A Project Report

Flight Delay Prediction for aviation Industry Using Machine Learning

Team ID : NM2023TMID17411

Team Leader : ARCHANA R

Team members : ABISHA A

: ABISHA KUMARI T

: AJITHA N

Flight Delay Prediction for aviation Industry Using Machine Learning

Introduction

1.1 Overview

Flight delay prediction is an important area of research in the aviation industry. Predicting flight delays accurately can help airlines and airports to improve their operations, reduce costs, and enhance customer satisfaction. Delay prediction models use various data sources, such as historical flight data, weather data, and airport congestion data, to estimate the probability of a flight being delayed.

There are two main types of delay prediction models: rule-based models and machine learning models. Rule-based models use a set of predefined rules to predict delays based on factors such as the time of day, the airline, and the airport. Machine learning models, on the other hand, use historical data to train predictive algorithms that can estimate the likelihood of delays based on a wider range of factors.

One of the challenges of building delay prediction models is the large number of variables that can affect flight delays, including weather conditions, air traffic congestion, aircraft maintenance issues, and crew scheduling problems. Another challenge is the need for accurate and timely data, which can be difficult to obtain in real-time.

Despite these challenges, several companies and research institutions have developed and deployed delay prediction models in the aviation industry. Some of the applications of these models include predicting delays for individual flights, optimizing flight schedules, and managing airport congestion.

In conclusion, flight delay prediction is a crucial area of research in the aviation industry, and it has the potential to improve operational efficiency and customer satisfaction. As technology advances and more data becomes available, we can expect to see further developments in this field in the coming years.

1.2 Purpose

The purpose of flight delay prediction in the aviation industry is to improve operational efficiency, reduce costs, and enhance customer satisfaction. Here are some specific purposes of flight delay prediction:

Optimize airline operations: By predicting flight delays in advance, airlines can make changes to their operations, such as re-routing flights, adjusting crew schedules, or providing advance notice to passengers. This can help to minimize the impact of delays on passengers and reduce costs associated with delays, such as compensation for passengers or crew.

Improve passenger experience: Delay prediction models can help airlines and airports to provide more accurate and timely information to passengers about flight delays. This can help to reduce passenger frustration and improve overall customer satisfaction.

Manage airport congestion: Delay prediction models can be used to anticipate periods of high congestion at airports and take steps to manage traffic flow. This can help to minimize delays and improve the efficiency of airport operations.

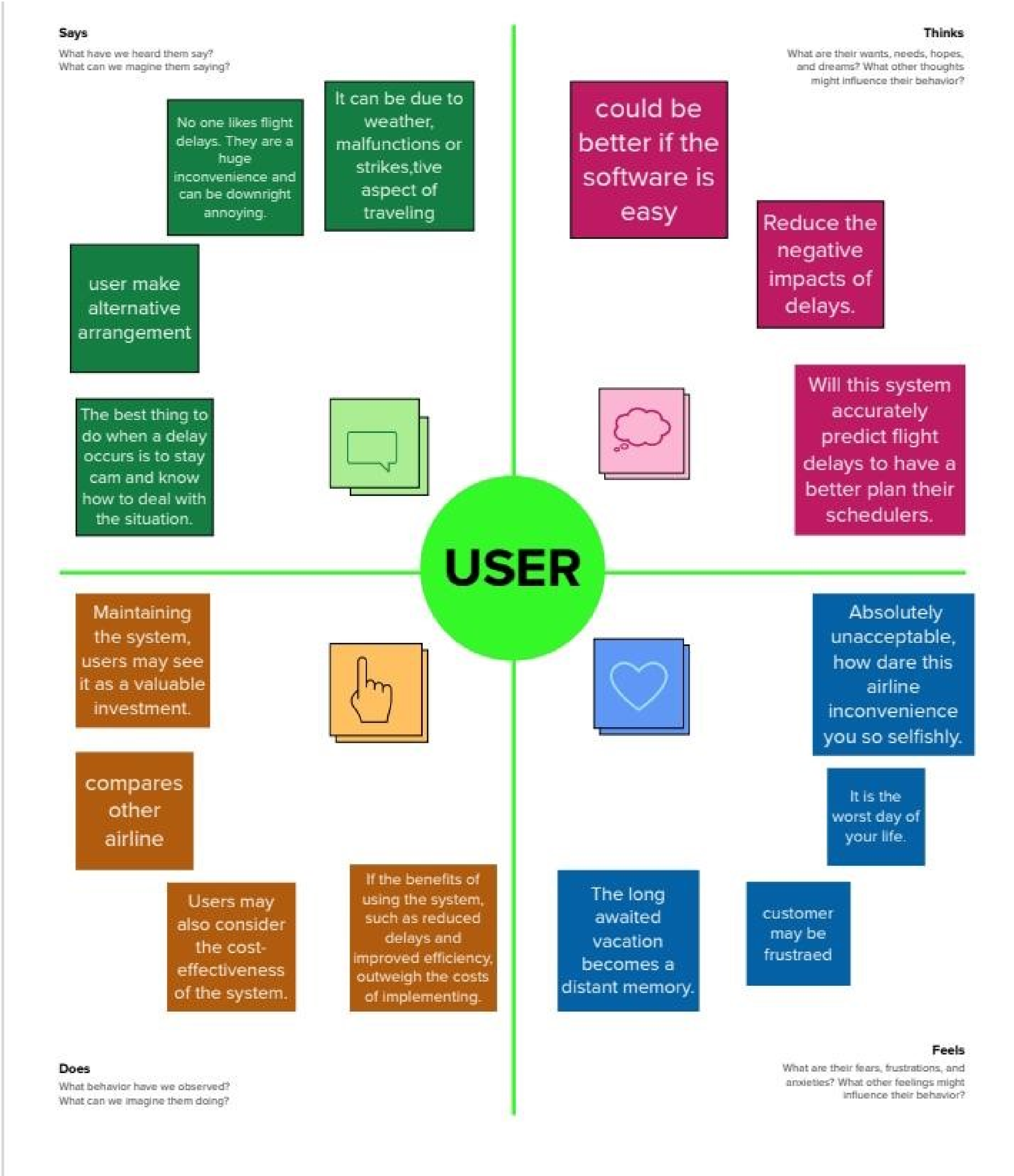
Reduce environmental impact: By minimizing delays and optimizing flight schedules, delay prediction models can help to reduce fuel consumption and emissions associated with air travel.

Increase safety: Delay prediction models can be used to help airlines and air traffic controllers to anticipate and manage potential safety issues, such as severe weather or runway closures.

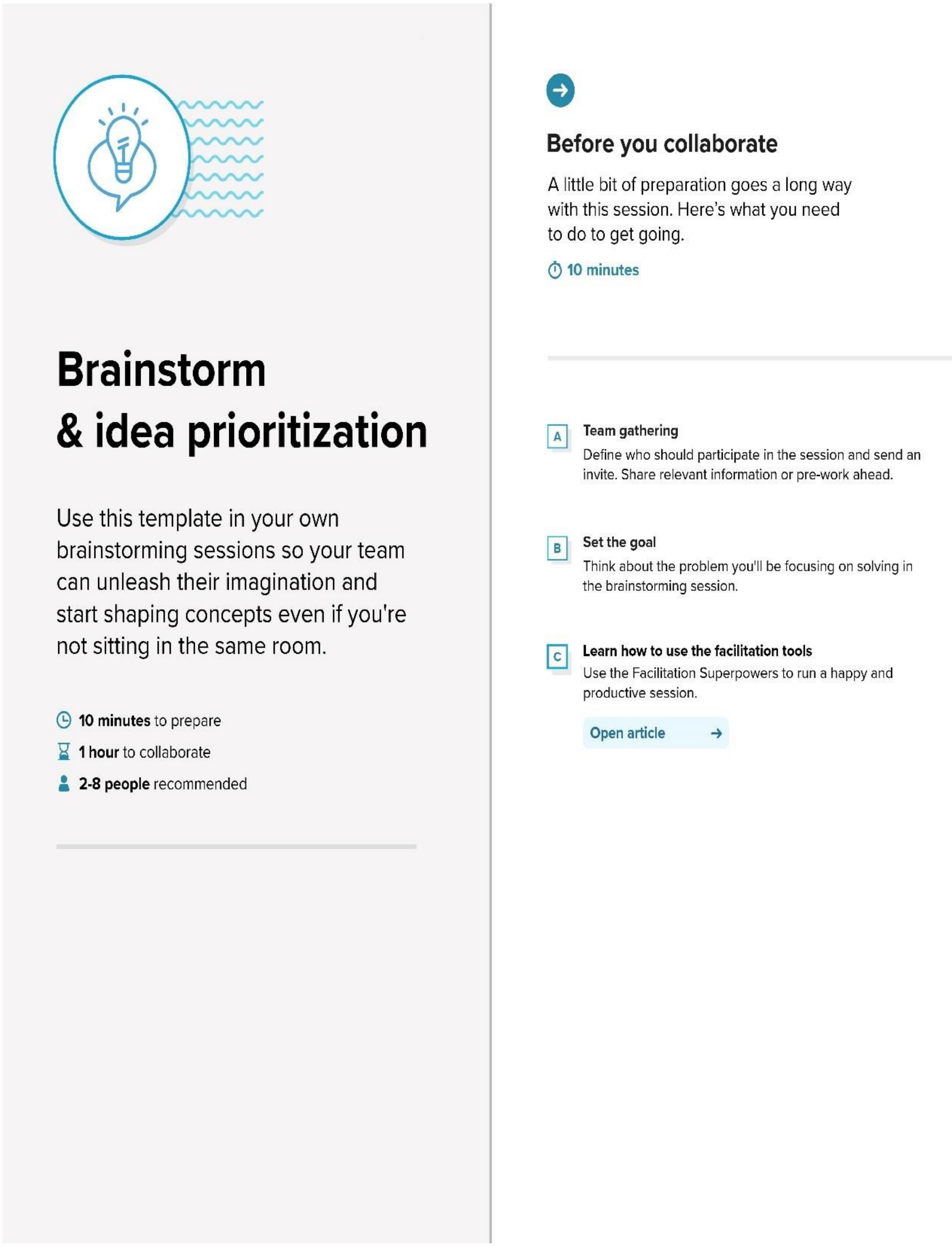
Overall, the purpose of flight delay prediction in the aviation industry is to improve the efficiency, safety, and customer experience of air travel. By predicting delays in advance, airlines and airports can take proactive measures to minimize their impact, and provide better service to passengers

Problem Definition & Design Thinking

2.1 Empathy Map



2.1 Problem Definition & Design Thinking



2

Brainstorm

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

10 minutes

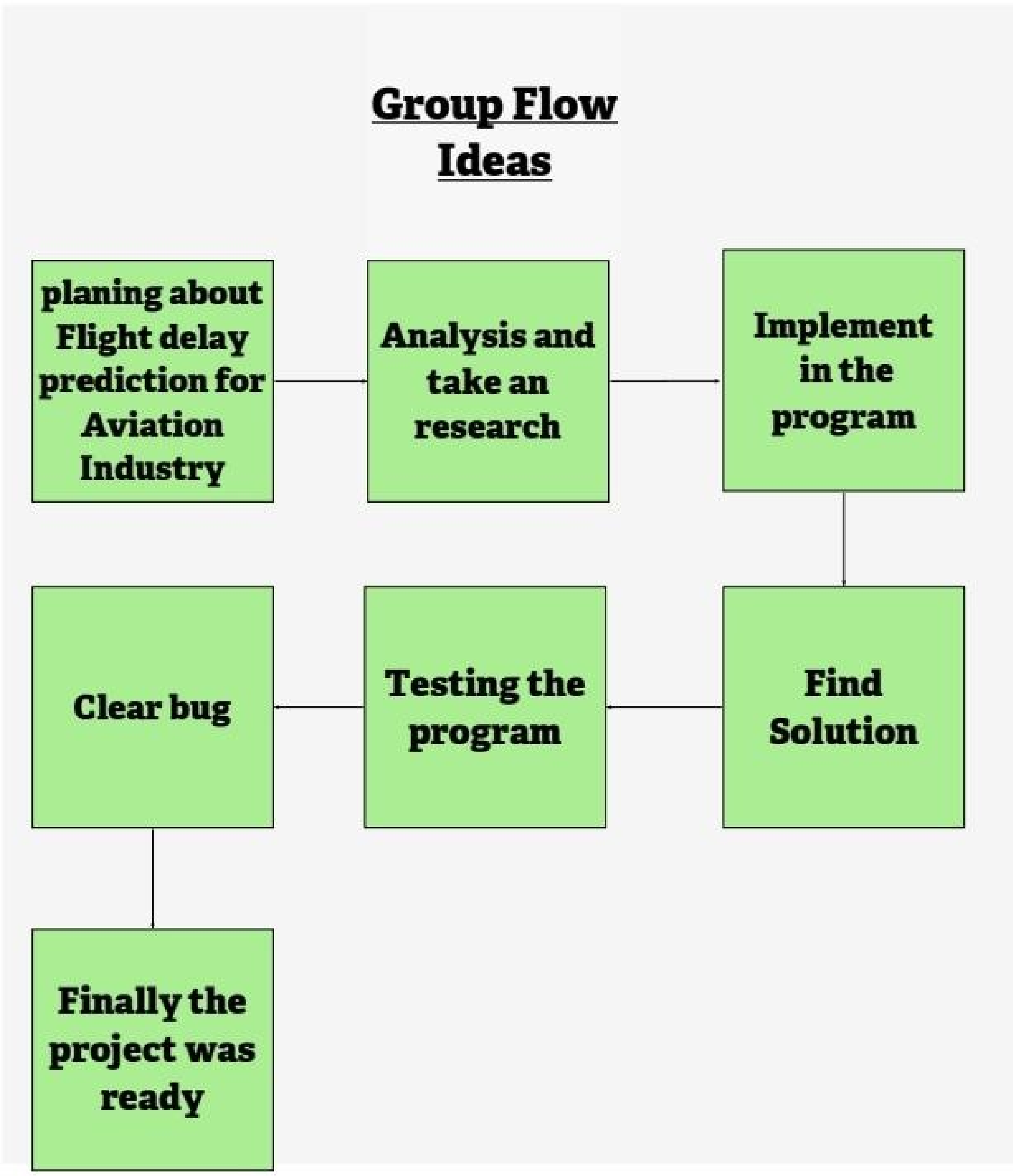
Archana R	Abisha A	Abisha Kumari T	Ajitha N
Data collection and cleaning: To develop an accurate predictive model, it is essential to have access to high-quality data on post flight delays as well as relevend weather and other factors. This data must be cleaned and pre-processed to remove any errors or inconsistencios.	Flight delay prediction is an important problem for the aviation and airports the travel expenence for passengers.	Choosing an appropriate algorithm: There are many machine learning algorithms that can be used for predicting flight delay including regression,decision trees, and neural networks.	Over fitting and generalization: A common problem in mechine learning is overfitting,where the model performs well on the training data.
Future selection and engineering: Once the data has been collected and cleaned, the next stop is to identify which features. (E.G. Weather variables, flight route,airline,etc..)core most relevant for predicting flight delays.	Machine Learning can be used to develop predittive models that can accurataly forecast flight delays,based on a range of factors such as weather conditions, air traffic congestion and mechanical issues.	Choosing the right algorithm for the problem at band requires careful consideration of factors such as the size and complexity of the dataset, the desired level of accurracy, and the computational resources available.	To avoid over fitting, techniques such as cross-validation and regularization can be used.
Feature engineering can also be used to create new variables that may improve the accuracy of the model.		Developing a flight delay prediction model using machine learning can be a chellenging task.	Deployment and maintence once a predictive model has been developed. It must be deployed and integrated into the airline or airports operations.Ongoing maintenance and monitoring is also required to ensure that the model continues to perform accurately over time.
			Overall, developing a flight delay prediction model using machine learning requires careful consideration of a range of factors, including data quality feature selection, algorithm choice, and deployment and maintenance.

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

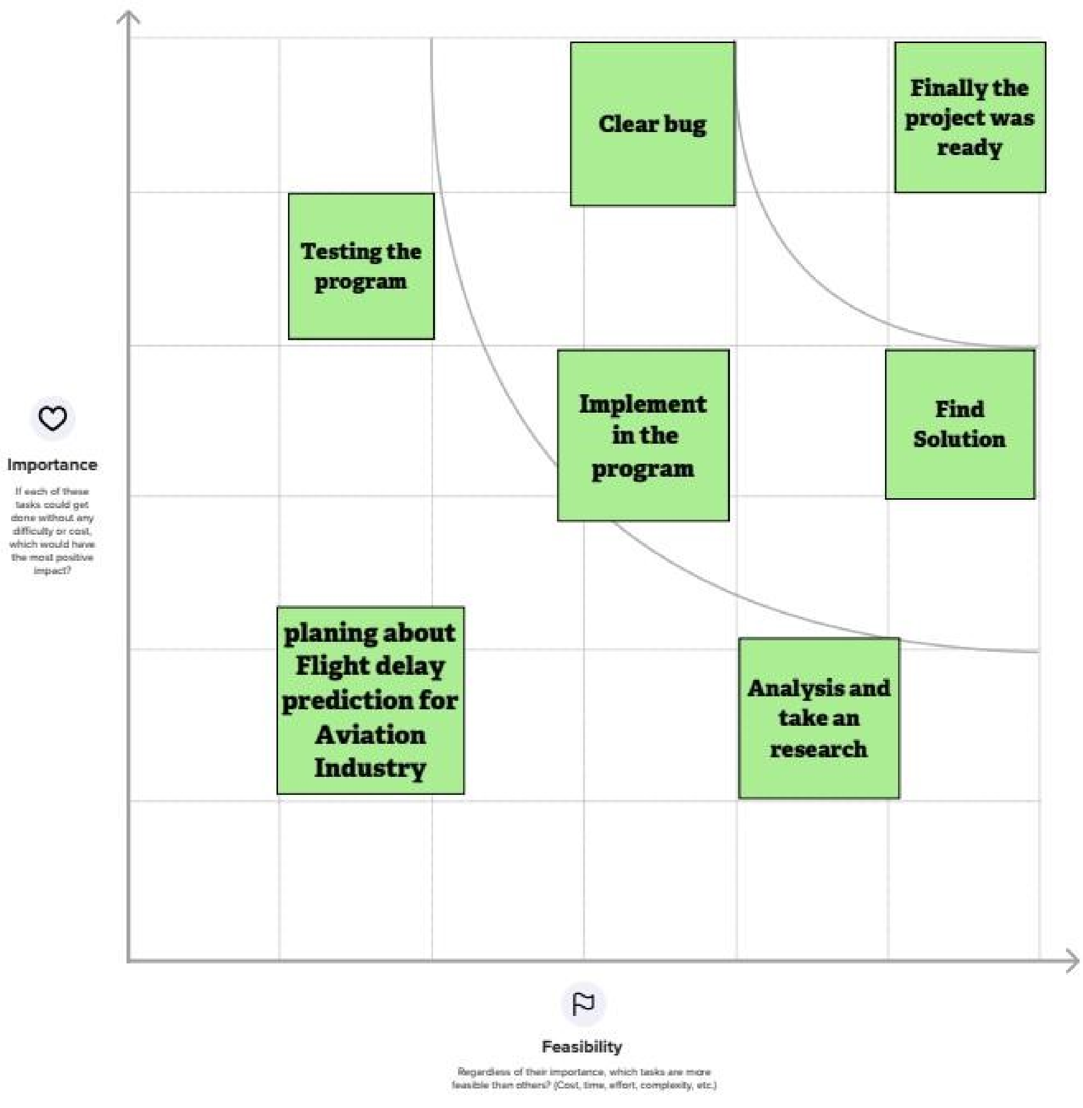
20 minutes



Prioritize

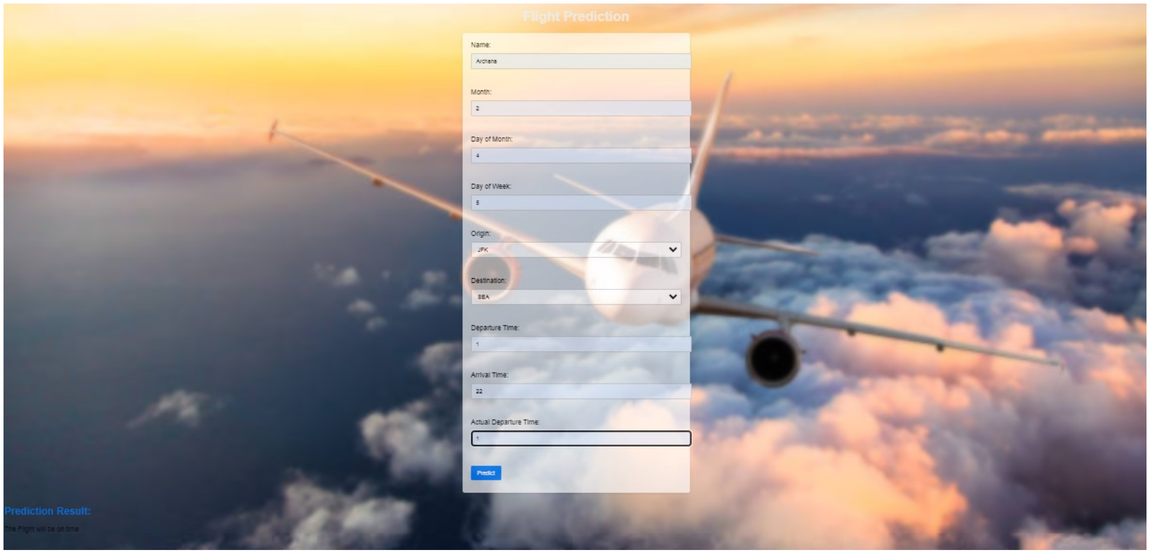
Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



Result

Output:



ADVANTAGES & DISADVANTAGES

Advantages

- Improved operational efficiency: Flight delay prediction can help airlines and airports to optimize their operations by anticipating potential delays and making proactive changes to flight schedules, crew assignments, and other resources. This can help to reduce costs and improve the efficiency of air travel.
- Enhanced customer satisfaction: Flight delays can be a major source of frustration for passengers, and accurate delay predictions can help airlines and airports to provide more accurate and timely information to passengers about their flights. This can help to reduce passenger stress and improve overall customer satisfaction.
- Better management of airport congestion: Delay prediction models can be used to anticipate periods of high congestion at airports and take steps to manage traffic flow. This can help to reduce delays and improve the efficiency of airport operations.
- More accurate planning and scheduling: Delay prediction can help airlines and airports to plan and schedule their operations more accurately by providing insights into potential delays and their causes. This can help to improve the accuracy of flight schedules and crew assignments.
- Improved safety: Delay prediction models can help airlines and air traffic controllers to anticipate and manage potential safety issues, such as severe weather or runway closures. This can help to improve the safety of air travel.

Disadvantages

- Accuracy limitations: While delay prediction models can be very accurate, they are not perfect and can sometimes make errors. These errors can lead to incorrect predictions, which can cause confusion or frustration for passengers and additional costs for airlines.
- Cost of implementation: Developing and implementing delay prediction models can be expensive, especially for smaller airlines or airports. There may be significant costs associated with acquiring and analyzing data, developing models, and integrating the models into existing systems.
- Overreliance on technology: Overreliance on delay prediction models and other technology can sometimes lead to complacency among airline and airport staff. If staff members assume that the models will always be accurate, they may be less likely to make critical decisions or take action when necessary.
- Privacy concerns: Delay prediction models rely on large amounts of data, including passenger and flight data. There may be concerns among passengers and privacy advocates about how this data is collected, used, and protected.
- Overall, while delay prediction models offer many benefits for the aviation industry, there are also potential drawbacks that must be considered. These include accuracy limitations, costs, overreliance on technology, privacy concerns, and ethical considerations

APPLICATION

Application of Job Prediction

- Airline operations: Airlines can use delay prediction models to optimize their operations by re-routing flights, adjusting crew schedules, or providing advance notice to passengers in the event of a delay. This can help to minimize the impact of delays on passengers and reduce costs associated with delays.
- Air traffic management: Air traffic controllers can use delay prediction models to anticipate potential congestion or safety issues and take steps to manage traffic flow or issue warnings to pilots. This can help to improve the safety and efficiency of air travel.
- Airport operations: Airports can use delay prediction models to anticipate periods of high congestion and take steps to manage traffic flow or allocate resources more effectively. This can help to reduce delays and improve the efficiency of airport operations.
- Flight planning and scheduling: Delay prediction models can be used to help airlines and airports plan and schedule flights more accurately, by providing insights into potential delays and their causes. This can help to improve the accuracy of flight schedules and crew assignments.
- Environmental impact: By minimizing delays and optimizing flight schedules, delay prediction models can help to reduce fuel consumption and emissions associated with air travel, reducing the environmental impact of the aviation industry.

CONCLUSION

Conclusion:

Flight delay prediction has become an increasingly important tool for the aviation industry, offering a range of benefits to airlines, airports, air traffic controllers, and passengers. By using data analytics and machine learning techniques, delay prediction models can accurately forecast flight delays and provide valuable insights into the causes of delays. This can help airlines and airports to optimize their operations, reduce costs, and improve the overall passenger experience. However, there are also potential drawbacks to consider, including accuracy limitations, implementation costs, and ethical concerns. Despite these challenges, the benefits of flight delay prediction make it an essential tool for the aviation industry, helping to improve safety, efficiency, and customer satisfaction, while reducing costs and minimizing the environmental impact of air travel. As technology continues to advance, it is likely that flight delay prediction will become even more sophisticated, providing even greater value to the aviation industry and its customers.

FUTURE SCOPE

Future Scope:

The future scope of flight delay prediction for the aviation industry is vast and promising. As technology continues to advance, new opportunities and challenges will emerge, leading to new applications and innovations. Here are some potential future developments in the field of flight delay prediction:

- Integration with other technologies: Delay prediction models could be integrated with other technologies, such as drones, to improve the accuracy and efficiency of flight operations.
- Real-time updates: Real-time updates could be provided to passengers, allowing them to adjust their travel plans or make alternative arrangements in the event of a delay.
- Personalized predictions: Delay prediction models could be personalized to individual passengers, taking into account factors such as past travel behavior, preferences, and booking history.

Overall, the future scope of flight delay prediction for the aviation industry is promising, with many potential applications and innovations on the horizon. As technology continues to advance, it is likely that delay prediction models will become even more accurate, efficient, and personalized, helping to improve safety, reduce costs, and enhance the overall passenger experience

APPENDIX

Source Code

1. Home.html

```
<!DOCTYPE html>
<html>
<head>
<title>Flight Prediction</title>
<style>
        body {
            font-family: Arial, sans-serif;
            background-color: blue;
            background: url('flight_bg.jpg') center center/cover no-repeat;
            backdrop-filter: blur(5px);
            opacity: 0.8;
        }

        h1 {
            text-align: center;
            color: #eee;
        }
```

```css
form {
    max-width: 500px;
    margin: 0 auto;
    padding: 20px;
    background-color: rgba(255, 255, 255, 0.8);
    border-radius: 5px;
    box-shadow: 0 2px 6px rgba(0, 0, 0, 0.1);
}

label, input, select {
    display: block;
    margin-bottom: 10px;
}

input[type="text"], select {
    width: 100%;
    padding: 10px;
    border: 1px solid #ccc;
    border-radius: 3px;
}

input[type="submit"] {
    background-color: #007bff;
    color: #fff;
    padding: 10px 15px;
    border: none;
```

```
            border-radius: 3px;
            cursor: pointer;
        }

        input[type="submit"]:hover {
            background-color: #0056b3;
        }

        h2 {
            margin-top: 30px;
            color: #007bff;
        }

        p {
            margin-top: 10px;
        }
</style>
</head>
<body>
<h1>Flight Prediction</h1>
<form action="/prediction" method="post">
<label for="name">Name:</label>
<input type="text" id="name" name="name"><br><br>
<label for="month">Month:</label>
<input type="text" id="month" name="month"><br><br>
<label for="dayofmonth">Day of Month:</label>
```

```html
<input type="text" id="dayofmonth" name="dayofmonth"><br><br>
<label for="dayofweek">Day of Week:</label>
<input type="text" id="dayofweek" name="dayofweek"><br><br>
<label for="origin">Origin:</label>
<select id="origin" name="origin">
<option value="msp">MSP</option>
<option value="dtw">DTW</option>
<option value="jfk">JFK</option>
<option value="sea">SEA</option>
<option value="alt">ALT</option>
</select><br><br>
<label for="destination">Destination:</label>
<select id="destination" name="destination">
<option value="msp">MSP</option>
<option value="dtw">DTW</option>
<option value="jfk">JFK</option>
<option value="sea">SEA</option>
<option value="alt">ALT</option>
</select><br><br>
<label for="dept">Departure Time:</label>
<input type="text" id="dept" name="dept"><br><br>
<label for="arrtime">Arrival Time:</label>
<input type="text" id="arrtime" name="arrtime"><br><br>
<label for="actdept">Actual Departure Time:</label>
<input type="text" id="actdept" name="actdept"><br><br>
<input type="submit" value="Predict">
```

```html
</form>

<h2>Prediction Result:</h2>
<p>{{ showcase }}</p>
</body>
</html>
```

2. app.py

```python
from flask import Flask, request, render_template
import joblib

# Load trained machine learning model
clf = joblib.load('flight_delay_prediction_model.joblib')

# Initialize Flask app
app = Flask(__name__)

# Define app routes
@app.route('/')
def home():
    return render_template('index.html')

@app.route('/predict', methods=['POST'])
def predict():
```

```
    # Get user input from web form
    departure_time = request.form.get('departure_time')
    arrival_time = request.form.get('arrival_time')
    carrier = request.form.get('carrier')
    distance = request.form.get('distance')

    # Make prediction using machine learning model
    prediction = clf.predict([[departure_time, arrival_time, carrier,
distance]])

    # Render prediction result in HTML template
    return render_template('prediction.html', prediction=prediction[0])

# Run app in debug mode
if __name__ == '__main__':
    app.run(debug=True)
```

3.flight.ipynb

```
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import RandomizedSearchCV
import imblearn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix,f1_score
dataset= pd.read_csv("flightdata.csv")
```

```
dataset.head()
```

	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	UNIQUE_CARRIER	TAIL_NUM	FL_NU
0	2016	1	1	1	5	DL	N836DN	139
1	2016	1	1	1	5	DL	N964DN	147
2	2016	1	1	1	5	DL	N813DN	159
3	2016	1	1	1	5	DL	N587NW	176
4	2016	1	1	1	5	DL	N836DN	182

5 rows × 26 columns

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11231 entries, 0 to 11230
Data columns (total 26 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   YEAR                 11231 non-null  int64
 1   QUARTER              11231 non-null  int64
 2   MONTH                11231 non-null  int64
 3   DAY_OF_MONTH         11231 non-null  int64
 4   DAY_OF_WEEK          11231 non-null  int64
 5   UNIQUE_CARRIER       11231 non-null  object
 6   TAIL_NUM             11231 non-null  object
 7   FL_NUM               11231 non-null  int64
 8   ORIGIN_AIRPORT_ID    11231 non-null  int64
 9   ORIGIN               11231 non-null  object
 10  DEST_AIRPORT_ID      11231 non-null  int64
 11  DEST                 11231 non-null  object
 12  CRS_DEP_TIME         11231 non-null  int64
 13  DEP_TIME             11124 non-null  float64
 14  DEP_DELAY            11124 non-null  float64
 15  DEP_DEL15            11124 non-null  float64
 16  CRS_ARR_TIME         11231 non-null  int64
 17  ARR_TIME             11116 non-null  float64
 18  ARR_DELAY            11043 non-null  float64
 19  ARR_DEL15            11043 non-null  float64
 20  CANCELLED            11231 non-null  float64
 21  DIVERTED             11231 non-null  float64
 22  CRS_ELAPSED_TIME     11231 non-null  float64
 23  ACTUAL_ELAPSED_TIME  11043 non-null  float64
 24  DISTANCE             11231 non-null  float64
 25  Unnamed: 25          0 non-null      float64
dtypes: float64(12), int64(10), object(4)
memory usage: 2.2+ MB
```

```
dataset = dataset.drop('Unnamed: 25', axis=1)
dataset.isnull().sum()
```

```
YEAR                    0
QUARTER                 0
MONTH                   0
DAY_OF_MONTH            0
DAY_OF_WEEK             0
```

```
UNIQUE_CARRIER           0
TAIL_NUM                 0
FL_NUM                   0
ORIGIN_AIRPORT_ID        0
ORIGIN                   0
DEST_AIRPORT_ID          0
DEST                     0
CRS_DEP_TIME             0
DEP_TIME               107
DEP_DELAY              107
DEP_DEL15              107
CRS_ARR_TIME             0
ARR_TIME               115
ARR_DELAY              188
ARR_DEL15              188
CANCELLED                0
DIVERTED                 0
CRS_ELAPSED_TIME         0
ACTUAL_ELAPSED_TIME    188
DISTANCE                 0
dtype: int64
```

```
dataset = dataset[[ "MONTH", "DAY_OF_MONTH", "DAY_OF_WEEK", "ORIGIN", "DEST", "CRS_ARR_TIME", "ARR_DEL15"]]
dataset.isnull().sum()
```

```
MONTH             0
DAY_OF_MONTH      0
DAY_OF_WEEK       0
ORIGIN            0
DEST              0
CRS_ARR_TIME      0
ARR_DEL15       188
dtype: int64
```

```
dataset[dataset.isnull().values.any(axis=1)].head()
```

	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	ARR_DEL15
177	1	9	6	MSP	SEA	852	NaN
179	1	10	7	MSP	DTW	1632	NaN
184	1	10	7	MSP	DTW	912	NaN
210	1	10	7	DTW	MSP	1303	NaN
478	1	22	5	SEA	JFK	723	NaN

```
dataset = dataset.fillna({'ARR_DEL15': 1})
dataset.iloc[177:185]
```

	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	ARR_DEL15
177	1	9	6	MSP	SEA	852	1.0
178	1	9	6	DTW	JFK	1724	0.0
179	1	10	7	MSP	DTW	1632	1.0
180	1	10	7	DTW	MSP	1649	0.0
181	1	10	7	JFK	ATL	1600	0.0
182	1	10	7	JFK	ATL	849	0.0
183	1	10	7	JFK	SEA	1945	0.0
184	1	10	7	MSP	DTW	912	1.0

```
dataset.head()
```

```
import math
```

0 1 1 5 ATL SEA 2143 0.0

```
for index, row in dataset.iterrows():
  dataset.loc[index,'CRS_ARR_TIME'] = math.floor(row['CRS_ARR_TIME']/100)
```

```
dataset.head()
```

	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	ARR_DEL15
0	1	1	5	ATL	SEA	21	0.0
1	1	1	5	DTW	MSP	14	0.0
2	1	1	5	ATL	SEA	12	0.0
3	1	1	5	SEA	MSP	13	0.0
4	1	1	5	SEA	DTW	6	0.0

```
dataset = pd.get_dummies(dataset, columns=['ORIGIN', 'DEST'])
dataset.head()
```

	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	CRS_ARR_TIME	ARR_DEL15	ORIGIN_ATL	ORIGIN_D
0	1	1	5	21	0.0	1	
1	1	1	5	14	0.0	0	
2	1	1	5	12	0.0	1	
3	1	1	5	13	0.0	0	
4	1	1	5	6	0.0	0	

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test = train_test_split(dataset.drop('ARR_DEL15',axis=1),dataset['ARR_DEL15'],test_size=0.2,random_state=42)
```

```
x_train.shape
```

```
(8984, 14)
```

```
x_test.shape
```

```
(2247, 14)
```

```
from  sklearn.ensemble import RandomForestClassifier
```

```
model = RandomForestClassifier(random_state=13)
model.fit(x_train, y_train)
```

```
▾        RandomForestClassifier
RandomForestClassifier(random_state=13)
```

```
predicted = model.predict(x_test)
model.score(x_test, y_test)
```

```
0.8633733867378727
```

```
from sklearn.metrics import roc_auc_score
```

```
probabilities = model.predict_proba(x_test)
```

```
roc_auc_score(y_test, probabilities[:, 1])
```

```
0.6847853166272488
```

```
from sklearn.metrics import confusion_matrix
```

```
confusion_matrix(y_test, predicted)
```

```
array([[1906,   30],
       [ 277,   34]])
```

```
from sklearn.metrics import precision_score
```

```
train_predictions = model.predict(x_train)
precision_score(y_train, train_predictions)
```

```
1.0
```

```
from sklearn.metrics import recall_score
```

```
recall_score(y_train, train_predictions)
```

```
0.9984025559105432
```

```
import pickle

pickle.dump(model,open('flight.pkl','wb'))
```

```
model_loaded = pickle.load(open('flight.pkl', 'rb'))
```

```
model_loaded.predict(x_test)
```

```
array([0., 0., 0., ..., 0., 0., 0.])
```

```
import joblib
```

```
joblib.dump(model,'flight2')
```

```
['flight2']
```