# PROJECT REPORT

# Developing A Flight Delay Prediction Model

# Using Machine Learning


Submitted By

**PNT2022TMID04987**

KEERTHANAAANAND - 921319104097
INDHUJA B - 921319104064
INDHUMATHI S - 921319104065
JEYANTHI C - 921319104076

**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 & 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** **18**

6.1 SPRINT PLANNING AND ESTIMATION 18

6.2 SPRINT DELIVERY SCHEDULE 21

**7 CODING & SOLUTIONING** **22**

**8 TESTING** **28**

8.1 TEST CASES 28
8.2 USER ACCEPTANCE TESTING 29
8.2.1 DEFECT ANALYSIS 29
8.2.2 TEST CASE ANALYSIS 29

**9 RESULTS** **30**

9.1 PERFORMANCE METRICS 30

**10 ADVANTAGES & DISADVANTAGES** **31**

ADVANTAGES 31
DISADVANTAGES 31

**11 CONCLUSION** **32**

**12 FUTURE SCOPE** **33**

**APPENDIX** **36**

SOURCE CODE 34

PROJECT DEMO

# 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 up losing 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)).

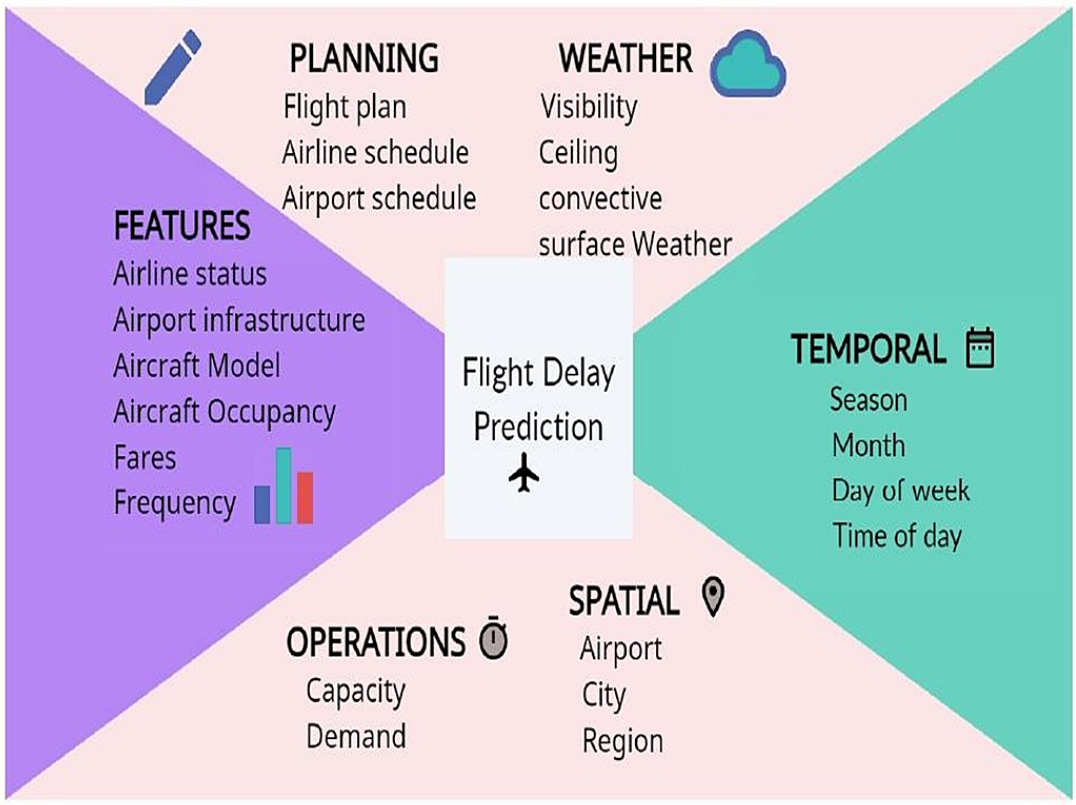
## 2.3 PROBLEM STATEMENT DEFINITION

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

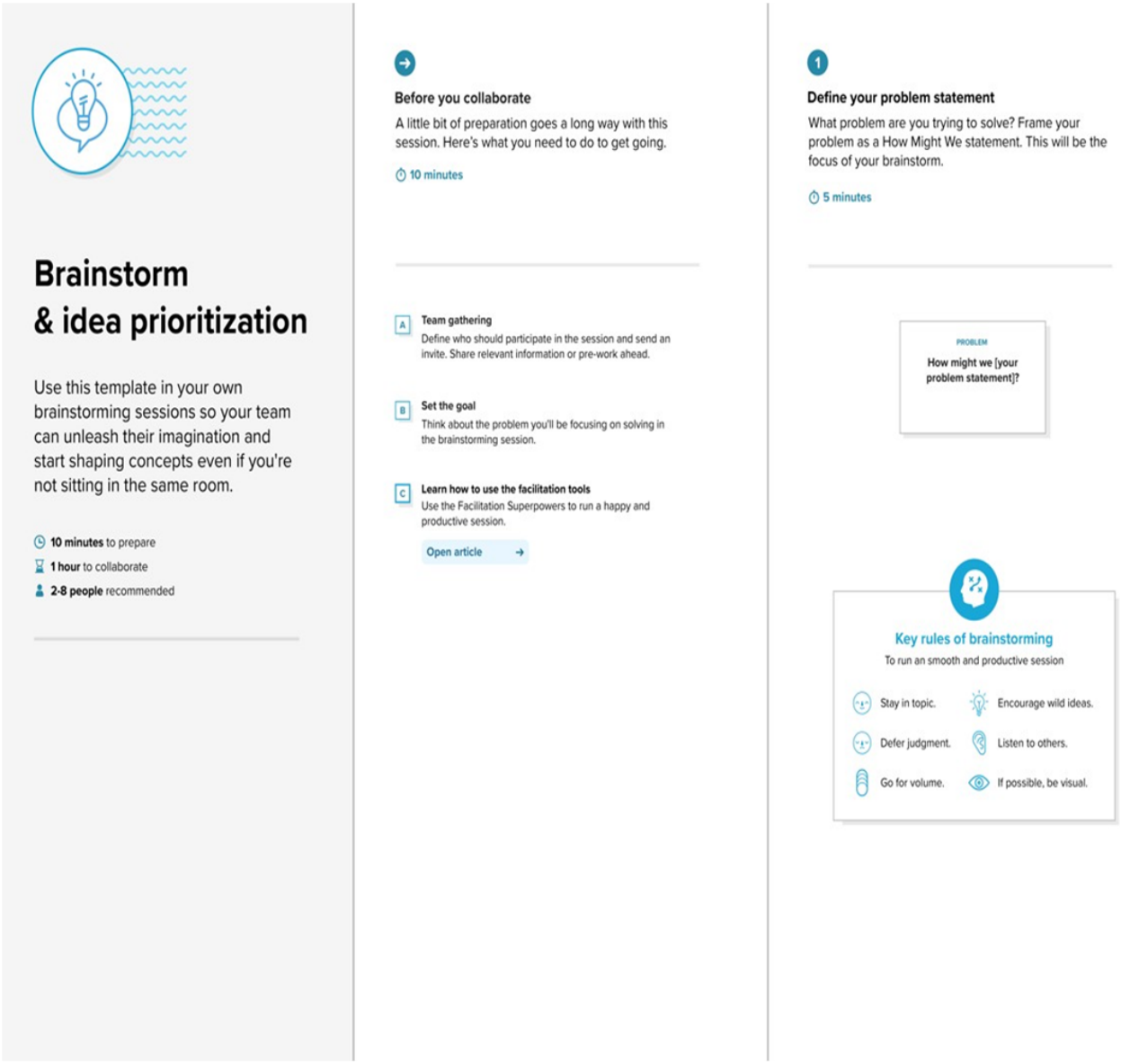
## 3.1 EMPATHY MAP CANVAS




## 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



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

2

### Brainstorm

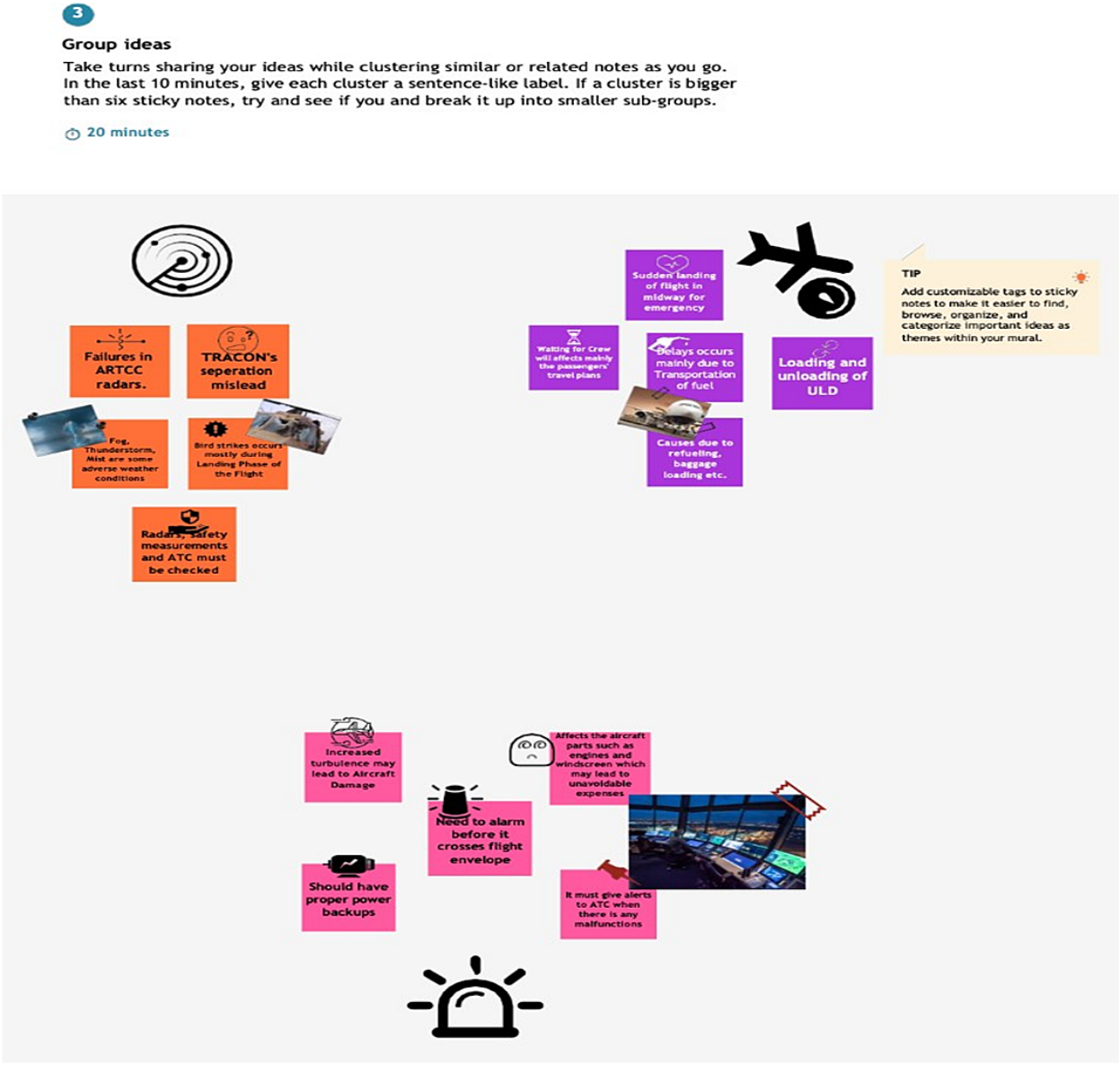
Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

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

# GROUPING IDEAS:



## 3.3 PROPOSED SOLUTION

| S.No | Parameters | Description |
|---|---|---|
| 1. | Problem Statement (Problem to be solved) | Predict theflight delay forthe efficient transportation |
| 2. | Idea / Solution description | To develop an ML model using deep learning to predict the delayed flight time and optimize flight operation |
| 3. | Novelty / Uniqueness | We are predicting the time varying parameters to predict the delayed flight with maximum accuracy using deep learning |
| 4. | Social Impact/ Customer Satisfaction | It makesthe air transportation more efficient and saves moretime for the passengers |
| 5. | Business Model(Revenue Model) | Using this model, we can create a revenue by giving appropriate solution about thedelayto the people |
| 6. | Scalability of the Solution | This makes the people to take the action according to the delay and it improves time management, businessvalue and more |

## 3.4 PROBLEM SOLUTION FIT

| 1.CUTOMER SEGMENT(S)<br>**Who is yourcustomer?** | 6. CUSTOMER CONSTRAINT.<br>**What constraint prevents your customerfrom taking action or limiting their choice of solution?** | 5. AVAILABLE SOLUTION<br>**Which solutions are available to the customer when they face the problem.** |
|---|---|---|
| Customers will be able to know the flight delay in priorso that they can planaccordingly. | Time is precious in a everyhuman's life. | We gather resources from the international and domestic airlines. |

| 1. TRIGGERS<br>**What triggers customers to act.** | 1. YOUR SOLUTION | |
|---|---|---|
| a. Customers get to know the absolute recommendation to their need | Our solution involves autonomous systemwhich doesthe following:<br>a. A personal Help desk which can be accessed through allthe devices whichare compatible with browser.<br>b. Customers can post theirqueries in the new threadsection. | ○ For a new query they need an online connectivity to post and receive recommendation from our team.<br>○ They can also use our chatbot 24/7W |

d.

b. Fast Response.

1. **EMOTIONS: BEFORE/AFTER**

**How do customers feel when they facea problem or a job and afterwards.**

a. Enables Customers to Trust to their agent about posting theirpersonal informations.

b. Feeling comfortab le with thesolution and the company's service.

c. They can also access the FAQ's Section to see if the problem is already listed

d. They can also view their results progress through theirmails.

e. They will get support from the team until the problem getsresolved.

hile they are in online.

**OFFLINE**

- They can Read themessa ges once it is received through the cloud app.
- They can access FAQ's whileth ey are offline.

# 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 | User Registration | Registration through Form<br>Registration throughGmail<br>Registration through LinkedIN |
| FR-2 | User Confirmation | Confirmation via Email Confirmationvia OTP |
| FR-3 | User requirements | Collecting informations like date of travel, departing & arrival destination, flight number or bookingnumber, etc for providing the status of the flight. |
| FR-4 | User friendliness | This system is easy to learn and understand. |

## 4.2 NON FUNCTIONAL REQUIREMENTS

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

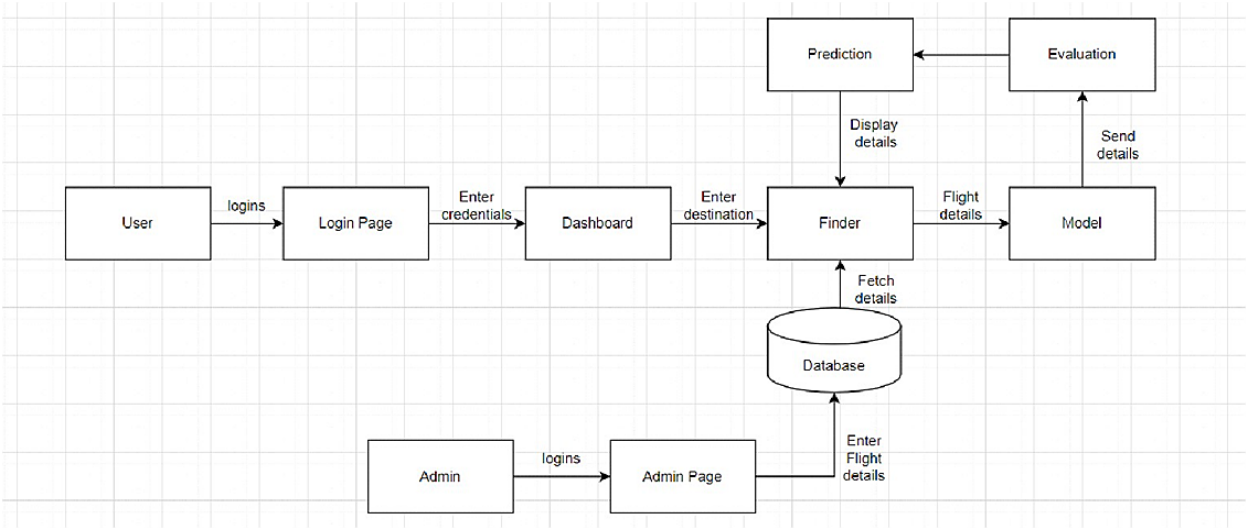
| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | **Usability** | How easyis it for a customer to use thesystem? |
| NFR-2 | **Security** | **Security**'s part will be protected againstmalware attacks or unauthorized access. But there's a catch. The lion's share of security non-functional requirements can be translatedinto concrete functional counterparts. If you want to protect the admin panel from unauthorized access, you would define the login flow and different user roles as system behavior or user actions. |
| NFR-3 | **Reliability** | **Reliability s**pecifies how likely the system orits element would run without a failure for a given period of time underpredefined conditions.<br>Traditionally, this probability is expressed in percentages. For instance, if the system has 85 percent reliability for a month, this means that during this month, under normal usage conditions, there's an 85 percent chance that thesystem won't experience critical failure. |

| | | |
|---|---|---|
| NFR-4 | **Performance** | **Performance** defines how fast a software system or a particular piece of it responds to certain users'actions under a certain workload.In most cases, this metric explains how long a user must wait before the target operation happens (the page renders, a transaction is processed, etc.) given the overall numberof users at the moment.But it's not always like that. Performance requirements may describe background processes invisible to users, e.g. backup. But let's focus on user-centricperformance. |

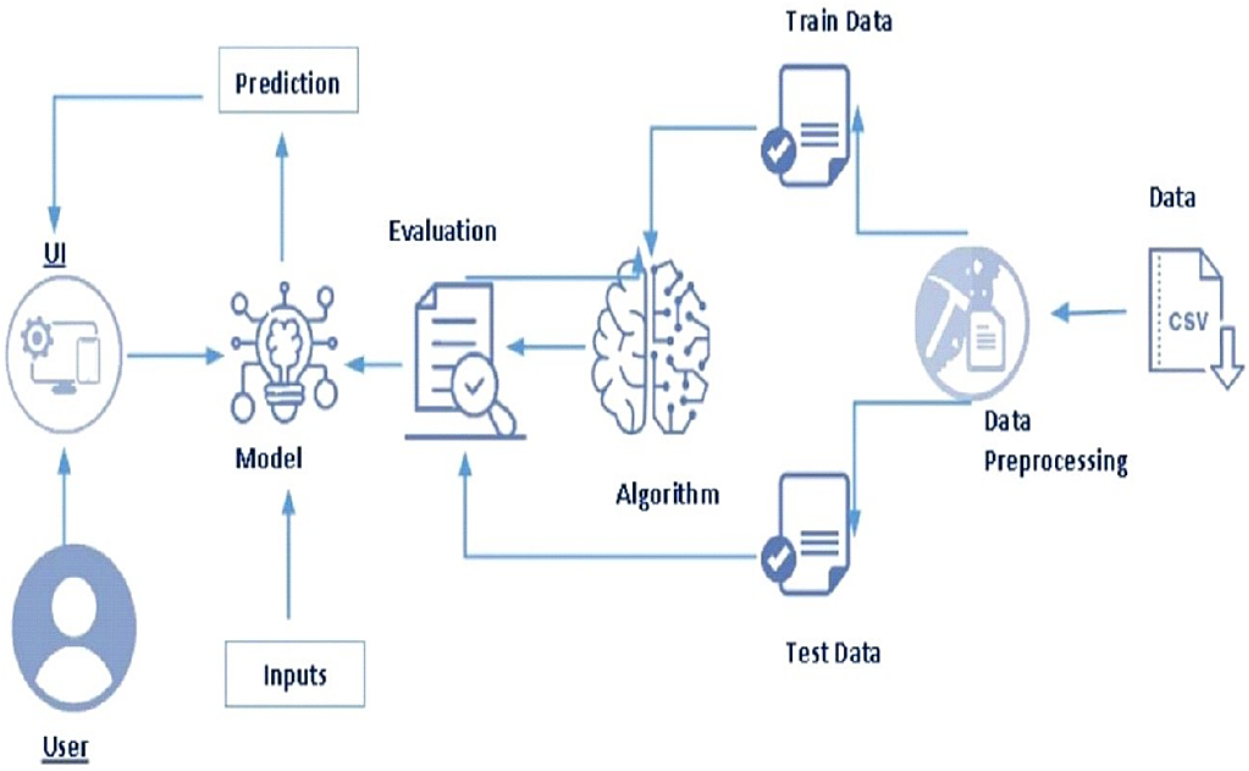
# CHAPTER 5

# 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 | Keerthaaanand Indhuja B |
| Sprint 1 | Data Preprocessing | USN-2 | As a user, I can load the dataset, handling the missingdata,scaling and split data into train and test. | 10 | Medium | Indhumathi S Jeyanthi C |
| Sprint 2 | Model Building | USN-3 | As a user, I will get an application with MLmodel which provides high accuracy of recognized handwritten digit. | 5 | High | Keerthaaanand Indhuja B Indhumathi S Jeyanthi C |
| Sprint 2 | Add CNN layers | USN-4 | Creating the modeland adding the input,hidden, and outputlayers to it. | 5 | High | Keerthaaanand Indhuja B Indhumathi S Jeyanthi C |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint 2 | Compiling the model | USN-5 | With both the trainingdata defined andmodel defined, it'stime to configure thelearning process. | 2 | Medium M | Keerthaaanand Indhuja B Indhumathi S Jeyanthi C |

| **Functional Requirement (Epic)** | **User Story Number** | **User Story / Task** | **Story Points** | **Priority** | **Team Members** |
|---|---|---|---|---|---|
| Train & test the model | USN-6 | As a user, let us train our model with our image dataset. | 6 | Medium | Keerthaaanand Indhuja B Indhumathi S |
| Save the model | USN-7 | As a user, the model is saved integrated with an android application or web application in order to predict something. | 2 | Low | Jeyanthi C |

| | | | | | |
|---|---|---|---|---|---|
| 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 | Indhuja B<br>Indhumathi S<br>Jeyanthi C |
| | USN-9 | As a user, I can know the details of the fundamental usage of the application. | 5 | Low | Keerthaaanand |
| | | er, I can see cted / zeddigits in ication. | 5 | Medium | Indhuja B<br>Indhumathi S |
| Train the model on IBM | USN-11 | As a user, I train the model on IBM and integrate flask with scoring end point. | 10 | High | Keerthaaanand<br>Indhuja B<br>Indhumathi S<br>Jeyanthi C |

## 6.2 SPRINT DELIVERYSCHEDULE

| Sprint | Total Story Points | Du | Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Released Date (Actual) |
|---|---|---|---|---|---|---|
| Sprint1 | 20 | 6 Days | 2022 | 29 Oct 2022 | 20 | 29 Oct 2022 |
| Sprint2 | 20 | 6 Days | 2022 | 05 Nov 2022 | 20 | 05 Nov 2022 |
| Sprint3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 20 | 12 Nov 2022 |
| Sprint4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 20 | 19 Nov 2022 |

# CHAPTER 7

# CODING & SOLUTIONING

```
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1">

<style>
h2 {text-align: center;}
* {
box-sizing: border-box;
}
body{
font-family: Arial, Helvetica, sans-serif;
background-image: url('https://www.teahub.io/photos/full/17-171442_airplane-desktop-wallpaper-data-src-airplane-wallpapers-transparent.jpg');
background-size: cover;
background-attachment: fixed;
}
input[type=text], select, textarea {
width: 100%;
padding: 12px;
border: 1px solid #ccc;
border-radius: 4px;
resize: vertical;
}
label {
padding: 12px 12px 12px 0;
```

```
display: inline-block;
}
input[type=submit] {
background-color: #04AA6D;
color:white;
padding: 12px 20px;
border: none;
border-radius: 4px;
cursor: pointer;
float: right;
}
input[type=submit]:hover {
background-color: #45a049;
}
.container {
border-radius: 5px;
background-color:transparent;
padding: 20px;
}
.col-25 {
float: left;
width: 25%;
margin-top: 6px;
}
.col-75 {
float: left;
width: 75%;
margin-top: 6px;
}
/* Clear floats after the columns */
.row:after {
content: "";
```

```
display: table;
clear: both;
}
</style>
</head>
<body>
<h2 style="color: #ebf7f3">Prediction of flight delay</h2>
<div class="container">
<form action="/action_page.php">
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Enter flight Number</label>
</div>
<div class="col-75">
<input type="text" id="fname" numbers="flight number" placeholder="flight
number..">
</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Month</label>
</div>
<div class="col-75">
<input type="text" id="" name="month" placeholder="month..">
</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Date of month</label>
</div>
<div class="col-75">
```

```
<input type="text" id="" name="Month" placeholder="Date of month..">

</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Day of week</label>
</div>
<div class="col-75">
<input type="calender" id="fname" numbers="Day of week"
placeholder="Day of week..">
</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Origin</label>
</div>
<div class="col-75">
</div>
<div class="col-25">
<select id="country" name="origin">
<option value="region">click here</option>
<option value="region">COK</option>
<option value="region">MAA</option>
<option value="region">JFK</option>

<option value="region">DXB</option>
</select>
</div>
</div>
<div class="row">
<div class="col-25">
```

```
<label style="color: #ebf7f3">Destination</label>
</div>
<div class="col-75">
</div>
<div class="col-25">
<select id="region" name="origin">
<option value="region">click here</option>
<option value="region">COK</option>
<option value="region">MAA</option>
<option value="region">JFK</option>

<option value="region">DXB</option>
</select>
</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Scheduled Departure Time</label>
</div>
<div class="col-75">
<input type="time" id="fname" numbers="predict" placeholder="scheduled

Departure Time..">
</div>

<div class="row">
<div class="col-25">
<label style="color: #ebf7f3"> Scheduled Arrival Time</label>
</div>
<div class="col-75">
<input type="time" id="fname" numbers="predict" placeholder="Arrival

Departure Time..">
```

```
</div>
</div>
<div class="row">
<div class="col-25">
<label style="color: #ebf7f3">Actual Departure Time</label>
</div>
<div class="col-75">
<input type="time" id="fname" numbers="predict" placeholder="..">

</div>
</div>
<div class="row">
<a href ="exe.html"><h4>SUBMIT</h4></a>
</div>
</form>
</div>
</body>
</html>
```

# CHAPTER 8
# TESTING

## 8.1 TEST CASES

| Test case ID | Feature | Component | Scenario | Expected Result | Actual Result | Status |
|---|---|---|---|---|---|---|
| HP_TC_001 | UI | Home Pag | Elements in the Home Page | The Home page must be displayed properly | Working as expected | PASS |
| BE_TC_001 | Functional | Backend | Check if all the routes are working properly | All the routes should properly work | Working as expected | PASS |
| RP_TC_001 | UI | Result Page | 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 | Total |
|---|---|---|---|---|---|
| By Design | 1 | 0 | 1 | 0 | 2 |
| Duplicate | 0 | 0 | 0 | 0 | 0 |
| External | 0 | 0 | 2 | 0 | 2 |
| Fixed | 4 | 1 | 0 | 1 | 6 |
| Not Reproduced | 0 | 0 | 0 | 1 | 1 |
| Skipped | 0 | 0 | 0 | 1 | 1 |
| Total | 5 | 1 | 3 | 3 | 12 |

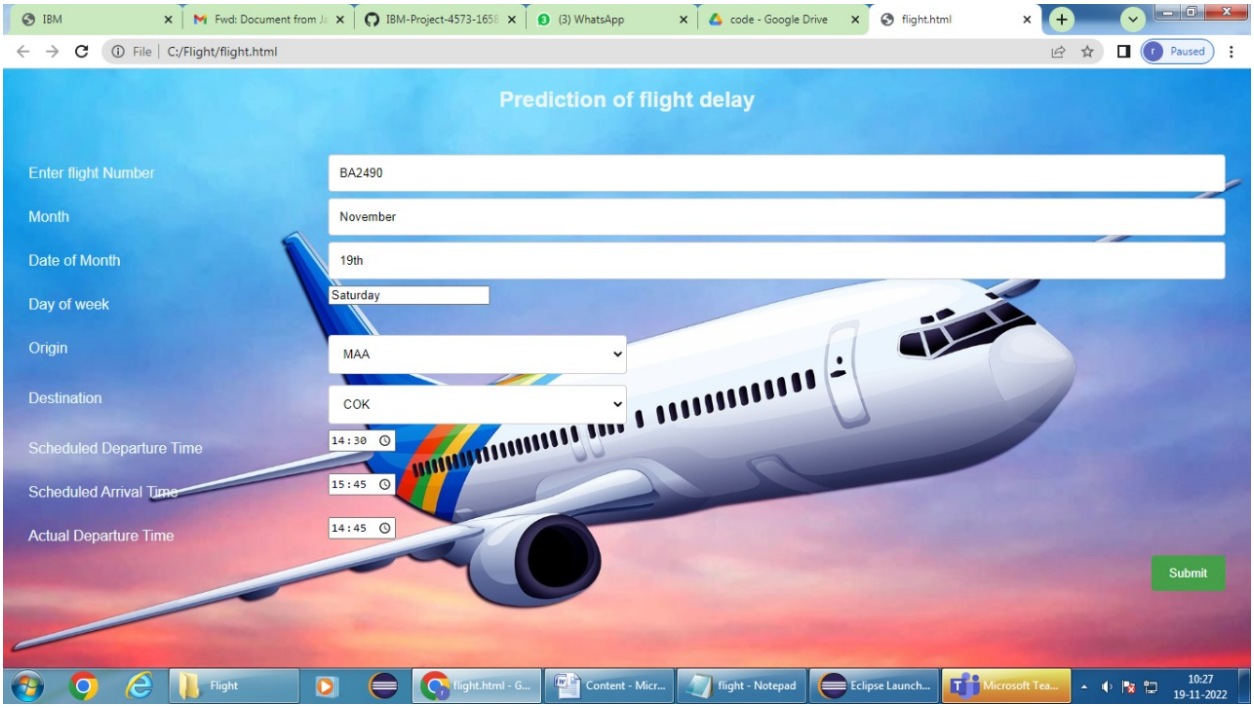
### 8.2.2 TEST CASE ANALYSIS

| Section | Total Cases | Not Tested | Fail | Pass |
|---|---|---|---|---|
| Client Application | 10 | 0 | 3 | 7 |
| Security | 2 | 0 | 1 | 1 |
| Performance | 3 | 0 | 1 | 2 |
| Exception Reporting | 2 | 0 | 0 | 2 |

# CHAPTER 9

# RESULTS

## 9.1 PERFORMANCE METRICS



# CHAPTER 10

# ADVANTAGES & 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

```
import numpy as np
import tensorflow #open source used for both ML and DL for computation
from tensorflow.keras.datasets import mnist #mnist dataset
from tensorflow.keras.models import Sequential #it is a plain stack of layers
from tensorflow.keras import layers #A Layer consists of a tensor- in tensor-out computat ion funct ion
from tensorflow.keras.layers import Dense, Flatten #Dense-Dense Layer is the regular deeply connected r
#faltten -used fot flattening the input or change the dimension
from tensorflow.keras.layers import Conv2D #convolutional layer
from keras.utils import np_utils #used for one-hot encoding
import matplotlib.pyplot as plt  #used for data visualization
```

```
(x_train, y_train), (x_test, y_test)=mnist.load_data () #splitting the mnist data into train and test
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz
11490434/11490434 [==============================] - 0s 0us/step
```

```
print (x_train.shape)  #shape is used for give the dimens ion values #60000-rows 28x28-pixels
print (x_test.shape)
```

```
(60000, 28, 28)
(10000, 28, 28)
```

```
#Reshaping to format which CNN expects (batch, height, width, channels)
x_train=x_train.reshape (60000, 28, 28, 1).astype('float32')
x_test=x_test.reshape (10000, 28, 28, 1).astype ('float32')
```

# Project Demo Video

https://drive.google.com/file/d/1e3Uix6v6VNxaIr5CpLPDKz-ewPfZDW7e/view?usp=share_link