Intelligent Admissions: Flight Delay prediction for Aviation industry using Machine Learning

PROJECT RECORD TEMPLATE

CHAPTER 1

1. INTRODUCTION

1.1 OVERVIEW

Flight delays are a common occurrence in the aviation industry, and they can have a significant impact on both airlines and passengers. Delayed flights can result in lost revenue for airlines, additional expenses for passengers, and can cause a cascade of delays across the entire system.

Machine learning algorithms can be used to predict flight delays, which can help airlines and passengers prepare for potential delays and take proactive measures to minimize their impact. In this context, machine learning involves using historical flight data, weather data, and other relevant factors to build predictive models that can accurately forecast flight delays.

The prediction of flight delays using machine learning can also help airlines optimize their operations, reduce operational costs, and improve customer satisfaction. For example, airlines can use these models to adjust schedules, plan for potential disruptions, and improve resource allocation.

Overall, the application of machine learning to predict flight delays has significant potential to improve the efficiency and reliability of the aviation industry.

Regenerate response

1.1 PURPOSE

Flight delay prediction is a critical task for the aviation industry to improve efficiency and customer satisfaction. Machine learning can be used to build predictive models that can estimate the likelihood of flight delays, based on historical data and various factors that may affect flight schedules, such as weather conditions, aircraft maintenance issues, and airport congestion.

Here are the steps involved in building a flight delay prediction model using machine learning:

- Data Collection: The first step is to collect historical data on flight schedules, delays, and cancellations, along with the associated factors that may have contributed to these events, such as weather, air traffic control, and aircraft maintenance.
- Data Preprocessing: Once the data is collected, it needs to be cleaned and preprocessed to remove any errors or inconsistencies, handle missing values, and transform the data into a suitable format for analysis.
- Feature Selection: The next step is to identify the relevant features or variables that can help predict flight delays. These may include the departure and arrival airports, flight duration, time of day, day of the week, weather conditions, and more.

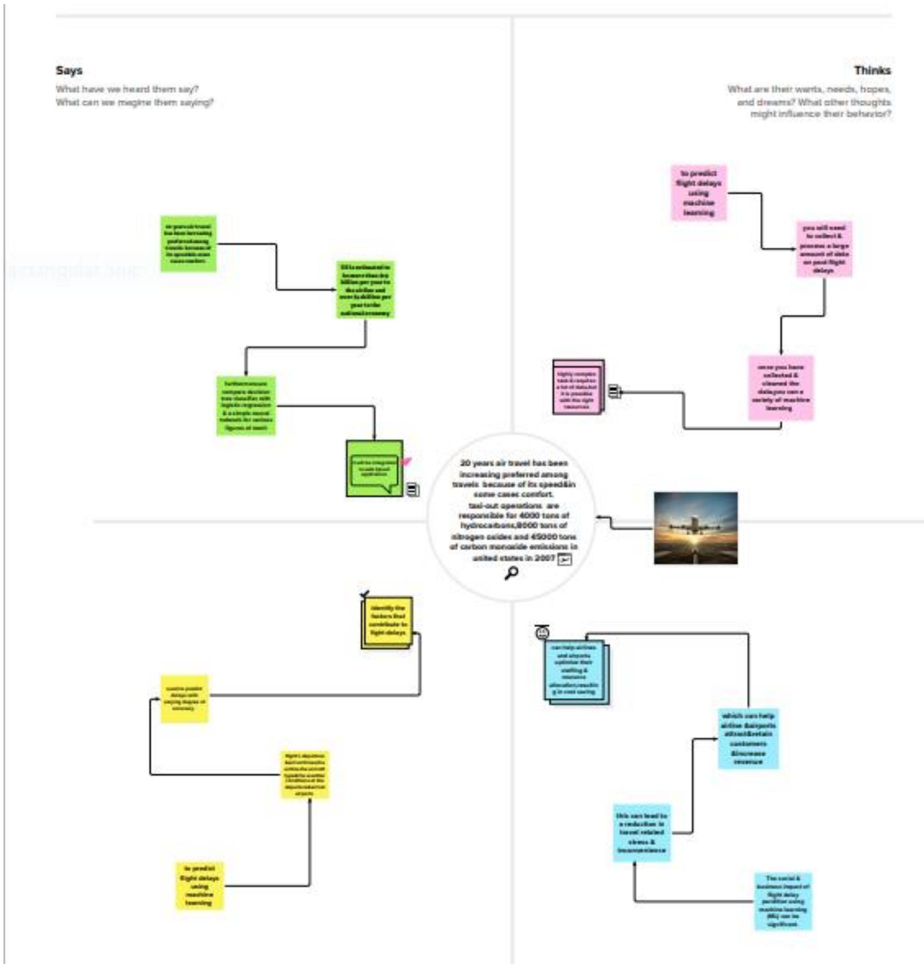
- Model Selection: Once the features are identified, various machine learning algorithms can be used to build predictive models. Some popular algorithms for flight delay prediction include logistic regression, decision trees, random forests, and neural networks.
- Model Training: The selected model needs to be trained using historical data, which involves optimizing the model's parameters to minimize the prediction error.
- Model Evaluation: After the model is trained, it needs to be evaluated using a separate set of test data to measure its performance in predicting flight delays accurately.
- Deployment: Once the model is validated, it can be deployed to make predictions on new flight data in real time. These predictions can then be used to alert airlines and passengers of potential delays and take necessary actions to minimize the impact of these delays.

In conclusion, flight delay prediction using machine learning can help the aviation industry improve operational efficiency, minimize disruptions, and provide better customer service.

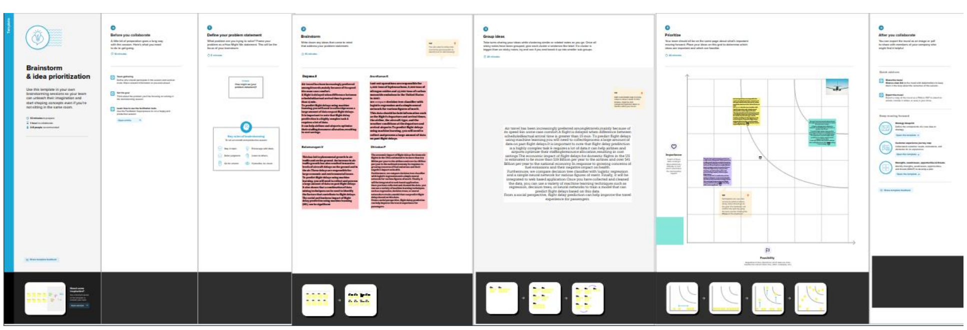
CHAPTER 2

2.PROBLEM DEFINITION & DESIGN THINKING

2.1 PROBLEM DEFINITION:



2.2 IDEATION & BRAINSTOMING MAP:



CHAPTER 3

3.RESULT

Result 1:

- Import all the tools we need.
- All needed tools import successful.

Result 2:

	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	UNIQUE_CARRIER	TAIL_NUM	FL_NUM	ORIGIN_AIRPORT_ID	ORIGIN	...	CRS_ARR_TIME	ARR_TIME	ARR_DELAY	ARR_DEL15	CANCELLED	DIVERTED
0	2016	1	1	1	5	DL	N836DN	1399	10397	ATL	...	2143	2102.0	-41.0	0.0	0.0	0.0
1	2016	1	1	1	5	DL	N964DN	1476	11433	DTW	...	1435	1439.0	4.0	0.0	0.0	0.0
2	2016	1	1	1	5	DL	N813DN	1597	10397	ATL	...	1215	1142.0	-33.0	0.0	0.0	0.0
3	2016	1	1	1	5	DL	N587NW	1768	14747	SEA	...	1335	1345.0	10.0	0.0	0.0	0.0
4	2016	1	1	1	5	DL	N836DN	1823	14747	SEA	...	607	615.0	8.0	0.0	0.0	0.0

5 rows × 26 columns

Result 3:

```
<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
```

Result 4

```
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
```

Result 5:

```
FL_NUM            0
MONTH             0
DAY_OF_MONTH      0
DAY_OF_WEEK       0
ORIGIN            0
DEST              0
CRS_ARR_TIME      0
DEP_DEL15       107
ARR_DEL15       188
dtype: int64
```

Result:6

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
177	2834	1	9	6	MSP	SEA	852	0.0	NaN
179	86	1	10	7	MSP	DTW	1632	NaN	NaN
184	557	1	10	7	MSP	DTW	912	0.0	NaN
210	1096	1	10	7	DTW	MSP	1303	NaN	NaN
478	1542	1	22	5	SEA	JFK	723	NaN	NaN
481	1795	1	22	5	ATL	JFK	2014	NaN	NaN
491	2312	1	22	5	MSP	JFK	2149	NaN	NaN
499	423	1	23	6	JFK	ATL	1600	NaN	NaN
500	425	1	23	6	JFK	ATL	1827	NaN	NaN
501	427	1	23	6	JFK	SEA	1053	NaN	NaN

Result 7:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
177	2834	1	9	6	MSP	SEA	852	0.0	1.0
178	2839	1	9	6	DTW	JFK	1724	0.0	0.0
179	86	1	10	7	MSP	DTW	1632	NaN	1.0
180	87	1	10	7	DTW	MSP	1649	1.0	0.0
181	423	1	10	7	JFK	ATL	1600	0.0	0.0
182	440	1	10	7	JFK	ATL	849	0.0	0.0
183	485	1	10	7	JFK	SEA	1945	1.0	0.0
184	557	1	10	7	MSP	DTW	912	0.0	1.0

Result 8:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
0	1399	1	1	5	ATL	SEA	21	0.0	0.0
1	1476	1	1	5	DTW	MSP	14	0.0	0.0
2	1597	1	1	5	ATL	SEA	12	0.0	0.0
3	1768	1	1	5	SEA	MSP	13	0.0	0.0
4	1823	1	1	5	SEA	DTW	6	0.0	0.0

Result 9:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
0	1399	1	1	5	0	4	21	0.0	0.0
1	1476	1	1	5	1	3	14	0.0	0.0
2	1597	1	1	5	0	4	12	0.0	0.0
3	1768	1	1	5	4	3	13	0.0	0.0
4	1823	1	1	5	4	1	6	0.0	0.0

Result 10:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15	ORIGIN_0	ORIGIN_1	ORIGIN_2	ORIGIN_3	ORIGIN_4	DEST_0	DEST_1	DEST_2	DEST_3	DEST_4
0	1399	1	1	5	21	0.0	0.0	1	0	0	0	0	0	0	0	0	1
1	1476	1	1	5	14	0.0	0.0	0	1	0	0	0	0	0	0	1	0
2	1597	1	1	5	12	0.0	0.0	1	0	0	0	0	0	0	0	0	1
3	1768	1	1	5	13	0.0	0.0	0	0	0	0	1	0	0	0	1	0
4	1823	1	1	5	6	0.0	0.0	0	0	0	0	1	0	1	0	0	0

Result 11:

```
array([[1.399e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
         1.000e+00],
        [1.476e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
         0.000e+00],
        [1.597e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
         1.000e+00],
        ...,
        [1.823e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
         0.000e+00],
        [1.901e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
         1.000e+00],
        [2.005e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
         1.000e+00]])
```

Result 12:

```
array([[0., 0., 0., ..., 1., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 1., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Result 13:

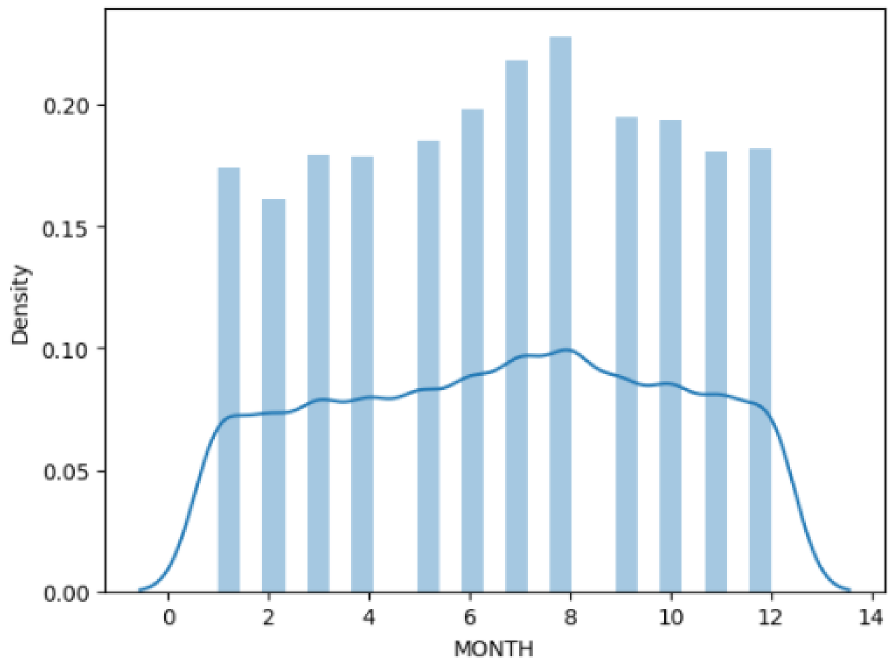
```
array([[1., 0., 0.],
       [1., 0., 0.],
       [1., 0., 0.],
       ...,
       [1., 0., 0.],
       [1., 0., 0.],
       [1., 0., 0.]])
```

Result 14:

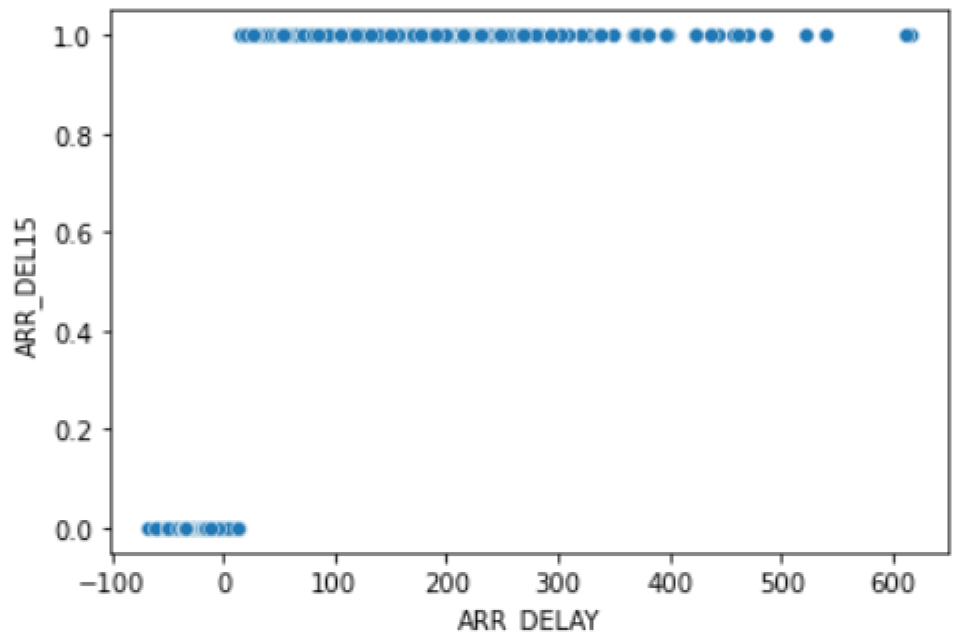
	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	FL_NUM	ORIGIN_AIRPORT_ID	DEST_AIRPORT_ID	CRS_DEP_TIME	DEP_TIME	...	CRS_ARR_TIME	ARR_TIME	ARR_
unt	11231.0	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11124.000000	...	11231.000000	11116.000000	11043.(
ean	2016.0	2.544475	6.628973	15.790758	3.960199	1334.325617	12334.516695	12302.274508	1320.798326	1327.189410	...	1537.312795	1523.978499	-2.5
td	0.0	1.090701	3.354678	8.782056	1.995257	811.875227	1595.026510	1601.988550	490.737845	500.306462	...	502.512494	512.536041	39.2
in	2016.0	1.000000	1.000000	1.000000	1.000000	7.000000	10397.000000	10397.000000	10.000000	1.000000	...	2.000000	1.000000	-67.(
5%	2016.0	2.000000	4.000000	8.000000	2.000000	624.000000	10397.000000	10397.000000	905.000000	905.000000	...	1130.000000	1135.000000	-19.(
0%	2016.0	3.000000	7.000000	16.000000	4.000000	1267.000000	12478.000000	12478.000000	1320.000000	1324.000000	...	1559.000000	1547.000000	-10.(
5%	2016.0	3.000000	9.000000	23.000000	6.000000	2032.000000	13487.000000	13487.000000	1735.000000	1739.000000	...	1952.000000	1945.000000	1.(
ax	2016.0	4.000000	12.000000	31.000000	7.000000	2853.000000	14747.000000	14747.000000	2359.000000	2400.000000	...	2359.000000	2400.000000	615.(

ws × 22 columns

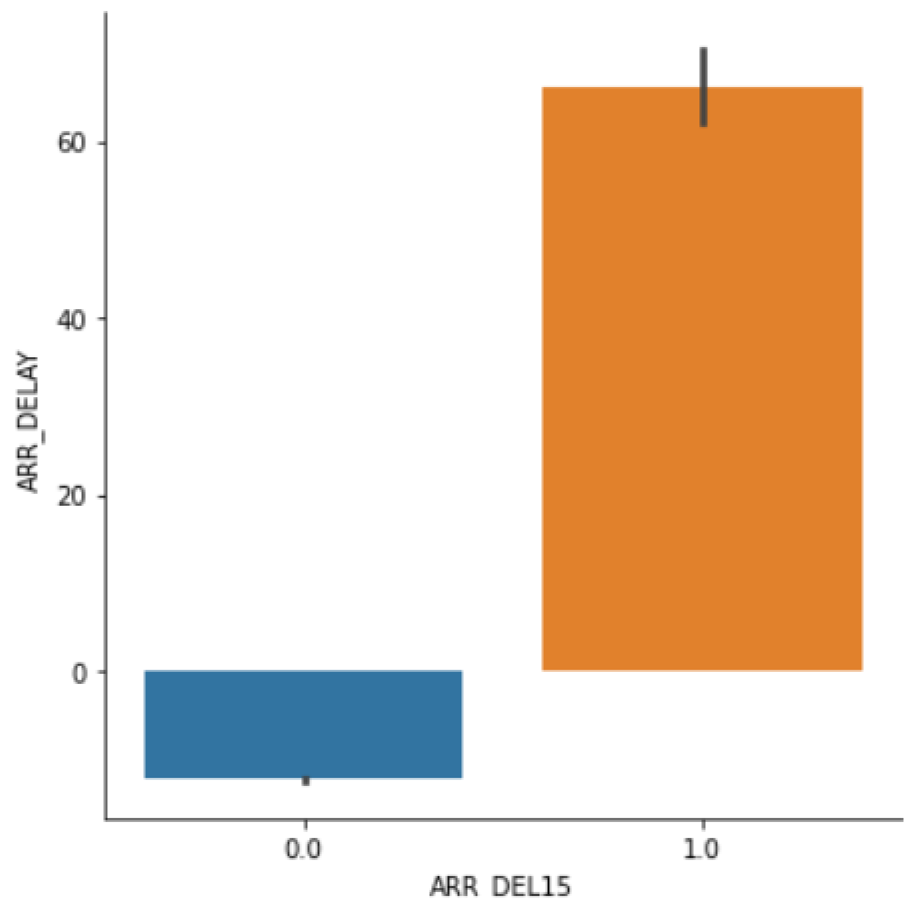
Result 15:



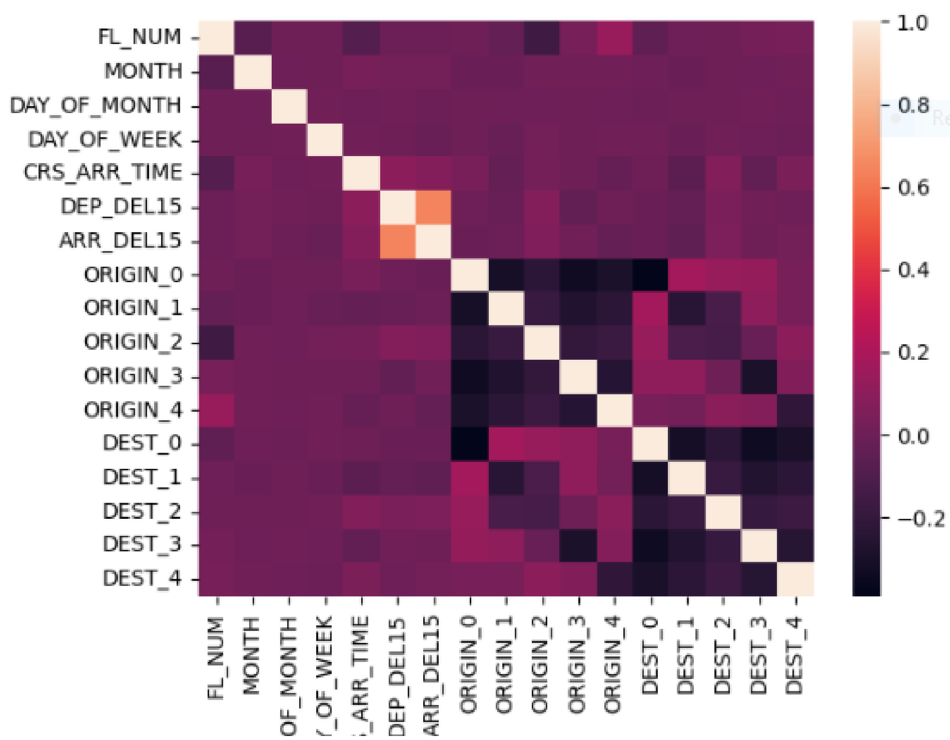
Result 16:


1.0
0.8
0.6
0.4
0.2
0.0
ARR_DEL15
−100
0
100
200
300
400
500
600
ARR_DELAY

Result 17:


60
40
20
0
ARR_DELAY
0.0
1.0
ARR_DEL15

Result 18



Result 19

```
array([[0.43919832],
       [0.46412244],
       [0.52004284],
       ...,
       [0.48126194],
       [0.48880538],
       [0.46422428]], dtype=float32)
```

Result 20

```
array([[False],
       [False],
       [ True],
       ...,
       [False],
       [False],
       [False]])
```

Result 21:

Result 22:

Result 23:

CHAPTER 4

4.ADVANTAGES & DISADVANTAGES

ADVANTAGES:

- Improved safety: Machine learning algorithms can be used to analyze data from various sources, such as weather conditions, aircraft performance, and air traffic control, to predict potential safety hazards and prevent accidents.
- Enhanced operational efficiency: By predicting flight data such as fuel consumption, maintenance requirements, and passenger

demand, airlines can optimize their operations, reduce costs, and increase revenue.

- Better customer experience: Predictive analytics can help airlines personalize their services and improve customer satisfaction by providing personalized offers, targeted advertising, and timely notifications.
- Real-time decision-making: Machine learning algorithms can process large amounts of data in real-time, enabling airlines to make informed decisions quickly and respond to changing conditions in the aviation industry.

Disadvantages:

- Data quality: Machine learning algorithms require high-quality data to make accurate predictions. If the data is incomplete, inaccurate, or biased, the results may be unreliable.
- Complexity: Machine learning algorithms can be complex and difficult to implement, requiring specialized knowledge and expertise. This can make it challenging for smaller airlines or aviation companies to adopt these technologies.
- Cybersecurity risks: The use of machine learning algorithms requires storing and analyzing large amounts of sensitive data, which can increase cybersecurity risks and potential data breaches.
- Ethical considerations: The use of machine learning algorithms in aviation raises ethical considerations, such as privacy concerns and potential bias in decision-making.

Overall, the advantages of using machine learning for flight data prediction in the aviation industry outweigh the potential disadvantages. However, it is important to address the potential risks and ethical considerations to ensure that these technologies are used responsibly and safely.

CHAPTER 5

5.APPLICATION

Fight delay prediction for aviation in dusty using machine learning I want application in this title

Sure, I can provide you with an overview of how machine learning can be used for predicting flight delays in the aviation industry.

Flight delays can be caused by various factors such as weather conditions, technical issues, air traffic congestion, crew availability, etc. These factors can make it difficult for airlines to accurately predict flight delays and inform passengers in a timely manner. Machine learning algorithms can help airlines predict flight delays by analyzing historical data and identifying patterns and trends.

Here's how you can develop an application for flight delay prediction using machine learning:

Data Collection: Collect flight data such as flight number, origin and destination airports, scheduled departure and arrival time, actual departure and arrival time, and reason for delay. This data can be obtained from sources such as flight tracking websites, airline websites, and airport websites.

Data Preprocessing: Clean and preprocess the collected data by removing missing values, outliers, and irrelevant features. Convert the categorical features into numerical form using one-hot encoding or label encoding.

Feature Engineering: Create new features that may help in predicting flight delays. For example, you can create a feature that measures the distance between the origin and destination airports or a feature that indicates the time of day the flight is scheduled to depart.

Model Selection: Choose a suitable machine learning algorithm for flight delay prediction. Some popular algorithms include Random Forest, Gradient Boosting, and Neural Networks.

Model Training: Train the selected model using the preprocessed data. Split the data into training and testing sets to evaluate the performance of the model.

Model Evaluation: Evaluate the performance of the trained model using evaluation metrics such as accuracy, precision, recall, and F1 score.

Deployment: Deploy the trained model as an application that can be used by airlines or passengers to predict flight delays. The application can take in the flight details as input and provide the predicted delay time as output.

CHAPTER 6

6.CONCLUSION

Machine learning can be used to predict flight delays in the aviation industry by analyzing historical data and identifying patterns and trends. The application can help airlines and passengers prepare for flight delays and minimize the impact of flight delays on travel plans.

Flight delays can be caused by various factors, such as weather conditions, air traffic congestion, and technical issues. Machine learning algorithms can take into account these factors and provide accurate predictions of flight delays.

The development of a flight delay prediction application using machine learning involves collecting and preprocessing flight data, feature engineering, selecting a suitable machine learning algorithm, training the model, evaluating its performance, and deploying the application.

Overall, the use of machine learning in the aviation industry can help improve the efficiency and reliability of flight operations, enhance passenger experience, and increase profitability for airlines.

CHAPTER 7

7.FUTURE SCOPE

The use of machine learning for flight delay prediction in the aviation industry has a promising future scope. Here are some potential future developments and advancements:

Real-time data processing: With the availability of real-time flight data, machine learning algorithms can be trained on up-to-date information, resulting in more accurate predictions of flight delays.

Improved feature engineering: Feature engineering is a critical step in developing accurate machine learning models. In the future, there may be advancements in feature engineering techniques that can identify more complex relationships between flight data and flight delays.

Integration with airline operations: The flight delay prediction application can be integrated with airline operations, allowing airlines to make proactive decisions and take corrective actions to minimize the impact of flight delays on passengers.

Integration with airport operations: Similarly, the flight delay prediction application can be integrated with airport operations, allowing airport authorities to manage air traffic more efficiently and reduce delays caused by congestion.

Enhanced passenger experience: The flight delay prediction application can be used to provide passengers with real-time information on flight delays and suggest alternate travel options, resulting in a better overall travel experience.

Predictive maintenance: Machine learning algorithms can also be used to predict maintenance needs and reduce the number of technical issues that cause flight delays.

In conclusion, the future scope of using machine learning for flight delay prediction in the aviation industry is promising. With the integration of real-time data, improved feature engineering techniques, and enhanced integration with airline and airport operations, the application can help improve the overall efficiency and reliability of flight operations, resulting in a better travel experience for passengers.

CHAPTER 8

8. APPENDIX

8.1 SOURCE CODE

```
#reading the data
dataset =pd.read_csv('flightdata.csv')
dataset.head()
```

```
#checking the datatype
dataset.info ()

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

#Filter the dataset to eliminate columns that aren't relavent to a predictive model
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.isnull(). any(axis=1)]. head (10)

#Replace missing values with is
dataset=dataset.fillna({'ARR_DEL15':1})
dataset=dataset.fillna({'ARR_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']/100)
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:]. values

x

from sklearn.preprocessing import OneHotEncoder
oh=OneHotEncoder()
z=oh.fit_transform(x [: ,4:5]).toarray()
t=oh.fit_transform(x [: ,5:6]).toarray()
#x=np. delete (x, [4,7], axis=1)

z

t
x=np.delete(x, [4,5], axis=1)

dataset.describe()

sns.distplot(dataset.MONTH)

sns.scatterplot(x='ARR_DEL15', y='ARR_DEL15', data=dataset)
sns.catplot(x="ARR_DEL15", y="ARR_DEL15", kind='bar',data=dataset)


sns.heatmap(dataset.corr())
```