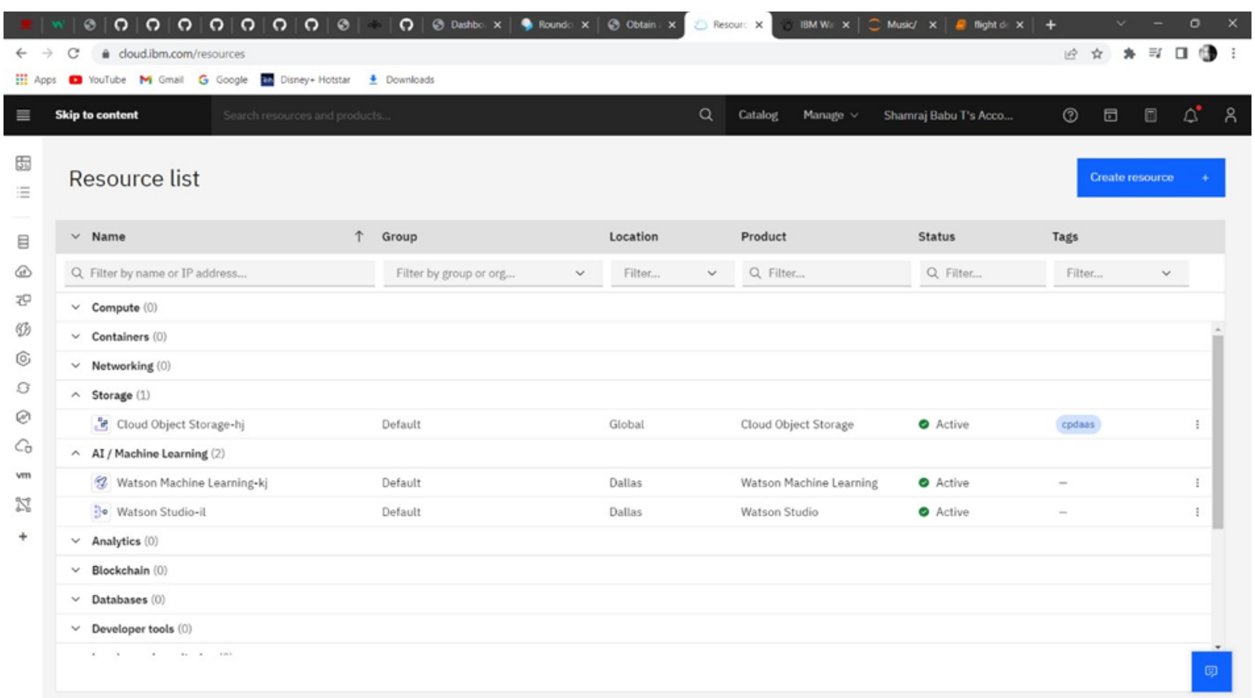
During Sprint 4 we have planned for training the model on IBM where we will register for IBM cloud, train the ML model on IBM and integrate flask with scoring end point.

Registered on IBM cloud and activated Watson machine learning, cloud storage and Watson studio then trained the ML model on IBM using API KEY




.

## IBM Cloud Resources:



## Deploy the Model on IBM Watson Studio:

In [63]:

```
pip install-U ibm-watson-machine-learning
from ibm_watson_machine_learning import APIClient
import json
```

In [64]:

```
wml_credentials = {
    "apikey":"4fJbepuok7oCO1UkeKU831Sq5rz5-JP0R_hljeB2oaEL","url":"https://us-
    south.ml.cloud.ibm.com"

}
```

In [65]:

```
wml_client = APIClient(wml_credentials)wml_client.spaces.list()
SPACE_ID="deaaa6e0-4843-467d-94d8-71d0272de83b"
wml_client.set.default_space(SPACE_ID)
wml_client.software_specifications.list(500)
```

In [66]:

```
import sys
sys.version

pip install ibm-watson-machine-learning
import sklearn sklearn.__version_____
```

In [67]:

```
MODEL_NAME = "RandomForestClassifier()"DEPLOYMENT_NAME = 'flight delay'
DEMO_MODEL = model
software_spec_uid = wml_client.software_specifications.get_id_by_name('runtime-22.1-py3.9')
```

In [68]:

In [69]:

```
model_props = {
    wml_client.repository.ModelMetaNames.NAME: MODEL_NAME,
    wml_client.repository.ModelMetaNames.TYPE: 'scikit-learn_1.0',
    wml_client.repository.ModelMetaNames.SOFTWARE_SPEC_UID:software_spec_uid
```

In [70]:

```
}
model_details = wml_client.repository.store_model(model=DEMO_MODEL,
meta_props=model_props, training_data=x_train,
    training_target=y_train
```

In [71]:

model_details
model_id = wml_client.repository.get_model_id(model_details)model_id In [72]:

deployment_props = wml_client.deployments.ConfigurationMetaNames.NAME:DEPLOYMENT_NAME,

In [73]:

wml_client.deployments.ConfigurationMetaNames.ONLINE: {}

}

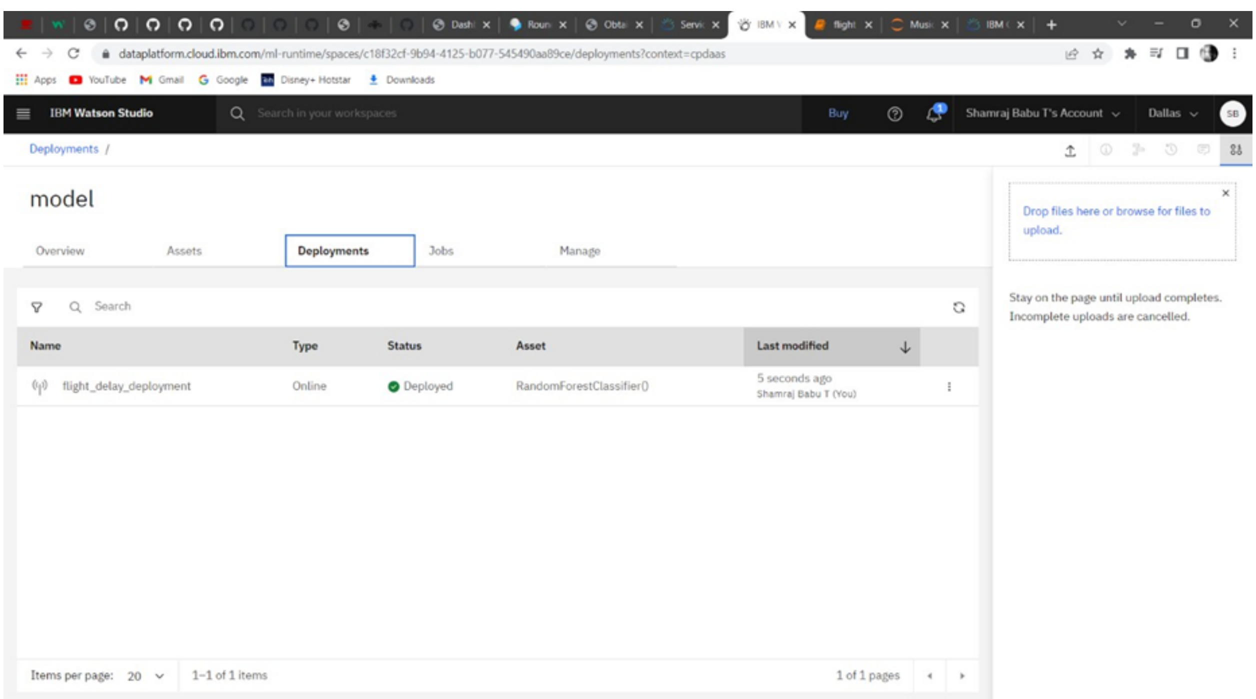
deployment_props = {

wml_client.deployments.ConfigurationMetaNames.NAME:DEPLOYMENT_NAME,

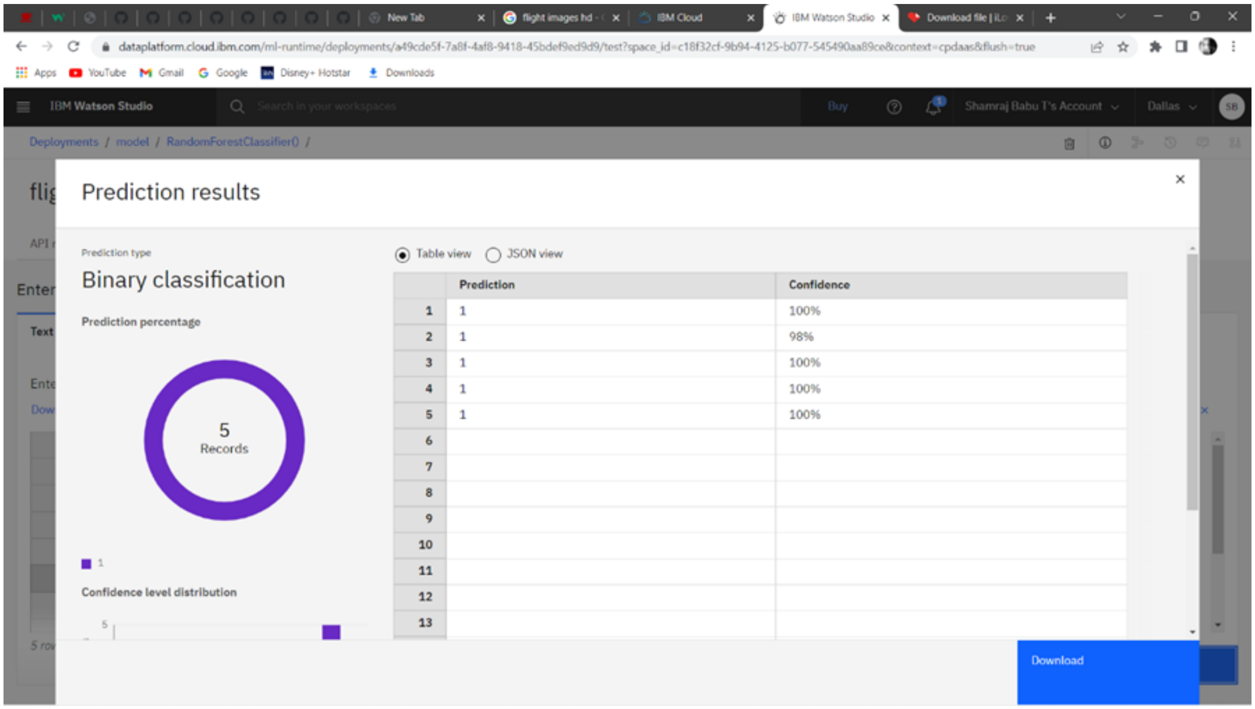
Iwml_client.deployments.ConfigurationMetaNames.ONLINE: {}

}

**Deployment The Model:**



**Testing the Deployment Model:**




## 7.3 Sprint 3

During Sprint2 we have planned for Creating HTML files, Build Python code and run the app

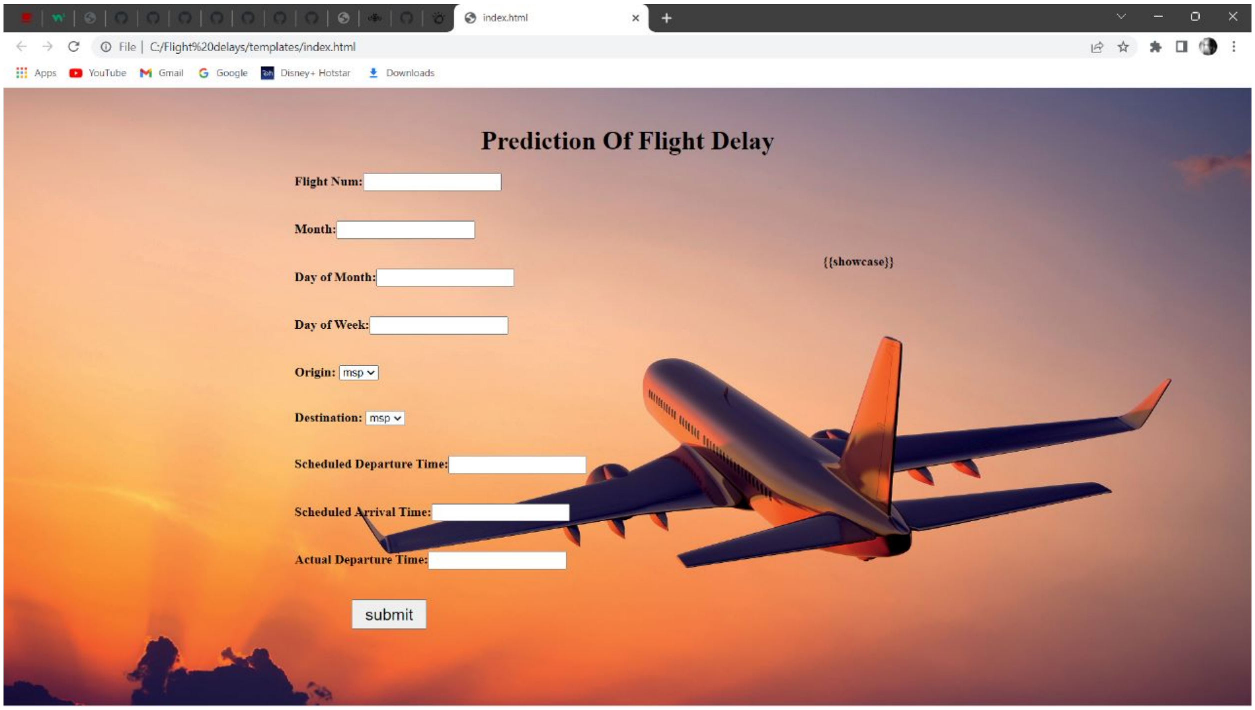



Figure 7.5 Website for flight delay prediction

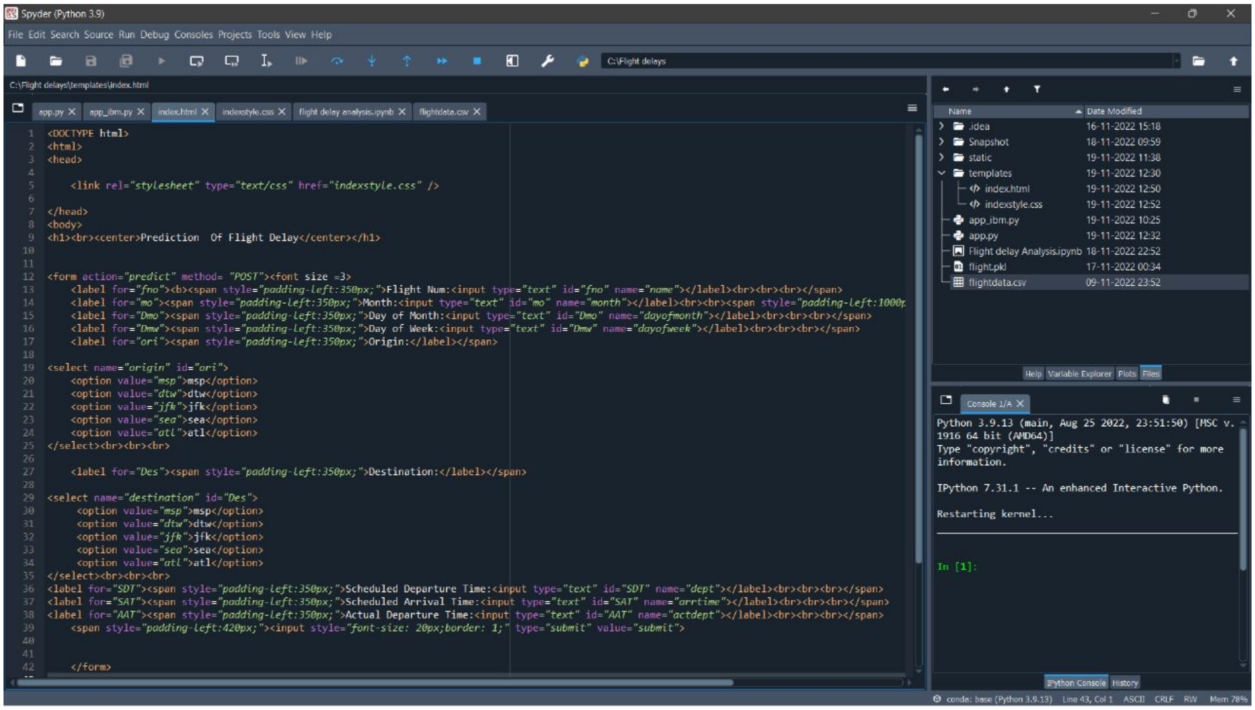



Figure 7.6 index.html

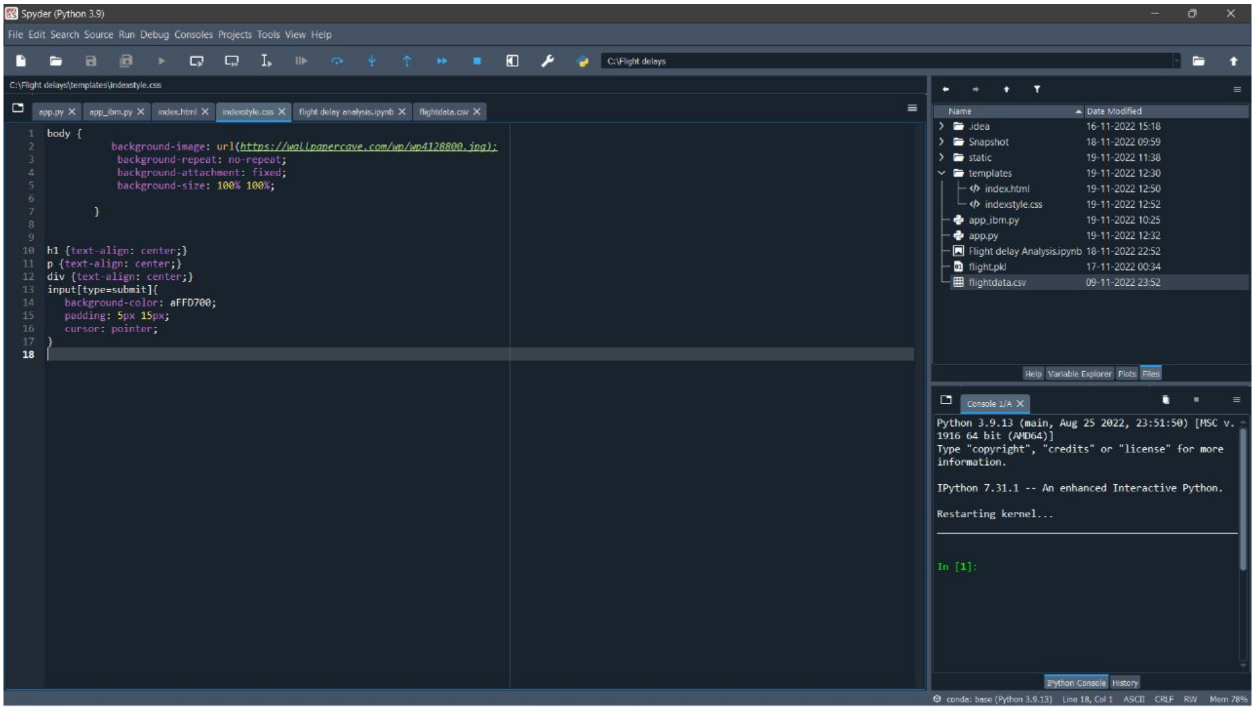


Figure 7.7 index style.html

## Building the flask file:

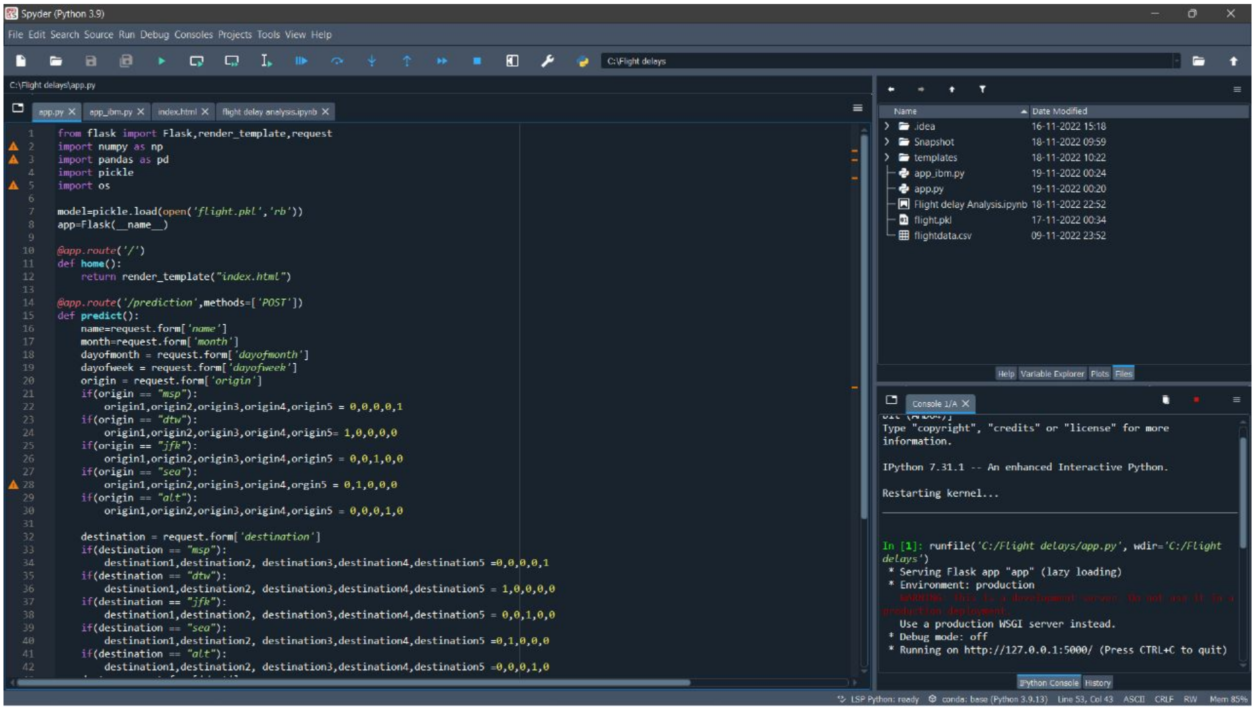
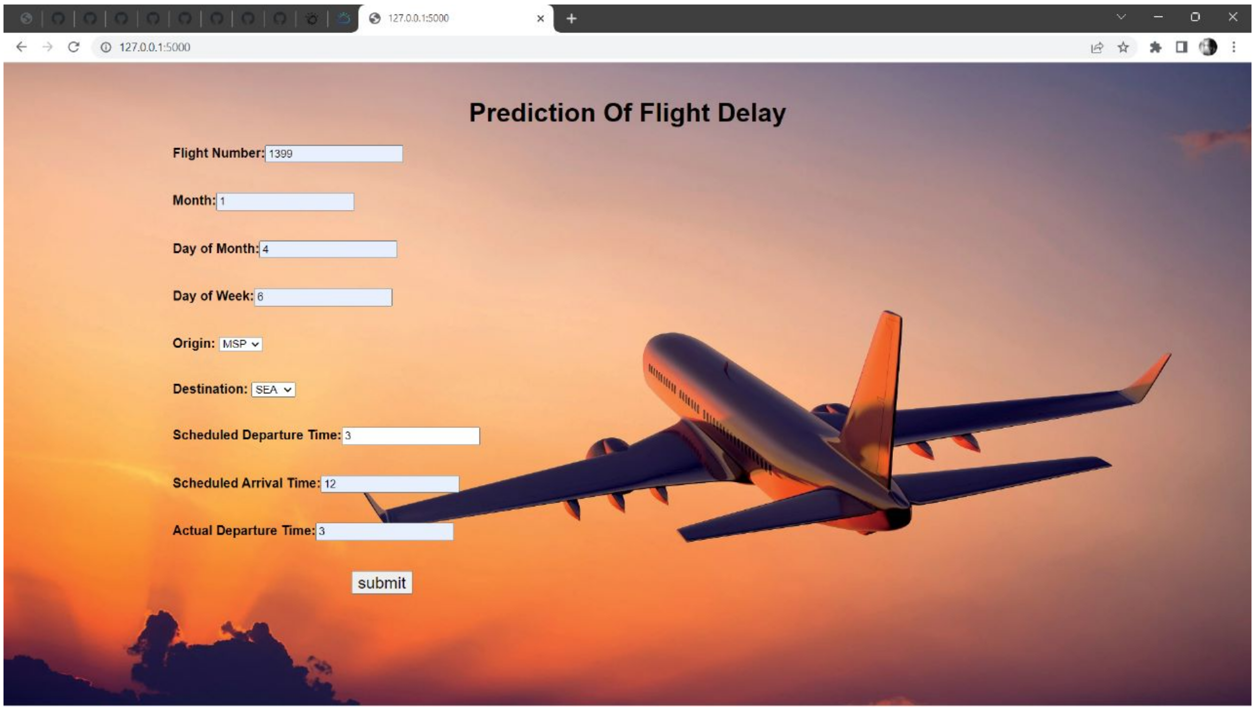
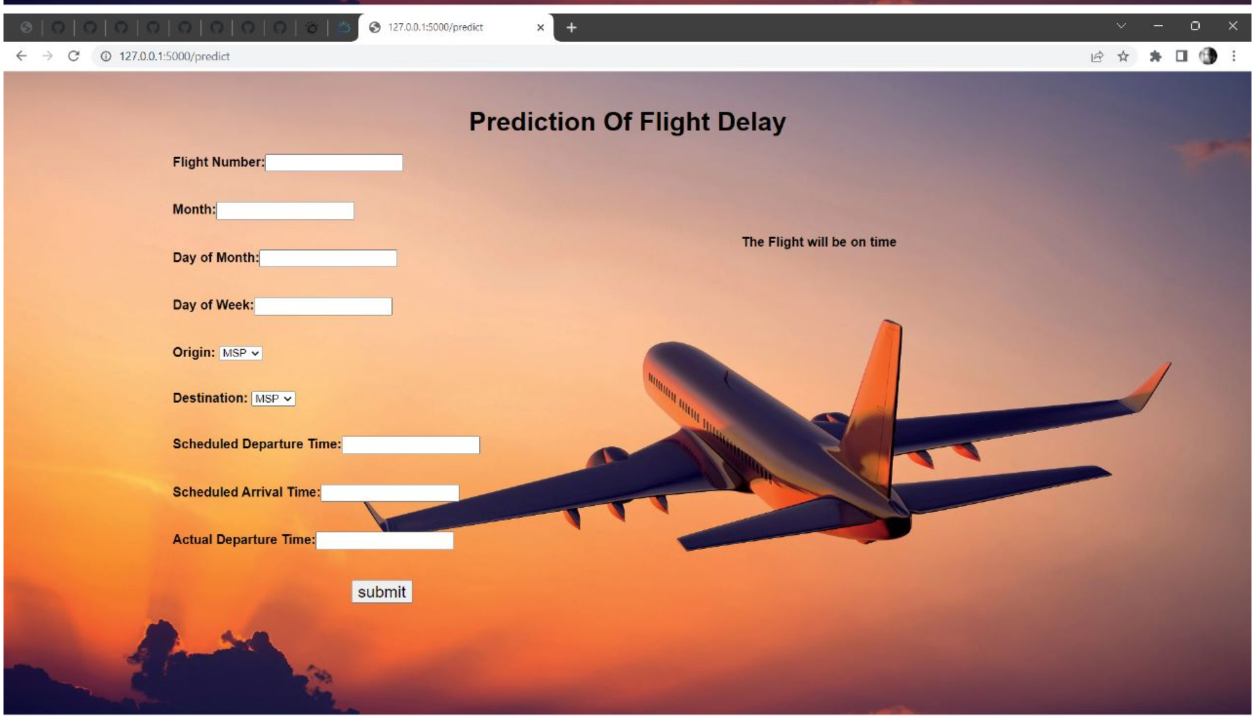


Figure 7.8  app.py

## SPRINT-4

During Sprint3 we have planned for asking users to enter numerical and selection data and tested for many inputs and checked the correctness of the result.



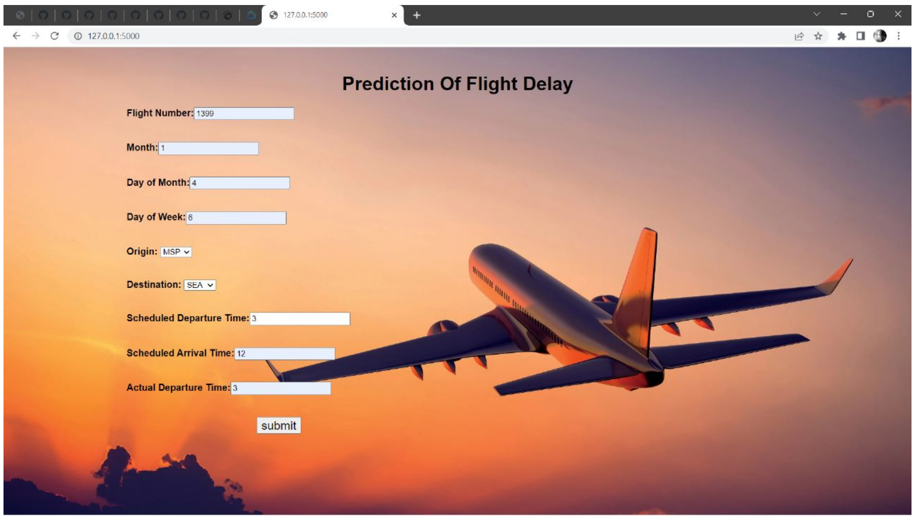
# CHAPTER 8

# TESTING

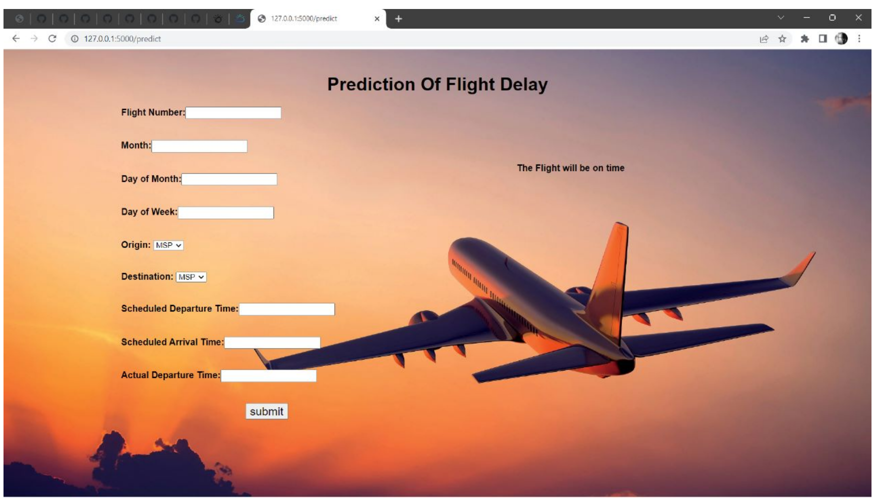
## 8.TESTING

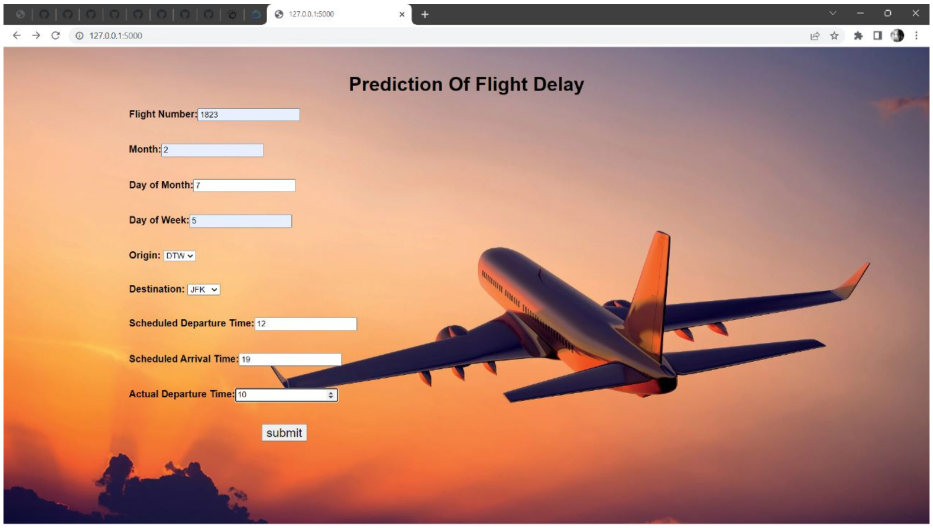
### 8.1 Test Cases

The website predicts if the flight is delayed or not using the given values entered by the user. This helps the user to find the alternative to travel.

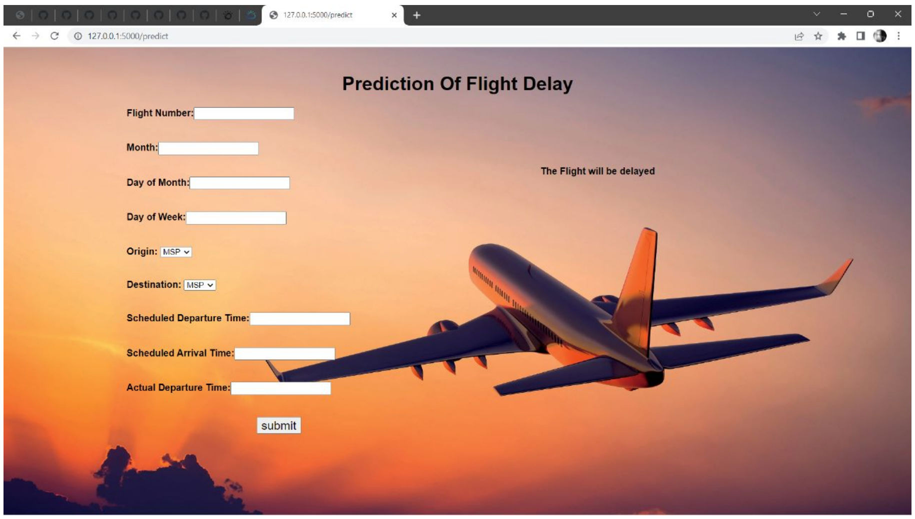



There is no delay in flight so the user can travel in the same flight.

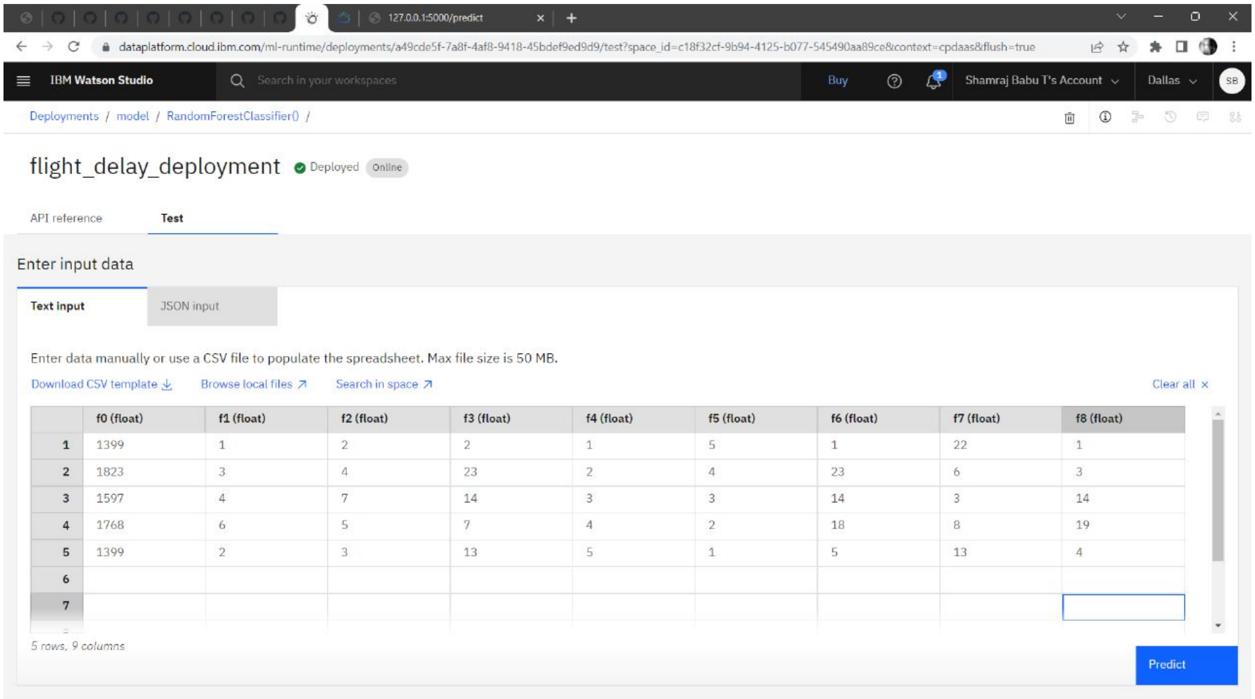


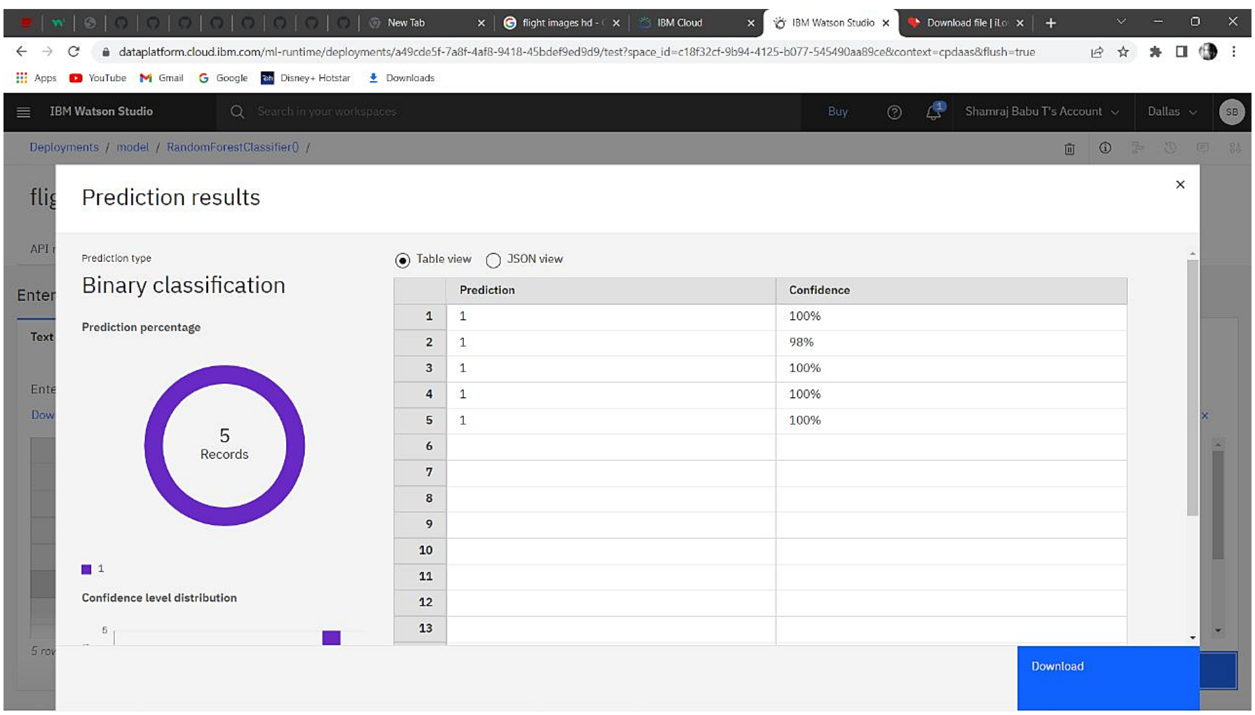



There is a delay in flight so the user can travel in the alternate flights.



## Deployment Testing:



### 8.2 User Acceptance Testing

This report shows the number of test cases that have passed, failed, and untested

| Section | Total Cases | Not Tested | Fail | Pass |
|---|---|---|---|---|
| Home Screen | 1 | 0 | 0 | 1 |
| User Input | 3 | 0 | 0 | 3 |
| Flight delay  testing | 2 | 0 | 0 | 2 |
| No Flight delay testing | 2 | 0 | 0 | 2 |
| Version Control | 2 | 0 | 0 | 2 |

# CHAPTER 9

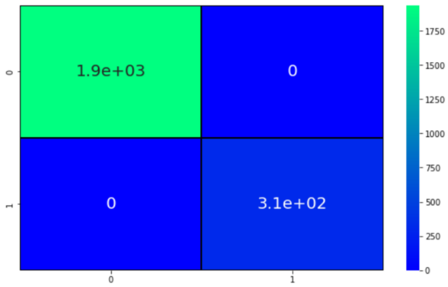
# RESULTS

## 9. RESULTS

### 9.1 Performance Metrics

Model: Random Forest Classification performance values

```
Testing Sensitivity for Random Forest 1.0
Testing Specificity for Random Forest 1.0
Testing Precision for Random Forest 1.0
Testing accuracy for Random Forest 1.0
```




In [37]:

```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00      1936
         1.0       1.00      1.00      1.00       311

    accuracy                           1.00      2247
   macro avg       1.00      1.00      1.00      2247
weighted avg       1.00      1.00      1.00      2247
```

Figure 9.1 Random forest classification metrics

Model: Decision Tree performance values

```
Testing Accuracy for Decision Tree 1.0
Testing Sensitivity for Decision Tree 1.0
Testing Specificity for Decision Tree 1.0
Testing Precision for Decision Tree 1.0
Testing accuracy for Decision Tree 1.0
```


In [39]:

```
print(classification_report(y_test,pred1))
```

```
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00      1936
         1.0       1.00      1.00      1.00       311

    accuracy                           1.00      2247
   macro avg       1.00      1.00      1.00      2247
weighted avg       1.00      1.00      1.00      2247
```

Figure 9.2 Decision tree metrics

Hence, we tested with Decision Tree and Random Forest Classification wherein the accuracy of Random Forest classification is 80% compared with Logistic Regression.

# CHAPTER 10

# CONCLUSION

**10. CONCLUSION**

The Flight Delay Prediction model focuses on predicting the delay in the aircrafts due to the increasing number of travellers in recent times. In order to build the flight prediction model, several steps are followed .The database for the flight delay prediction model is taken and analysed in a proper fashion. The required libraries needed for implementing the model are imported. The importance of each attribute is then calculated. The unwanted attributes of the databases are identified and are removed. The missing values in the remaining dataset are handled properly. One hot encoding is performed and then finally the dataset is splitted for training and testing purposes. This is how the dataset is pre-processed. Then, the Random Forest Classifier model is selected. The random forest classifier model is trained on the pre-processed dataset and is evaluated using different metrics. Once the model is evaluated, the model is saved using pickle for further predictions. The accuracy of the Random Forest classifier is around 80%. The appropriate pages of the application are created using HTML and Python. The saved model is executed and tested with appropriate test cases. Finally, the model is made to train in the IBM cloud. Flask framework (Python) is used to integrate the model and the application. Each user can view the flight delay using this application. The end user will be asked to enter certain values .The end user might be able to view the delay of the flights after giving the appropriate inputs.

# CHAPTER 11

# FUTURE SCOPE

**11. FUTURE SCOPE**

The proposed flight delay prediction model has an accuracy of around 80%. The model performs relatively well. Yet, the accuracy of the model can be improved by using advanced machine learning algorithms. The application in the proposed system can be improved by using a user authentication module.

# CHAPTER 12

# APPENDIX

## index.html

```
<DOCTYPE html>
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1">
<style>
h2 {text-align: center;}
* {
box-sizing: border-box;
}
body{
```

```
font-family: Arial, Helvetica, sans-serif;

background-image: url(https://wallpapercave.com/wp/wp4128800.jpg);

background-size: cover;

background-attachment: fixed;

}

</style>

</head>

<body>

<h1><br><center>Prediction  Of Flight Delay</center></h1>

<form action="predict" method= "POST"><font size =3>

   <label for="fno"><b><span style="padding-left:200px;">Flight Number:<input
type="number" id="fno" name="name"></label><br><br><br></span>

   <label for="mo"><span style="padding-left:200px;">Month:<input type="number"
id="mo" name="month"></label><br><br><span style="padding-
left:900px;"><b>{{showcase}}</b><br></span>

   <label for="Dmo"><span style="padding-left:200px;">Day of Month:<input
type="number" id="Dmo" name="dayofmonth"></label><br><br><br></span>

   <label for="Dmw"><span style="padding-left:200px;">Day of Week:<input
type="number" id="Dmw" name="dayofweek"></label><br><br><br></span>

   <label for="ori"><span style="padding-left:200px;">Origin:</label></span>

<select name="origin" id="ori">

   <option value="msp">MSP</option>

   <option value="dtw">DTW</option>

   <option value="jfk">JFK</option>

   <option value="sea">SEA</option>

   <option value="atl">ATL</option>

</select><br><br><br>
```

```
   <label for="Des"><span style="padding-left:200px;">Destination:</label></span>
<select name="destination" id="Des">
    <option value="msp">MSP</option>
    <option value="dtw">DTW</option>
    <option value="jfk">JFK</option>
    <option value="sea">SEA</option>
    <option value="atl">ATL</option>
</select><br><br><br>
<label for="SDT"><span style="padding-left:200px;">Scheduled Departure Time:<input
type="number" id="SDT" name="dept"></label><br><br><br></span>
<label for="SAT"><span style="padding-left:200px;">Scheduled Arrival Time:<input
type="number" id="SAT" name="arrtime"></label><br><br><br></span>
<label for="AAT"><span style="padding-left:200px;">Actual Departure Time:<input
type="number" id="AAT" name="actdept"></label><br><br><br></span>
   <span style="padding-left:420px;"><input style="font-size: 20px;border: 1;"
type="submit" value="submit">
</form>
</body>
</html>
```

## app.py

```
from flask import Flask,render_template,request
from werkzeug.utils import secure_filename
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 index():
   return render_template("index.html")
@app.route('/predict',methods=['POST','GET'])
def predict():
   name=request.form['name']
   month=request.form['month']
   dayofmonth = request.form['dayofmonth']
   dayofweek = request.form['dayofweek']
   origin = request.form['origin']
   if(origin == "msp"):
      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,orgin5 = 0,1,0,0,0
   if(origin == "alt"):
      origin1,origin2,origin3,origin4,origin5 = 0,0,0,1,0
destination = request.form['destination']
   if(destination == "msp"):
      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)
    print(dept15)
total =
[[name,month,dayofmonth,dayofweek,origin1,origin2,origin3,origin4,origin5,destination1,
destination2,destination3,destination4,destination5,dept15,arrtime]]
    y_pred = model.predict(total)
    print(y_pred)
    if (dept15 == 0):
        ans = "The Flight will be on time"
    else:
        ans = "The Flight will be delayed"
    return render_template("index.html",showcase = ans)
if __name__=='__main__':
    app.run(debug = False)
```

# CHAPTER 13

**GitHub & Project Demo Link**

**GitHub link:**

https://github.com/hari000007/IBM-Project-1277-1658382839

**Demo link:**

https://youtu.be/oCiaLRYk1DA