

# PROJECT REPORT

**Project Title** - Developing a Flight Delay Prediction Model using Machine Learning

## **Team Members**

- RAVI TEJA N G (Team Leader)
- RAKESH S
- RAVI CHANDRAN C
- VISHWA T

# INDEX

## 1. INTRODUCTION

1. Project Overview
2. Purpose

## 2. LITERATURE SURVEY

1. Existing problem
2. References
3. Problem Statement Definition

## 3. IDEATION & PROPOSED SOLUTION

1. Empathy Map Canvas
2. Ideation & Brainstorming
3. Proposed Solution
4. Problem Solution fit

## 4. REQUIREMENT ANALYSIS

1. Functional requirement
2. Non-Functional requirements

## 5. PROJECT DESIGN

1. Data Flow Diagrams
2. Solution & Technical Architecture
3. User Stories

## 6. PROJECT PLANNING & SCHEDULING

1. Sprint Planning & Estimation
2. Sprint Delivery Schedule
3. Reports from JIRA

## 7. CODING & SOLUTIONING

1. Feature 1
2. Feature 2

## 8. TESTING

1. Test Cases
2. User Acceptance Testing

## 9. RESULTS

1. Performance Metrics

## 10. ADVANTAGES & DISADVANTAGES

## 11. CONCLUSION

## 12. FUTURE SCOPE

## 13. APPENDIX

Source Code , GitHub & Project Demo Link

# 1. Introduction

## 1. Project Overview

Over the last twenty years, air travel has been increasingly preferred among travelers, 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.

## 2. Purpose

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 15 minutes. Furthermore, we compare decision tree classifier with logistic regression and a simple neural network for various figures of merit.

# 2. LITERATURE SURVEY

## 1. Existing solution

Since flight delays cause multiple problems across the world, there has been a significant improvement in delay prediction models right from the 1990s. The quantity of the delay decreased the quality of marketing strategies. A delay in the departure or arrival of a domestic flight affects the operation of an international flight. A small amount of change in the delay value can be a massive amount of success for airport sectors. The models developed during this system can be applied to predict the incidence of flight delay at airports. Such prognosticative capabilities would help traffic managers and airline dispatchers to organize mitigation strategies for lowering traffic disruptions. Nowadays, service quality plays an important role in attracting customers. Among these, air travels have their special customers and the most important matter in these travels is the flight time, on-time arrival at destination for passengers such as those who have an important meeting, that has been leading to high expenses for the passenger until they get to their destination on time. Flight delay has negative economic effects on the passenger, agencies and airport. Therefore, any reduction of these effects requires decreasing postponed flight price, so that

prediction or estimation has a great significance and numerous studies have been dedicated to this subject. Correspondingly, all the scientists have tried to design a model that understands effective factors and computes effect of each factor and their relation. Overall, the prediction methods are classified into five groups including Statistical Methods, Probability methods, network-based methods, operational methods and machine learning methods. In one of the best studies that has been performed based on statistics delay time has been considered to be reduced. Their study has investigated important factors before fly and those which occur on the ground. In the next step, it has predicted the delay at destination based on factors that occur in the vicinity of arrival time at destination.

Eventually, results have shown that whenever, the delay is correctly predicted, passenger disaffection and fuel consumption decrease and consequently number of flight increases. Moreover, it is possible to increase the agencies benefits through reducing number of passengers who wrongly selected their routes or specifying the probabilities for some flights and optimizing delay time prediction. Another prominent investigation based on Probability has been done and the author believes that huge storm in U.S.A has highly affected the flight delay. This study has been devoted to predict delay based on mathematical calculations and through considering delay time duration of the flights that had been engaged to storm in the same day. Metrological reports have shown the effect of storm one hour before and after event cause ephemeral climate at the region. In the next step, Monte-Carlo simulation has been used to estimate the airport runway capacity, so that traffic of each runway would have been estimated. As the research has employed only one factor, the model has not enough accuracy, but it is possible to increase region air capacity path structure.

## 2. References

1. Flight delay prediction based on deep learning and LevenbergMarquart algorithm  
Yazdi, M.F., Kamel, S.R., Chabok, S.J.M. et al. Flight delay prediction based on deep learning and Levenberg-Marquart algorithm. J Big Data 7, 106 (2020). <https://doi.org/10.1186/s40537-020-00380-z>
2. Study of Flight Departure Delay and Causal Factor Using Spatial Analysis  
Shaowu Cheng, Yaping Zhang, Siqi Hao, Ruiwei Liu, Xiao Luo, Qian Luo, "Study of Flight Departure Delay and Causal Factor Using Spatial Analysis", Journal of Advanced Transportation, vol. 2019, Article ID 3525912, 11 pages, 2019. <https://doi.org/10.1155/2019/3525912>
3. Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data  
Yuemin Tang. 2021. Airline Flight Delay Prediction Using Machine Learning Models. In 2021 5th International Conference on E-Business and Internet (ICEBI 2021), October 15-17, 2021, Singapore, Singapore. ACM, New York, NY, USA, 7 Pages. <https://doi.org/10.1145/3497701.3497725>

## 3.Problem Statement Definition

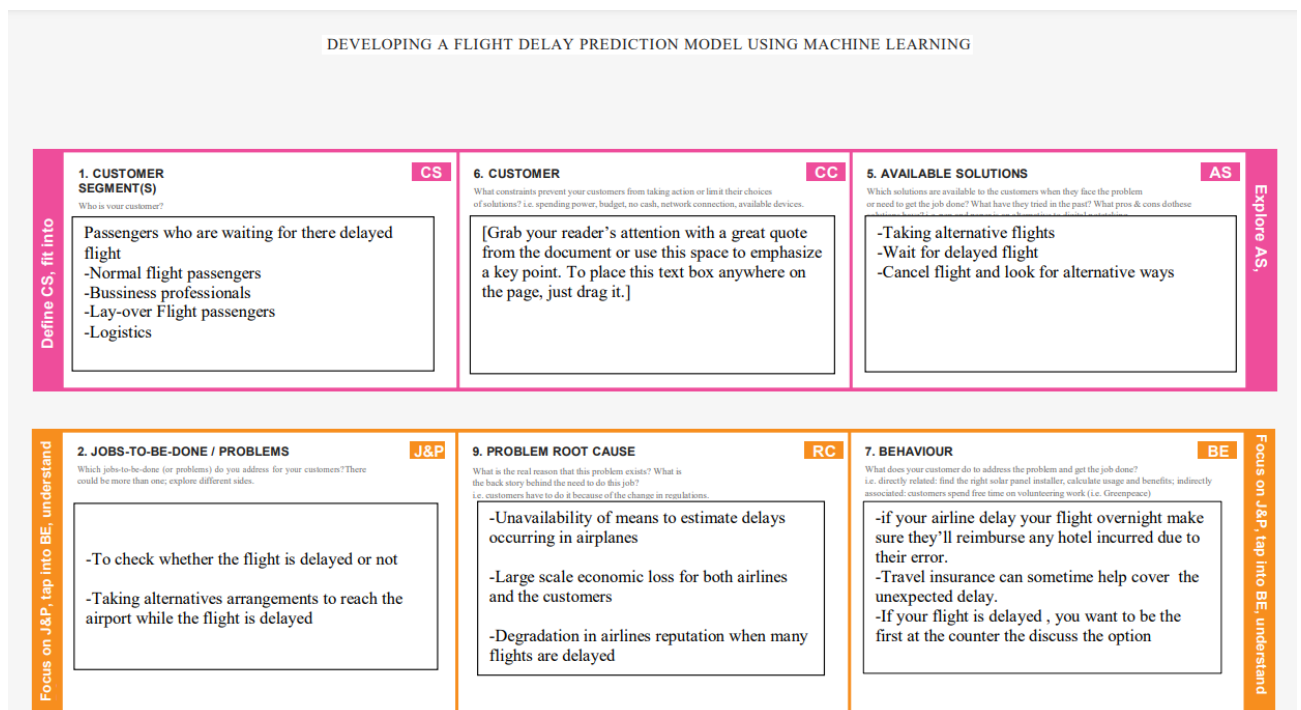
Nowadays, the aviation industry plays a crucial role in the world's transportation sector, and a lot of businesses rely on various airlines to connect them with other parts of the world. But, extreme weather conditions may directly affect the airline services by means of flight delays. To solve this issue, accurately predicting these flight delays allows passengers to be well prepared for the deterrent caused to their journey and enables airlines to respond to the potential causes of the flight delays in advance to diminish the negative impact.



### 3. Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Due to poor weather condition, some of technical problems occurred in aircraft leads the flight delay. so ,the travellers hates flying late. Due to this problem the air travellers count will decrease day by day. We need to fix this problem to improve airline operations and passenger satisfaction, which will result in a positive impact on the economy.
2.	Idea / Solution description	<b>Idea:</b> Collect the Passengers flight on-time data and process the collected data, and using some required Data Science algorithm to predict the delay of the flight's.
3.	Novelty / Uniqueness	<b>Uniqueness:</b> To collect the data's of flights and weather conditions to train our model to predict the results
4.	Social Impact / Customer Satisfaction	<b>Customer Satisfaction:</b> Passenger should be able to go to the correct his/her destination by correct time
5.	Business Model (Revenue Model)	♣ Application ♣ Website
6.	Scalability of the Solution	By using this type of application or a website we would know about the flight's delay. By adding some extra features to our home page to know the details about the flight and where it's been flying and when will we reach the destination

### 4. Problem Solution Fit



## 3. TRIGGERS

TR

- Cancellation of flights
- Getting Boring
- Guilty of wasting time
- Missing layover flight
- Uncertainty in deciding if the flight is delayed when they start late from the airport

## 4. EMOTIONS: BEFORE / AFTER

## Before:

- Worried about missing important events and missing layover flights
- If the flight is gonna be canceled
- Frustrated
- About the unexpected delay/cancellation
- Not knowing the news of delay beforehand
- About the weather

## After:

- Gets to enjoy the airline benefits
- Stay relaxed after getting a proper update from the airline
- Relieved if an alternate solution is found

## 10. YOUR SOLUTION

SL

The main aim is to develop an application that predicts the flights delay using a machine learning model with the data of flights and delays so far and estimate the time of delay taking spatial dependencies of flights into account.

## 8. CHANNELS of BEHAVIOUR

CH

## 8.1 ONLINE

- Check if a particular flight will be delayed and the estimated time of arrival
- Giving ratings and feedbacks for various flights so as to improve the app's performance in predicting further delays

## 8.2 OFFLINE

- Finding alternate flights in the airport / alternative travel routes
- Hotels near the airport can be visit for overnight stays during delay at night/midnight

# 4. REQUIREMENT ANALYSIS

## 1. Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub -Task)
FR-1	User Registration	<ul style="list-style-type: none"><li>Registration through User ID/Password</li><li>Registration through Gmail</li><li>Registration through Phone number / OTP</li></ul>
FR-2	User Confirmation	<ul style="list-style-type: none"><li>Confirmation via Email</li><li>Confirmation via OTP</li><li>Confirmation via Phone call</li></ul>
FR-3	User Login	<ul style="list-style-type: none"><li>Login with UserID/Password</li><li>Login with gmail</li><li>Login with phone number/OTP</li></ul>
FR-4	Support	<ul style="list-style-type: none"><li>Support option provided for queries and contact customer support</li></ul>
FR-5	Predication of delay	<ul style="list-style-type: none"><li>Requesting for prediction by providing details of delayed flight's.</li><li>Shows prediction results</li></ul>
FR-6	Trust ability of prediction	<ul style="list-style-type: none"><li>Gives the accuracy percentage about the prediction data</li></ul>
FR-7	Notify User	<ul style="list-style-type: none"><li>Notify user with the delay time, if they wish</li><li>Notify user about flight arrival before 45 mintues</li></ul>
FR-8	Get feed back	<ul style="list-style-type: none"><li>Get feedback about their user experience about the prediction data</li><li>Request give rating / support</li></ul>
FR-9	Log Out	<ul style="list-style-type: none"><li>Log out from the application</li></ul>

## 2. Non-functional Requirements:

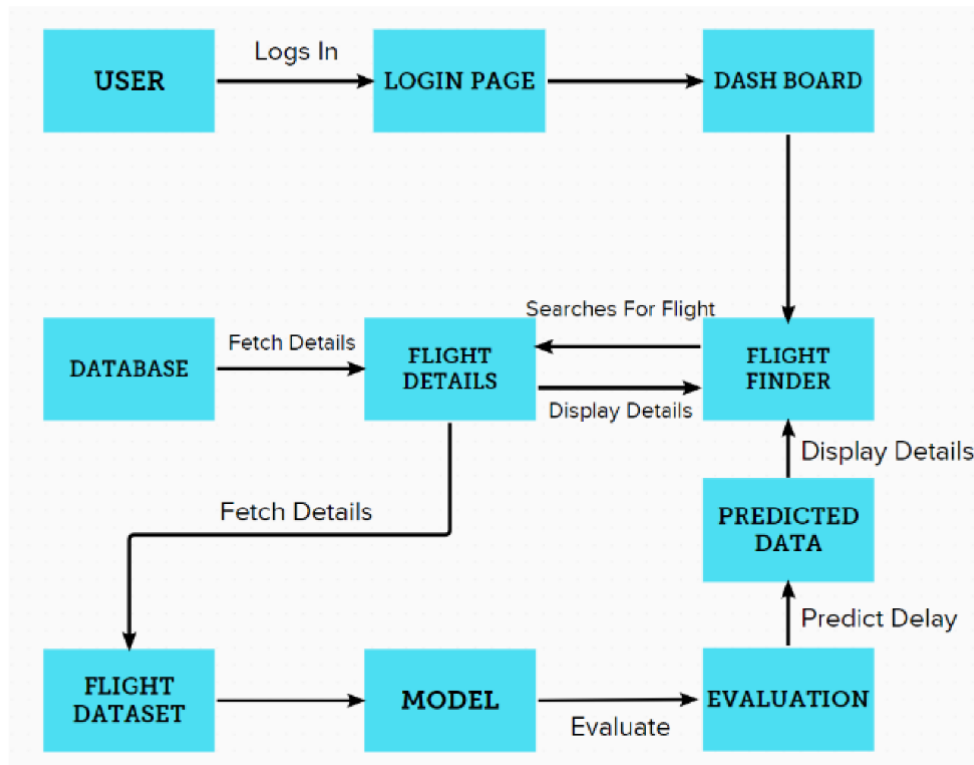
Following are the non-functional requirements of the proposed solution.

FR No.	Non -Functional Requirement	Descript ion
NFR-1	<b>Usability</b>	Web app is provided with smooth and user-friendly GUI.
NFR-2	<b>Security</b>	Data security of user is ensured with IBM Cloud security, login with your secured login details
NFR-3	<b>Reliability</b>	This web app have reliability by deploying in IBM Watson
NFR-4	<b>Performance</b>	50 request per second is handled.
NFR-5	<b>Availability</b>	99% avail with the help of IBM Cloud.
NFR-6	<b>Scalability</b>	It had high scalability by having ability to extend there computational resource when request came



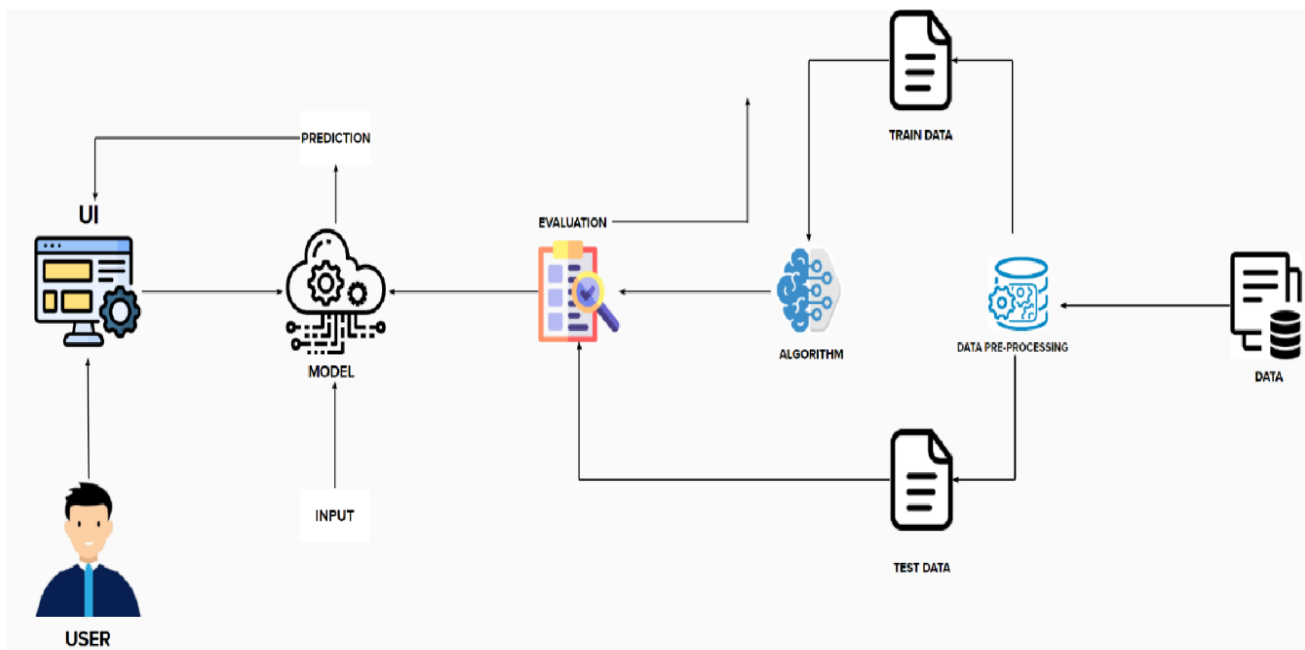
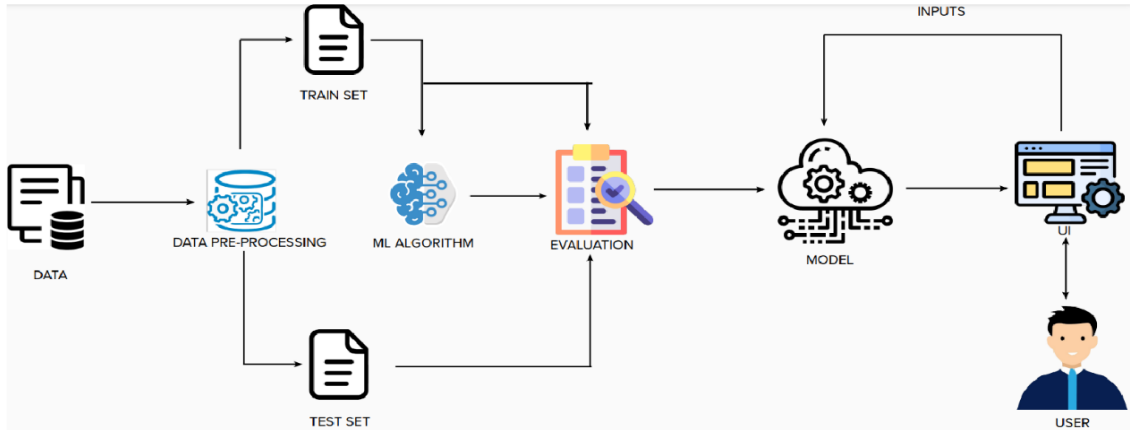
# 5. PROJECT DESIGN

## 1. Data Flow Diagram



## 2. SOLUTION AND TECHNICAL ARCHITECTURE

**Solution Architecture Diagram:**



### 3.User Stories

Use the below template to list all the user stories for the product.

#### User Stories

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, confirmation of password.	I can access my account/dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email after I register 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, Instagram, other social media	I can register & access the dashboard with Facebook/Instagram Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register and access the dashboard	Medium	Sprint-1

	Login	USN-5	As a user, I can log into the application by entering email & password	I can access the dashboard	High	Sprint-1
	Dashboard	USN-6	As a user, I can navigate through different pages using the dashboard	I can access various pages	High	Sprint-1
	Search	USN-7	As a user, I can search for flights for various locations	I can receive information on different flights for various locations	High	Sprint-2

	View	USN-8	As a user, I can view the details of flights	I will get the information such as flight no, departure and arrival time etc.,	High	Sprint-2
	Receive notifications	USN-9	As a user, I will receive notifications about the flight	I will get frequent updates of the flight's location	Low	Sprint-3
	Track	USN-10	As a user, I can track the location of my flight	I can track my flight	Medium	Sprint-3,4
Admin	GPS	USN-11	As an admin, I will need the location of flights	I can track my flight	High	Sprint-3,4
	Analyze	USN-12	As an admin, I will analyze the given dataset	I can analyze the dataset	High	Sprint-2
	Predict	USN-13	As an admin, I will predict the delays	I can predict the flight delays	High	Sprint-2
Customer (Mobile or Web user)	Arrival and Departure time of flights	USN-14	As a user, I can search for the details of a specific flight with flight number or name	I can find all the details of a flight	Medium	Sprint-2

		USN-15	As a user, I can find the accurate arrival and departure time of flights	I can find the actual timings of the flight	High	Sprint-3
Customer (Mobile or Web user)	Real time flight delay	USN-16	As a user, I can find exactly how long the flight will be delayed	I can get the accurate delayed time	High	Sprint-3
		USN-17	As a user, I can get real time timings that are updated every few seconds.	I can check the updated time	High	Sprint-3
Customer Care Executive	Helpdesk	USN-18	I can provide other alternative flights to the passenger's destination	I can check for alternative flights	High	Sprint-2
		USN-19	As a customer care executive, I can provide the contact details of the airlines to help the passenger to contact them in case of any query	I can give the airlines' phone number	Medium	Sprint-4
Customer Care Executive	Feedback	USN-20	I can collect all the feedback and suggestions that are given by the passengers, after using this application	I can record the feedbacks	Medium	Sprint-4
Administrator	Authentication	USN-21	As an admin, I can authenticate the registration and login credentials of the passengers.	I can validate the passengers' login	High	Sprint-1
		USN-22	As an admin, I ensure the security of the passengers' details	I maintain the security of user details	High	Sprint-4

## 6. PROJECT PLANNING AND SCHEDULING

## 1.Sprint Planning and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration and Login	USN-1	As a new user, I can register for the application by entering my email and my password.	2	High	Ravi Teja N.G, Rakesh S
Sprint-2	Confirmation email	USN-2	As a user, I will receive confirmation email once I have registered for the application	2	Medium	Vishwa , Ravi Chandran
Sprint-1	User login	USN-3	As a user, I can login into the application by entering the registered email-id and password	2	High	Ravi Teja N.G, Vishwa
Sprint-2	Admin Panel	USN-4	As an admin, I can authenticate the registration and login credentials of the passengers	2	High	Ravi Teja N.G, Rakesh S, Vishwa
Sprint-3	Arrival and Departure time of flights	USN-5	As a user, I can find all the details of a specific flight with its number or name	2	High	Ravi Teja N.G, Ravi Chandran
Sprint-3		USN-6	As a user, I can find exactly how long the flight will be delayed	2	High	Ravi Teja N.G, Rakesh S, Vishwa , Ravi Chandran
Sprint-4	Helpdesk	USN-7	As a customer care executive, I can provide the contact details of the airlines	1	Medium	Ravi Teja N.G, Rakesh S, Vishwa , Ravi Chandran

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4		USN-8	As a passenger, I can find alternative flights to the destination that are available	1	High	Ravi Teja N.G, Rakesh S, Vishwa , Ravi Chandran
Sprint-4	Feedback	USN-9	As a user, I can provide my suggestions and feedback for the improvement of the application	2	Medium	Ravi Teja N.G, Rakesh S, Vishwa , Ravi Chandran

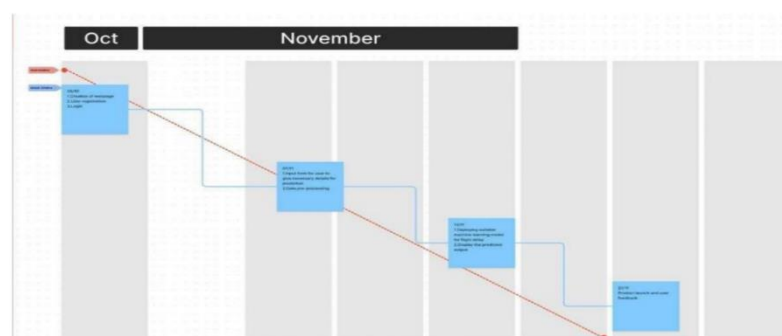
## 2. Sprint Schedule

### 3.

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	4	5 Days	30 October 2022	03 November 2022	4	03 November 2022
Sprint-2	4	5 Days	04 November 2022	08 November 2022	4	08 November 2022
Sprint-3	4	5 Days	09 November 2022	13 November 2022	4	13 November 2022
Sprint-4	4	6 Days	14 November 2022	19 November 2022	4	19 November 2022

## 4. Reports for JIRA

Burndown Chart:



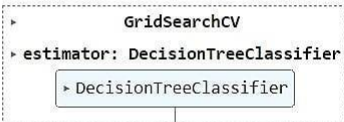
## 7. CODING AND SOLUTIONING

We started by preprocessing the data and then visualized the data to see patterns. Then we used Decision Tree as our model as it provided better prediction accuracy. We further tuned the model by hyperparameter tuning for our model using Grid Search CV. In Grid Search CV we can see that max depth of 6 and min samples split of 2 provided the best accuracy and then we cross validated the model using KFold function with the k value as 6 which gave an accuracy of 92%.

```
In [31]: from sklearn.model_selection import cross_val_score, KFold, GridSearchCV
kf = KFold(n_splits = 6, shuffle = True, random_state = 25)
params = {'max_depth': [4,5,6],
          'min_samples_split': [2,3,4],
          'criterion': ['gini', 'entropy', 'log_loss']}
```

```
In [32]: grid_cv = GridSearchCV(clf, params, cv = kf)
grid_cv.fit(X_train,y_train)
```

```
Out[32]:
```



```
  > GridSearchCV
  > estimator: DecisionTreeClassifier
    > DecisionTreeClassifier
```

```
In [35]: grid_cv.best_params_
```

```
Out[35]: {'criterion': 'entropy', 'max_depth': 6, 'min_samples_split': 2}
```

```
In [34]: cv_results = cross_val_score(clf, X_train,y_train, cv = kf)
print(cv_results)

[0.92552366 0.9193173 0.91925466 0.92934783 0.92313665 0.9060559 ]
```

---



# 8. TESTING

## 1.Test Cases

Test case ID	Feature Type	Component	Test Scenario	Pre-Requlite	Steps To Execute	Test Date	Expected Result	Actual Result	Status	Comments
PredictionPage_TC_D01	Functional	Home Page	Verify user is able to see the Prediction input page when user clicked on URL.	Any Latest Browser	1.Enter URL, and click go 2.Verify Prediction input displayed or not	Flask App	Prediction input page should display	Working as expected	Pass	
PredictionPage_TC_D02	UI	Home Page	Verify the UI elements in Prediction page.	Any Latest Browser	1.Enter URL, and click go 2.Verify Prediction page with below UI elements: a.Flight number b.Date c.Origin and Destination dropdown d.Flight Timings e.Prediction button	Flask App	Application should show below UI elements: a.Flight Number b.Date c.Origin and Destination dropdown d.Flight Timings e.Prediction button	Working as expected	Pass	
PredictionPage_TC_D03	Functional	Home page	Verify user is able to predict the flight with the proper details	Any Latest Browser	1.Enter URL, and click go 2.Enter valid flight number 3.Enter Valid date 4.Enter valid origin and destination 5.Enter valid flight timings 6.Click predict	Flight Number: 23587 Month: 12 Day: 12 Origin: ALT Destination: SEA Scheduled Dept Time: 1215 Actual Dept Time: 1236 Scheduled Arr Time: 1420	User should navigate to result page and input details are recieved properly	Working as expected	Pass	
PredictionPage_TC_D04	Functional	Home Page	Verify user is able to log into application with invalid input	Any Latest Browser	1.Enter URL, and click go 2.Enter valid flight number 3.Enter Valid date 4.Enter valid origin and destination 5.Enter valid flight timings 6.Click predict	Flight Number: 23587 Month: 12 Day: 12 Origin: ALT Destination: ALT Scheduled Dept Time: 1215 Actual Dept Time: 1236 Scheduled Arr Time: 1420	Application should show 'origin and destination airport cant be same airport' validation message.	Working as expected	Fail	The origin airport and the destination airport cannot be the same
PredictionPage_TC_D05	Functional	Home Page	Verify user is able to log into application with invalid input	Any Latest Browser	1.Enter URL, and click go 2.Enter valid flight number 3.Enter Valid date 4.Enter valid origin and destination 5.Enter valid flight timings 6.Click predict	Flight Number: 23587 Month: 14 Day: 12 Origin: ALT Destination: SEA Scheduled Dept Time: 1215 Actual Dept Time: 1236 Scheduled Arr Time: 1420	Application should show 'month value cant be more than 12' validation message.	Working as expected	Fail	The month value can't be more than 12
ResultPage_TC_D01	UI	Result page	Verify user is able to view the predictde results	Any Latest Browser	1.Enter URL, and click go 2.Enter the correct input values 3.Click the predict button 4.View Results page	Flight Number: 23587 Month: 12 Day: 12 Origin: ALT Destination: SEA Scheduled Dept Time: 1215 Actual Dept Time: 1236 Scheduled Arr Time: 1420	Application should show 'Flight is on time or Flight is delayed' message.	Working as expected	Pass	

## 2 .User Acceptance Testing

### 2. Defect Analysis

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	7	5	3	1	16
Duplicate	0	0	0	0	0
External	5	0	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	0	1	1
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	23	7	7	23	60

### 3. Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Model Evaluation	10	0	0	10
Client Application	20	0	0	20
Exception Reporting	5	0	0	5
Final Report Output	2	0	0	2

## 9.

# RESULTS

## 1. Performance Metrics

```
In [24]: #Model Evaluation
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
print(accuracy_score(y_test, pred))

0.9163899788711138

In [26]: print(confusion_matrix(y_test, pred))

[[2732  164]
 [ 113  304]]

In [27]: print(classification_report(y_test, pred))
```

	precision	recall	f1-score	support
0.0	0.96	0.94	0.95	2896
1.0	0.65	0.73	0.69	417
accuracy			0.92	3313
macro avg	0.80	0.84	0.82	3313
weighted avg	0.92	0.92	0.92	3313

## 10. ADVANTAGES AND DISADVANTAGES

The model can predict with an accuracy of 91% which will help the passengers to plan their trip accordingly. It is user friendly and easy to interact with.

The disadvantage is that the model is trained only for five airports in the USA. This model can't be used to predict the flight delay for any other airports.

## 11.

## FUTURE WORKS

We can expand the projects by adding more airports so that the flight delay can be predicted all around the world. We can also take in the weather conditions as input so that we can predict the delay even accurately. We can also predict the time of the delay.

## 12. CONCLUSION

In this project, we used flight data such as dates, origin and destination airport, scheduled and actual departure time to predict flight delay. Our result shows that the Decision Tree method yields the best performance. In the end, our model provides an accuracy of 92%. As a result, there can be additional features related to the causes of flight delay that are not yet discovered using our existing data sources. So more data sources will help us to create a better model and will help us to predict flight delays based on it.

# 13.APPENDIX

## SOURCE CODE

```
#Importing the necessary librariesimport
numpy as np
import pandas as pd
import matplotlib.pyplot as pltimport
seaborn as sns
%matplotlib inline
#Loading the dataset
df = pd.read_csv("flightdata.csv")
pd.set_option('display.max_columns', None) df.head()
#Dropping unnecessary columns df.drop('Unnamed: 25',
axis = 1, inplace = True)#Dataset Info
df.info()
#Handling missing values
df.isnull().sum()
#Dropping the missing values
df.dropna(subset=['DEP_TIME','ARR_DELAY', inplace = True) #Data
Visualization
ax = sns.countplot(y = df['ORIGIN'], order = df['ORIGIN'].value_counts().index);
ax.set_title(" Airports w.r.t Depature Flights", fontsize = 16); ax.set_xlabel("Number
of Flights", fontsize = 14);
ax.set_ylabel("Airport Code", fontsize = 14);
ax.bar_label(ax.containers[0], label_type = 'center', color = 'white', size = 14); ax =
sns.countplot(y = df['DEST'], order = df['DEST'].value_counts().index);
ax.set_title(" Airports w.r.t Arrival Flights", fontsize = 16); ax.set_xlabel("Number
of Flights", fontsize = 14);
ax.set_ylabel("Airport Code", fontsize = 14);
ax.bar_label(ax.containers[0], label_type = 'center', color = 'white', size = 14); fig, ax =
plt.subplots(1, 2, figsize = (10,10))
ax[0].pie(df['DEP_DEL15'].value_counts(), labels = ['On Time', 'Delayed'], autopct =
'%1.2f%%',startangle = 90, explode = (0,0.1));
ax[0].title.set_text("Ratio of Delayed Departure Flights"); ax[1].pie(df['ARR_DEL15'].value_counts(),
labels = ['On Time','Delayed'], autopct = '%1.2f%%',startangle = 90, explode = (0,0.1));
ax[1].title.set_text("Ratio of Delayed Arrival Flights");
sns.heatmap(df.corr());
new_df = pd.get_dummies(df, columns = ['ORIGIN','DEST'])
```

```

#Splitting into independant and dependant values X=
new_df[['MONTH','DAY_OF_MONTH','DAY_OF_WEEK','ORIGIN_ATL','ORIGIN_DTW','ORIGIN_JFK',
'ORIGIN_MSP','ORIGIN_SEA','DEST_ATL','DEST_DTW','DEST_JFK','DEST_MSP','DEST_SEA','CRS
_DEP_TIME','DEP_TIME','DEP_DEL15','CRS_ARR_TIME']]
y = new_df['ARR_DEL15']
#Splitting into training and testing data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30) #Model
Building
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(max_depth = 4, min_samples_split = 4, random_state = 25) #Model
Training
clf.fit(X_train, y_train)
pred = clf.predict(X_test)
#Model Evaluation
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
print(accuracy_score(y_test, pred))
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred)) #Saving the
model
import pickle
pickle.dump(clf, open('flightclf.pkl','wb'))

```

## **GITHUB**

<https://github.com/IBM-EPBL/IBM-Project-1751-1658411466>

## **DEMO LINK**

<https://drive.google.com/file/d/1dLp9BjoBfYFn8b1589Zfq4gr2J3b84WV/view?usp=sharing>