## PROJECT REPORT

# Developing A Flight Delay Prediction Model

# **Using Machine Learning**

Submitted By

PNT2022TMID32340

ROSHINI BANU L -810019104057 PRIYADHARSHINII S - 810019104052 RAJ CLINTON G - 810019104053 SAKTHIVEL B -810019104060

### TABLE OF CONTENTS

1 INTRODUCTION	4
1.1 PROJECT OVERVIEW	4
1.2 PURPOSE	4
2 LITERATURE SURVEY	5
2.1 EXISTING PROBLEM	5
2.2 REFERENCES	5
2.3 PROBLEM STATEMENT DEFINITION	5
3 IDEATION AND PROPOSED SOLUTION	6
3.1 EMPATHY MAP CANVAS	6
3.2 IDEATION & amp; BRAINSTORMING	6
3.3 PROPOSED SOLUTION	10
3.4 PROBLEM SOLUTION FIT	11
4 REQUIREMENT ANALYSIS	13
4.1 FUNCTIONAL REQUIREMENTS	13
4.2 NON FUNCTIONAL REQUIREMENTS	14
5 PROJECT DESIGN	16
5.1 DATA FLOW DIAGRAM	16
5.2 SOLUTION & TECHNICAL ARCHITECTURE	17
6 PROJECT PLANNING AND SCHEDULING	17
6.1 SPRINT PLANNING AND ESTIMATION	17
6.2 SPRINT DELIVERY SCHEDULE	18
7 CODING & SOLUTIONING	21

8 TESTING	24
8.1 TEST CASES	24
8.2 USER ACCEPTANCE TESTING	24
8.2.1 DEFECT ANALYSIS	25
8.2.2 TEST CASE ANALYSIS	25
9 RESULTS	26
9.1 PERFORMANCE METRICS	26
10 ADVANTAGES & DISADVANTAGES	28
ADVANTAGES	28
DISADVANTAGES	28
11 CONCLUSION	32
12 FUTURE SCOPE	33
	36
APPENDIX	31
SOURCE CODE	34
DEMO VIDEO	

## CHAPTER-1 INTRODUCTION

### 1.1 PROJECT OVERVIEW

Over the last twenty years, air travel has been increasingly preferred among travelers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. An increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. These delays are responsible for large economic and environmental losses. The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays.

Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vectors like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use a decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when the difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifiers with logistic regression and a simple neural network for various figures of merit.

### 1.2 PURPOSE

Prior prediction of flight arrival delays is necessary for both travelers and airlines because delays in flights not only trigger huge economic loss but also airlines end uplosing their reputation that was built for several years and passengers lose their valuable time.

## CHAPTER - 2 LITERATURE SURVEY

### 2.1 EXISTING PROBLEM

Flight scheduling, is one of the challenging issue in the Modern world, which is faced with many uncertain conditions. One of such conditions is delay, which comes from various factors and inflicts considerable cost on airlines, operators, and travelers. So, we are implemented flight delay prediction through some proposed approaches based on machine learning algorithms.

### 2.2 REFERENCES

- [1] https://doi.org/10.24200/sci.2017.20020 (Khaksar, H., & Sheikholeslami, A. (2017)).
- [2] https://doi.org/10.1177/0361198120930014 (Esmaeilzadeh, E., & Mokhtarimousavi, S. (2020)).
- [3] https://doi.org/10.5121/ijdkp.2018.8301(M. Al-Tabbakh, S., M. Mohamed, H., & H. El, Z. (2018)).
- [4]https://doi.org/10.3390/su12072749(Ye, B., Liu, B., Tian, Y., & Wan, L. (2020)).
- [5]https://doi.org/10.1016/j.tre.2019.03.013(Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019)).
- [6]https://doi.org/10.5772/9385 (Oladipupo, T. (2010)).
- [7] https://doi.org/10.1186/s40537-020-00355-0(Nibareke, T.,& Laassiri, J. (2020)).

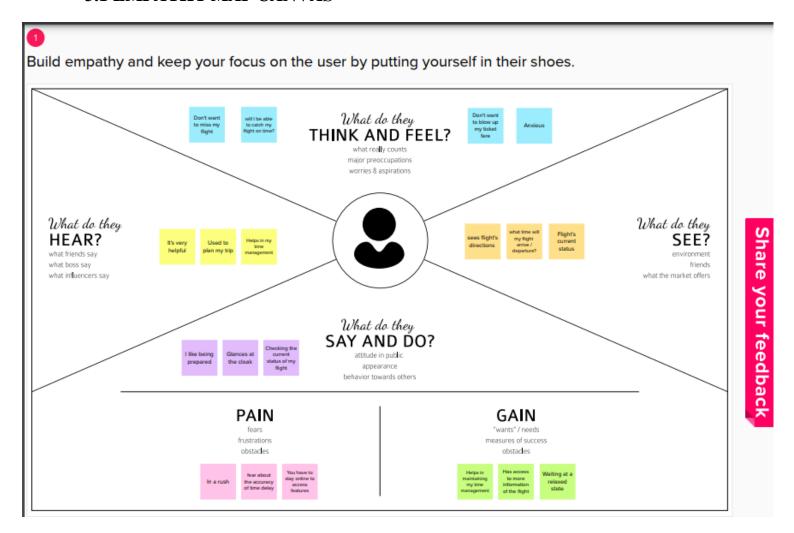
### 2.3 PROBLEM STATEMENT DEFINITION

Throughout the year 2015, there has been over 5,4 million domestic flights within the US. All of their metadata are recorded and saved in the Department of Transportation's (DOT) Bureau of Transportation Statistics. Flight delays cause significant financial and other losses to airlines, airports, and passengers. Their prediction is crucial during the decision-making process for all players of American aviation industry. Therefore, predicting the likelihood of delay based on flights' features bridges an important information asymmetry between airlines and passengers.

## **CHAPTER 3**

### IDEATION AND PROPOSED SOLUTION

### 3.1 EMPATHY MAP CANVAS



### 3.2 IDEATION & BRAINSTORMING

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcomeand built upon, and all participants are encouraged to collaborate, helpingeach other developa rich amount of creativesolutions.

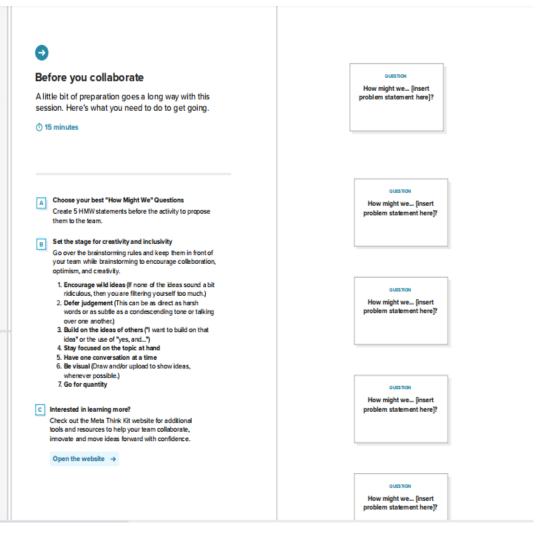
Step-1: Team Gathering, Collaboration and Select the Problem Statement

# Conducting a brainstorm

Executing a brainstorm isn't unique; holding a productive brainstorm is. Great brainstorms are ones that set the stage for fresh and generative thinking through simple guidelines and an open and collaborative environment. Use this when you're just kicking-off a new project and want to hit the ground running with big ideas that will move your team forward.

- ( 15 minutes to prepare
- 30-60 minutes to collaborate
- 3-8 people recommended





## Step-2: Brainstorm, Idea Listing and Grouping



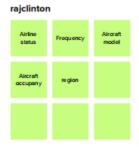
### **Brainstorm solo**

Have each participant begin in the "solo brainstorm space" by silently brainstorming ideas and placing them into the template. This "silent-storming" avoids group-think and creates an inclusive environment for introverts and extroverts alike. Set a time limit. Encourage people to go for quantity.



priyadharshinii						
Flight plan	Flight schedule	Airport schedule				
connective weather						





sakthivel		
operational coinditions	cancelaion	new delay
Time of day	city	

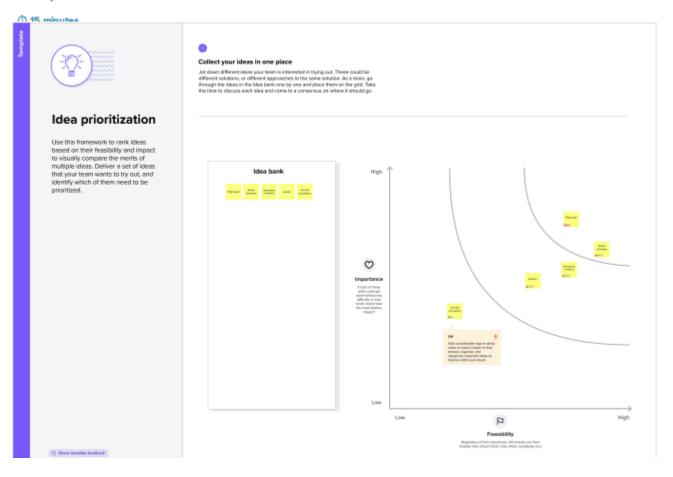
### **GROUPING IDEAS:**

#### Brainstorm as a group

Have everyone move their ideas into the "group sharing space" within the template and have the team silently read through them. As a team, sort and group them by thematic topics or similarities. Discuss and answer any questions that arise. Encourage "Yes, and..." and build on the ideas of other people along the way.

TIP

You can use the Voting session tool above to focus on the strongest ideas.



### 3.3 PROPOSED SOLUTION

S.No	Parameters	Description
1.	Problem Statement (Problem to be solved)	To find the Flight delay using its given information.
		To give the result in an efficient way through the Flight current status
3.	Novelty / Uniqueness	To give accuracy in the Flight Delay prediction.
4.	Social Impact/ Customer Satisfaction	It will be very useful and convenient to foreseen the Flight timings.
5.	Business Model(Revenue Model)	It will also have a high growth in the business related areas.
6.	Scalability of the Solution	Its scalability would be in wide range.Increase the accuracy in Flight delay prediction.

## 3.4 PROBLEM SOLUTION FIT

Project Title: Flight Delay prediction model	Project Design Phase-I - Solution Fit Template	TeamID: PNT2022TMID34340
CUSTOMER SEGMENT  Flight Passengers.  CS  CS	6. CUSTOMER CONSTRAINTS  Network connections, available devices.	5. AVAILABLE SOLUTIONS  From manual to digital updated accurate information and can foreseen the delay.  Explore AS, differentiate
2. JOBS-TO-BE-DONE/PROBLEMS  To find the Flight delay using its given information by the user.	9. PROBLEM ROOT CAUSE  Passengers can forecast the delay, if their Flight will be delay and can see the current status of flight too.	7. BEHAVIOUR  Directly-related:fill the Flight's information . Indirectly-related:waiting for the accurate flight current status and its delay.
3. TRIGGERS  reading about a more efficient solution.  4. EMOTIONS: BEFORE / AFTER  M	10. YOUR SOLUTION Giving more accuracy in the Flight Delay prediction.	8.CHANNELS of BEHAVIOUR  8.1 ONLINE Extract online info from #7  8.2 OFFLINE Extract offline channels from #7 and use them for customer development.

## **CHAPTER 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	HomePage	·Flight delay prediction · Information about Flight information required for prediction · If new User, <b>REGISTER</b> · If Already exist, <b>SIGN IN</b>
FR-2	User Registration	Enters Mail ID and other personal details required for Registering
FR-3	User Login	Uses Mail ID and Password for login
FR-4	Flight Info	Flight Info Flight number,name,destination and timing should be entered for prediction

## **4.2 NON FUNCTIONAL REQUIREMENTS**

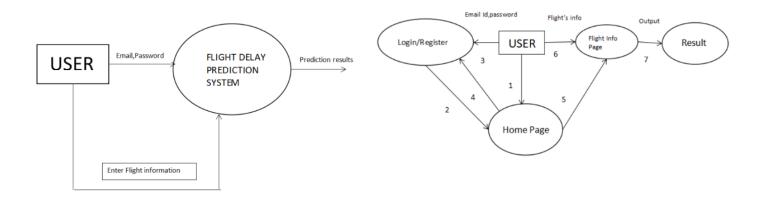
Following are the non-functional requirements of the proposed solution.

FR No.	Non- Functional Requirement	Description
NFR-1	Usability	It describes the context of When, Where and Why. The specific activities the requirements describe should reflect the both range of goals that the system must support
NFR-2	Security	Security functionality that ensures one of many different security properties of software is being satisfied. Security requirements are derived from industry standards, applicable laws, and a history of past vulnerabilities.
NFR-3	Reliability	It is the measure of the prediction based on the Flights given information and its current status.
NFR-4	Performance	It defines the accuracy of the delay of fligh

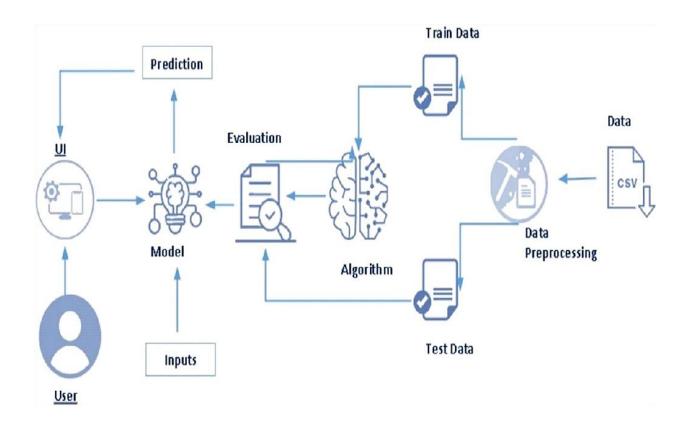
## **CHAPTER - 5**

## **PROJECT DESIGN**

## **5.1 DATA FLOW DIAGRAM**



## 5.2 SOLUTION & TECHNICAL ARCHITECTURE



# CHAPTER 6 PROJECT PLANNING AND SCHEDULING

## **6.1 SPRINT PLANNING AND ESTIMATION**

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint 1	Data Collection	USN-1	As a user, I can collect the dataset from various resources with different flight models	10	Low	Priyadharshinii S Roshini banu L
Sprint 1	Data Pre- processing	USN-2	As a user, I can load the data set, handling the missing data, scale and split data into train and test.	10	Medium	Priyadharshinii S Roshini banu L
Sprint 2	Model Building	USN-3	As a user, I will get an application with ML model which provides high accuracy of recognized handwritten digit.	5	High	Priyadharshinii S Roshini banu L Rajclinton Sakthivel
Sprint 2	Add CNN layers	USN-4	Creating the modeland adding the input,hidden, and outputlayers to it.	5	High	Priyadharshinii S Roshini banu L Rajclinton Sakthivel

Sprint 2	Compiling the model	USN-5	With both the training data and model it's time to configure the learning process.	2	Medium	Priyadharshinii S Roshini banu L Rajclinton Sakthivel

Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priorit y	Team Members
Train & test the model	USN-6	As a user, let us train our model with our image Data set.	6	Medium	Roshini banu L Rajclinton Sakthivel
Save the model	USN-7	As a user, the modelis saved integrated with an android application or web application in order to predict something.	2	Low	Priyadharshnii S
Building UI Application	USN-8	As a user I will upload the flight model image to the application by clicking a upload button.	5	High	Priyadharshinii S Sakthivel

	USN-9	As a user, I can know the details of the fundamental usage of the application.	5	Low	Rajclinton
Train the model on IBM	USN-10	As a user, I train the model on IBM and integrate flask with scoring end point.	10	High	Priyadharshinii S Roshini banu L Rajclinton Sakthivel

## **6.2 SPRINT DELIVERYSCHEDULE**

Sprint	Total Story Points	Due	Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Released Date (Actual)
Sprint1	20	3 Days	04 Nov 2022	07 Nov 2022	20	07 Nov 2022
Sprint2	20	3 Days	07 Nov 2022	10 Nov 2022	20	10 Nov 2022
Sprint3	20	3 Days	10 Nov 2022	13 Nov 2022	20	13 Nov 2022
Sprint4	20	3 Days	13 Nov 2022	16 Nov 2022	20	16 Nov 2022

### **CHAPTER-7**

### **CODING AND SOLUTION**

### FLIGHTDELY.HTML:

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta http-equiv="X-UA-Compatible">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="{{ url_for('static',filename='styles.css') }}">
    <script src="{{url_for('static', filename='styles/delaypredict.js')}}"></script>
    <title>Flight Delay Prediction</title>
</head>
<body id="flight-form">
<h2 id="main-head" class="centered-head">FLIGHT DELAY PREDICTION</h2>
<img src="{{url_for('static', filename='styles/images/Flight.png')}}" id="bgimg">
<form name="flightForm" action="/result" method="POST" target=" blank">
    <div id="form-content">
        <div id="block1">
            <div class="detail-container">
                <label for="fno" class="label-item">Enter the Flight Number</label>
                <input type="number" id="fno" name="fno" class="text-input">
            </div>
            <div class="detail-container">
                <label for="month" class="label-item">Month</label>
                <input type="number" id="month" name="month" class="text-input"</pre>
onblur="checkValid('month');" placeholder="Enter the Month Number">
                <div class="alert-text" id="month-valid">Enter a valid month between 1 to
12.</div>
            </div>
            <div class="detail-container">
                <label for="daym" class="label-item">Day of Month</label>
                <br>
                <input type="number" id="daym" name="daym" class="text-input"</pre>
onblur="checkValid('daym');">
                <div class="alert-text" id="daym-valid">Enter a valid day of month.</div>
            </div>
            <div class="detail-container">
                <label for="dayw" class="label-item">Day of Week</label>
                <br>
```

ZΤ

```
<input type="number" id="dayw" name="dayw" class="text-input"</pre>
onblur="checkValid('dayw');">
                <div class="alert-text" id="dayw-valid">Enter a valid day between 1 to 7.</div>
            <div class="detail-container">
                <label for="org" class="label-item">Origin</label>
                <select id="org" name="org" class="select-input">
                    <option value="ATL" class="option-item">ATL</option>
                    <option value="SEA" class="option-item">SEA</option>
                    <option value="DTW" class="option-item">DTW</option>
                    <option value="MSP" class="option-item">MSP</option>
                    <option value="JFK" class="option-item">JFK</option>
                </select>
            </div>
            <div class="detail-container">
                <label for="dest" class="label-item">Destination</label>
                <br>
                <select id="dest" name="dest" class="select-input" onblur="checkValid('dest');">
                    <option value="ATL" class="option-item">ATL</option>
                    <option value="SEA" class="option-item">SEA</option>
                    <option value="DTW" class="option-item">DTW</option>
                    <option value="MSP" class="option-item">MSP</option>
                    <option value="JFK" class="option-item">JFK</option>
                </select>
                <div class="alert-text" id="dest-valid">Enter different Origin and
Destination.</div>
            </div>
        </div>
        <div id="block2">
            <div class="detail-container">
                <label for="sdt" class="label-item">Scheduled Departure Time</label>
                <br>
                <input type="number" id="sdt" name="sdt" class="text-input"</pre>
onblur="checkValid('sdt'); placeholder="Enter in the format HHMM">
                <div class="alert-text" id="sdt-valid">Enter a valid time between 500 to
2359.</div>
            </div>
            <div class="detail-container">
                <label for="sat" class="label-item">Scheduled Arrival Time</label>
                <input type="number" id="sat" name="sat" class="text-input"</pre>
onblur="checkValid('sat');" placeholder="Enter in the format HHMM">
                <div class="alert-text" id="sat-valid">Enter a valid time between 500 to
2359.</div>
            </div>
            <div class="detail-container">
```

### **RESULT.HTML**:

```
<!doctype html>
<html>
<head>
    <title>Flight Delay Prediction - Result</title>
    <link rel="stylesheet" href="{{ url_for('static',filename='styles/result_styles.css') }}">
</head>
<body>
    <img src="{{url_for('static', filename='styles/images/Flight.png')}}" id="bgimg">
    {% if prediction[0]== 0.0 %}
    <div class="pred_result" id="result_0">There is a good chance that your flight will be on
time</div>
    {% endif %}
    {% if prediction[0] == 1.0 %}
    <div class="pred_result" id="result_1">There is a possibility that your flight will be
delayed</div>
    {% endif %}
</body>
</html
```

# CHAPTER 8 TESTING

## **8.1 TEST CASES**

Test	Feature	Component		Expected	Actual	Status
case ID			Scenario	Result	Result	
HP_TC_0 01	UI	Home Page	Elements in the home page	The Homepage must be displayed properly	Working as expected	PASS
BE_TC_0 01	Functional	Backend	Check if all the routes are working properly	All theroutes should work properly	Working as expected	PASS
RP_TC_001	UI	Result	Verify UI elements in the Result Page	The Result page must be display properly	Working as expected	PASS

## 8.2 USER ACCEPTANCE TESTING

## **8.2.1 DEFECT ANALYSIS**

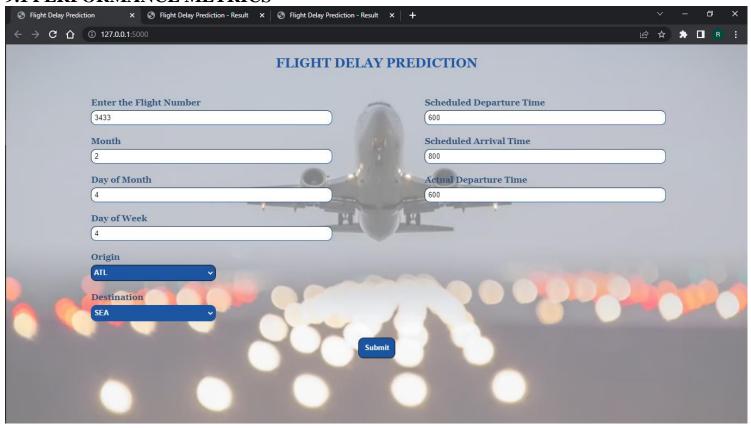
Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	1	4	2	3 10	
Duplicate	0	0	3	0	3
External	0	1	0	0	1
Fixed	1	3	2	6	12
Not Reproduced	0	0	0	0	0
Skipped	1	1	0	0	2
Won't Fix	1	0	0	0	1
Totals	4	9	7	9	29

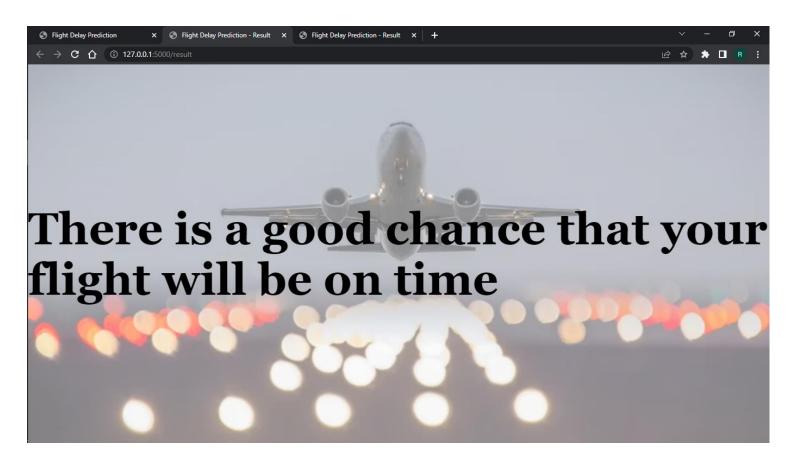
### 8.2.2 TEST CASE ANALYSIS

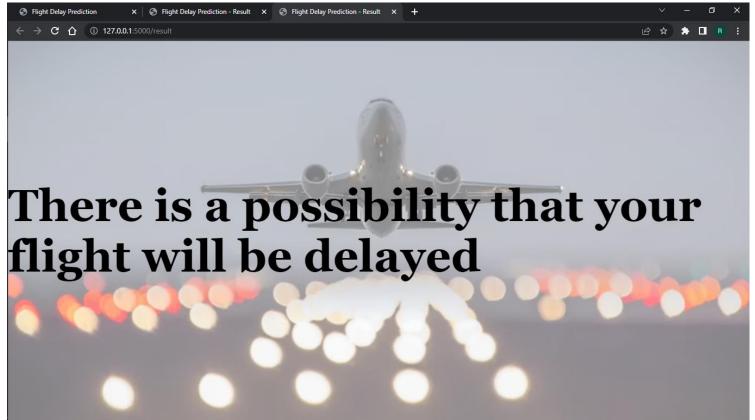
Section	<b>Total Cases</b>	Not Tested	Fail	Pass
Model Evaluation	10	0	0	10
Client Application	20	0	0	20
Exception Reporting	2	0	0	2
Final Report Output	4	0	0	4

## CHAPTER - 9 RESULTS

### 9.1 PERFORMANCE METRICS







## CHAPTER - 10 ADVANTAGES AND DISADVANTAGES

### **ADVANTAGES**

- Reduces manual work
- More accurate than average human
- Capable of handling a lot of data
- Can be used anywhere from any device

### **DISADVANTAGES**

- Cannot handle complex data
- All the data must be in digital format
- Requires a high performance server for faster predictions
- Prone to occasional errors

### **CHAPTER 11**

## **CONCLUSION**

This project demonstrated a web application that uses machine learning to recognise handwritten numbers. Flask, HTML, CSS, JavaScript, and a few other technologies were used to create this project. The model predicts the handwritten digit using a CNN network. During testing, the model achieved a 99.61% recognition rate. The proposed project is scalable and can easily handle a huge number of users. Since it is a web application, it is compatible with any device that can run a browser. This project is extremely useful in real-world scenarios such as recognizing number plates of vehicles, processing bank cheque amounts, numeric entries in forms filled up by hand (tax forms) and so on. There is so much room for improvement, which can be implemented in subsequent versions.

### **CHAPTER - 12**

### **FUTURE SCOPE**

This project is far from complete and there is a lot of room for improvement. Some of the improvements that can be made to this project are as follows:

- Add support to detect from digits multiple images and save the results
- Add support to detect multiple digits
- Improve model to detect digits from complex images
- Add support to different languages to help users from all over the world

This project has endless potential and can always be enhanced to become better. Implementing this concept in the real world will benefit several industries and reduce the workload on many workers, enhancing overall work efficiency.

### **APPENDIX**

#### **SOURCE CODE**

#### MODEL CREATION

```
from flask import Flask, render_template, request
import pandas as pd
import joblib
import numpy as np
app = Flask(__name__)
```

```
@app.route('/')
def home():
    return render_template('Flightdelay.html')
```

```
@app.route('/result', methods = ['POST'])
def predict():
    fl_num = int(request.form.get('fno'))
    month = int(request.form.get('month'))
    dayofmonth = int(request.form.get('daym'))
    dayofweek = int(request.form.get('dayw'))
    sdeptime = request.form.get('sdt')
    adeptime = request.form.get('adt')
    arrtime = int(request.form.get('sat'))
    depdelay = int(adeptime) - int(sdeptime)
    inputs = list()
    inputs.append(fl_num)
    inputs.append(month)
    inputs.append(dayofmonth)
    inputs.append(dayofweek)
    if (depdelay < 15):</pre>
        inputs.append(0)
    else:
        inputs.append(1)
    inputs.append(arrtime)
    origin = str(request.form.get("org"))
    dest = str(request.form.get("dest"))
    if(origin=="ATL"):
        a=[1,0,0,0,0]
        inputs.extend(a)
    elif(origin=="DTW"):
        a=[0,1,0,0,0]
```

```
inputs.extend(a)
    elif(origin=="JFK"):
        a=[0,0,1,0,0]
        inputs.extend(a)
    elif(origin=="MSP"):
        a=[0,0,0,1,0]
        inputs.extend(a)
    elif(origin=="SEA"):
        a=[0,0,0,0,1]
        inputs.extend(a)
    if(dest=="ATL"):
        b=[1,0,0,0,0]
        inputs.extend(b)
    elif(dest=="DTW"):
        b=[0,1,0,0,0]
        inputs.extend(b)
    elif(dest=="JFK"):
        b=[0,0,1,0,0]
        inputs.extend(b)
    elif(dest=="MSP"):
        b=[0,0,0,1,0]
        inputs.extend(b)
    elif(dest=="SEA"):
        b=[0,0,0,0,1]
        inputs.extend(b)
    prediction = preprocessAndPredict(inputs)
    #Pass prediction to prediction template
    print(inputs)
    return render template('/result.html', prediction = prediction)
def preprocessAndPredict(inputs):
    test data = np.array(inputs).reshape((1,16))
    model file = open('model.pkl', 'rb')
    trained model = joblib.load(model file)
    df = pd.DataFrame(data=test_data[0:, 0:], columns=['FL_NUM', 'MONTH', 'DAY_OF_MONTH',
'DAY_OF_WEEK', 'DEP_DEL15', 'CRS_ARR_TIME', 'ORIGIN_ATL', 'ORIGIN_DTW', 'ORIGIN_JFK',
'ORIGIN MSP', 'ORIGIN SEA', 'DEST ATL', 'DEST DTW', 'DEST JFK', 'DEST MSP', 'DEST SEA'])
    data = df.values
    result = trained_model.predict(data)
```

```
print(result)
return result
```

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

### **DEMO VIDEO LINK:**

 $https://drive.google.com/drive/folders/1D719\_M82 juSzpD7 hpCJhrtCWKD4OUdPO? usp=share\_link$