A project report on

Developing a Flight Delay Prediction Model using Machine Learning

BE.ELECTRONIC & COMMUNICATION ENGINEERING

BY

AJAY KUMAR.P(511319106002)

SANJAY.S(511319106027)

SUBASH.C(511319106030)

SUNDARA MOORTHY.GP(511319106031)

ABSTRACT:

Flight delays hurt airlines, airports, and passengers. Their prediction is crucial during the decision making process for all players of commercial aviation. Moreover, the development of accurate prediction models for flight delays became cumbersome due to the complexity of air transportation system, the number of methods for prediction, and the deluge of flight data.

In this context, this paper presents a thorough literature review of approaches used to build flight delay prediction models from the Data Science perspective. We propose a taxonomy and summarize the initiatives used to address the flight delay prediction problem, according to scope, data, and computational methods, giving particular attention to an increased usage of machine learning methods

Besides, we also present a timeline of significant works that depicts relationships between flight delay prediction problems and research trends to address them

Project Report Format

1. INTRODUCTION

- Project Overview
- Purpose

2. LITERATURE SURVEY

- 2.1 Existing problem
- 2.2 References
- 2.3 Problem Statement Definition

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

1.INTRODUCTION:

Delay is one of the most remembered performance indicators of any transportation system. Notably commercial aviation players understand delay as the period by which a flight is late or postponed. Thus delay may be represented by the difference between scheduled and real times of departure or arrival of a plane . Country regulator authorities have a multitude of indicators related to tolerance thresholds for flight delays. Indeed, flight delay is an essential subject in the context of air transportation systems. In 2013, 36% of flights delayed by more than five minutes in Europe, 31.1% of flights delayed by more than 15 minutes in the United States, and 16.3% of flights were canceled or suffered delays greater than 30 minutes in Brazil . This indicates how relevant this indicator is and how it affects no matter the scale of airline meshes."Flight Delay Predictions" is a supervised machine learning project. The primary goal of this project is to predict airline delays caused by various factors. Flight delays lead to negative impacts, mainly economical for commuters, airline industries and airport authorities. Furthermore, in the domain of sustainability, it can even cause environmental harm by the rise in fuel consumption and gas emissions. Here we train a regression model to predict if a flight will be delayed by more than 15 minutes. The model was trained using features of the flights known at the time of booking such as the airline, month, week, and hour of departure. This report will help to understand basic concepts of supervised machine learning and the concepts used in making this project.

Project Overview

This dataset reports flights in the United States, including carriers, arrival and departure delays, and reasons for delays, for 2016. This dataset was obtained from the **RITA** (Research and Technology Bureau of Transportation Statistics)

A flight is considered delayed when it arrived 15 or more minutes than the schedule . Delayed minutes are calculated for delayed flights only.

When multiple causes are assigned to one delayed flight, each cause is

prorated based on delayed minutes it is responsible for. The displayed numbers are rounded and may not add up to the total.

Purpose:

Therefore, predicting flight delays can improve airline operations and passenger satisfaction, which will result in a positive impact on the economy. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.

2.LITERATURE SURVEY:

Nowadays, service quality plays an important role in attracting customers. Among these, air travels have their special customers and the most important matter
in these travels is the flight time, on time arrival at destination for passengers such those who have an important meeting, that has been leading to high expenses for the passenger until they get to their destination on time. Flight delay has negative economic effects on the passenger, agencies and airport. Therefore, any reduction of
these effect requires decreasing postponed flight price, so that prediction or estimation has a great significance and numerous studies has been to dedicated this subject. Correspondingly, all the scientists have tried to design a model that understands effective factors and computes effect of each factor and their relation. Overall, the prediction methods are classified into five groups including Statistical Methods, Probability methods, network-based methods, operational methods and machine learning methods .
In one of the best studies that has been performed based on statistics delay time has been considered to be reduced. Their study has investigated important factors before fly and those which occur on the ground. In the next step, it has predicted the delay at destination based on factors that occur in the vicinity of arrival

time at destination. Eventually, results have shown that whenever, the delay is correctly predicted, passenger disaffection and fuel consumption decrease and consequently number of flight increases. Moreover, it is possible to increase the agencies benefits through reducing number of passengers who wrongly selected their routes or specifying the probabilities for some flights and optimizing delay time
prediction.

Another prominent investigation based on Probability has been done and the author believes that huge storm in U.S.A has highly affected the flight delay. This study has been devoted to predict delay based on mathematical calculations and through considering delay time duration of the flights that had been engaged to storm in the same day. Metrological reports have shown the effect of storm one hour
before and after event cause ephemeral climate at the region. In the next step, Monte-Carlo simulation has been used to estimate the airport runway capacity, so that traffic of each runway would have been estimated. As the research has employed only one factor, the model has not enough accuracy, but it is possible to increase region air capacity path structure.

REFERENCES:

Paper1 : **DOI: https://doi.org/10.1145/3497701.3497725**
ICEBI 2021: 2021 5th International Conference on E-Business and Internet, Singapore, Singapore, October 2021

Paper2: 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)

Paper3: IOP Conference Series: Earth and Environmental Science, Volume 81, 2nd International Conference on Materials Science, Energy Technology and Environmental Engineering (MSETEE 2017) 28–30 April 2017, Zhuhai, China**Citation** Yi Ding 2017 *IOP Conf. Ser.: Earth Environ. Sci.* **81** 012198**DOI** 10.1088/1755-1315/81/1/012198

EXISTING PROBLEMS:

- **Air Carrier:** The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- **Extreme Weather:** Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
- **National Aviation System (NAS):** Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- **Late-arriving aircraft:** A previous flight with same aircraft arrived late, causing the present flight to depart late.
- **Security:** Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

PROBLEM STATEMENT DEFINITION:

Problem is the core feature in domain taxonomy. There are three major concerns regarding the flight delay prediction problem

1.**delay propagation**

2.root delay and

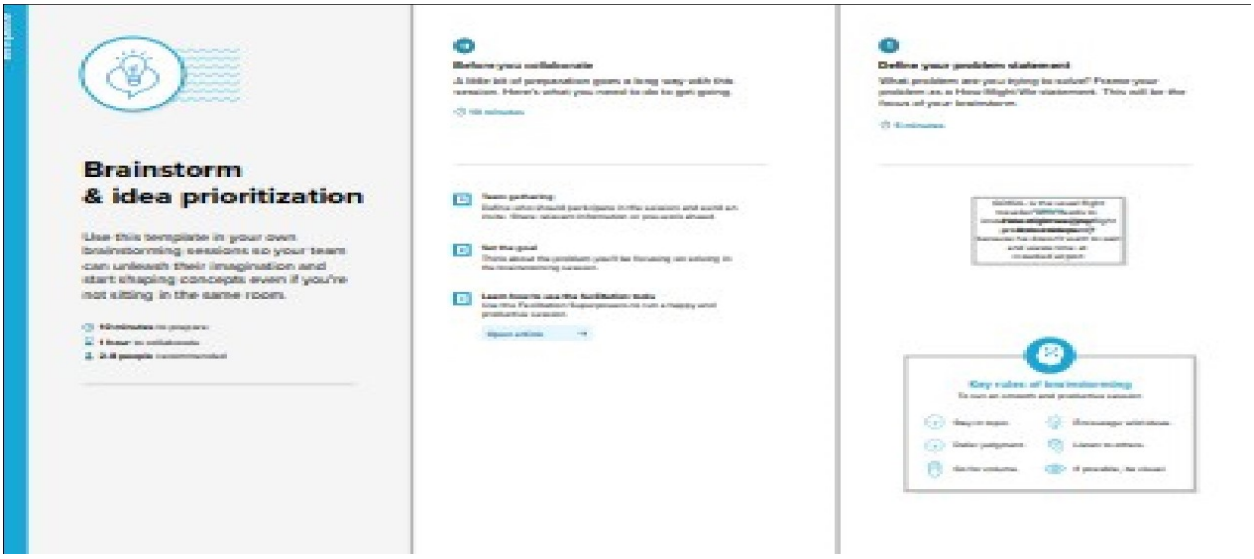
3.cancellation.

Depending on the emphasis of the research, authors select one of these lines to develop their models.

IDEATION AND PROPOSED SOLUTION:

EMPATHY MAP CANVAS:

IDEATION AND BRAINSTORMING:



2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

Ajay Kumar

- Can improve airline operations and passenger satisfaction,
- By using Machine Learning (ML) Algorithms you can try to predict if your flight will be delayed in many ways
- By proper communication to passengers
- By predicting the past Flight Delays

Sanjay

- Check wether conditions
- Reducing air traffic
- Creating a delay model using previous delays
- Developing a AI model to predict

Sundara Moorthy

- Check flight mechanism
- Increasing aviation signal detection strength
- Creating proper airways
- Checking aviation detection

Subash

- Be ready for the process
- Check the weather before you leave
- Research your flight's on -time performance
- Be informed about your delays

3

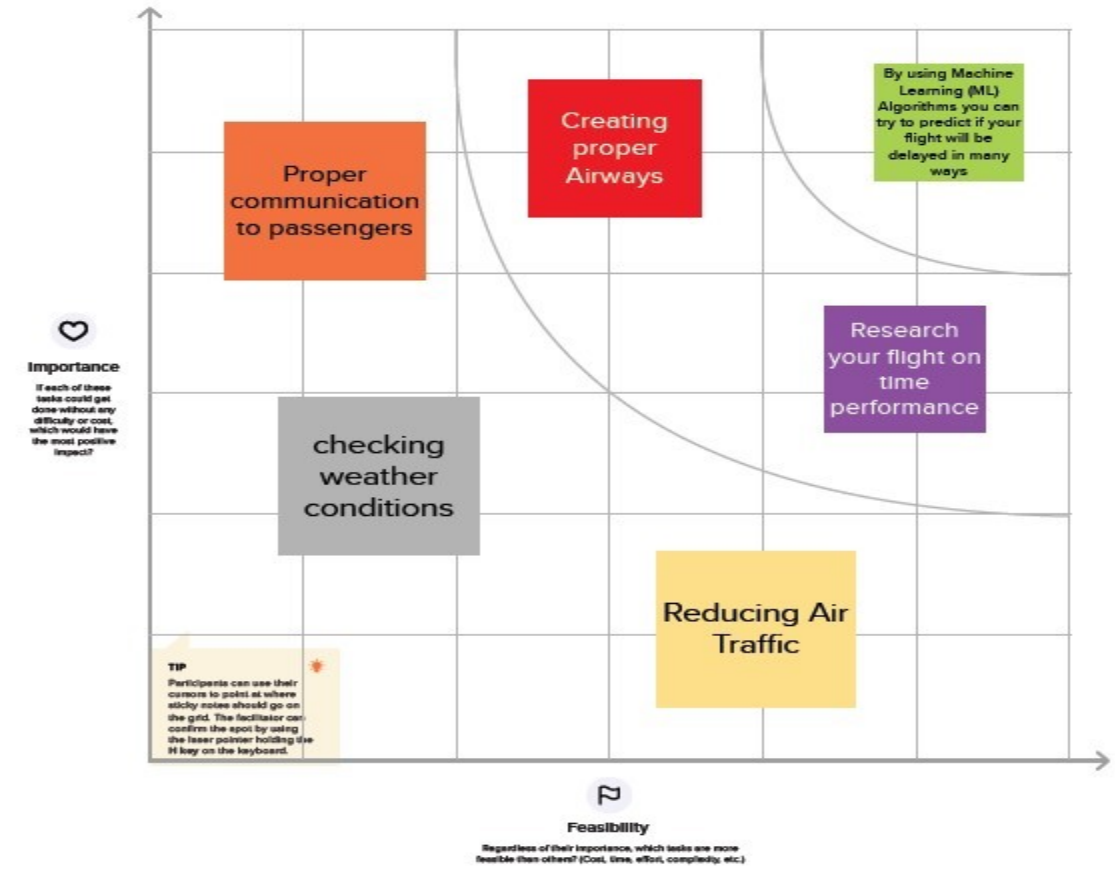
Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, 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

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.



PROPOSED SOLUTION:

Proposed Solution Template:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	GOKUL is the usual flight traveller who needs to know about his traveling flight arrival delays because he doesn't want to wait and waste time at crowded airport
2.	Idea / Solution description	Creating a Prediction model using Machine Learning to predict the flight delays
3.	Novelty / Uniqueness	The accuracy of the prediction will be high and it will be usefull for the passengers to predict their flight delays
4.	Social Impact / Customer Satisfaction	The customer can be able to predict their flight delays from home not to going by airport.
5.	Business Model (Revenue Model)	It can be expanded and be converted to business model for the usual flight travellers.
6.	Scalability of the Solution	The scalability of the project is high. It can be expanded to perfom on increased work load.

REQUIREMENT ANALYSIS:

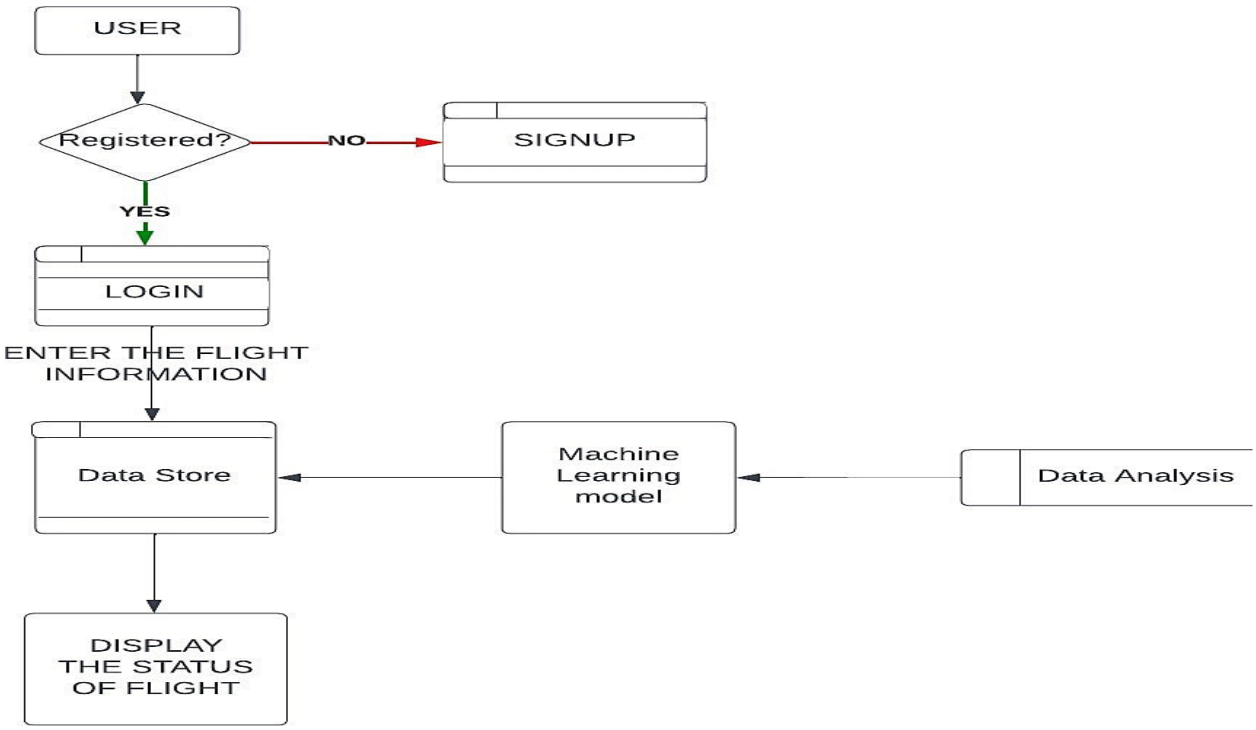
FUNCTIONAL REQUIREMENT:

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via MOBILE OTP
FR-3	User location verification	Confirmation using gps
FR-4	User nationality	Confirmation of nationality Identity verification

NON FUNCTIONAL REQUIREMENT:

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Precise. Easy to access.
NFR-2	**Security**	Datas are integrated. Informations are hidden from third parties.
NFR-3	**Reliability**	Flight informations are true to extent. Better prediction.
NFR-4	**Performance**	High speed performance. Uninterrupted services.

4.PROJECT DESIGN:



USER STORIES:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-4	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-5	As a user I can view my predictions on my dashboard		High	Sprint-1

Customer (Web user)	Registration	USN-1	As a web user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Spint-1
		USN-2	As a web user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a web user, I can log into the application by entering email & password		High	Sprint-1
	Wireframe	USN-4	As a web user , I can have a better vision about the website	I experience a good content placement and workflow.	High	Sprint-1
	Screen space	USN-5	As a web user , I can have a better illumination	I can have a good interaction with website	Medium	Sprint-1
Customer Care	Knowledge	USN-1	As a customer care executive , I must have	User will have a good	High	Sprint-2

PROJECT PLANNING AND SCHEDULING:

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.	3	High	Ajay Kumar
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Subash C
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	Sundaramoorthy
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Sanjay S
Sprint-1	Profile Page	USN-6	As a user, I can view my profile	1	High	Ajay Kumar
Sprint-2		USN-3	As a user, I can register for the application through Facebook, Instagram, other social media	2	Low	Sanjay S
Sprint-2	Search	USN-7	As a user, I can search for flights for different locations	2	High	Subash C
Sprint-2	View	USN-8	As a user, I can view the details of flights	1	High	Sundaramoorthy
Sprint-2	Analyse	USN-12	As an admin, I will analyse the given dataset	5	High	Ajay Kumar
Sprint-2	Predict	USN-13	As an admin, I will predict the delays	8	High	Sanjay S

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-3	Visualisation		Visualize the predicted data	5	High	Ajay Kumar
Sprint-3, 4	Receive notifications	USN-9	As a user, I will receive notifications about the flights	3	Low	Subash C
Sprint-3, 4			Backend for notifications	5	Low	SundaraMoorhty
Sprint- 3, 4	Track	USN-10	As a user, I can track the location of my flight	3	Medium	Sanjay S

Spint-3, 4	GPS	USN-11	As an admin, I will need the location of flights	3	High	Sanjay S

7.CODING AND SOLUTION:

FEATURE 1:

FLASK APP CODE:

```
from flask import Flask, request, jsonify, render_template, url_for, request
import pickle
app = Flask(__name__)
model = pickle.load(open('flightclf.pkl', 'rb'))
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict',methods=['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 == "map"):
        origin1, origin2, origin3, origin4, origin5 = 0, 0, 0, 0, 1
    if (origin == "dtw"):
        origin1, origin2, origin3, origin4, origin5 = 1, 0, 0, 0, 0
    if (origin == "jfk"):
        origin1, origin2, origin3, origin4, origin5 = 0, 0, 1, 0, 1
    if (origin == "sea"):
        origin1, origin2, origin3, origin4, origin5 = 0, 1, 0, 0, 1
    if (origin == "alt"):
        origin1, origin2, origin3, origin4, origin5 = 0, 0, 0, 1, 0


    destination = request.form['destination']
    if (destination == "map"):
        destination1, destination2, destination3, destination4, destination5 = 0, 0, 0, 0, 1
```

```
    if (destination == "dtw"):
        destination1, destination2, destination3, destination4, destination5 = 1, 0, 0, 0, 0
    if (destination == "jfk"):
        destination1, destination2, destination3, destination4, destination5 = 0, 0, 1, 0, 0
    if (destination == "sea"):
         destination1, destination2, destination3, destination4, destination5 = 0, 1, 0, 0, 0
    if (destination == "alt"):
        destination1, destination2, destination3, destination4, destination5 = 0, 0, 0, 1, 0
    dept = request.form['dept']
    arrtime = request.form['arrtime']
    actdept = request.form['actdept']
    dept15 = int(dept) - int(actdept)
    total = [[name, month, dayofmonth, dayofweek, origin1, origin2, origin3, origin4, origin5,
destination1, destination2,
          destination3, destination4, destination5]]
    y_pred = model.predict(total)
    print(y_pred)

    if (y_pred == [0.]):
        ans = "The flight will be on time"
    else:
        ans = "the flight will be delayed"
    return render_template("index.html", showcase=ans)
if __name__== '__main__':
    app.run(debug=False)
```

HTML CODE:

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <!-- The above 3 meta tags *must* come first in the head; any other head content must come
*after* these tags -->

    <title>Flight Delay Prediction</title>
```

```
    <!-- Google font -->
    <link href="https://fonts.googleapis.com/css?family=Lato:400,700" rel="stylesheet">

    <!-- Bootstrap -->
    <link type="text/css" rel="stylesheet" href="{{ url_for('static', filename='css/bootstrap.min.css')
}}" />

    <!-- Custom stlylesheet -->
    <link type="text/css" rel="stylesheet" href="{{ url_for('static', filename='css/style.css') }}" />
    <style>
    #booking
    {
            font-family: 'Lato', sans-serif;
    background: url(../static/background.jpg);
            background-size: cover;
            background-position: center;
            color: #191a1e;
    }
    </style>
    <!--<img src="{{url_for('static', filename='background.png')}}" />-->

    <!-- HTML5 shim and Respond.js for IE8 support of HTML5 elements and media queries -->
    <!-- WARNING: Respond.js doesn't work if you view the page via file:// -->
    <!--[if lt IE 9]>
                        <script
src="https://oss.maxcdn.com/html5shiv/3.7.3/html5shiv.min.js"></script>
                        <script src="https://oss.maxcdn.com/respond/1.4.2/respond.min.js"></script>
                      <![endif]-->

</head>

<body>
    <div id="booking" class="section">
        <div class="section-center">
            <div class="container">
                <div class="row">
                    <div class="col-md-4">
                        <div class="booking-cta">
                            <h1>Flight Delay Prediction</h1>
                            <p></p>
```

```
                <div class="container">
                  <!-- {{ prediction_text }} -->
                  {% if prediction_text == 0 %}
                  <h2>The flight is not delayed</h2>

                  {% elif prediction_text == 1 %}
                  <h2>The flight is delayed</h2>
                  {% endif %}

                </div>
              </div>
            </div>

            <div class="col-md-7 col-md-offset-1">
              <div class="booking-form">
                <form action={{ url_for("predict") }} method="post">
                  <div class="row">

                    <div class="col-md-4">
                      <div class="form-group">
                        <span class="form-label">Year</span>
                          <input type="text" class="form-control" name="year"
placeholder="Enter year" required="true">
                        <span class="select-arrow"></span>
                      </div>
                    </div>
                    <!Year>

                    <div class="col-md-4">
                      <div class="form-group">
                        <span class="form-label">Month</span>
                        <input type="text" class="form-control" name="month"
placeholder="Enter month" required="true">
                        <span class="select-arrow"></span>
                      </div>
                    </div>
                    <!Month>

                    <div class="col-md-4">
                      <div class="form-group">
                        <span class="form-label">Date</span>
```

```
                                <input type="text" class="form-control" name="day"
placeholder="Enter date" required="true">
                                    <span class="select-arrow"></span>
                                </div>
                            </div>
                            <! Date>

                    </div>
                    <!--Year,Month,Date end-->

                    <div class="row">

                        <div class="col-md-6">
                            <div class="form-group">
                                <span class="form-label">Select an Airline</span>
                                <!-- <input class="form-control" type="date" required> -->
                                <select class="form-control" name="carrier">
                                    <option value="UA">United Air Lines Inc.(UA)</option>
                                    <option value="AA">American Airlines Inc.(AA)</option>
                                    <option value="US">US Airways Inc.(US)</option>
                                    <option value="F9">Frontier Airlines Inc.(F9)</option>
                                    <option value="B6">JetBlue Airways(B6)</option>
                                    <option value="OO">Skywest Airlines Inc.(OO)</option>
                                    <option value="AS">Alaska Airlines Inc.(AS)</option>
                                    <option value="WN">Southwest Airlines Co.(WN)</option>
                                    <option value="DL">Delta Air Lines Inc.(DL)</option>
                                    <option value="EV">Atlantic Southeast Airlines(EV)</option>
                                    <option value="HA">Hawaiian Airlines Inc.(HA)</option>
                                    <option value="MQ">American Eagle Airlines Inc.(MQ)</option>
                                    <option value="VX">Virgin America(VX)</option>
                                    <option value="9E">Endeavor Air(9E)</option>
                                    <option value="FL">AirTran Airways(FL)</option>
                                    <option value="YV">Mesa Airlines(YV)</option>
                                </select>
                            </div>
                        </div>
                        <!Airline>

                    </div>
                    <!--Airline end-->
```

```
                    <div class="row">

                      <div class="col-md-6">
                        <div class="form-group">
                          <span class="form-label">Flying from</span>
                          <!--<input class="form-control" type="text" placeholder="City or
airport">-->
                          <select class="form-control" name="origin">
                            <option value="EWR">Newark Liberty International
Airport(EWR)</option>
                            <option value="JFK">John F. Kennedy International Airport(New York
International Airport)(JFK)</option>
                            <option value="LGA">LaGuardia Airport(Marine Air
Terminal)(LGA)</option>
                          </select>
                        </div>
                      </div>
                      <!Flying from>

                      <div class="col-md-6">
                        <div class="form-group">
                          <span class="form-label">Flying to</span>
                          <!-- <input class="form-control" type="text" placeholder="City or
airport"> -->
                          <select class="form-control" name="dest">
                            <option value="ATL">Hartsfield-Jackson Atlanta International
Airport(ATL)</option>
                            <option value="ORD">Chicago O'Hare International
Airport(ORD)</option>
                            <option value="LAX">Los Angeles International
Airport(LAX)</option>
                            <option value="BOS">Gen. Edward Lawrence Logan International
Airport(BOS)</option>
                            <option value="MCO">Orlando International Airport(MCO)</option>
                            <option value="CLT">Charlotte Douglas International
Airport(CLT)</option>
                            <option value="SFO">San Francisco International
Airport(SFO)</option>
                            <option value="FLL">Fort Lauderdale-Hollywood International
Airport(FLL)</option>
                            <option value="MIA">Miami International Airport(MIA)</option>
```

<option value="DCA">Ronald Reagan Washington National Airport(DCA)</option>
<option value="DTW">Detroit Metropolitan Airport(DTW)</option>
<option value="DFW">Dallas/Fort Worth International Airport(DFW)</option>
<option value="RDU">Raleigh-Durham International Airport(RDU)</option>
<option value="TPA">Tampa International Airport(TPA)</option>
<option value="DEN">Denver International Airport(DEN)</option>
<option value="IAH">George Bush Intercontinental Airport(IAH)</option>
<option value="MSP">Minneapolis-Saint Paul International Airport(MSP)</option>
<option value="PBI">Palm Beach International Airport(PBI)</option>
<option value="BNA">Nashville International Airport(BNA)</option>
<option value="LAS">McCarran International Airport(LAS)</option>
<option value="SJU">Luis MuÃ±oz MarÃ-n International Airport(SJU)</option>
<option value="IAD">Washington Dulles International Airport(IAD)</option>
<option value="PHX">Phoenix Sky Harbor International Airport(PHX)</option>
<option value="BUF">Buffalo Niagara International Airport(BUF)</option>
<option value="CLE">Cleveland Hopkins International Airport(CLE)</option>
<option value="STL">St. Louis International Airport at Lambert Field(STL)</option>
<option value="MDW">Chicago Midway International Airport(MDW)</option>
<option value="SEA">Seattle-Tacoma International Airport(SEA)</option>
<option value="CVG">Cincinnati/Northern Kentucky International Airport(CVG)</option>
<option value="MSY">Louis Armstrong New Orleans International Airport(MSY)</option>
<option value="RSW">Southwest Florida International Airport(RSW)</option>

<option value="CMH">Port Columbus International Airport(CMH)</option>
<option value="CHS">Charleston International Airport/Charleston AFB(CHS)</option>
<option value="PIT">Pittsburgh International Airport(PIT)</option>
<option value="SAN">San Diego International AirportÂ (Lindbergh Field)(SAN)</option>
<option value="MKE">General Mitchell International Airport(MKE)</option>
<option value="JAX">Jacksonville International Airport(JAX)</option>
<option value="BTV">Burlington International Airport(BTV)</option>
<option value="SLC">Salt Lake City International Airport(SLC)</option>
<option value="AUS">Austin-Bergstrom International Airport(AUS)</option>
<option value="ROC">Greater Rochester International Airport(ROC)</option>
<option value="RIC">Richmond International Airport(RIC)</option>
<option value="PWM">Portland International Jetport(PWM)</option>
<option value="HOU">William P. Hobby Airport(HOU)</option>
<option value="IND">Indianapolis International Airport(IND)</option>
<option value="MCI">Kansas City International Airport(MCI)</option>
<option value="SYR">Syracuse Hancock International Airport(SYR)</option>
<option value="BWI">Baltimore-Washington International Airport(BWI)</option>
<option value="MEM">Memphis International Airport(MEM)</option>
<option value="PHL">Philadelphia International Airport(PHL)</option>
<option value="GSO">Piedmont Triad International Airport(GSO)</option>
<option value="ORF">Norfolk International Airport(ORF)</option>
<option value="DAY">James M. Cox Dayton International Airport(DAY)</option>
<option value="PDX">Portland International Airport(PDX)</option>

<option value="SRQ">Sarasota-Bradenton International Airport(SRQ)</option>
<option value="SDF">Louisville International AirportÂ (Standiford Field)(SDF)</option>
<option value="XNA">Northwest Arkansas Regional Airport(XNA)</option>
<option value="MHT">Manchester-Boston Regional Airport(MHT)</option>
<option value="BQN">Rafael HernÃ¡ndez Airport(BQN)</option>
<option value="CAK">Akron-Canton Regional Airport(CAK)</option>
<option value="OMA">Eppley Airfield(OMA)</option>
<option value="SNA">John Wayne AirportÂ (Orange County Airport)(SNA)</option>
<option value="GSP">Greenville-Spartanburg International Airport(GSP)</option>
<option value="SAV">Savannah/Hilton Head International Airport(SAV)</option>
<option value="GRR">Gerald R. Ford International Airport(GRR)</option>
<option value="HNL">Honolulu International Airport(HNL)</option>
<option value="LGB">Long Beach AirportÂ (Daugherty Field)(LGB)</option>
<option value="SAT">San Antonio International Airport(SAT)</option>
<option value="TYS">McGhee Tyson Airport(TYS)</option>
<option value="MSN">Dane County Regional Airport(MSN)</option>
<option value="DSM">Des Moines International Airport(DSM)</option>
<option value="STT">Cyril E. King Airport(STT)</option>
<option value="ALB">Albany International Airport(ALB)</option>
<option value="BDL">Bradley International Airport(BDL)</option>
<option value="BUR">Bob Hope AirportÂ (Hollywood Burbank Airport)(BUR)</option>
<option value="PVD">Theodore Francis Green State Airport(PVD)</option>
<option value="BGR">Bangor International Airport(BGR)</option>
<option value="PSE">Mercedita Airport(PSE)</option>
<option value="SJC">Norman Y. Mineta San JosÃ© International Airport(SJC)</option>

```
                                <option value="OKC">Will Rogers World Airport(OKC)</option>
                                <option value="AOK">Oakland International Airport(OAK)</option>
                                <option value="TUL">Tulsa International Airport(TUL)</option>
                                <option value="SMF">Sacramento International
Airport(SMF)</option>
                                <option value="BMH">Birmingham-Shuttlesworth International
Airport(BHM)</option>
                                <option value="ACK">Nantucket Memorial Airport(ACK)</option>
                                <option value="AVL">Asheville Regional Airport(AVL)</option>
                                <option value="ABQ">Albuquerque International
Sunport(ABQ)</option>
                                <option value="MVY">Martha's Vineyard Airport(MVY)</option>
                                <option value="EGE">Eagle County Regional Airport(EGE)</option>
                                <option value="CRW">Yeager Airport(CRW)</option>
                                <option value="ILM">Wilmington International
Airport(ILM)</option>
                                <option value="CAE">Columbia Metropolitan
Airport(CAE)</option>
                                <option value="TVC">Cherry Capital Airport(TVC)</option>
                                <option value="MYR">Myrtle Beach International
Airport(MYR)</option>
                                <option value="CHO">Charlottesville-Albemarle
Airport(CHO)</option>
                                <option value="BZN">Bozeman Yellowstone International AirportÂ
(Gallatin Field Airport)(BZN)</option>
                                <option value="JAC">Jackson Hole Airport(JAC)</option>
                                <option value="PSP">Palm Springs International
Airport(PSP)</option>
                                <option value="EYW">Key West International
Airport(EYW)</option>
                                <option value="HDN">Yampa Valley AirportÂ (Yampa Valley
Regional)(HDN)</option>
                                <option value="MTJ">Montrose Regional Airport(MTJ)</option>
                                <option value="SBN">South Bend International AirportÂ (South
Bend Regional)(SBN)</option>
                                <option value="ANC">Ted Stevens Anchorage International
Airport(ANC)</option>
                                <option value="LEX">Blue Grass Airport(LEX)</option>
                            </select>
                        </div>
                    </div>
```

```
                        <!Flying to>

                    </div>
                    <!--Flying to,from end-->

                    <div class="form-btn">
                       <button class="submit-btn">Predict</button>
                    </div>
                    <!Button>

                  </form>
                  <!--Form end-->

                </div>
                <!Booking form>
              </div>

            </div>
          </div>
        </div>
    </div>
</body>

</html>
```

FEATURE-2:

DATA PREPROCESSING:

```
ile=(r'C:\Users\HP\OneDrive\Documents\ML MODEL DEPLOYMENT\ML MODEL
DEPLOYMENT\FLIGHTDELAYDATA.CSV')
import sys
import pandas as pd
import numpy as np
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.metrics import accuracy_score
import sklearn.metrics as metrics
dataset=pd.read_csv(file)
dataset.info()
dataset.describe()
dataset.isnull().sum()
dataset['DEST'].unique()
sns.heatmap(dataset.corr())
dataset=dataset[["FL_NUM","MONTH","DAY_OF_MONTH","DAY_OF_WEEK","ORIGIN","DEST","CRS_
ARR_TIME","DEP_DEL15","ARR_DEL15"]]
dataset.isnull().sum()
dataset = dataset.fillna({'ARR_DEL15':1})
dataset =dataset.fillna({'DEP_DEL15':0})
dataset.iloc[177:185]
import math
for index,row in dataset.iterrows():
  dataset.loc[index,'CRS_ARR_TIME'] = math.floor(row['CRS_ARR_TIME'])
dataset.head()
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset['DEST']=le.fit_transform(dataset['DEST'])
dataset['ORIGIN']=le.fit_transform(dataset['ORIGIN'])
dataset.head(5)
dataset =pd.get_dummies(dataset,columns=['ORIGIN','DEST'])
dataset.head()
x=dataset.iloc[:,0:8].values
y=dataset.iloc[:,8:9].values
from sklearn.preprocessing import OneHotEncoder
oh=OneHotEncoder()
z=oh.fit_transform(x[:,4:5]).toarray()
t=oh.fit_transform(x[:,5:6]).toarray()
z
t
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
x_test.shape
x_train.shape
y_test.shape
```

```
y_train.shape
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(max_depth = 4, min_samples_split = 4, random_state = 0)
clf.fit(x_train, y_train)
pred = clf.predict(x_test)
decisiontree = clf.predict(x_test)
decisiontree
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, decisiontree))
pickle.dump(clf, open('flightclf.pkl','wb'))
```

8.TESTING:

TESTCASES:

TEST CASES

USER ACCEPTANCE:

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the Flight Delay Prediction project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	1	4	2	3	10
Duplicate	0	0	3	0	3
External	0	1	0	0	1
Fixed	1	3	2	6	12
Not Reproduced	0	0	0	0	0
Skipped	1	1	0	0	2
Won't Fix	1	0	0	0	1
Totals	4	9	7	9	29

3. Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Model Evaluation	10	0	0	10
Client Application	20	0	0	20
Exception Reporting	2	0	0	2
Final Report Output	4	0	0	4

RESULTS:

CLASSIFICATION RESULTS:

```
In [90]: report = classification_report(decisiontree, y_test)
         print(report)
```

```
              precision    recall  f1-score   support

           0       1.00      0.82      0.90      2186
           1       0.13      0.93      0.23        61

    accuracy                           0.83      2247
   macro avg       0.56      0.88      0.56      2247
weighted avg       0.97      0.83      0.88      2247
```

CONFUSION MATRIX:

```
In [91]: cm = confusion_matrix(y_test, decisiontree)
         print(cm)
```

```
[[1798    4]
 [ 388   57]]
```

ACCRUACY:

```
In [89]: from sklearn.metrics import accuracy_score,confusion_matrix,\
          classification_report
         print(accuracy_score(y_test, decisiontree))
```

```
0.8255451713395638
```

ADVANTAGES AND DIS ADVANTAGES:

Therefore, predicting flight delays can **improve airline operations and passenger satisfaction**, which will result in a positive impact on the economy. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays

Delay in flight is inevitable which has *too much negative economic effects on passengers, agencies and airport*.

CONCLUSION:

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. In this context, researchers created flight delay models for delay prediction over the last years, and this work contributes with an analysis of these models from a Data Science perspective. We developed a taxonomy scheme and classified models in respect of detailed components. 10 preprint Figure 8: Trends in machine learning for flight delay prediction Mainly, the taxonomy includes domain and Data Science branches. The former branch categorizes the problem (flight delay prediction) and the scope. The last branch groups methods and data handling. It was observed that the flight delay prediction is classified into two main categories, such as delay propagation and root delay and cancellation. Besides, the scope determines one of the three specific extents: airline, airport, en-route airspace or an ensemble of them. Additionally, considering Data Science branch, we aimed at the datum, by categorizing data sources, dimensions that can be used in the models, and data management techniques to preprocess data and improve prediction models efficiency. We also studied and divided the main methods into five categories: statistical analysis, probabilistic models, network representation, operations research, and machine learning. Those categories have been grouped as their use on specific forecast models for flight delays. Besides the

taxonomic scheme, we also presented a timeline with all articles to spot trends and relationships involving the main elements in the taxonomy. In the light of the domain-problem classification, this timeline showed a dominance of delay propagation and root delay over cancellation analysis. Researchers used to focus on statistical analysis and operational research approaches in the past. However, as the data volume grows, we noticed the use of machine learning and data management is increasing significantly. This clearly characterizes a Data Science trend. Researchers from airlines, airports, and academia will require a combination of skills of both domain specialists and data scientists to enable knowledge discovery from flight Big Data.

FUTURE SCOPE:

Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. An accurate estimation of flight delay is critical for airlines because **the results can be applied to increase customer satisfaction and incomes of airline agencies**

APPENDIX:

GITHUB LINK: https://github.com/IBM-EPBL/IBM-Project-40017-1660612967