DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

An IBM NALAIYA THIRAN PROJECT 2022

TEAM ID – PNT2022TMID01346

TEAM MEMBERS

MONISHA D ***(211419205111)***

DEEPIKA E ***(211419205036)***

MOHANA PRIYA B ***(211419205105)***

SAI KRITHIKA R ***(211419205139)***

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

PANIMALAR ENGINEERING COLLEGE, POONAMALLE

CHENNAI – 600 025

TABLE OF CONTENT

1. INTRODUCTION

1.1 PROJECT OVERVIEW

The number of planes that fail to take off on time likewise rises as more people opt to travel by air. The airline industry experiences financial challenges as a result of this growth, which also worsens the overcrowding at airports. Delays in air travel are a sign of the aviation system's inefficiency. Both airline firms and their customers pay a hefty price for it. The Entire Delay Impact Study estimated that the total cost of air transportation delays to passengers and the airline sector in the US in 2007 was $32.9 billion, resulting in a $4 billion decline in GDP. As a result, anticipating delays can enhance airline operations and passenger pleasure, which will benefit the economy.

In this project we are using machine learning by predicting the flight arrival and departure time in the destination location and to help the passenger to save their time and can plan accordingly. The passenger can predict the delay of flight by entering their flight details in the given website and checking whether the flight is delay or not. This process will be easy to passenger to be relax and managing the time accordingly. The algorithm used for developing the flight delay prediction using machine learning is decision tree.

1.2 PURPOSE

- Flight delays are unavoidable, and they significantly affect the carriers' profits and losses. For airlines, estimating flight delays correctly is essential since the data may be used to boost client happiness and revenue for airline agencies.
- It can save time for passenger to be prepared and can plan their schedules to reach the airport.
- We set out to develop a model to forecast long-term airline delays because travellers might benefit greatly from knowing the likelihood of a flight delay or cancellation.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

[1] Flight Delay Prediction System

Authors: Mrs Yogita Borse , Dhruvin Jain, Shreyash Sharma, Viral Vora, Aakash Zaveri

Year: 2020

One of the key business issues that airlines face is that the vital prices that are related to flights being delayed because of natural occurrences and operational shortcomings. Delays in departure can occur due to bad weather conditions, seasonal and holiday demands, airline policies, technical issue such as problems in airport facilities, luggage handling and mechanical apparatus, and accumulation of delays from preceding flights.

The models developed during this system may be applied to predict the incidence of flight delay at airports. This issue can be reduced by developing the flight delay prediction tool which can be developed using statistical analysis, probabilistic models and classification approach or methods like Naïve Bayes Classification, Bayesian Network Algorithm, decision tree, logistic regression etc. We propose a flight delay prediction system which focuses mainly on predicting delay of a flight based on the weather situation .This paper presented the need to develop a system to predict the delay in flights along with its methodology.

The paper gives details about the range of different methodology that is used or can be used to find out the delay in flights. As flight delay cost a lot to the airlines as well as passengers in financial and environmental terms, flight delay is the talk of the hour.

[2] Flight Delay Prediction Based with Machine Learning.

Authors: Irmak Hatipoglu, Omur Tosun, Nedret Tosun

Year: 2022

The delay of a planned flight causes many undesirable situations such as cost, customer satisfaction, and environmental pollution. There is only one way to prevent these problems before they occur, and that is to know which flights will be delayed. The aim of this study is to predict delayed flights. For this, the use of machine learning techniques, which have become widespread with the development of computer capacities and data storage systems, is preferred.

Estimations are made with three up-to-date techniques XGBoost, LightGBM, and CatBoost techniques based on Gradient Boosting from machine learning techniques. The bayesian technique is used for hyper-parameter settings. The results are analyzed and shared with and without SMOTE. Hence, these are the following methods we use in this paper.

The application of machine learning techniques to anticipate flight delays is new, but it has a lot of potential. Companies will be able to avert problems before they develop if delays are correctly estimated, which can generate plenty of issues. As a result, concrete advantages such as lower costs and higher customer satisfaction will emerge. Improvements will be made at the most vulnerable place in the aviation business. This paper has developed a new approach for airline companies to detect delayed flights. In order to achieve these different approaches, which are XGBoost, LightGBM, and CatBoost, were used.

[3] Flight Delay Predictions and the Study of Its Casual Factors Using Machine Learning Algorithms.

Authors: Cho Yin Yiu, Kam K.H. Ng, Kin Chung Kwok, Wing Tung Lee, Ho Tung Mo

Year: 2021

The term ‘flight delay’ is the measure of actual arrival/departure time compared to the scheduled arrival/departure time, while the actual time is later than the scheduled time. However, different stakeholders may have different interpretations. The Federal Aviation Administration defines a flight delay as having an actual arrival/departure time that is 15 minutes later than the scheduled arrival/departure time. Much research attempted to deal with flight delay issue by formulating various models to predict their occurrence.

In this paper we adopted several machine learning algorithms to predict flight delay and compared their performances in the case of the HKIA. The analysis concluded that the ANNs algorithm is the most effective in predicting flight delay. During flight planning, these important contributing factors could be emphasized.

A dataset with a longer duration might aid in further development. Some data is also missing due to flight cancellation, etc., causing reduction in accuracy. The current model could be improved to provide a comprehensive analysis and accurate prediction of flight delay. The number of take-off and landing flights shall also be further balanced to enhance the robustness of the results.

[**4] Machine Learning Model – Based Prediction of Flight Delay.**

Authors: N Lakshmi Kalyani, Jeshmitha G, Bindu Sri Sai U, Samanvitha M, Mahesh J, Dr B.V. Kiranmayee

Year: 2019

With the air travel increasing rapidly there is a serious problem of flight delays for both airlines and passengers. Passengers not only lose their time but also their trust in airlines. This will result in a huge economic loss to the airline companies and Airlines lose their reputation as well. Thus, proper monitoring and prediction of flight delays are very important.

A more precise prediction model can aid in optimizing flight operations which benefit both passengers and airlines equally. Considering all the parameters that are the cause for the delay, weather found to affects the delay to a great extent and hence used it as a contributing aspect to predict the delay of the flight.

The algorithm that was used here was XGBoost classification algorithm as its speed of execution and model performance are very good. For knowing the delay time, domestic flight data is trained using Linear Regression algorithm which then predicts by how much time the flight will get delayed.

In This paper the aims are to predict the flight's delay along with the estimation of delay time in minutes using machine learning algorithms namely Decision Tree Algorithm (XGBoost) and Linear regression. Data set of both flights and weather will be taken to compare with the given inputs and validate them by applying classification and Regression concepts of Machine Learning.

[5] Applying Machine Learning To Aviation Big Data For Flight Delay Prediction.

Authors: Yushan Jiang , Yongxin Liu ,Dahai Liu ,Houbing Song

Year:2020

Nowadays big data analytics and machine learning have been applied successfully in many domains, their applications in aviation are limited. This paper presents a comprehensive study of flight delay spanning data pre-processing, data visualization and data mining, in which we develop several machine learning models to predict flight arrival delays. Two data sets were used, namely Airline OnTime Performance (AOTP) Data and Quality Controlled Local Climatological Data (QCLCD). Data visualization is a process which is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts. It helps users in analyzing a large amount of data in a simpler way. It makes complex data more accessible, understandable, and usable. This paper aims to recognize useful patterns of the flight delay from aviation data and perform accurate delay prediction. The best result for flight delay prediction (five classes) using machine learning models is 89.07% (Multilayer Perceptron). A Convolution neural network model is also built which is enlightened by the idea of pattern recognition and success of neural network method, showing a slightly better result with 89.32% prediction accuracy. This paper provides a comprehensive aviation data analytic regarding flight delay. QCLCD and AOTP data are used to construct a new dataset with both flight information and weather condition. Then this dataset is further explored and some useful pattern toward flight delay is shown.

[6] A Novel Integration Platform to Reduce Flight Delays in The National Airspace System.

Authors: Chuyang Yang, Zachary A. John H. Mott

Year: 2021

Roughly 20% of passengers' total travel time is due to such delays, causing $35 billion annually in lost revenue and impacting not only the airline industry, but the retail, lodging, restaurant, and tourism industries, as well. The Federal Aviation Administration's effort in aiding decision-making at airports is readily apparent in the Next Generation Air Traffic Control (NextGen) System's System-Wide Information Management (SWIM) program, and in-flight delay information from the FAA Air Traffic Control System Command Center (ATCSCC), Artificial Neural Networks (ANN), Cloud Computing The term "cloud" was used to refer to the platform for distributed computing which allows users to minimize up-front information technology (IT) infrastructure costs but also get their applications up and running faster. . By integrating various databases with existing NextGen's SWIM and FAA CDM and GDP programs and harnessing remote cloud computing of deep learning algorithms, precise and accurate flight delay forecasts are generated. Allowing for the full realization of potentials in schedule optimization, emission reduction, and resource utilization, the delay predictions provided by the proposed system could significantly grow airports' capabilities through improved operational efficiency.

[7] Airline Delay Prediction Using Machine Learning and Deep Learning Techniques.

Authors: Devansh Shah, Ayushi Lodaria, Danish Jain, Lynette D'Mello.

Year: 2020

In this paper, they have tried to predict flight delays using different machine learning and deep learning techniques. By using such a model, it can be easier to predict whether the flight will be delayed or not. Factors like 'WeatherDelay', 'NASDelay', 'Destination', 'Origin' play a vital role in this model. Using machine learning algorithms like Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), the f1-score, precision, recall, support and accuracy have been predicted. They have employed Long Short-Term Memory RNN architecture trying to prove that the accuracy increases with deeper architectures. To train the model, stochastic gradient descent (SGD) algorithm is utilized.

In this paper they have shown that machine learning and deep learning algorithms can be efficaciously used to predict flight delays. Therefore, in this paper, by using Random Forest based model the flight delay can be predicted, which will be beneficial for all the entities involved i.e., airport, airline and passengers. Therefore, the analysis of flight delay carried out through this paper is based solely on scientific parameters and is of paramount importance in the aviation industry.

[8] Airline delay prediction by machine learning algorithms.

Authors: H. Khaksar, A. Sheikholeslami

Year: 2019

In this paper, they have implemented fight delay prediction through the proposed approaches that were based on machine learning algorithms. The parameters that enabled effective estimation of delay were identified and then, Bayesian modeling, decision tree, cluster classification, random forest, and hybrid method were applied to estimate the occurrences and magnitude of delay in a network. These methods were tested on a US fight dataset and then, referred for a large Iranian airline network. Results showed that the parameters affecting delay in US networks were visibility, wind, and departure time, whereas those affecting delay in the Iranian airline flights were f eet age and aircraft type.

The analysis and design of complicated and large-scale systems with many variables require new methods that can identify, classify, and analyze voluminous data. Accordingly, researchers put forward data mining approaches to identifying, collecting, classifying, and ranking or generating and storing valuable information from such databases. FDP methods, namely decision tree, cluster, Bayesian, random forest, and hybrid classification, were proposed in this research. These approaches were examined on the basis of real datasets on US and Iranian flight networks. The results indicated that the hybrid approach exhibited a performance superior to those of the other methods and was therefore adopted as the FDP model.

2.2 REFERENCES

[1] Flight Delay Prediction System

Authors: Mrs Yogita Borse , Dhruvin Jain, Shreyash Sharma, Viral Vora, Aakash Zaveri

Year: 2020

[2] Flight Delay Prediction Based with Machine Learning.

Authors: Irmak Hatipoglu, Omur Tosun, Nedret Tosun

Year: 2022

[3] Flight Delay Predictions and the Study of Its Casual Factors Using Machine Learning Algorithms.

Authors: Cho Yin Yiu, Kam K.H. Ng, Kin Chung Kwok, Wing Tung Lee, Ho Tung Mo

Year: 2021

[4] Machine Learning Model – Based Prediction of Flight Delay.

Authors: N Lakshmi Kalyani, Jeshmitha G, Bindu Sri Sai U, Samanvitha M, Mahesh J, Dr B.V. Kiranmayee

Year: 2019

[5] Applying Machine Learning To Aviation Big Data for Flight Delay Prediction.

Authors: Yushan Jiang, Yongxin Liu, Dahai Liu, Houbing Song

Year: 2020

[6] A Novel Integration Platform to Reduce Flight Delays in The National Airspace System.

Authors: Chuyang Yang, Zachary A. John H. Mott

Year: 2021

[7] Airline Delay Prediction Using Machine Learning and Deep Learning Techniques.

Authors: Devansh Shah, Ayushi Lodaria, Danish Jain, Lynette D'Mello.

Year: 2020

[8]Airline delay prediction by machine learning algorithms.

Authors: H. Khaksar, A. Sheikholeslami

Year: 2019

2.3PROBLEM STATEMENT

Create a problem statement to understand your customer's point of view. The Customer Problem Statement template helps you focus on what matters to create experiences people will love.

A well-articulated customer problem statement allows you and your team to find the ideal solution for the challenges your customers face. Throughout the process, you'll also be able to empathize with your customers, which helps you better understand how they perceive your product or service.

PROBLEM STATEMENT 1

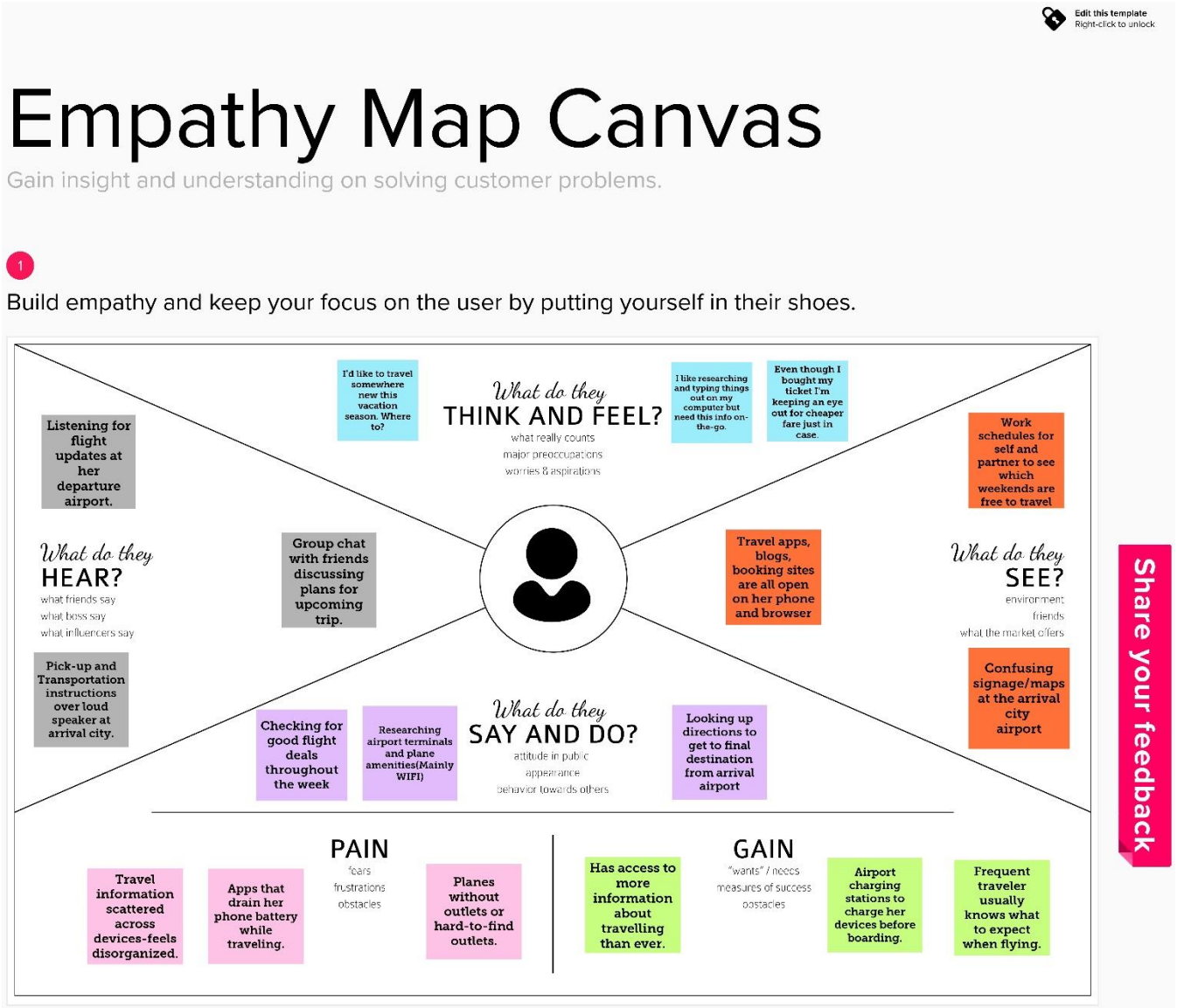
PROBLEM STATEMENT 2

Problem Statement (PS)	Iam (Customer)	I’m trying to	But	Because	Which makes me feel
PS-1	Traveler	Boarding Flight	Takes Long Time	Flight Delay Due to Several Reasons	Waste of Time and Feeling Irritated
PS-2	Airline	Schedule Flight	Delay of Few Flights	Flight Delay Due to Several Reasons (Air Traffic Control, Bad weather)	Difficult to Manage Schedules

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAY CANVAS

Teams can utilise an empathy map as a collaborative tool to learn more about their clients. An empathy map can depict a group of users, such as a consumer segment, in a manner similar to user personas. Teams can better understand a principal user's motivations, issues, and user experience by using the empathy map, which depicts that person. Empathy mapping is a straightforward yet powerful workshop that can be used with a wide range of users, including stakeholders, specific use cases, or entire teams.



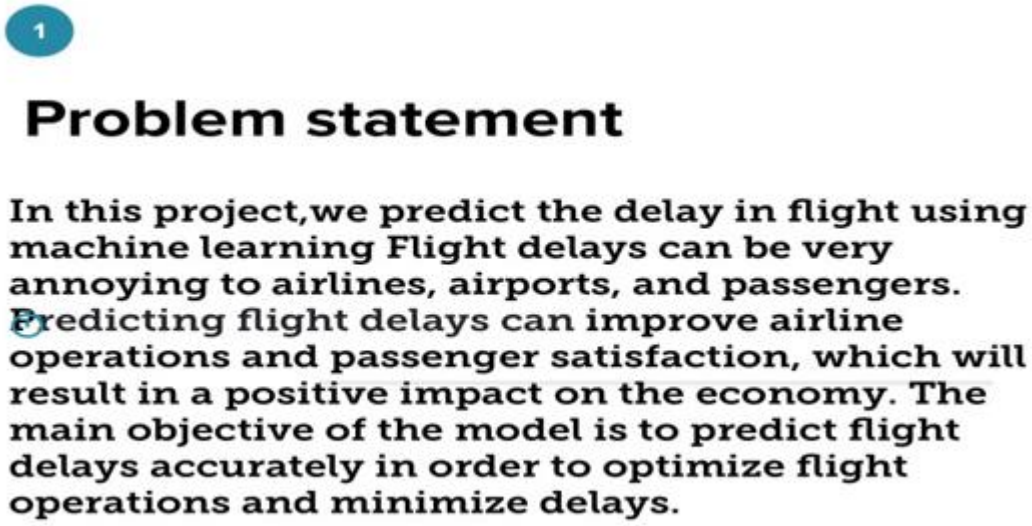
3.2 IDEATION & BRAINSTROMING

Brainstorm is nothing but to suggest ideas for the project before starting the project. The process of brainstorming can assist the group focus its ideas and find solutions.

Project Ideas are where you begin documenting proposals for future research grant applications. During this stage, you are recording important project-related details as well as locating collaborators, potential funders, budget details, and project-related metadata. You can also make tasks and assign them to project-related people.

An administrative grouping of projects is known as a project group. Project groups make it possible for administrative operations to have an impact on several projects and users at once.

STEP-1: PROBLEM STATEMENT



STEP-2: BRAINSTORM, IDEA LISTING AND GROUPING

3

Group ideas

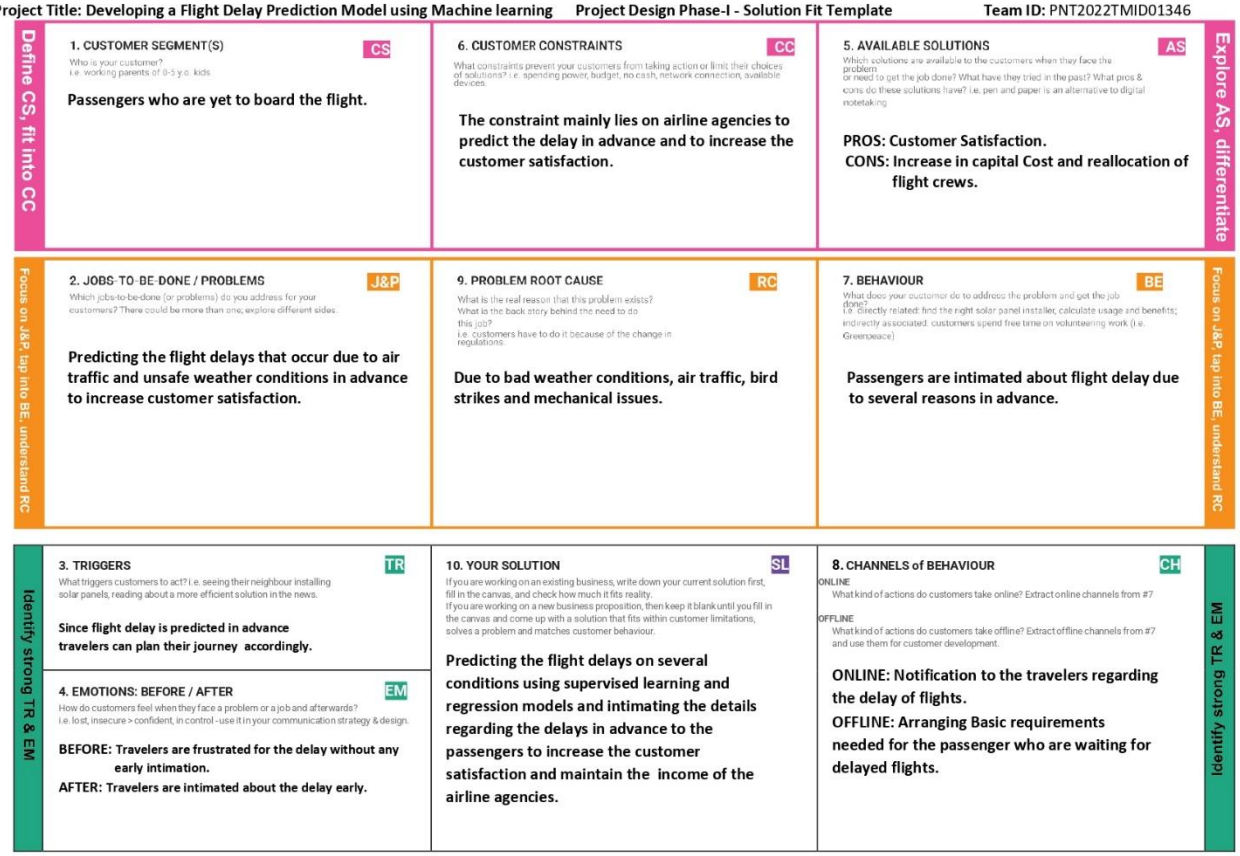
STEP-3: IDEA PRIORITIZATION

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	In this project, we predict the delay in flight using machine learning Flight delays can be very annoying to airlines, airports, and passengers. Predicting flight delays can improve airline operations and passenger satisfaction, which will result in a positive impact on the economy. The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays.
2.	Idea / Solution description	To predict flight delay that occurs due to bad weather conditions and air traffic using supervised learning in machine learning.
3.	Novelty / Uniqueness	Since large number of data is being trained using supervised learning and regression models the flight delay prediction is made precise and accurate.
4.	Social Impact / Customer Satisfaction	By predicting the flight delay in advance the passengers can plan accordingly.
5.	Business Model (Revenue Model)	Operating flights according to the schedule reflect an efficient airline since flight delay is predicted in advance the cost of cancellation reduces.
6.	Scalability of the Solution	It increase the customer satisfaction and maintaining the schedules by the airline agencies and increases the income.

3.4 PROBLEM SOLUTION FIT

Project Title: Developing a Flight Delay Prediction Model using Machine learning **Project Design Phase-I - Solution Fit Template** **Team ID:** PNT2022TMID01346



4. REQUIREMENT ANALYSIS

4.1 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	Registration through Form Registration through Gmail Registration through LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User login	Login through form
FR-4	Forgot password	TP via email
FR-5	Book Flights	The flight ticket booking is done and receipt of bookingis sent to email of user
R-6	Request Cancellation	The user wants to cancel the ticket reserved due to Delay

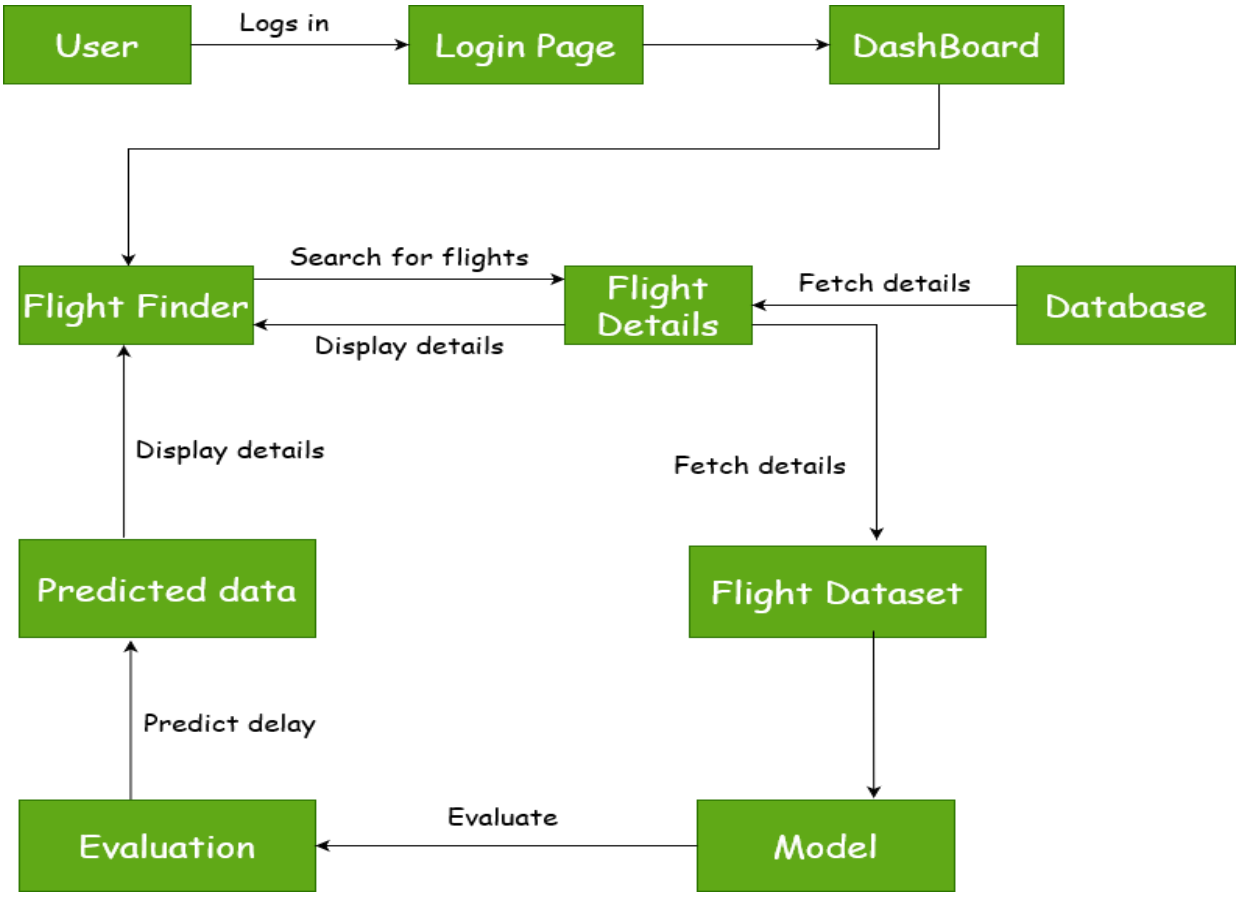
4.2 NON-FUNCTIONAL REQUIREMENTS

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

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	this project, we use flight data, weather and demand data to predict flight departure delay
NFR-2	**Security**	flight is about to leave and a passenger is still at security the airline decides whether to wait for the passenger or not
NFR-3	**Reliability**	The reason you want to arrive to the airport before your flights original time because flight delays are usually not reliable estimates
FR-4	**Performance**	Performance defines how fast a software system or a particular piece of it responds to certain users' actions under a certain workload. The system should provide accurate delays of the Flight
NFR-5	**Availability**	Availability describes how likely the system is accessible to a user at a given point in time. 24/7 available
NFR-6	**Scalability**	Scalability assesses the highest workloads under which the system will still meet the performance Requirements. Can handle multiple users at the same time.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM



5.2 TECHNICAL & SOLUTION ARCHITECTURE

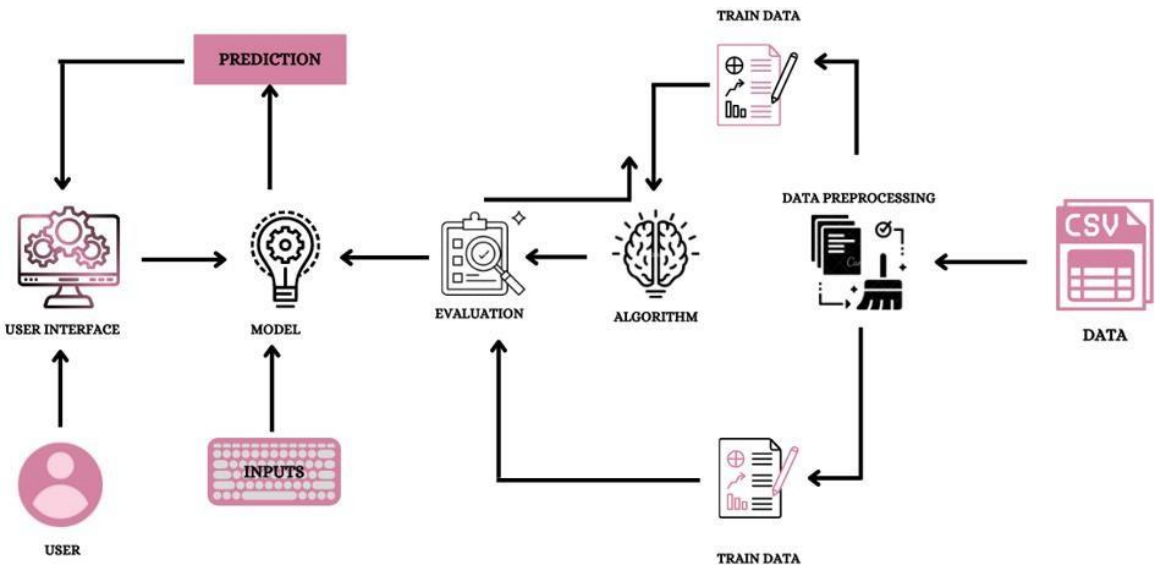
The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

TABLE-1: COMPONENTS & TECHNOLOGIES

S.No	Component	Description	Technology
1.	User Interface	User interacts using web application Web UI	HTML, CSS, JS, Flask
2.	Application Logic-1	Logic for a process in the application	Python (numpy libraries)
3.	Application Logic-2	Logic for a process in the application	Python, java
4.	Application Logic-3	Logic for model building	Machine learning model Decision tree classifier .
5.	Data Preprocessing	To clean data	Panadas, numpy
6.	Database	Database contains the user information andflight details	MySQL
7.	File Storage	File storage requirements	IBM Block Storage or Other storage
8.	External API-1	Time Door is a REST API for statistical insightsinto time series data.	IBM Weather API, Http, Timedoor.
9.	External API-2	Purpose of External API used in theapplication	Email API, etc
10.	Machine Learning Model	Ensemble of multiple decision trees provide better classification accuracy. Random forest obtains a class vote for decision tree.	Random forest, Decision treeclassifier.
11.	Infrastructure (Server / Cloud)	Application Deployment on Local System	IBM cloud

TABLE-2: APPLICATION CHARACTERISTICS

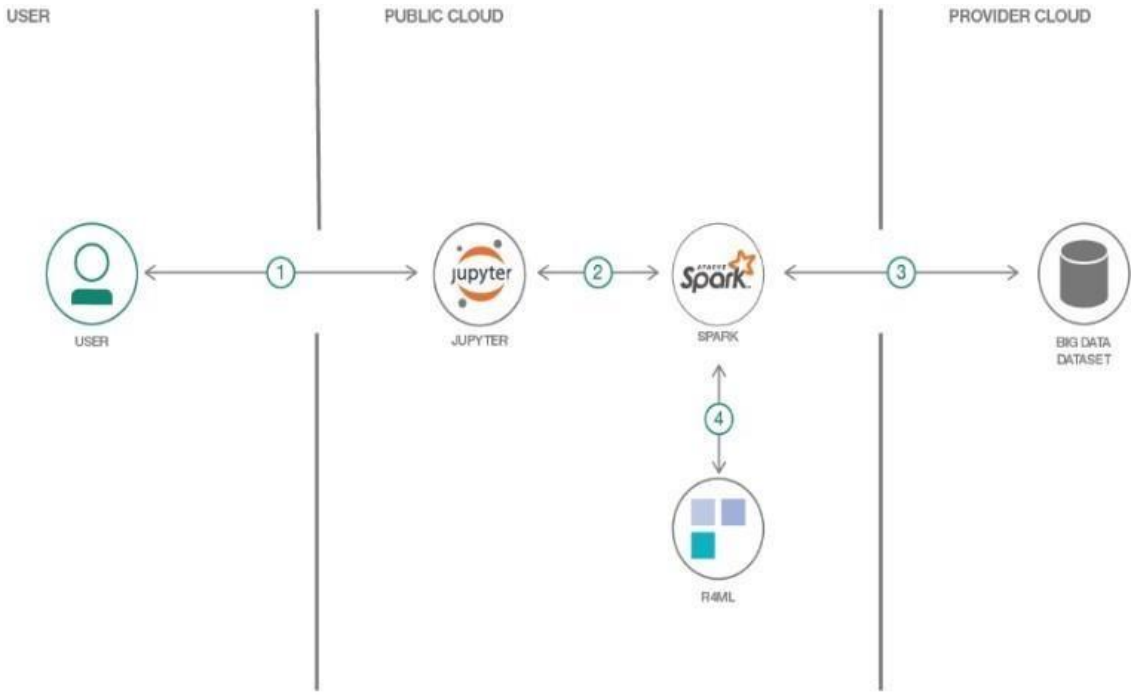
S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	It is used to build almost any type of website — from content management systems and wikis, through to social networks and newssites	Python-Flask
2.	Security Implementations	It is used to identify the threats in the system. To measure the potential vulnerabilities of thesystem	Encryptions, Risk assessment, Authentication
3.	Scalable Architecture	To determine the user limit for the webapplication.	Scalability testing
4.	Availability	Running an application for a planned period oftime, collecting failure events and repair times	Load testing, endurance testing.
5.	Performance	Determines the behaviour of web application when it receives extreme variations in Traffic.	Spike testing, stress testing.



SOLUTION ARCHITECTURE

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.



5.3 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, and confirming my password.	I can accessmy account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the Application.	can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application	I can register & access the dashboard with	Low	Sprint-2

		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the app by entering email and 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 various pages	High	Sprint-2
	Search	USN-7	As a user, I can search for flights with destination location.	I can receive information on various flights.	High	Sprint-2
	View	USN-8	As a user, I can view the details of the flights.	get the information such as flight no, departure and arrival time, etc.	High	Sprint-3
	Receive notifications	USN-9	As a user, I will receive notifications about the flight.	I get frequent updates of the flight's location.	Low	Sprint-3

Admin	GPS	USN-10	As an admin, I need the location of Flights	I can track the flights	High	Sprint-4
	Analyse	USN-11	As an admin, I will analyse the given dataset.	I can analyse the dataset.	High	Sprint-2

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

The definition of a sprint is a dedicated period of time in which a set amount of work will be completed on a project. Its part of the agile methodology, and an agile project will be broken down into a number of sprints, each sprint taking the project closer to completion. In the scrum process, sprint planning marks the beginning of the sprint. Sprint planning's goal is to specify what can be completed in a sprint and how it will be done. The entire scrum team collaborates on sprint planning.

Use the below template to create product backlog and sprint schedule.

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	User Registration	USN-1	As a user, I can register for the application by entering email, password, and confirming password.	2	High	1. SAI KRITHIKAR 2. DEEPIKA E
Sprint-2	User Confirmation	USN-2	As a user, I will receive confirmation email oncelhave registered for the application	1	High	1. MONISHA D 2. MOHANA PRIYA
Sprint-2	User login	USN-3	As a user, I Can login with the credentials Ihave used during registration.	2	Medium	1. MONISHA 2. MOHANA PRIYA B
Sprint-3	Forgot password	USN-4	As a user, I can reset the password via registered email or mobile number.	2	Medium	1. MONISHA 2. SAI KRITHIKAR
Sprint-3	Book Flights	USN-5	As a user, I can log into the application by entering email & password and book flights.	1	High	1. DEEPIKA E 2. MOHANA PRIYA B
Sprint -4	Request Cancellation	USN -6	As a user I can request cancellation of flight by Using the login credentials which is used during registration.	2	Medium	1. SAI KRITHIKAR 2. DEEPIKA E

6.2 SPRINT DELIVERY SCHEDULE

A sprint schedule is a written description of the entire sprint planning process. It's one of the initial steps in the agile sprint planning process, and it calls for sufficient investigation, preparation, and coordination. It centres on a product backlog, which is a list of open requests for development and iteration. A burn down chart, which displays how rapidly a team is progressing through a customer's user stories, is a project management chart. This agile tool records the description of a feature from the viewpoint of the end user and compares the overall effort to the quantity of work for each agile sprint.

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	2	6 Days	24 Oct 2022	29 Oct 2022	2	29 Oct 2022
Sprint-2	3	6 Days	31 Oct 2022	05 Nov 2022	3	1 Nov 2022
Sprint-3	3	6 Days	07 Nov 2022	12 Nov 2022	3	13 Nov 2022
Sprint-4	2	6 Days	14 Nov 2022	19 Nov 2022	2	19 Nov 2022

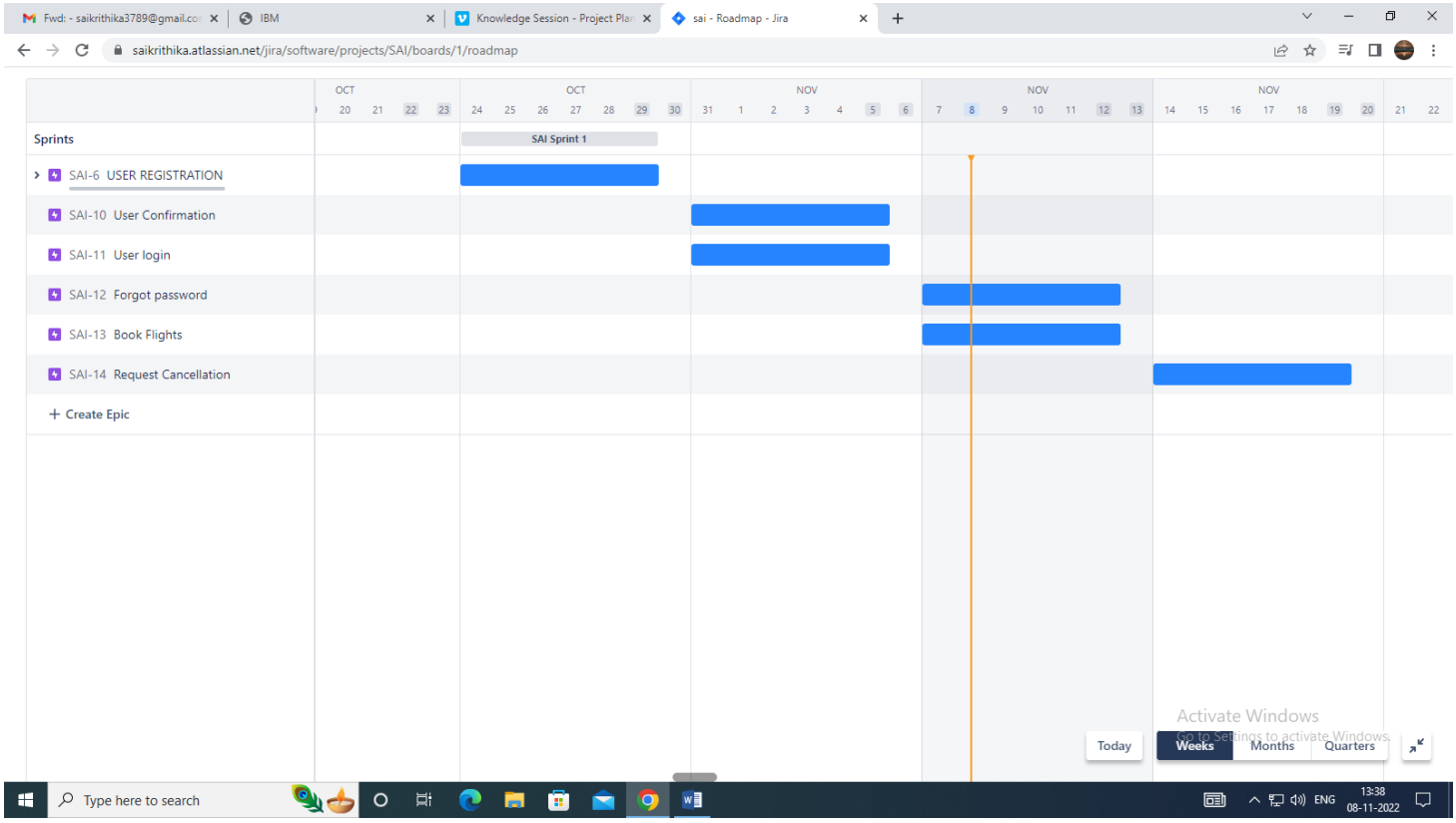
VELOCITY

AV = SPRINT DURATION VELOCITY

= 24/10

= 2.4

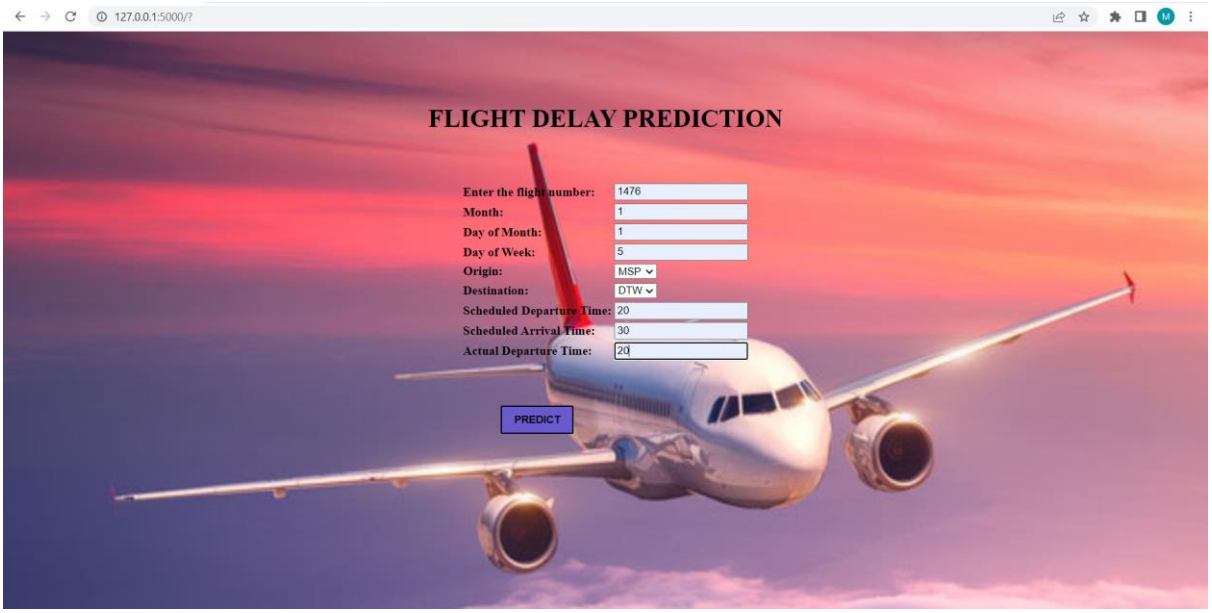
6.3 REPORTS FROM JIRA



7. CODING & SOLUTION

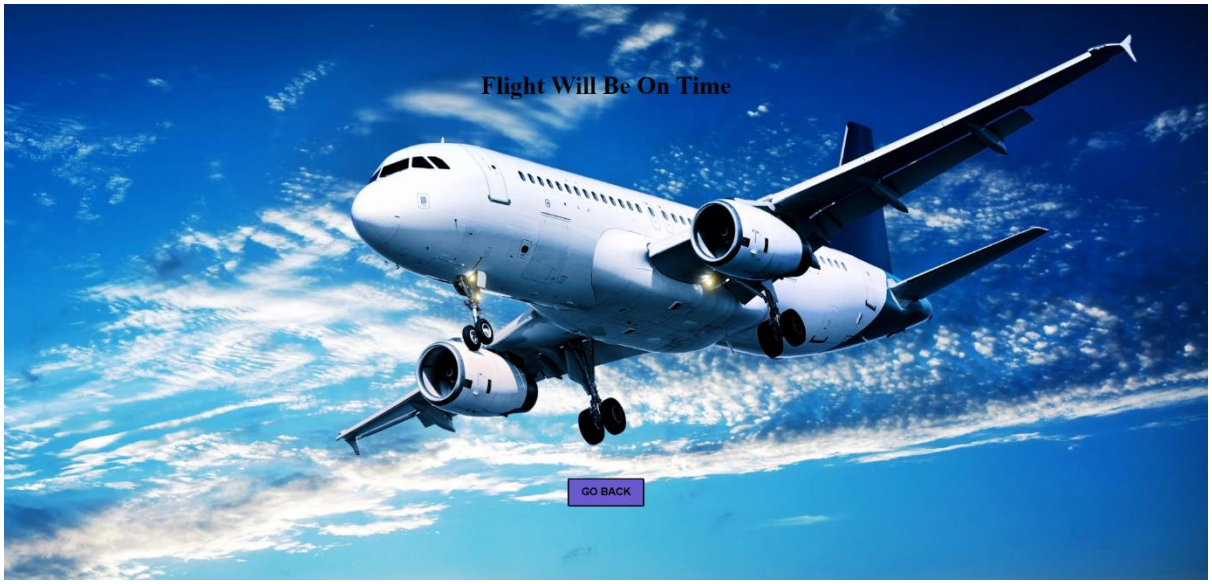
7.1 FEATURE 1

Flight Delay.html



7.2 FEATURE 2

fd.html



8. TESTING

8.1 TEST CASES

Activity Number	Activity Name	Detailed Activity Description	Assigned To	Date
1	Prerequisites	Prerequisites areall the needs at the requirement level needed for the execution of the different phases of a project.	Monisha D Deepika E Mohana Priya B Sai Krithika R	Completed
2	Data Collection	IBM provides the dataset. It is the actual datasetused to train the model for performing various actions. In this activity let's focus on gathering thedataset.	Monisha D Deepika E	Completed
3	Data Preprocessing	Data Pre processing is a technique that is used to convert the raw data into a clean data set.	Monisha D Deepika E	Completed

4	Ideation Phase			
4.1	Literature Survey & Information Gathering	Literature survey on theselected project & gathering information by referring the, technical papers, research publications etc.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	23 September 2022
4.2	Prepare Empathy Map	Prepare Empathy Map Canvas to capture the userPains & Gains, Prepare listof problem statements.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	30 September 2022
4.3	Brainstorming-idea generation	List the by organizing the brainstorming session and prioritize the top3 ideas based onthe feasibility & importance.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	14 October 2022
5	Project Design Phase 1			
5.1	Propose d Solution	Preparation of proposed solution document, whichincludes the novelty, feasibility of idea, business model, social impact, scalability of solution.	Deepika E Sai Kirthika R	14 October 2022

5.2	Problem Solution Fit	Prepared problem is analyzed and make effective solutions for the problem.	Deepika E Sai Kirthika R	14 October 2022

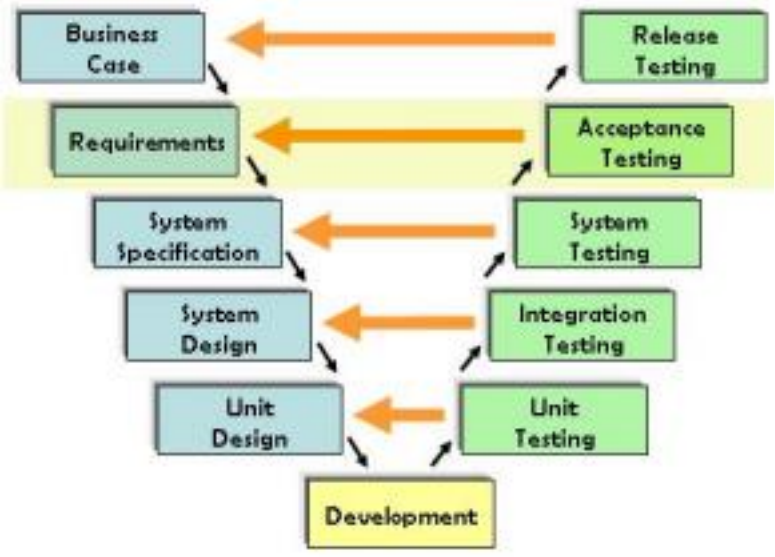
5.3	Solution Architecture	Prepare an architecture forsolution.	Monisha D Mohana Priya B	14 October 2022
6	Project Design Phase 2			
6.1	Solution Requirements	Prepare the Functional Requirement andNon Functional Document.	Monisha D	16 October 2022
6.2	Customer Journey Map	Preparation of customer journey maps tounderstand the user interactions& experiences with the application (entry to exit).	Deepika E	16 October 2022
6.3	Data Flow Diagrams	Prepare a Data Flow Diagram for Project use level 0(Industry Standard).	Sai Krithika R	16 October 2022
6.4	Technology Stack	Prepare Technology Architecture of the solution.	Mohana Priya B	16 October 2022
7	Project Planning Phase			
7.1	Milestones & Tasks	Prepare Milestone & Activity List	Mohana Priya B	6 November 2022
7.2	Sprint Schedules	Prepare Sprint Delivery Plan	Sai Krithika R	6 November 2022

8	Project Developmen tPhase			
8.1	Coding & Solutioning	Sprint-1 Delivery: Develop the Code, Test and push it to GitHub.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	In Progress
8.2	Acceptance Testing	Sprint-2 Delivery: Develop the Code, Test and push it to GitHub. Sprint-3 Delivery: Develop the Code, Test and push it to GitHub.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	In Progress
8.3	Performanc eTesting	Sprint-4 Delivery: Develop the Code, Test and push it to GitHub.	Monisha D Deepika E Mohana Priya B Sai Kirthika R	In Progress

8.2 USER ACCENPTANCE TESTING

Section	Total cases	Not tested	Fail	Pass
Home screen	1	0	0	1
User input	4	0	0	4
Flight delay testing	2	0	0	2
No flight delay testing	2	0	0	2
Version control	1	0	0	1

User Acceptance Testing (UAT) is a type of testing performed by the end user or the client to verify/accept the software system before moving the software application to the production environment. UAT is done in the final phase of testing after functional, integration and system testing is done. The main Purpose of UAT is to validate end to end business flow. It does not focus on cosmetic errors, spelling mistakes or system testing. User Acceptance Testing is carried out in a separate testing environment with production-like data setup. It is a kind of black box testing where two or more end-users will be involved.



9. RESULTS

9.1 PERFORMANCE METRICS

Train and Test The Model

In [30]:
```
from sklearn.preprocessing import StandardScaler
sc= StandardScaler()
x_train= sc.fit_transform(x_train)
x_test=sc.transform(x_test)
```

In [31]:
```
!pip install imblearn
```

```
Requirement already satisfied: imblearn in c:\users\monishanu\anaconda3\lib\site-packages (0.0)
Requirement already satisfied: imbalanced-learn in c:\users\monishanu\anaconda3\lib\site-packages (from imblearn) (0.9.1)
Requirement already satisfied: scipy>=1.3.2 in c:\users\monishanu\anaconda3\lib\site-packages (from imbalanced-learn->imblearn) (1.9.1)
Requirement already satisfied: scikit-learn>=1.1.0 in c:\users\monishanu\anaconda3\lib\site-packages (from imbalanced-learn->imblearn) (1.1.3)
Requirement already satisfied: numpy>=1.17.3 in c:\users\monishanu\anaconda3\lib\site-packages (from imbalanced-learn->imblearn) (1.21.5)
Requirement already satisfied: joblib>=1.0.0 in c:\users\monishanu\anaconda3\lib\site-packages (from imbalanced-learn->imblearn) (1.1.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\monishanu\anaconda3\lib\site-packages (from imbalanced-learn->imblearn) (2.2.0)
```

In [32]:
```
import imblearn
```

In [33]:
```
from imblearn.over_sampling import SMOTE
smote=SMOTE()
```

In [34]:
```
x_train_smote,y_train_smote =smote.fit_resample(x_train,y_train)
```

Decision Tree Classifier

In [35]:
```
from sklearn.tree import DecisionTreeClassifier
classifier =DecisionTreeClassifier(random_state=0)
classifier.fit(x_train_smote,y_train_smote)
```

Out[35]:
```
DecisionTreeClassifier(random_state=0)
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [36]:
```
decisiontree = classifier.predict(x_test)
```

In [42]:
```
decisiontree
```

Out[42]:
```
array([1, 0, 0, ..., 0, 0, 0], dtype=uint8)
```

Model Evaluation

In [37]:
```
from sklearn.metrics import accuracy_score
acc = accuracy_score(y_test,decisiontree)
acc
```

Out[37]:
```
0.9773030707610146
```

In [38]:
```
from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,decisiontree)
```

In [39]:
```
cm
```

Out[39]:
```
array([[1771,   31],
       [  20,  425]], dtype=int64)
```

Saving The Model

In [40]:
```
import pickle
pickle.dump(classifier,open('flight.pkl','wb'))
```

In [ ]:

10. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Flight delays are an important subject in the literature due to their economic and environmental impacts. They may increase costs to customers and operational costs to airlines. Apart from outcomes directly related to passengers, delay prediction is crucial during the decision-making process for every player in the air transportation system.
- Which causes delay in flights can be delay propagation, delay caused on the departure point or the root of the flight, and cancellation of flights. These problems cannot be eliminated forever, but a delay prediction tool will allow the operator and the administrators to take the concerned actions for smooth operation.
- Through this project, we created a machine learning model that can predict the flight departure delays. The model was able to catch 62% of the departure delays. Also, it was observed that the delay of the flights was heavily dependent on the departure airports.

DISADVANTAGES

- Passengers suffer a loss of time, missed business opportunities or leisure activities, and airlines attempting to make up for delays leads to extra fuel consumption and a larger adverse environmental impact.
- Airlines delays make immense loss for business field as well as in budget loss for a country.
- Flight delays end up hurting airports, passengers and airlines.

11. CONCLUSION

In this project, we anticipate airline departure delay using flight data, weather data, and demand data. Our findings demonstrate that, when compared to the decision tree model approach produces the greatest results. The decision tree model takes a long time to run and does not always result in better outcomes. In the end, 91% of the non-delayed flights were properly predicted by our algorithm. Only 41% of the time is the flight delay forecast accurate. As a result, there might be other characteristics of the factors that cause flight delays that haven't yet been found using our current data sources. It may be possible to routinely use tweets to ascertain an understanding of concurrent airline delays and traffic patterns, which could be useful in a variety of circumstances.

12. FUTURE SCOPE

Based on data analysis from the year 2008, this project. There is a sizable dataset accessible from 1987 to 2008, but managing a larger dataset necessitates extensive data pre-processing and cleaning. Therefore, adding a larger dataset is a part of this project's future effort. Pre-processing a bigger dataset can be done in a variety of methods, such as establishing a Spark cluster on a computer or using cloud services like AWS and Azure. Now that deep learning has advanced, we can employ neural networks algorithms to analyse aviation and meteorological data. Neural networks employ a methodology for pattern matching. For data modelling, it is organised into three fundamental sections: feed forward networks, feedback networks, and self-organization.

13. APPENDIX

SOURCE CODE

app_ibm.py

```
from flask import Flask,render_template,request
import numpy as np
import pandas as pd
import pickle
import os

model=pickle.load(open('flight.pkl','rb'))
app=Flask(__name__)

# NOTE: you must manually set API_KEY below using information
retrieved from your IBM Cloud account.
API_KEY = "9rgwTXNNVEK-K7ZBxOJ8ypezVmm8-
qZLspTCTlpEoejm"
token_response
= requests.post('https://iam.cloud.ibm.com/identity/token',
data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

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

app = Flask(__name__)

@app.route('/')
def home():
    return render_template('Flight Delay.html')

@app.route('/next')
def next():
    return render_template('fd.html')

@app.route('/fd',methods = ['GET','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 == "msp"):
        origin0,origin1,origin2,origin3,origin4 = 0,0,0,0,1
    if (origin == "dtw"):
        origin0,origin1,origin2,origin3,origin4 = 1,0,0,0,0
    if (origin == "jfk"):
        origin0,origin1,origin2,origin3,origin4 = 0,0,1,0,0
    if (origin == "sea"):
        origin0,origin1,origin2,origin3,origin4 = 0,1,0,0,0
    if (origin == "atl"):
        origin0,origin1,origin2,origin3,origin4 = 0,0,0,1,0
    destination = request.form['destination']
    if (destination == "msp"):
        destination0,destination1,destination2,destination3,destination4 =
0,0,0,0,1
    if (destination == "dtw"):
        destination0,destination1,destination2,destination3,destination4 =
1,0,0,0,0
    if (destination == "jfk"):
        destination0,destination1,destination2,destination3,destination4 =
0,0,1,0,0
    if (destination == "sea"):
        destination0,destination1,destination2,destination3,destination4 =
0,1,0,0,0
    if (destination == "atl"):
        destination0,destination1,destination2,destination3,destination4 =
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,arrtime,dept15,origin0,origin1,orig
in2,origin3,origin4,destination0,destination1,destination2,destination3,de
stination4,]]


    # NOTE: manually define and pass the array(s) of values to be scored
in the next line
    payload_scoring = {"input_data": [{"fields":
```

```
[['f0','f1','f2','f3','f4','f5','f6','f7','f8','f9','f10','f11','f12','f13','f14','f15']],
"values": total}]}

    response_scoring = requests.post('https://us-
south.ml.cloud.ibm.com/ml/v4/deployments/a270a246-e678-4ae8-9e7d-
1944f4f2b20c/predictions?version=2022-11-11', json=payload_scoring,
     headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    print(response_scoring.json())
    pred = response_scoring.json()
    y_pred = pred['predictions'][0]['values'][0][0]
    if(y_pred==0):
        ans = "The flight will be on time"
    else:
        ans = "The flight will be delayed"
    return render_template('fd.html',showcase = ans)

if __name__ == '__main__' :
     app.run(debug= False)
```

13.2 GITHUB LINK

Github link:

https://github.com/IBM-EPBL/IBM-Project-4867-1658741237

Project demo link:

https://drive.google.com/drive/folders/1mPNSjo70G9Ml_1hEXciuDqbotzypMQhR?usp=share_link