

Flight Delay Prediction Model using Machine Learning

TEAM ID: PNT2022TMID02170

**KESHIKA.A - 2116190701094
KEVIN AUSTIN- 2116190701095
JENU MARIA.J- 2116190701083
JYOTSNA JOY- 2116190701085**

Project Report Format

1. INTRODUCTION

1.1 Abstract

2. LITERATURE SURVEY

2.1 Existing problem

2.2 References

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

3.3 Proposed Solution

3.4 Problem Solution fit

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

4.2 Non-Functional requirements

5. PROJECT DESIGN

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

5.3 User Stories

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

6.2 Sprint Delivery Schedule

6.3 Reports from JIRA

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature

7.2 Database Schema (if Applicable)

8. TESTING

8.1 Test Cases

8.2 User Acceptance Testing

9. RESULTS

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

Source Code

GitHub & Project Demo Link

1.

INTRODUCTION ABSTRACT

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.

The purpose of this project is to look at the approaches used to build models for predicting flight delays that occur due to bad weather conditions.

In the first part of the project, we look at using Python based Logistic Regression along with Support Vector Machine and then plugging the dataset into our classifier for results.

In the second part of the project, we primarily focus on gathering a dataset from Twitter, breaking the dataset down and identifying relevant attributes. Upon examining the results, we compare the results with other models such as Random Forest Classifier and derive the best classifier to solve the problem.

2.

LITERATURE SURVEY

S.NO	TITLE OF THEPAPER	Authors and Year	PROBLEMS ADDRESSED BY THE PAPER	METHODOLOGY USED	LIMITATION OFTHE SYSTEM	Dataset used and source of Dataset (ifany)	Accuracy (Ifany
1	Flight Delay Prediction Based on Aviation Big Data and Machine Learning	Guan Gui , Fan Liu(2020)	This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks.	This approach where the key ADS-BOUT subsystem,flight transmitters periodically send their own information to other flights and ground stations. The ADS-BIN subsystem, the flight receivers receive out-message from other flights andthe ground stations. paper, random forest-based and LSTM-	It does not focus on collecting,generating more training data, integrating more information like airport traffic flow, airport visibility into our dataset, and designing more delicate networks.		

				based architectures have been implemented to predict individual flight delay.			
2	Flight delay prediction based on deep learning and Levenberg-Marquart algorithm.	Maryam Farshchian Yazdi ¹ , Seyed Reza Kamel (2020)	This paper proposes a model for predicting flight delay based on Deep Learning (DL). DL is one of the newest methods employed in solving problems with high level of complexity and massive amount of data.	The approach of noisy data requires utility of stack denoising autoencoder (SDA) in designing the model and structure of the flight delay forecast in a model with Levenberg-Marquart (LM) algorithm. In addition, in this paper by developing a deep learning-based model, the accuracy of flight delay predictions can be increased.	It does not contain enough data for recognize all the signs. Accuracy is very bad.		
3	Chained Predictions of Flight Delay Using Machine Learning	Jun Chen, Meng Li	This paper presented a	The combinations of Bureau of Transportation Statistics	The updation of the actual departure delay with the		

		(2019)	new machine learning based air traffic delay prediction model that combined multi-label random forest classification and approximated delay propagation model.	(BTS), the National Oceanic and Atmospheric Administration (NOAA) and Aviation System Performance Metrics (ASPM), we design an algorithm to select the optimal training features that predicts the departure delay and arrival delay with the highest accuracy.	iteration number along the itinerary, the models accuracy can be further improved.		
4	Probabilistic Flight Delay Predictions Using Machine Learning and Applications to the Flight-to-Gate Assignment Problem	Micha Zoutendijk, Mihaela Mitici(2021)	This paper is based on utility of the estimated delay distributions, we integrate these probabilistic predictions into a probabilistic flight-to-gate assignment problem.	The prediction algorithm is employed, of which AdaBoost performs best. find that the features based on trajectory data contribute the greatest to the predictive accuracy, and the best result is found using LightGBM. The classification and regression results obtained in these studies generate an estimate for individual flight delay in the form of a class or a point estimate.	There is no varying assignment costs and airline gate usage, and, secondly, the integration of probabilistic flight delay predictions into models for other airport operations.		
5	Flight Delay Prediction	Bhuvan Bhatia(2016)	This project goal is to use	They discuss the Random Forest method and the	This project does not include a larger dataset		

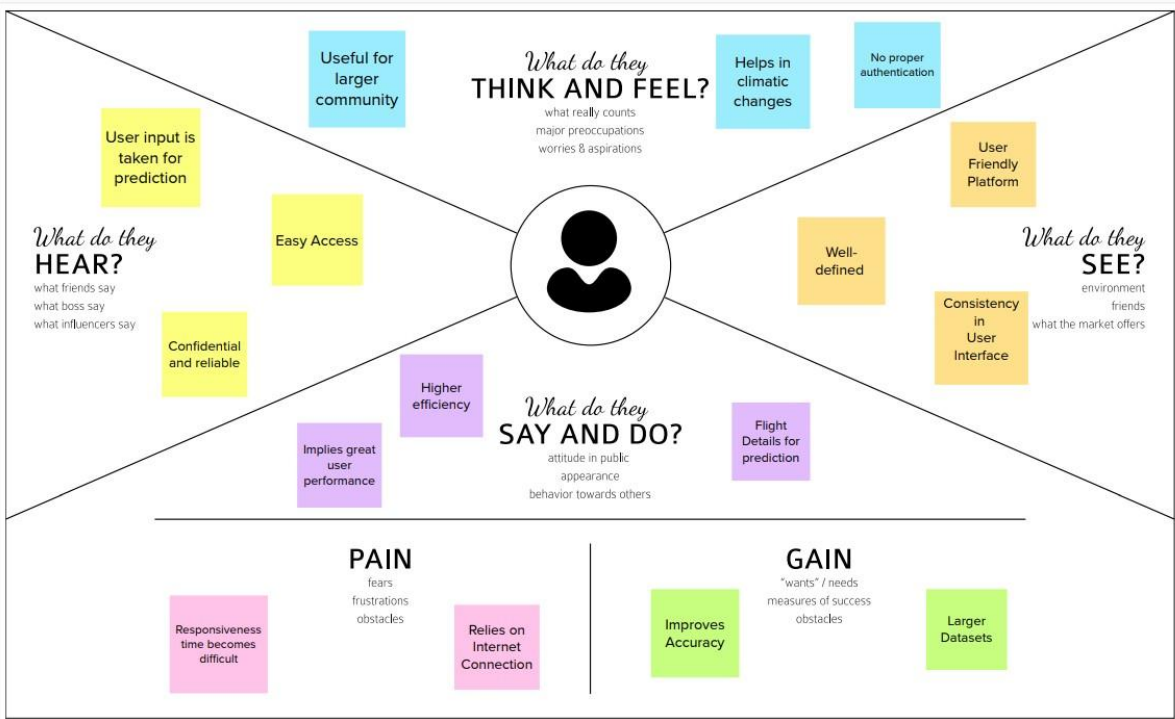
			exploratory analysis and to build machine learning models to predict airline departure and arrival delays.	SVM model.	where it has different ways to preprocess a larger dataset like running a Spark cluster over a server or using a cloud-based services like AWS and Azure to process the data.		
--	--	--	--	------------	---	--	--

3.

3.1.

IDEATION PHASE

EMPATHY MAP



3.3.

PROBLEM STATEMENT

Problem statement



3.4. PROBLEM SOLUTION FIT

Problem-Solution fit canvas 2.0

Purpose / Vision

Define CS, fit into CS	1. CUSTOMER SEGMENT(S) CS Business peoples and regular flight users face a important problem which was missing their flight due to un accurate prediction of arrival and delay. Many emergency patients whom have to fly for their treatment suffers due to flight delay.	6. CUSTOMER CONSTRAINTS CC Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. This project proposes a model for predicting flight delay based on Deep Learning (DL). DL is one of the newest methods employed in solving problems with high level of complexity and massive amount of data.	5. AVAILABLE SOLUTIONS AS Instead of linear regression prediction model, we used Gradient Boosted Algorithm. Merits: The better algorithm makes the prediction more accurate. Demerits: The delay of the flight causes time loss, financial and mental pressure on the passengers.	Explore AS, fit into AS
	2. PROBLEMS J&P The impact of flight delay can be a risk and this risk represents financial losses, the dissatisfaction of passengers, time losses, loss of reputation and bad business relations. If an airline doesn't deal with this problem immediately, it will cause other problems.	9. PROBLEM ROOT CAUSE RC <ul style="list-style-type: none"> Lack of or incorrect documentation Lack of or incorrect training Lack of management commitme 	7. BEHAVIOUR BE Due to delay of flight passenger looses his patience and his temper increases slightly	
Identify strong TR & EM	3. TRIGGERS TR Extreme Weather, Late Arriving Aircraft, Waiting for Connecting Passengers or Connecting Bags, Mechanical Delays or etc are the reasons why passengers face flight delays. It's important to notify the passengers about their flight is delayed or not which can make them, some flexibility into their schedule.	10. YOUR SOLUTION SL Data science based flight delay prediction uses gradient boosting algorithm is better and faster algorithm.Futher flight can be predicted.The prediction can be commonly found in web application using machine learning.	8. CHANNELS of BEHAVIOUR CH <div> 1. ONLINE We notify the information about of flight in web application. </div> <div> 8.2 OFFLINE You are offline in application show last information about the flight </div>	Extract online & offline CH of BE
	4. EMOTIONS: BEFORE / AFTER EM Many reasons 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 aspect. So here we propose a flight delay prediction model to predict if a flight will be delayed or not before it is even announced on the departure.			



Problem-Solution fit canvas is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license
 Created by Daria Nepriakhina / Amaltama.com

Architectural Workflow:

User view:

1. User enters flight details in the UI
2. Entered input is sent to the classifier model deployed through IBM Watson.
3. The model predicts the estimated time of departure/arrival delay and sends it to the UI.
4. The predicted value is displayed to the UI

Model view

1. The dataset is preprocessed for handling missing/categorical values.
2. Spatial and other features are extracted.
3. The features are split into training and test set.
4. A Random forest classifier is built and is trained with the training data.
5. The model is evaluated using testing data.
6. The trained model is deployed in IBM Watson.

CUSTOMER JOURNEY



Customer experience journey map

Use this framework to better understand your customer's needs, motivations, and "bottlenecks" by illustrating a key scenario or process from start to finish. When possible, use this map to document and summarize interviews and observations with real people rather than relying on your hunches or assumptions.

Product School

submitted by
19070074 - Kesha A (T)
19070105 - Kevin Austin (TM1)
190.01083 - Jenu Maria J (IM2)
190701085 - Jyotsna Joy (TM3)

Team ID: PNT2322TMD02.70

© 2019 Product School



Document an existing experience

Have your team focus on a specific scenario or process within a existing product or service. In the **Start** row, document the step-by-step process someone typically experiences, then add details to each of the other rows.

How do you feel about this experience? What do you like? What do you dislike? What do you want to change?

Start	Entice	Enter	Engage	Exit	Extend
Start How do you, as a customer, begin your journey? What is the first step?	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Steps What does the person (or group) typically experience?	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Interactions What interactions do they have with each step along the way? • "Hello, how are you?" • "Welcome, please follow me." • "Thank you for your purchase." • "We'll be in touch soon." • "We'll be in touch soon."	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Goals & motivation At each step, what is a person's primary goal or motivation? (What are they trying to do...?)	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Positive moments What steps does a typical person find enjoyable, surprising, fun, or satisfying, delightful, or exciting?	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Negative moments What steps does a typical person find frustrating, confusing, annoying, costly, or time-consuming?	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?
Areas of opportunity How might we make each step better? What ideas do we have? What have others suggested?	Entice How does someone get into the experience? What is the first step?	Enter What do people experience as they begin the process?	Engage In the core moments in the process, what happens?	Exit What do people typically experience as the process finishes?	Extend What happens after the experience is over?



Need a hand? We're here to help. Contact us at support@productschool.com or call 1-800-555-1234.



4. Project Planning Phase
4.1. SPRINT DELIVERY PHASE

Product Backlog, Sprint Schedule, and Estimation :

Sprint	Functional Requirement(Epic)	User Story Number	User Story/Task	Story point	Priority	Team Members
Sprint-1	Registration	USN-2	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	KESHIKA A
Sprint-1		USN-2	As a user, I will receive confirmation email once I have register for the application	1	High	KESHIKA A
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	KEVIN AUSTIN
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	JENU MARIA J
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	JYOTSNA JOY

Project Tracker :

Sprint	Total Story point	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Point Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	26 Oct 2022	31 Oct 2022	20	31 Oct 2022
Sprint-2	20	6 Days	1 Nov 2022	6 Nov 2022	20	6 Nov 2022
Sprint-3	20	6 Days	7 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	13 Nov 2022	18 Nov 2022	20	18 Nov 2022

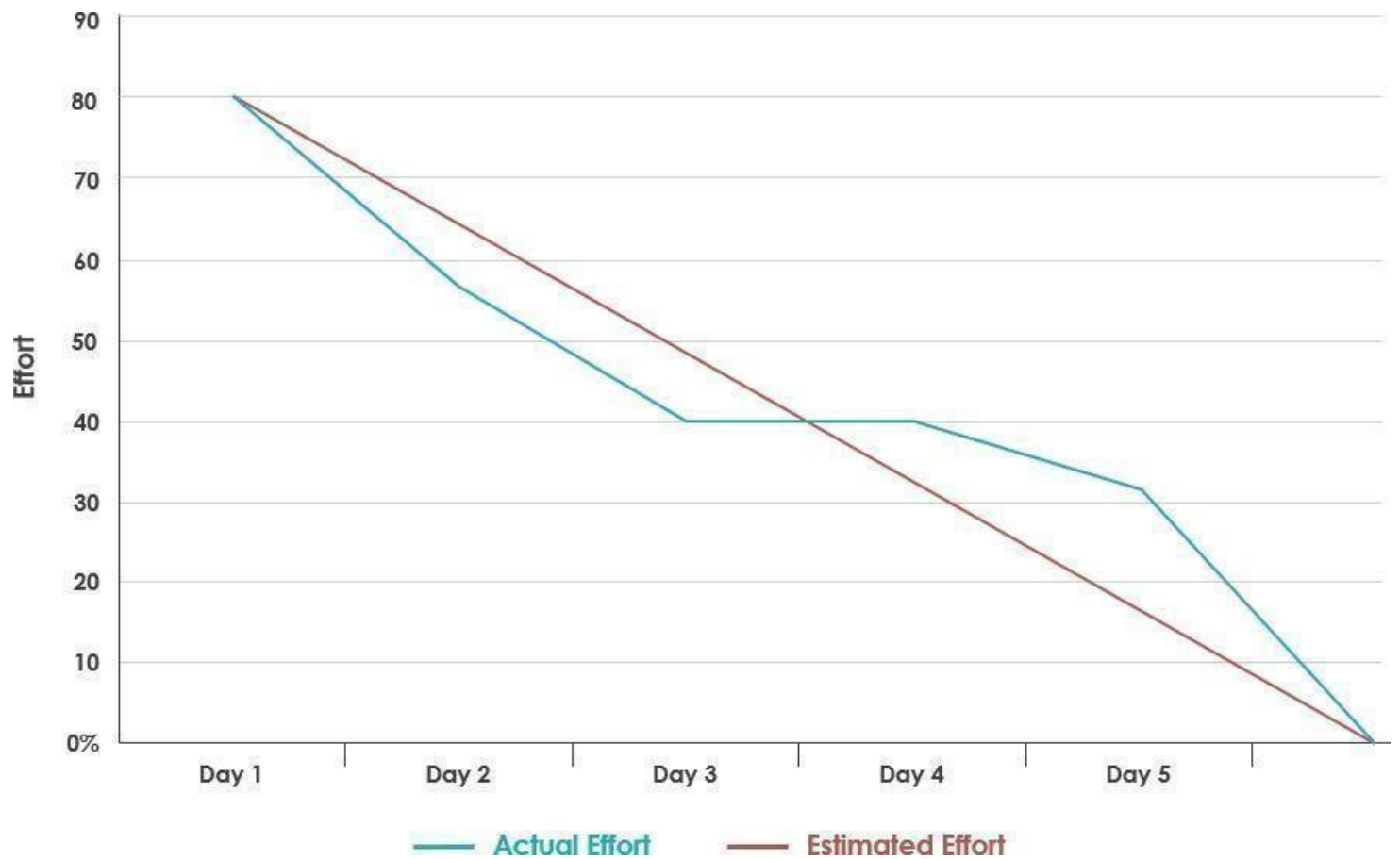
Velocity:

We have a 5-day sprint duration, and the velocity of the team is 15 (points per sprint). The team's average velocity (AV) per iteration unit (storypoints per day)

$$\text{Actual Velocity} = \text{Sprint} \\ \text{Duration/Velocity} = 15/5 = 3$$

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time



4.2. MILESTONE AND ACTIVITY LIST

S.NO	ACTIVITY TITLE	ACTIVITY DESCRIPTION	DURATION
1	Project preparation	Assign team members, Create repository in the GitHub, download rocket-chat essentials and join respective project channel.	1 WEEK
2	Attend class	Attend sessions on IBM, team leader assign task to each member of the project, attend quiz , submit assignment .	1 WEEK
3	Working on different phases of project	Ideation phase-literature survey, Project design phase-I-proposed solution, solution architecture, project design phase II-customer journey ,data flow ,technical architecture, planning phase-milestones, tasks, sprints schedule.	4 WEEK
4	Developing project	Develop the code, test and push it to GitHub, clarify queries.	2 WEEK
5	Budget and scope of project	Analyze and making the project budget and discuss with team for budget prediction .	1 WEEK

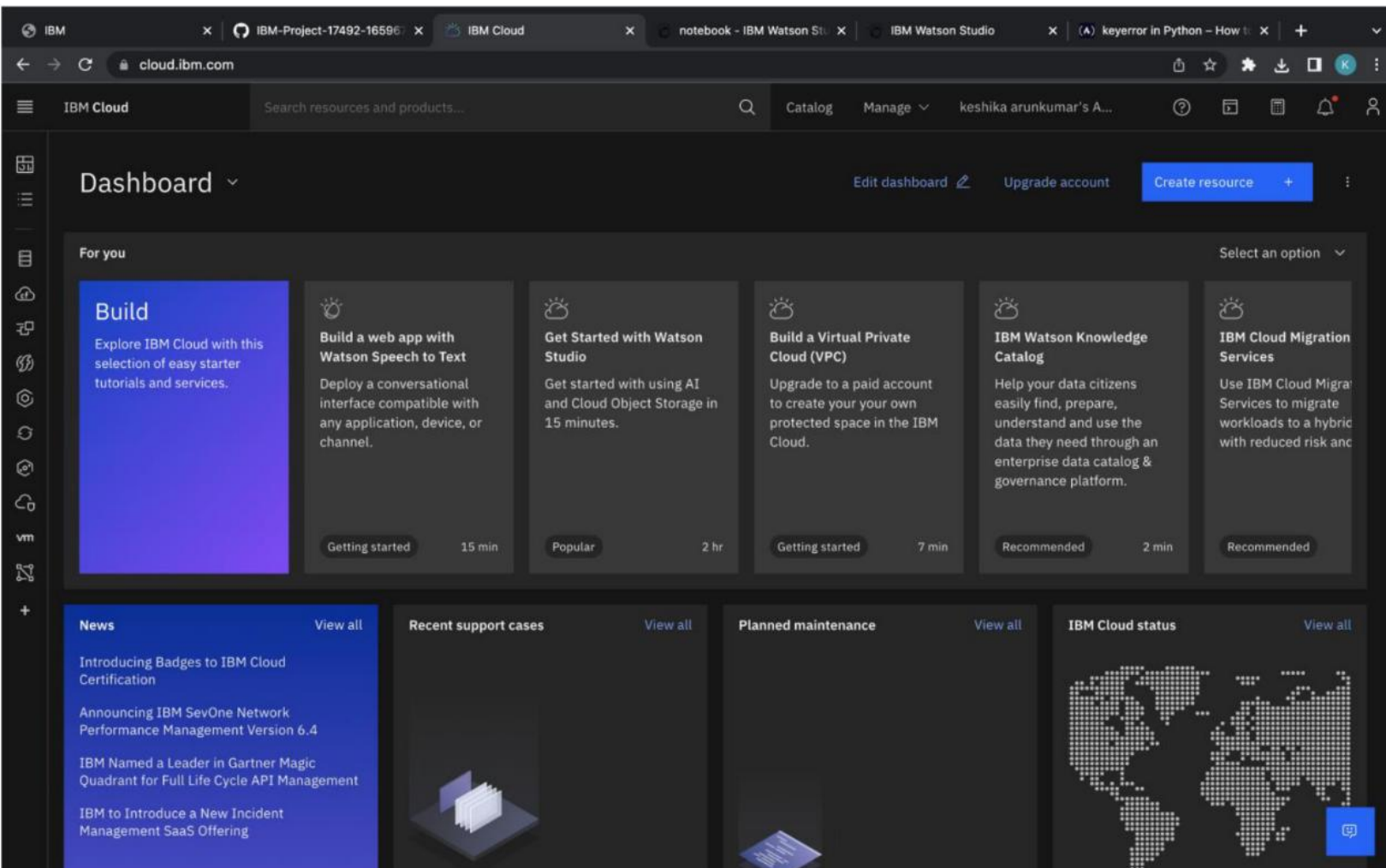
Milestone:

When project begins then it is expected that project related activities must be initiated. In project planning, series of milestones must be established. Milestone can be defined as recognizable endpoint of software project activity. At each milestone, report must be generated.

Milestone is distinct and logical stage of the project. It is used as signal post for project start and end date, need for external review or input and for checking budget, submission of the deliverable, etc. It simply represents

clear sequence of events that are incrementally developed or build until project gets successfully completed. It is generally referred to as task with zero-time duration because they are used to symbolize an achievement or point of time in project. It helps in signifying change or stage in development.

Result Screenshot:



IBM Project-17492-16596 IBM Cloud notebook - IBM Watson Studio IBM Watson Studio keyerror in Python - How to

dataplatform.cloud.ibm.com/analytics/notebooks/v2/fd757821-19f5-46f1-b8d0-225db55a93aa?projectId=ac053c85-7d36-4d15-839f-e9736bc40bfd&conte...

IBM Watson Studio Search in your workspaces Buy keshika arunkumar's Acco... Dallas KA

Projects / flight delay prediction / notebook

File Edit View Insert Cell Kernel Help Not Trusted Python 3.9

DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

Data Preprocessing

```
In [1]: # importing the required libraries
import sys
import numpy as np
import pandas as pd
import seaborn as sns
import pickle
%matplotlib inline
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import sklearn.metrics as metrics
import warnings
warnings.filterwarnings('ignore')
```

Importing Data

```
In [2]: import os, types
import pandas as pd
from botocore.client import Config
import boto3

def iter (self): return 0
```

IBM Project-17492-16596 IBM Cloud notebook - IBM Watson Studio IBM Watson Studio keyerror in Python - How to

dataplatform.cloud.ibm.com/ml-runtime/deployments/f26e6fbc-bb80-4bed-8366-278db09a3e06?space_id=a8bfc874-2391-4d38-bf7d-3d9d2a37b9cd&co...

IBM Watson Studio Search in your workspaces Buy keshika arunkumar's Acco... Dallas KA

Deployments / space / flightdelay /

flightdelay

Deployed Online

API reference Test

Direct link

Endpoint

<https://us-south.ml.cloud.ibm.com/ml/v4/deployments/f26e6fbc-bb80-4bed-8366-278db09a3e06/predict> IAM

Code snippets

cURL	Java	JavaScript	Python	Scala
<p># NOTE: you must set \$API_KEY below using information retrieved from your IBM Cloud account.</p> <pre>curl --insecure -X POST --header "Content-Type: application/x-www-form-urlencoded" --header "Accept: application/json" --data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey" --data-urlencode "apikey=\$API_KEY" "https://iam.cloud.ibm.com/identity/token"</pre> <p># the above CURL request will return an auth token that you will use as \$IAM_TOKEN in the scoring request below</p> <p># TODO: manually define and pass values to be scored below</p> <pre>curl -X POST --header "Content-Type: application/json" --header "Accept: application/json" --header "Authorization: Bearer \$IAM_TOKEN" -d '{"input_data": [{"fields": [ARRAY_OF_INPUT_FIELDS], "values": [ARRAY_OF_VALUES_TO_BE_SCORED, \$ANOTHER_ARRAY_OF_VALUES_TO_BE_SCORED]}]}' "https://us-south.ml.cloud.ibm.com/ml/v4/deployments/f26e6fbc-bb80-4bed-8366-278db09a3e06/predict"</pre>				

flightdelay

Created Nov 19, 2022, 11:15 PM

Updated Nov 19, 2022, 11:15 PM

Deployment ID f26e6fbc-bb80-4bed-8366-278db...

Software specification runtime-22.1-py3.9

Copies 1

Serving name No serving name.

Description No description provided.

Tags Add tags to make assets easier to find.

Associated asset flightdelay 7aaa81a8-e0f7-464...

Model ID 7aaa81a8-e0f7-4648-a9ab-28e4e...

Restricted Mode is intended for safe code browsing. Trust this window to enable all features. [Manage](#) [Learn More](#)

app_ibm.py x app.py

Users > keshikaarunkumar > Desktop > ntp 2 > app > app_ibm.py

```
1 import numpy as np
2 import os
3 from flask import Flask, request, jsonify, render_template, json
4 import pickle
5 import requests
6
7 # NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
8 API_KEY = "B_EVqatLL8m6PwFFYDpCpevQdnFSa0_NrmyPks-lUclUY"
9 token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
10 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
11 mltoken = token_response.json()["access_token"]
12
13 header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
14 app = Flask(__name__)
15 model = pickle.load(open('rfmodel.pkl', 'rb'))
16
17 @app.route('/')
18 def home():
19     return render_template('mainpage.html')
20
21 @app.route('/predict', methods=['POST'])
22 def predict():
23     ...
24     For rendering results on HTML GUI
25     ...
26     sm=[6,7,8]
27     wt=[9,10,11]
28     sp=[12,1,2,3]
29     fl=[4,5]
30     farr=[int(x) for x in request.form.values()]
31     if farr[1] in sm:
32         farr.append(0)
33     elif farr[1] in wt:
34         farr.append(1)
35     elif farr[1] in sp:
36         farr.append(2)
37     else:
38         farr.append(3)
39     farr += [1, 2, 3, 4]
40     final_features=[int(x) for x in farr]
41     print(final_features)
42     payload_scoring = {"input_data": [{"fields": [{"QUARTER", "MONTH", "DAY_OF_MONTH", "DAY_OF_WEEK", "FL_NUM", "ORIGIN", "DEST", "CRS_DEP_TIME.1", "CRS_ARR_TIME.1", "CRS_ELA
43
44     response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/f26e6fbc-bb80-4bed-8366-278db09a3e06/predictions?version=2022-11-19', json=
```

Ln 48, Col 32 Spaces: 4 UTF-8 CRLF Python

GitHub :

<https://github.com/IBM-EPBL/IBM-Project-17492-1659672629>

Project Demo Link:

<https://github.com/IBM-EPBL/IBM-Project-17492-1659672629/blob/main/project%20development/sprint4/demovideo.mov>