WEBPHISHINGDETECTION

NALAIYATHIRANPROJECTBASEDLEARNING

On

DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

A PROJECT REPORT

AKASHR 19106807

SHIVAS 19106139

SUNIL R 19106120

BACHELOR OF

ENGINEERING IN

ELECTRONICSANDCOMMUNICATIONENGINEERING

HINDUSTHANCOLLEGEOFENGINEERINGANDTECHNOLOGY

Approved by AICTE, New Delhi, Accredicted with 'A' Grade by NAAC**(An Autonomous Institution, Affiliated to Anna University, Chennai)COIMBATORE –641032**

November2022

TABLEOFCONTENTS

CHAPTERNO	TITLE	PAGE NO
	ABSTRACT	
1	INTRODUCTION	1
2	OBJECTIVE	2
3	IDEATIONPHASE	3
	LiteratureSurvey	3
	EmpathyMap	6
	Ideation	7
	ProblemStatement	9
4	PROJECTDESIGNPHASEI	11
	ProposedSolution	11
	ProblemSolutionFit	13
	SolutionArchitecture	14
5	PROJECTDESIGNPHASEII	16
	CustomerJourneyMap	16
	SolutionRequirements	19
	DataFlowDiagrams	20
	TechnologyStack	22
6	PROJECTPLANNINGPHASE	24
	PrepareMilestoneandActivityList	25
	SprintDeliveryPlan	26
7	PROJECTDEVELOPMENTPHASE	28
	ProjectDevelopment–DeliveryofSprint–1	28
	ProjectDevelopment–Delivery ofSprint–2	31
	ProjectDevelopment–Delivery ofSprint–3	34
	ProjectDevelopment–Delivery ofSprint–4	34
8	CONCLUSION	35
9	REFERENCES	36

ABSTRACT

Abstract— Flight Planning is one of the challenges in industrial world which faces many uncertain conditions. One such condition is delay occurrence, which stems from various factors and imposes considerable costs on airlines, operators, and travelers. 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. Here in flight delay prediction system based on the weather parameters which can result in delays. The system considers the temperature, humidity, rain in mm, visibility and month number as important parameters for prediction of delay.

Keywords— Flight delay, weather, supervised machine learning, Naive Bayes

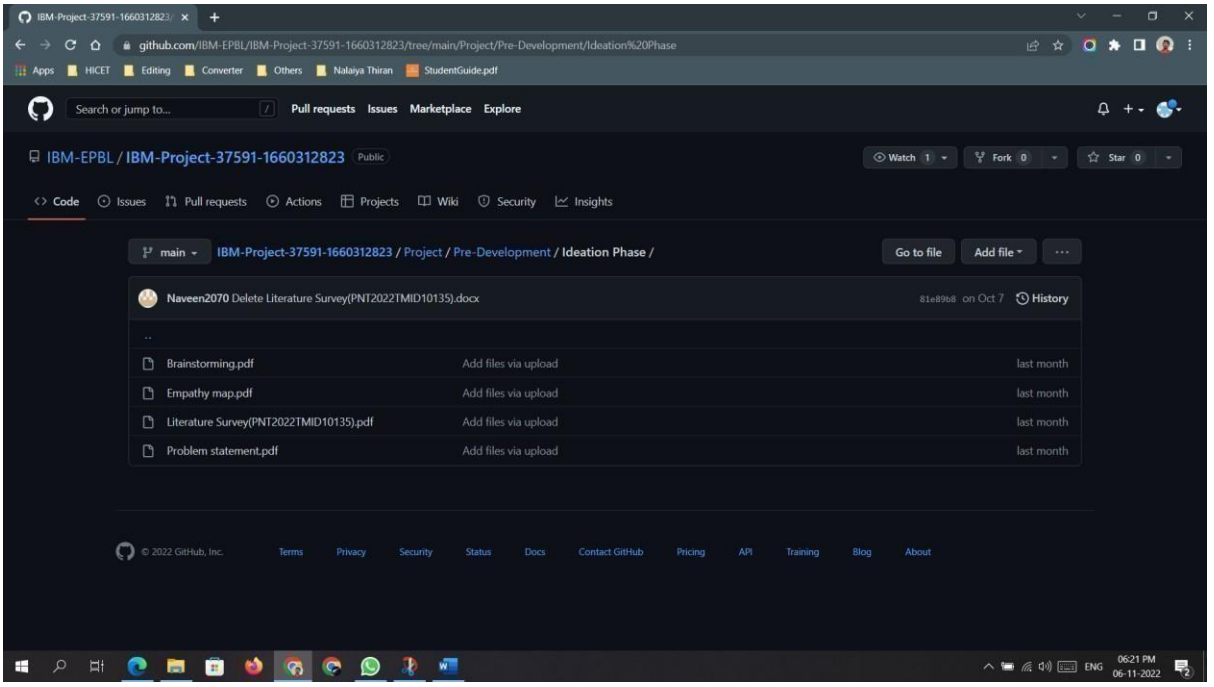
1. INTRODUCTION

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 that is an upscale affair for the airlines, making issues in scheduling and operations for the endusers therefore inflicting unhealthy name and client disco ntent. As we all know that we have a tendency to not get the flight delay before departure as customers of the Airline Company neither the airline company's ground staff gets the airline delay prediction supported varied conditions. However, we all know that one in all the most reasons for delay in flights is that the weather. This motivates us to use the live weather knowledge in conjunction with different metrics to calculate the delay on the wing before departure. Indian state of affairs, in 2017, in line with the reports by the directorate General of Civil Aviation (DGCA), between January and April, close to 5.12 hundred thousand domestic passengers in India faced issues because of airline corporations denying boarding, moreover as flight cancellations and delays [2]. Airline corporations had to pay the passengers compensations of over Rs. twenty five crore for varied inconveniences throughout the first four months of this year. Hence, the prediction analysis retrieved from this project can contribute within the form of a prototype in helping to identify operational variables that contribute to delays in any country scenario[2]

2. OBJECTIVE

- To improve airline operations and passenger satisfaction, which will result in a positive impact on the economy.
- To compare the performance of machine learning classification algorithms when predicting flight delays.
- To Develop a business model to predict flight delays.

3. IDEATIONPHASE



3.1 LiteratureSurvey

3.2

FLIGHT DELAY PREDICTION MODEL USING

MACHINE LANGUAGE

TEAMLEADER : AKASH.R

TEAM MEMBERS : SHIVA.R, SUNIL.R

GUIDE : POONGUZHALI. P K

. Passenger airlines, cargo airlines, and air traffic control systems are the main elements of any transportation system in the modern world.

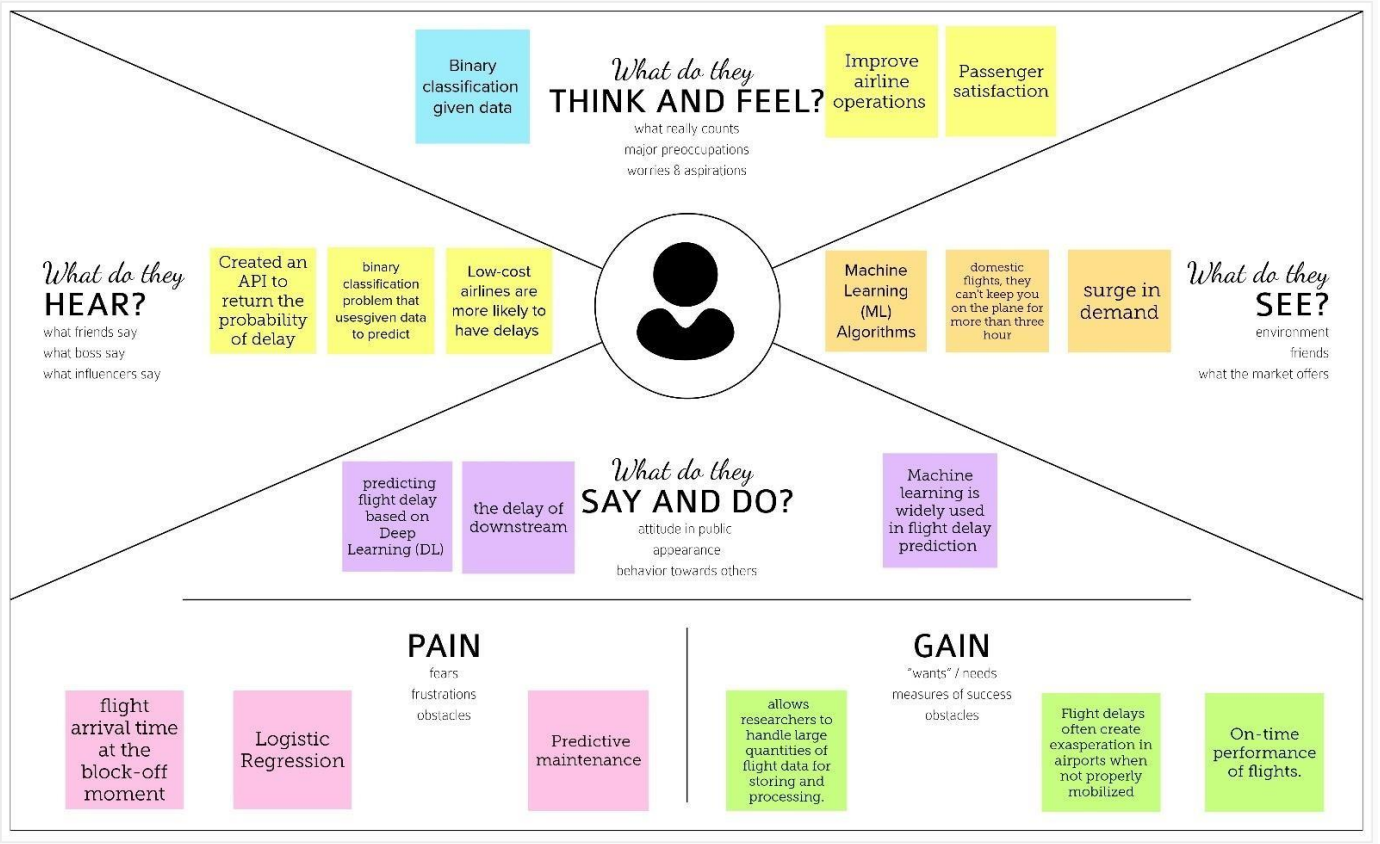
Nations all around the world have attempted to develop different methods over time to enhance the aeroplane transportation system. This has significantly altered how airlines operate. Modern travelers may experience inconvenience from flight delays. Around 20% of airline flights are cancelled or delayed annually, costing passengers more than $20 billion in lost time and money

The number of planes that don't take off on time likewise rises as more people prefer to travel by air. This expansion makes airports even more crowded and harms the airline industry's bottom line. The inefficiency of the aviation system is demonstrated by delays in air transportation. Both the airline firms and their customers pay a hefty price for it. The Total Delay Impact Study estimated, The entire cost of delays in air travel to passengers and the airline sector in the US in 2007 was $32.9 billion, which resulted in a $4 billion drop in GDP . Predicting delays can therefore enhance airline operations and passenger pleasure, which will have a positive effect on the economy.

The major objective of this study is to assess how well machine learning classification methods perform when forecasting flight delays. John F. indira gandhi International Airport in delhi served as the study's airport. Analysis was done on flight data departing from Delhi Airport during a one-year period. The study used a number of algorithms, and its predictions were assessed using a variety of metrics. provides an explanation of the theoretical underpinnings of several machine learning models and performance evaluation techniques. discusses relevant works by earlier researchers. Different models' empirical processes and outcomes are displayed and compared in presentthe results of the comparative analysis and suggests future research areas.

EmpathyMap

Figure1:EmpathyMap




2
Brainstorm
Write down any ideas that come to mind that address your problem statement.
10 minutes
AKASK
DELAY PROPOGATION
NEW ROOT DELAY
CANCELLATION
AIRLINE
AIRPORT
ENROUTE AIRSPACE
AIRPORT WEATHER
AIRLINE DEMAND
FLIGHT SCHEDULE
SUNIL
ORIGINAL LOCATION CODE
DEPARTURE DATE
ARRIVAL TIME
ARRIVAL DATE
CARRIER CODE
DEPATURE TIME
AIRCARFT CODE
DURATION
FLIGHT NUMBER
SHIVA
ARRIVAL
TAXIWAY
MAIL
CARGO PROCESSING
CATERING
DATA BANK
SCOPE
PROBABILISTIC MODEL
CLUSTERS

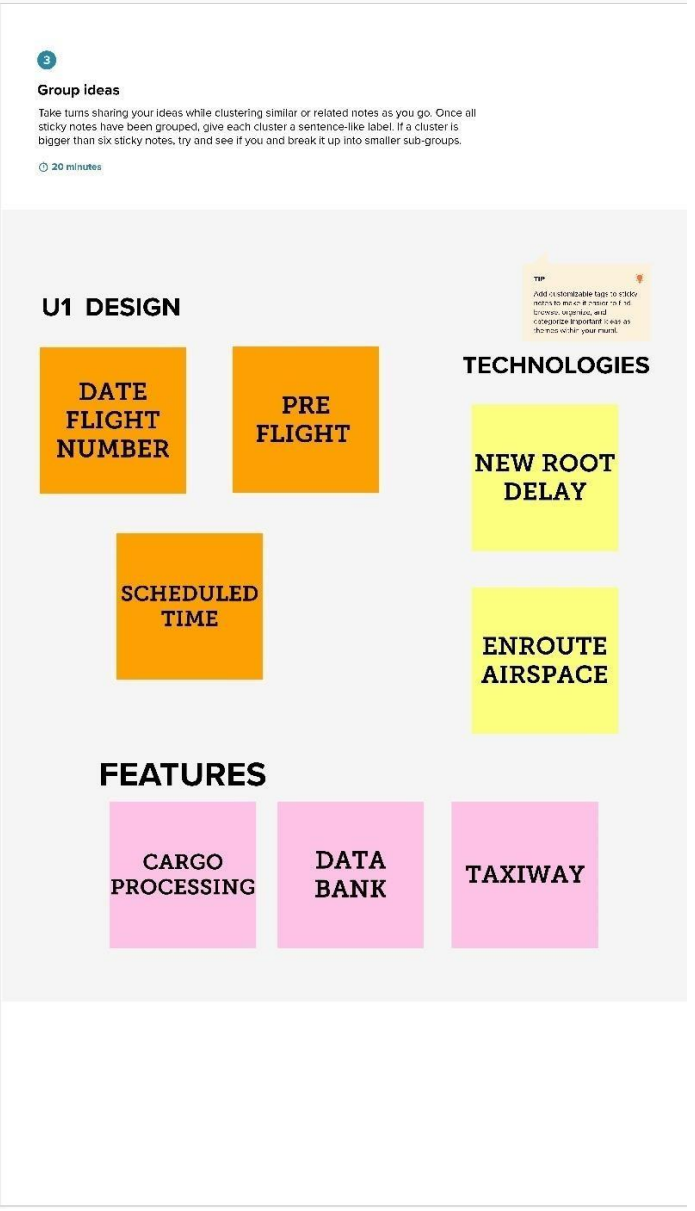

3
Group ideas
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
20 minutes
U1 DESIGN
DATE FLIGHT NUMBER
PRE FLIGHT
SCHEDULED TIME
TECHNOLOGIES
NEW ROOT DELAY
ENROUTE AIRSPACE
FEATURES
CARGO PROCESSING
DATA BANK
TAXIWAY

Figure3:Ideation(Continuation)

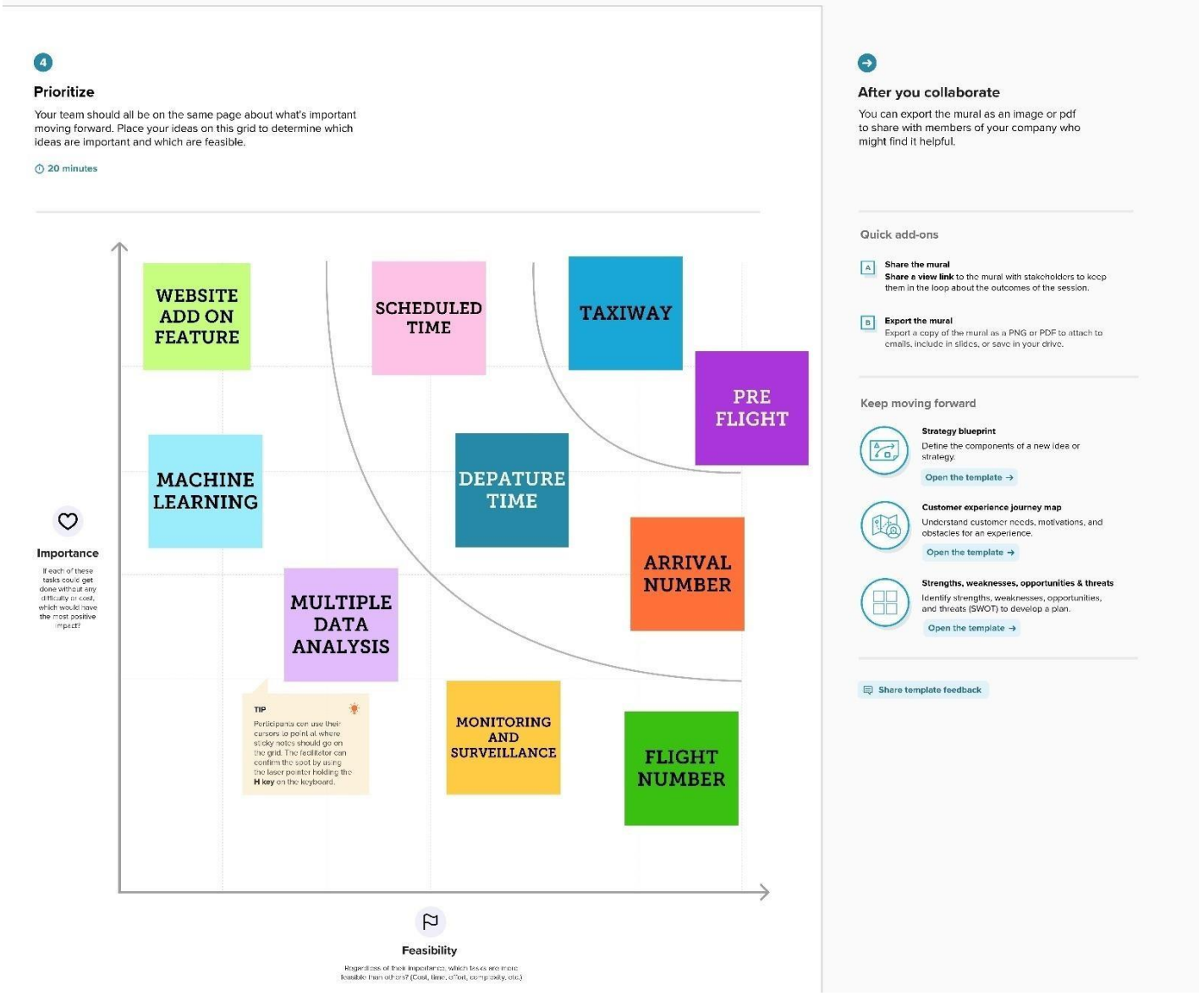


Figure4:Ideation(Continuation)

ProblemStatement

Ideation Phase
Define the Problem Statements

DATE	19 September 2022
TEAM ID	PNT2022TMID10226
TEAM LEAD	AKASH
TEAM MEMBERS	SUNIL SHIVA
PROJECT TITLE	DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

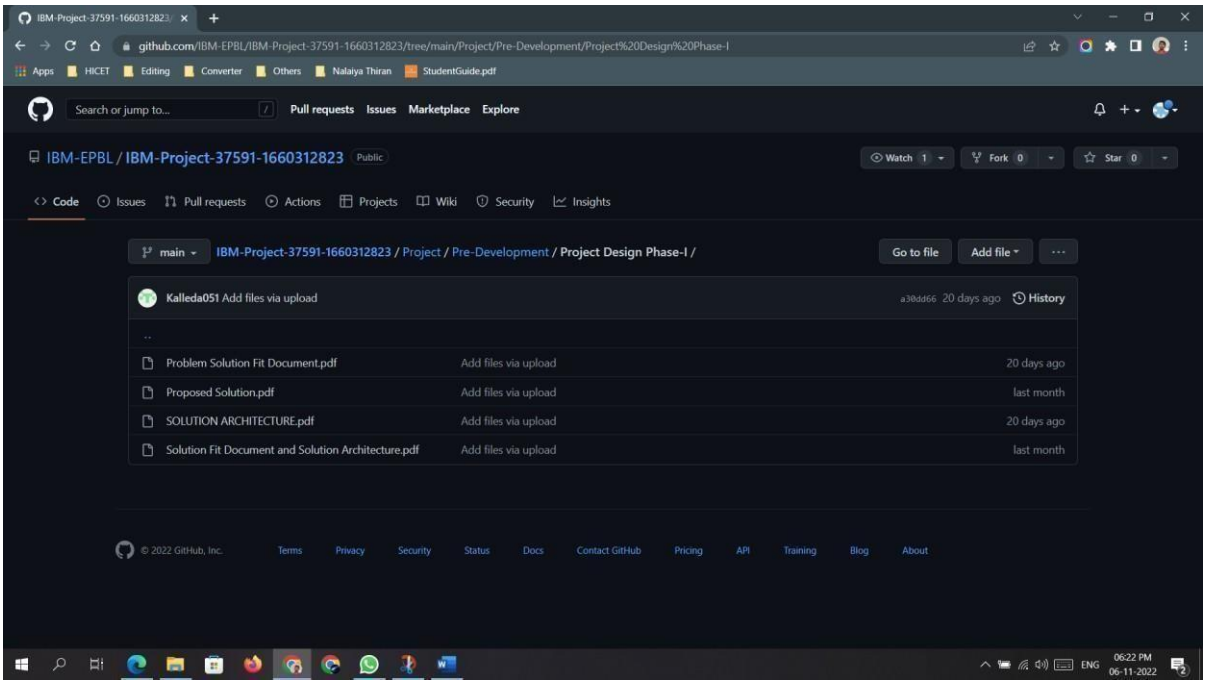
Figure5

Table1:ProblemStatement

ProblemStatement(PS):	Flight delay is inevitable and it plays an important role in both profits and loss of the airlines.an accurate estimation of flights is critical for airlines because the results can be applied to increase customer satisfaction and incomes of airline agencies .there have been many researches on modelling and predicting flights delays,where most of them have been trying to predict the delay through extracting important characteristics and most related features
Iam(US ER)	Flight Delay Prediction

I'mtryingto	Arrival delay prediction and departure delay prediction
But	One has to keep in mind that this can differ from arrival delays
Because	This suggests that a important variable in the modelling of delays will be take off time
Whichmakesmefeel	To achieve a goals and mastering visualization techniques thus seems important

4. PROJECTDESIGNPHASEI

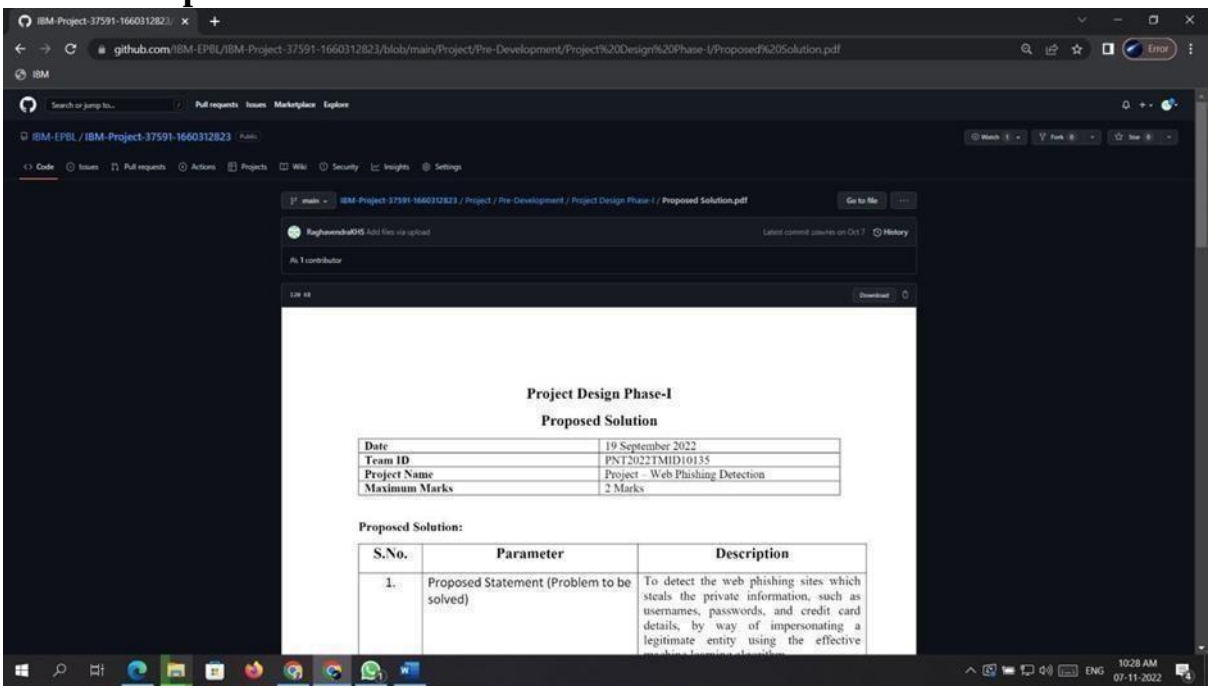


ProposedSolution

Project Design Phase-I
Proposed Solution Template

Date	24 September 2022
Team ID	PNT2022TMID10226
Project Name	Developing a flight delay prediction model using Machine learning
Maximum Marks	2 Marks

Table2:ProposedSolution



Project Design Phase-I

Proposed Solution

Date	19 September 2022
Team ID	PNT2022TMID10135
Project Name	Project – Web Phishing Detection
Maximum Marks	2 Marks

Proposed Solution:

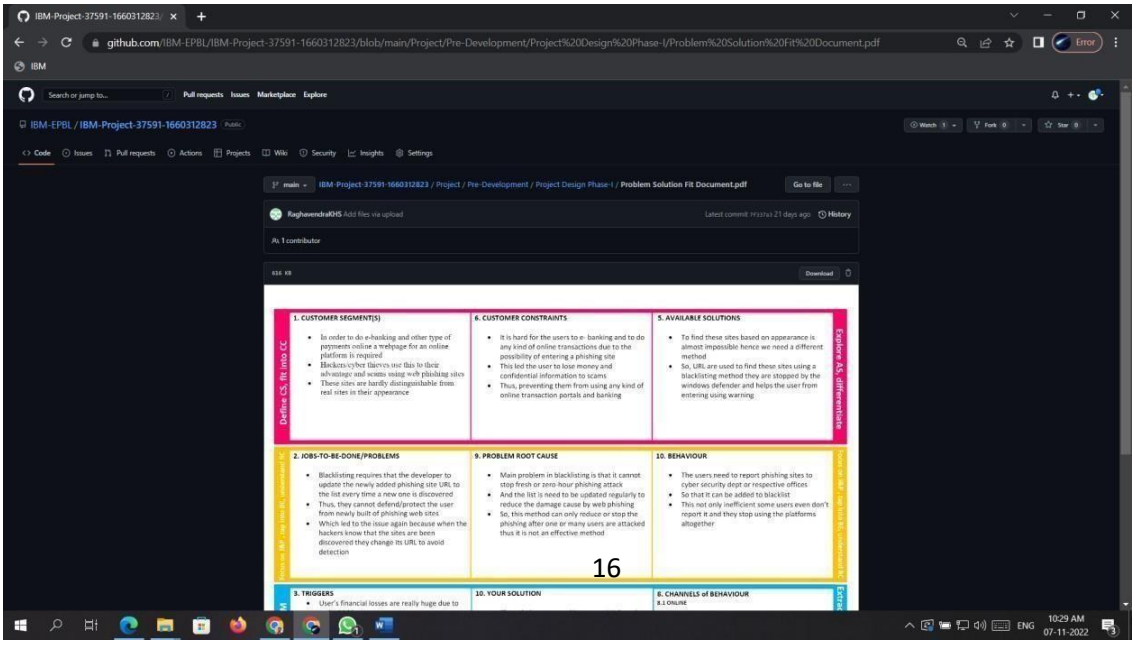
S.No.	Parameter	Description
1.	Proposed Statement (Problem to be solved)	To detect the web phishing sites which steals the private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity using the effective

S.No.	Parameter	Description
1.	ProposedStatement(Problemtobesolved)	Flight delays have been the most challenging area for airlines to improve. They have been affecting the air industry directly and indirectly causing unforeseen expenses thereby reducing the reputation of the industry and the airlines. Thus, knowing if a flight would be delayed beforehand can let passengers and airlines be prepared for the circumstances.This solution aims at making it possible by predicting arrival and departure delays using Machine learning
2.	Idea/SolutionDescription	Building an application interface for customers(passengers and airlines) to know if a flight is delayed by implementing a machine learning based model to predict departure and arrival delays of an aircraft considering spatial, temporal and other dependencies causing the delay

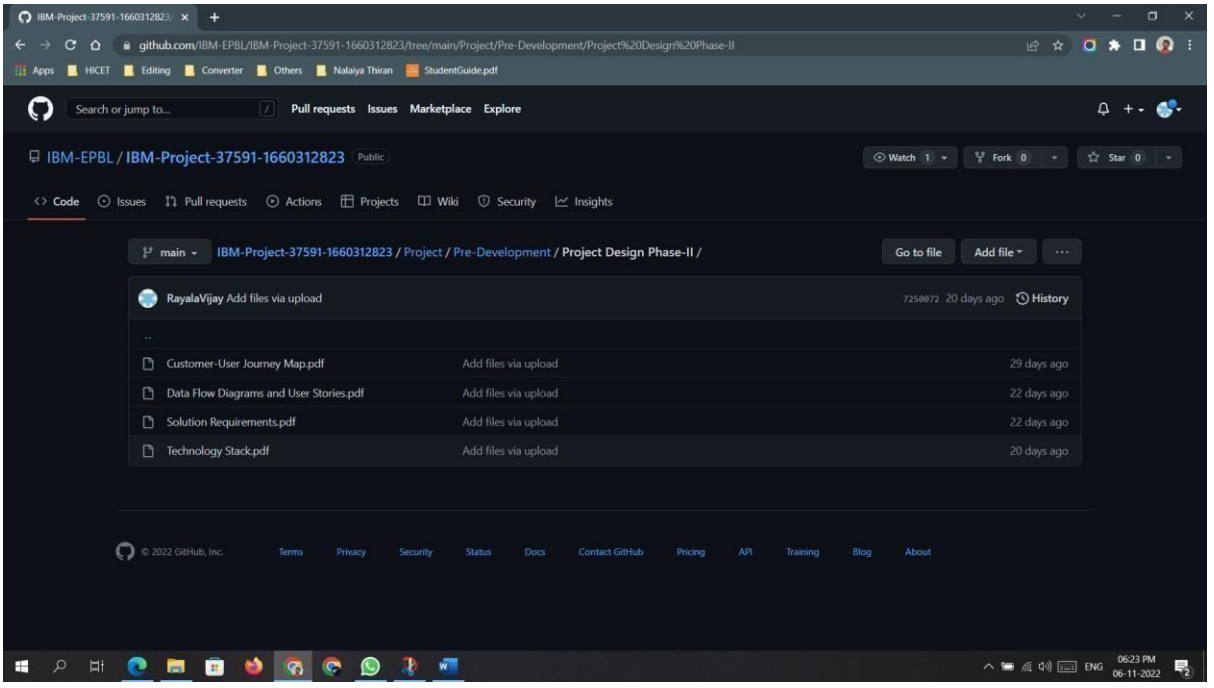
3.	Novelty/Uniqueness	The solution takesinto account all possible reasonsfordelay(crew delys, weather, air traffic, aircraft type) to provide an accurate prediction. Apartfrom predicting arrival delays, departure delays are also predicted in order for the passengers to prepare accordingly and for the airline to make arrangements suitably
4.	SocialImpact/CustomerSatisfaction	A lot of time and money can be saved for the customers and the loyalty and trust of customers towards the company increases.

		Improves airline operations by letting the company prepare in prior to adversaries (like crew illness, timeouts, rescheduling) leading to passenger satisfaction which will result positively on the economy and brand value
5.	BusinessModel(RevenueModel)	Business to Consumer model .The solution is a low-cost airline model planned to be created as an application with which the consumers can interact directly to know the details of their flight. It follows a non-monetary revenue model where the consumers aren't charged for what they get but are asked to provide their flight details and ratings which can be used to improve the model and shared with the airline in return for airline's flight data.
6.	Scalabilityofthesolution	The present solution is drafted with the aim of experimenting with airlines based out of the United States of America.If there is a possibility to acquire data of a broader region (say North America, other continents), then the solution can be developed to benefit a wider range of people. International flight dependenciesin both temporal and spatial focus can be derived from that data to provide more accurate predictions. Presence of ADS-B data can further increase the efficiency of system making it reach global audience and live time tracking of flights

ProblemSolutionFit



5. PROJECTDESIGNPHASEII



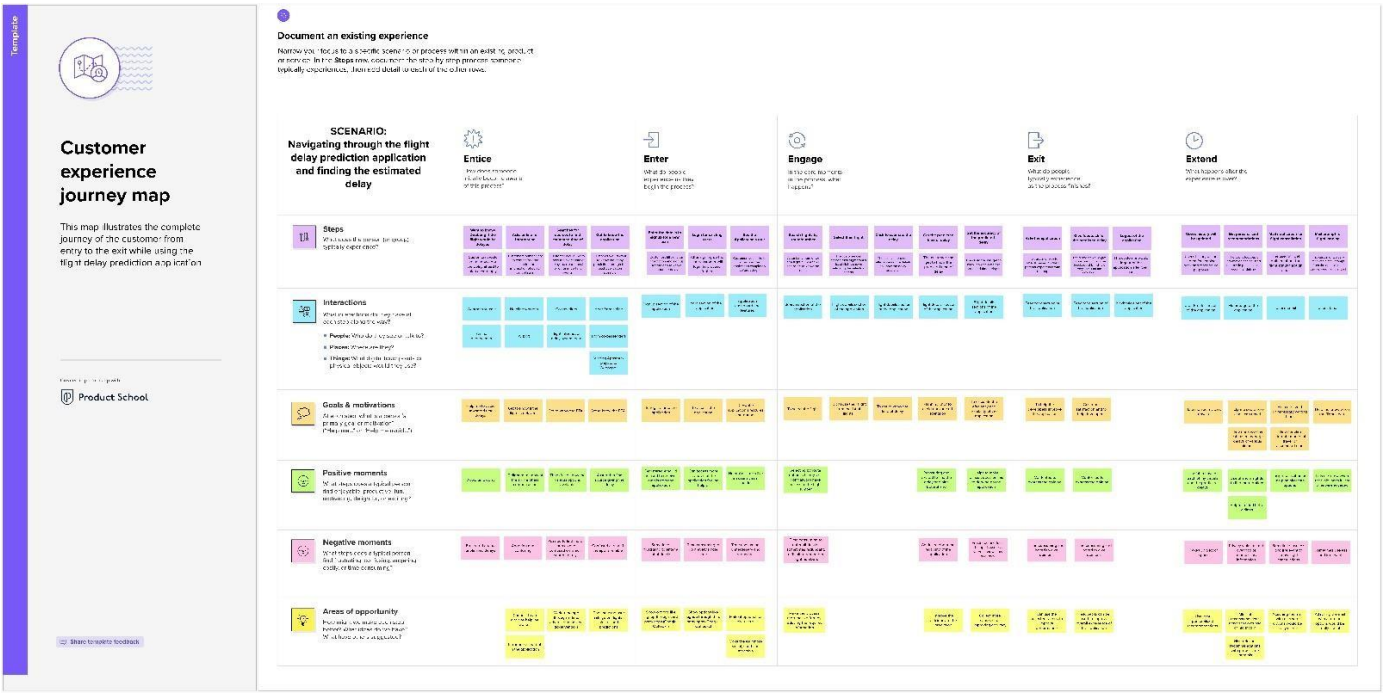
CustomerJourneyMap

PROJECT PHASE – II

CUSTOMER JOURNEY MAP

DATE	**03 OCTOBER 2022**
TEAM ID	**PNT2022MID10226**
PROJECT NAME	**FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING**

Figure8: CustomerJourneyMap



SolutionRequirements

Project Design Phase-II
Solution Requirements (Functional & Non-functional)

Date	15 October 2022
Team ID	PNT2022TMID10226
Project Name	Developing a flight delay prediction model using machine learning
Maximum Marks	4 Marks

Table3:FunctionalRequirements:

Followingarethefunctionalrequirementsoftheproposedsolution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Signup	• Registration through UserID/Password • Registration through Gmail • Registration through Phone number
FR-2	User Confirmation	• Confirmation via Email • Confirmation via OTP
FR-3	User Login	• Login with UserID/Password • Login with gmail • Login with phone number/OTP
FR-4	Search Flight	• Get the entered flight details
FR-5	Predict Delay Time	• Feed the details to the model and find prediction • Display the received prediction
FR-6	Predict Delay Accuracy	• Get the accuracy of delay • Display the accuracy
FR-7	Notify the user	• Send prediction results to mail • Notify 30 minutes before flight arrival/departure
FR-8	Get feedback	• Get descriptive feedback • Get ratings from user
FR-9	User Logout	• Logout of the application

Table4:Non-functionalRequirements:

Followingarethenon-functionalrequirementsoftheproposedsolution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	• An app tour would be shown to the users. • To guide new users who search flights, in the search box where the user needs to type the flight details, • a message such as Try "BOM MAA" or "Mumbai Chennai" will be displayed.
NFR-2	**Security**	• During registration, a 2 factor authentication through mail would confirm if the user is reliable. • The user would be able to login to the app only with his credentials. • He would be allowed to change the password only after a 2-factor authentication and a notification would also be sent to his mailbox to indicate the change.
NFR-3	**Reliability**	• There is a 75 percent chance under optimal condition that the application won't experience critical failure • There is 80 percent restoring capability even if the system fails.
NFR-4	**Performance**	• The application load time would take 3 seconds or less with a WiFi/LTE connection. • Time taken to predict the delay would be no more than 5 seconds.
NFR-5	**Availability**	• Duing any new update/maintenance, a message would be displayed in the application 48 hours before the scheduled time regarding the same. • The functional requirement 'Search flight' function may not be available when all the flights are canceled as in case of pandemic or in war stricken areas.. • The user gets the prediction result through mail. • If there is any problem with the model, the user would receive an alert that there is an issue in the prediction and the system would get back within 10 mins. • The system would be available to use during the other times.
NFR-6	**Scalability**	• Though it is out of scope keeping our implementation in mind, the system can be made scalable enough to support 1,000,000 visits at the same time while maintaining optimal performance. • It can also be scaled to predict delays with international flights and delays due to weather by training the model with appropriate data, given that it must be available.

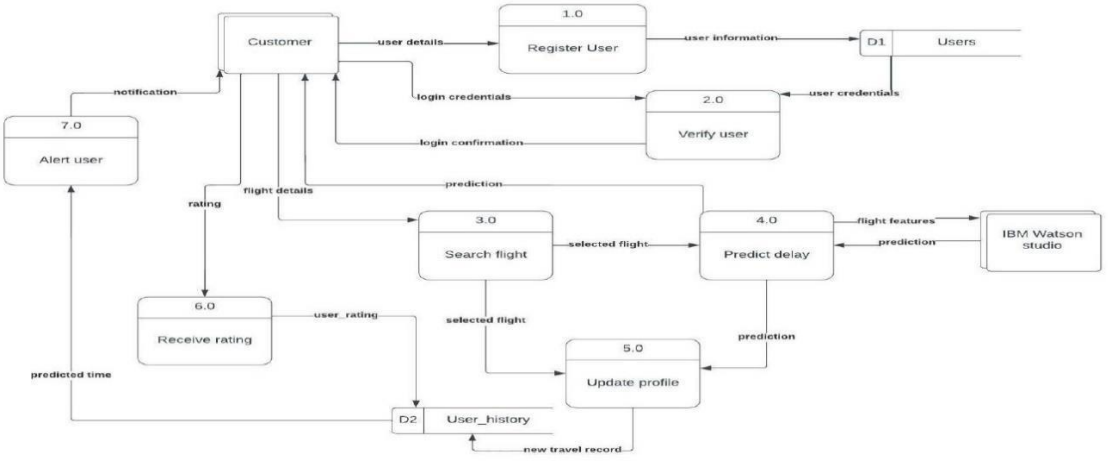
DataFlowDiagrams

Project Design Phase-II
Data Flow Diagram & User Stories

Date	15 October 2022
Team ID	PNT2022TID10226
Project Name	Developing a flight delay prediction model using Machine learning
Maximum Marks	4 Marks

Figure10:DataFlowDiagram

Data Flow Diagrams:



Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	03 October 2022
Team ID	PNT2022TMID10226
Project Name	Developing a Flight Delay Prediction Model using Machine Learning
Maximum Marks	4 Marks

TechnicalArchitecture:

Technical Architecture:

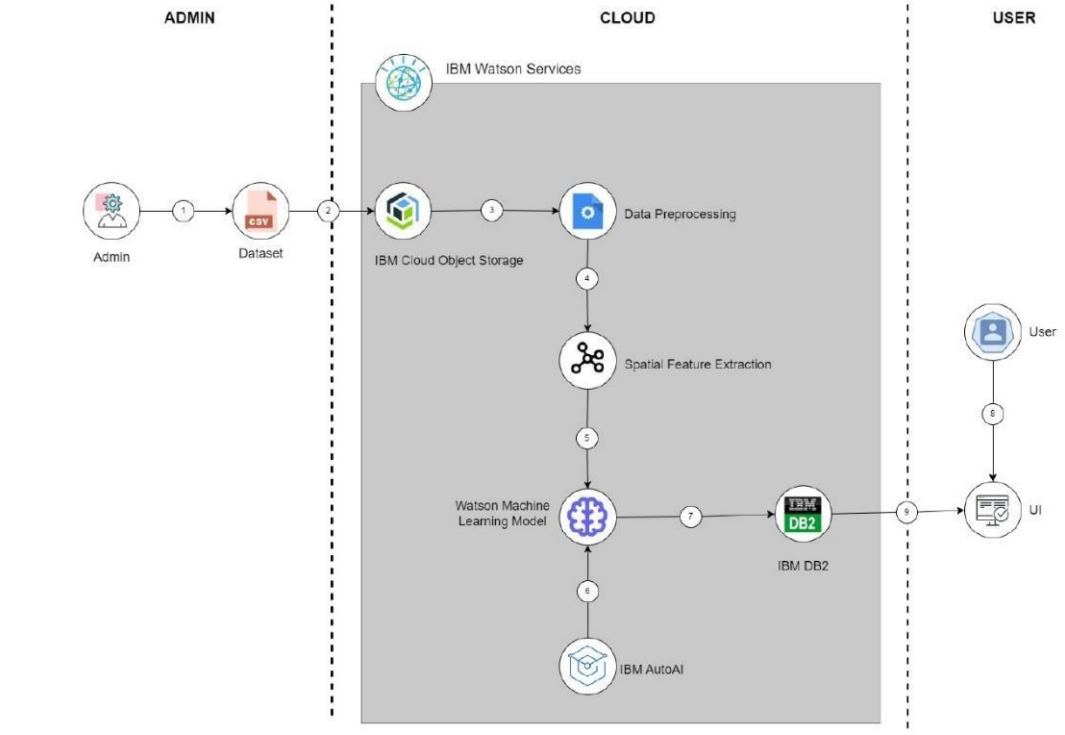


Table-5:Components&Technologies:

S.No	Component	Description	Technology
1.	User Interface - front end	To interact with the application - Login, Delay Requests	HTML, CSS, JS
2.	Spatial Feature Extraction	To calculate crowdedness at particular source/destination	Python, IBM Watson AI

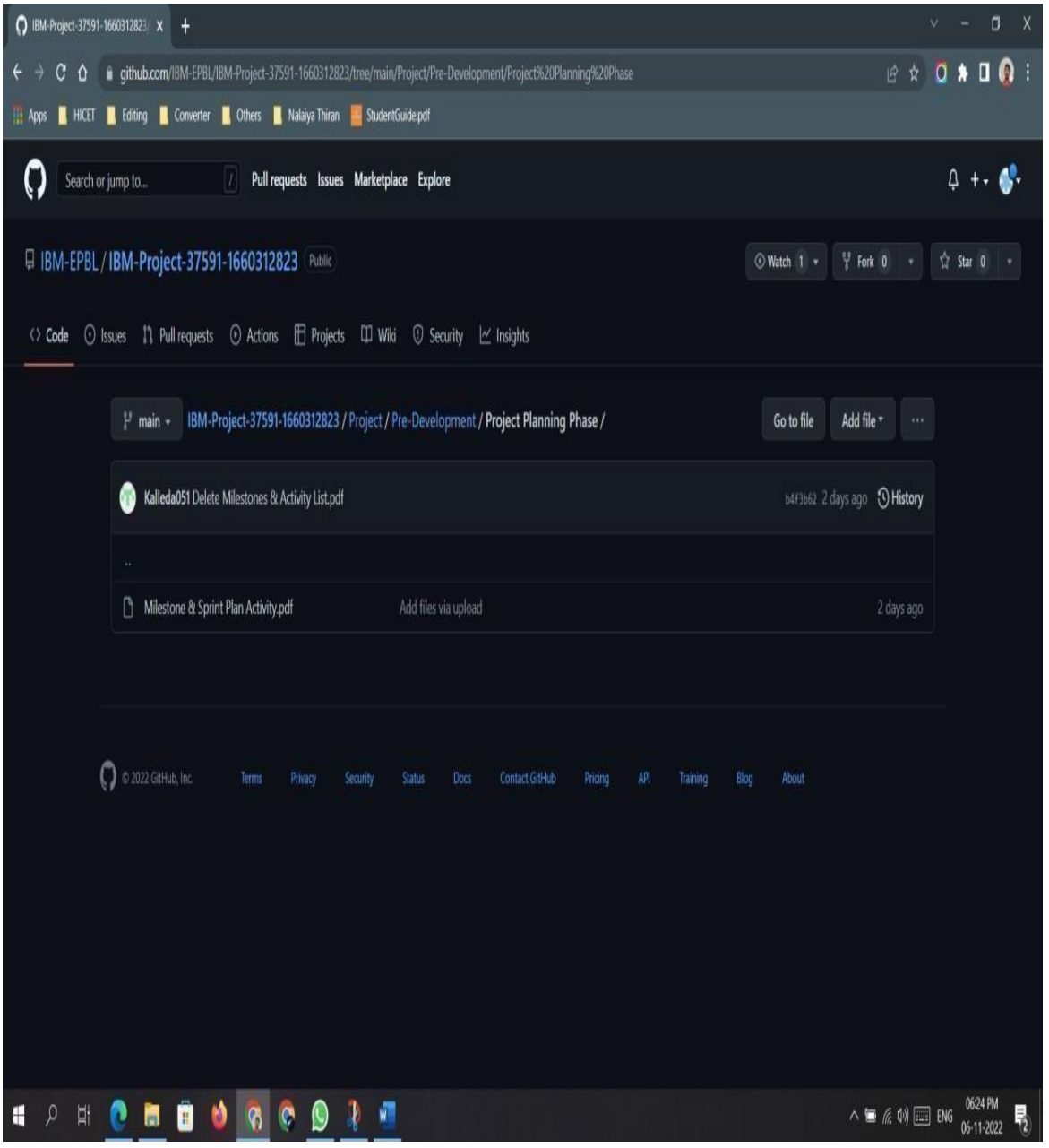
3.	Delay Prediction	Predict the delay for the given flight	BM Watson AI
4.	Cloud Database	To store the user details	IBM DB2
5.	File Storage	To store the delay dataset	IBM Cloud Object Storage

Table-6:ApplicationCharacteristics:

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	HTML, CSS, JavaScript, Bootstrap, Flask, Kubernetes
2.	Security Implementations	Native three step protection	Kubernetes IBM DB2
3.	Scalable Architecture	It can be scaled by adding master node and extra working nodes to the main cluster.	Kubernetes IBM Cloud
4.	Availability	IBM Kubernetes uses Kubernetes load balancers namely kube-proxy and ingress controllers.	IBM Kubernetes
5.	Performance	Performance can be enhanced by adding more working nodes to the master cluster.	IBM Kubernetes

6. PROJECTPLANNINGPHASE



6.1 SprintDeliveryPlan

Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	22 October 2022
Team ID	PNT2022TMID10226
Project Name	Project –Flight Delay Predication Model Using By Machine Learning.
Maximum Marks	8 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

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	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Akash
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Shiva
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	Sunilkumar
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	Akash
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Shiva
Sprint-3	Dashboard	USN_6	To view dashboard on our project and check Customer information.	2	High	Sunilkumar
Sprint-4	Review	USN-7	As a customer review our site.	1	High	Akash

Project Tracker, Velocity & Burndown Chart: (4 Marks)

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 Oct 2022
Sprint-3	25	6 Days	07 Nov 2022	12 Nov 2022	19	07 Oct 2022
Sprint-4	25	6 Days	14 Nov 2022	19 Nov 2022	20	14 Nov 2022

Velocity:

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

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

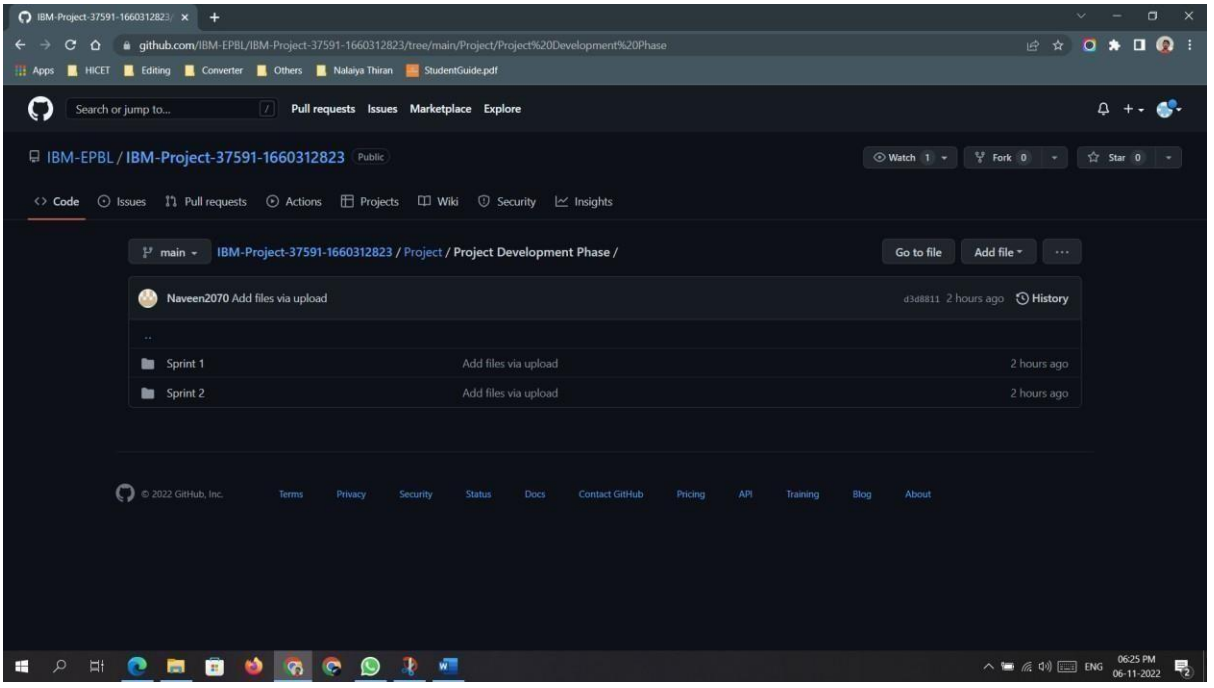
Sprint 1& Sprint 2:

AV= 20/6 =3.3

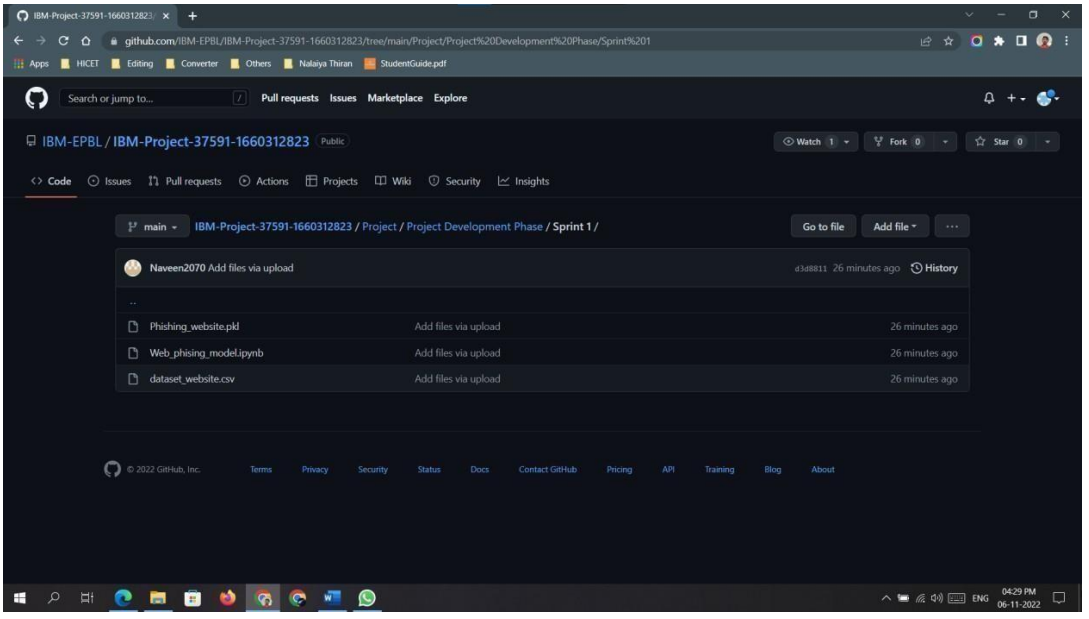
Sprint 3& Sprint 4:

AV=25/6=4.1

7. PROJECTDEVELOPMENTPHASE



ProjectDevelopment–DeliveryofSprint1

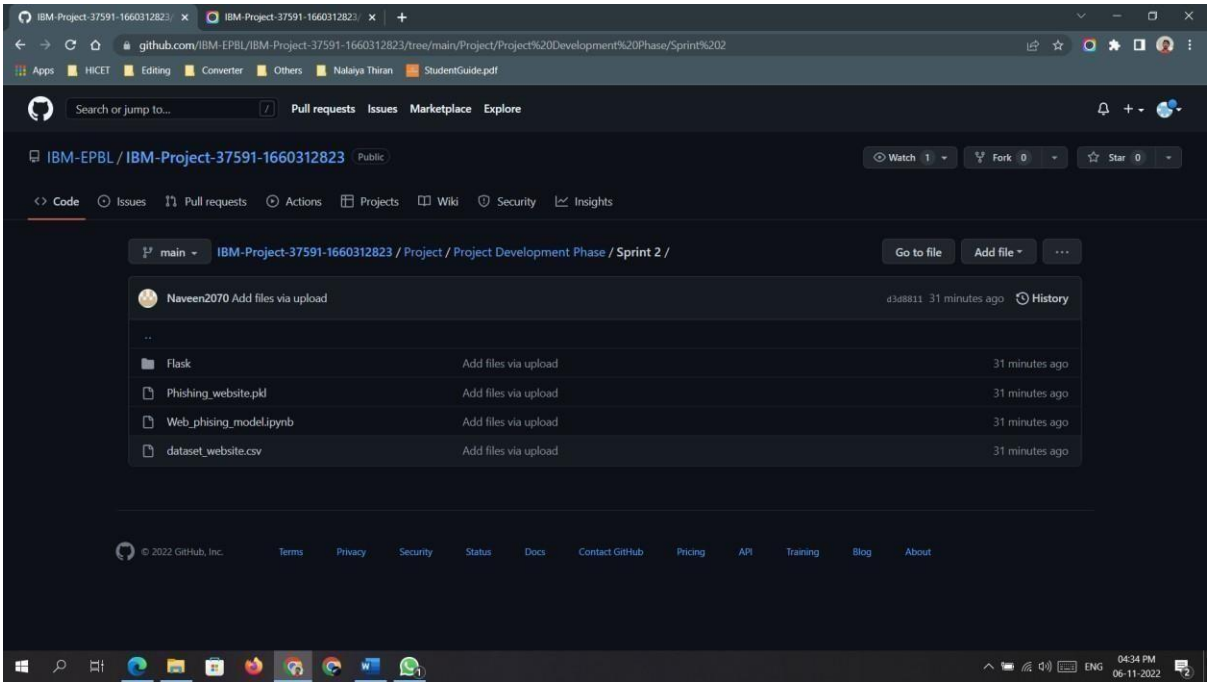


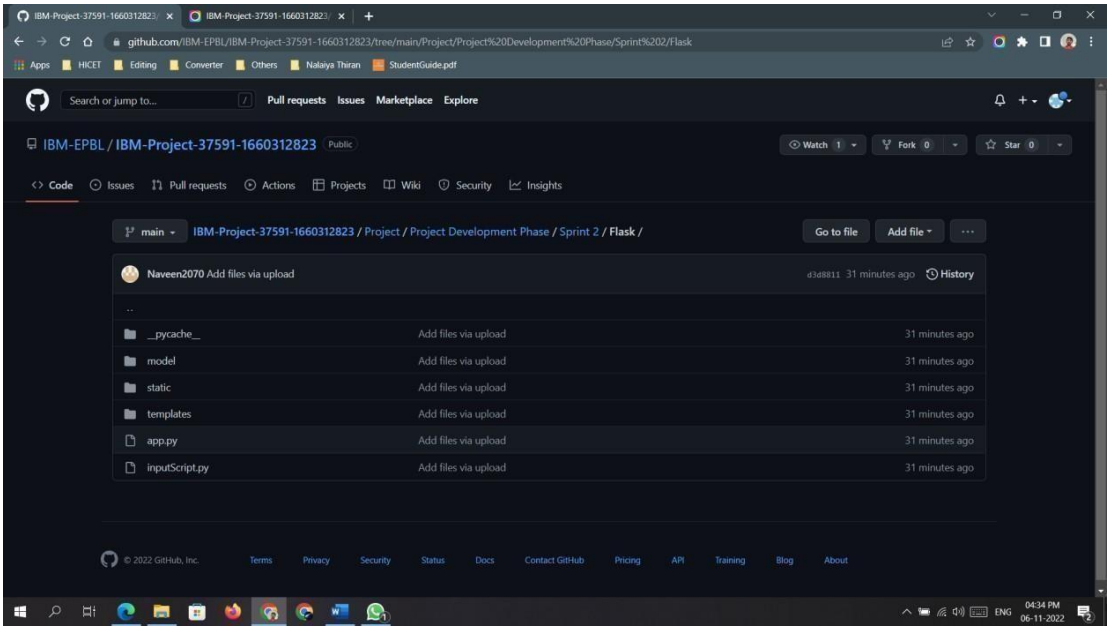
login.php

```
<?php
include 'config.php';
use PHPMailer\PHPMailer\PHPMailer;
use PHPMailer\PHPMailer\SMTP;
use PHPMailer\PHPMailer\Exception;
require 'vendor/autoload.php';
session_start();
error_reporting(0);
if (isset($_SESSION["user_id"]))
  {header("Location:
  welcome.php");
}
if (isset($_POST["signup"])) {
  $full_name = mysqli_real_escape_string($conn, $_POST["signup_full_name"]);
  $email = mysqli_real_escape_string($conn, $_POST["signup_email"]);
  $password = mysqli_real_escape_string($conn, md5($_POST["signup_password"]));
  $cpassword = mysqli_real_escape_string($conn, md5($_POST["signup_cpassword"]));
  $token = md5(rand());
  $check_email = mysqli_num_rows(mysqli_query($conn, "SELECT email FROM users
WHERE email='$email'"));
```

```
  if ($password !== $cpassword) {
    echo "<script>alert('Password did not match.');</script>";
  } elseif ($check_email > 0) {
    echo "<script>alert('Email already exists in out database.');</script>";
  } else {
    $sql = "INSERT INTO users (full_name, email, password, token, status) VALUES
('$full_name', '$email', '$password', '$token', '0')";
    $result = mysqli_query($conn, $sql);
    if ($result) {
      $_POST["signup_full_name"] = "";
      $_POST["signup_email"] = "";
      $_POST["signup_password"] = "";
      $_POST["signup_cpassword"] = "";
      $to = $email;
      $subject = "Email verification - Find Me";
      $message = "
      <html>
      <head>
      <title>{$subject}</title>
      </head>
      <body>
      <p><strong>Dear {$full_name},</strong></p>
      <p>Thanks for registration! Verify your email to access our website. Click below link to
verify your email.</p>
      <p><a href='{$base_url}verify-email.php?token={$token}'>Verify Email</a></p>
      </body>
      </html>
      ";
```

ProjectDevelopment–DeliveryofSprint2





ProjectDevelopment–DeliveryofSprint3

(INPROGRESS)

ProjectDevelopment–DeliveryofSprint4

(INPROGRESS)

8. CONCLUSION

The data set selected for this paper is imbalanced distributed, which may cause significant variation in the performance of each algorithm. In this paper, this problem was solved by the use of weighted evaluation measures. For future studies, using techniques such as SMOTE can better resolve this imbalance and improve the prediction. The result of algorithm comparison shows that tree-basedensemble algorithms tend to better predict flight delays of this data set. It will be valuable to repeat similar experimental processes using more tree-based ensemble algorithms to discover their significance in flight delay prediction.

REFERENCES

[1] Khaksar, H., &Sheikholeslami, A. (2017). Airline delay prediction by machine learning algorithms. *Scientia Iranica*.Ards, S., Chung, C., & Myers, S. (1999).

[2] Ye, B., Liu, B., Tian, Y., & Wan, L. (2020). A methodology for predicting aggregate flight departure delays in airports based on supervised learning. *Sustainability*, *12*(7), 2749

[3] ATLIOĞLU, M. C., BOLAT, M., ŞAHİN, M., TUNALI, V., & KILINÇ, D. (2020). Supervised learning approaches to flight delay prediction. *Sakarya University Journal of Science*

[4] Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 203–221

[5] Stefanovič, P., Štrimaitis, R., &Kurasova, O. (2020). Prediction of flight TIME deviation for Lithuanian airports using supervised machine learning model. *Computational Intelligence and Neuroscience*, *2020*, 1–10

[6] Oladipupo, T. (2010). Types of machine learning algorithms. *New Advances in Machine Learning*Hill, R. B. (2001). The role of race in foster care placements.

[7] Nibareke, T., &Laassiri, J. (2020). Using big Data-machinelearning models for Prediction and flight delay analytics.

[8] Hsiao C-Y, Hansen M. Air transportation network flows: equilibrium model. Transp Res Rec. 2005;1915(1):12–9.