Flight Delay Prediction for aviation Industry using Machine Learning

SUBMITTED BY …

S.Vinisa Bsc (cs)

GOVT ARTS AND SCIENCE COLLEGE,

JAYANKONDAM

INTRODUCTION:

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 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.

- **OVERVIEW:**

According to, taxi-out operations are responsible for 4,000 tons of hydrocarbons, 8,000 tons of nitrogen oxides and 45,000 tons of carbon monoxide emissions in the United States in 2007. Moreover, the economic impact of flight delays for domestic flights in the US is estimated to be more than $19 Billion per year to the airlines and over $41 Billion per year to the national economy In response to growing concerns of fuel emissions and their negative impact on health, there is active research in the aviation industry for finding techniques to predict flight delays accurately in order to optimize flight operations and minimize delays.

LITERATURE SURVEY:

Indian state of affairs, in 2017, in line with the reports by the directorate General of Civil Aviation (DGCA), between January and April, close to 5.12 hundred thousand domestic passengers in India faced issues because of airline corporations denying boarding, moreover as flight cancellations and delays [2] . Airline corporations had to pay the passengers compensations of over Rs. twenty five crore for varied inconveniences throughout the first four months of this year. Hence, the prediction analysis retrieved from this project can contribute within the form of a prototype in helping to identify operational variables that contribute to delays in any country scenario[2] The main issues associated with flight delay prediction are known and arranged in taxonomy. It includes the problem that causes the flight delay, the range of institution it affects, and ways that of handling flight delay prediction downside. It considers flight domain options, like problem and scope. Major problem. As discussed, considering the standard taxonomy of the flight delay and its problems, one will contemplate the scope of prediction to be one in every of these factors or combination of those factors[3]. The models developed during this system may be applied to predict the incidence of flight delay at airports. Such prognosticative capabilities would facilitate traffic managers and airline dispatchers to organize mitigation methods for reducing traffic disruptions. This issue can be

reduced by developing the flight delay prediction tool which can be developed using following methods.

- **THEORITICAL ANALYSIS:**

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

HARWARE/SOFTWARE DESIGNING

The Hardware required for the development for this project is :

Processor : Intel Core TM i5-9300H

Processor speed : 2.4GHz

RAM size : 8GB DDR

System Type : X64-based processor

- **SOFTWARE DESIGNING**:

The software required for the development of this project is:

Desktop GUI : Anaconda Navigator

Operating system : Windows 10

Front end : HTML, CSS,JAVA SCRIPT

Programming : PYTHON

EXPERIMENTAL INVESTIGATION IMPORTING AND READING THE DATASET

Importing the Libraries:

First step is usually importing the libraries that will be needed in the program.

Pandas:

It is a python library is used for data manipulation.

NumPy:

This python library is used for numerical analysis.

Pickle:

To serialize your machine learning algorithms and save the serialized format to a file.

Matplotlib and Seaborn:

Both are the data visualization library used for plotting graph which will help as for understanding the data.

Seaborn:

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. For a brief introduction to the ideas behind the library, you can read the introductory notes or the paper.

Sklearn:

Scikit-Learn, also known as sklearn is a python library to implement machine learning models and statistical modelling. Through scikit-learn, we can implement various machine learning models for regression, classification, clustering, and statistical tools for analyzing these models.

Reading the dataset:

For this project, we make use of three different datasets(Books_Ratings,Books, Users).We will be selecting the important features from these datasets that will help us in recommending the best results.

The next step is to read the dataset into a data structure that's compatible with pandas. There is a function for it, called read_csv(). We will need to locate the directory of the CSV file at first(it's more efficient to keep the dataset in the same directory as your program). If the dataset in same directory of your program,

you can directory read it, without any path. After the next Steps we made following bellow:

1.Ecploratory Date Analysis

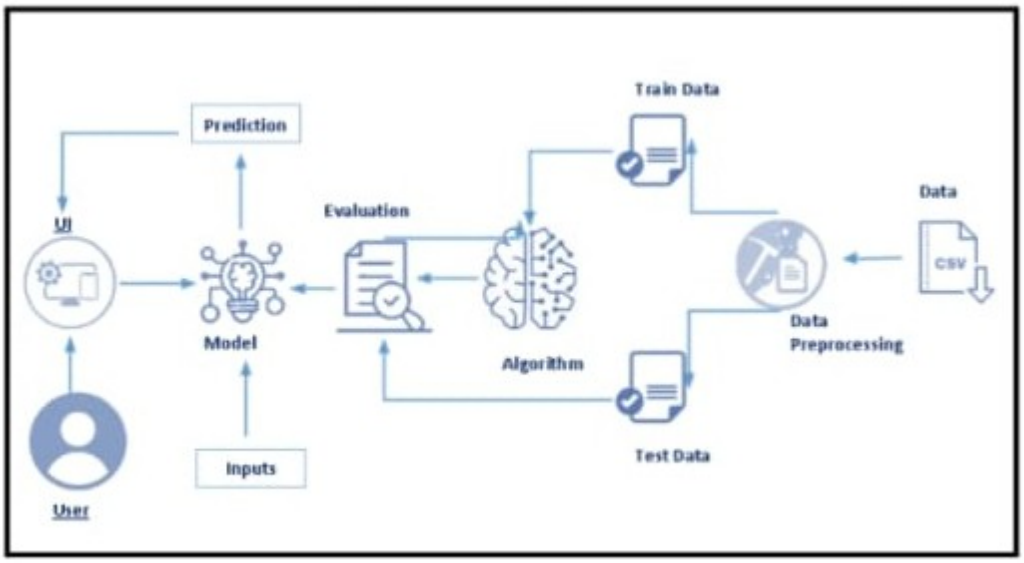
2.Model Building

3.Performance Testing & Hyperparameter Tuning

4.Model Deployment

5.Project Demonstration & Documentation

FLOWCHART:



Project flow:

- User interacts with the UI to enter the input.
- Entered input is analysed by the model which is integrated.
- Once model analyses the input the prediction is showcased on the UI

To a accomplish this,we have to complete all the activities listed below.

Define problem/problem Understanding

- Specify the business problem
- Bussiness requirements
- Literature Sirvey
- Social or Business impact

Data Collection & Preparation

- Collect the dataset
- Data Preparation

Exploratory Data Analysis

- Descriptive statistical
- Visual Analysis

Model Building

- Training the model in multiple algorithms
- Testing the mode

l Performance Testing & Hyperparameter Tuning

- Testing model with multiple evaluation metrics
- Comparing model accuracy before & after applying hyperparameter tuning

Model Deployment

- Save the best mode
- Integrate with Web Framework

Project Demonstration & Documentation

- Record explanation Video for project end to end solution
- Project Documentation-Step by step project development procedure .

RESULT:

Import the libraries:

```
From flask import Flask ,request,render_template
Import numpy as np
Import pandas as pd
Import pickle
Import os
```

Load the saved model. Importing the flask module in the project is mandatory. An object of Flask class is our WSGI application. Flask constructor takes the name of the current module (__name__) as argument.

```
Model = pickle.load (open ('flight.pkl',rb'))
App = Flask (_name_)
```

Render HTML page:

```
@app.route(‘/’)
Def  home();
Return rebder_template(“index.html”)
@app.route(‘/prediction’,methods =[‘POST’])
```

Here we will be using a declared constructor to route to the HTML page which we have created earlier.
In the above example, ‘/’ URL is bound with the home.html function. Hence, when the home page of the web server is opened in the browser, the html page will be rendered. Whenever you enter the values from the html page the values can be retrieved using POST Method.

Retrieves the value from UI:

```
Def predict();
Name = request.form[‘name’]
Month = request.form[‘month’]
Dayof month =request.form[‘deyofmonth’]
Dayofweek = request.form[‘dayofweek’]
Origin = request.form[‘origin’]
if(origin == “map”):
        Originl,origin2,origin3,origin4,origin5 = 0,0,0,0,1
if(origin == “dtw”):
         Originl,origin2,origin3,origin4,origin5 = 1,0,0,0,0
if(origin == “jfk”):
          Originl,origin2,origin3,origin4,origin5 =0,0,1,0,0
if(origin == “sea”):
           Originl,origin2,origin3,origin4,origin5 =0,1,0,0,0
If(origin == “alt”):
          Originl,origin2,origin3,origin4,origin5 =0,0,0,1,0
```

```
Destination = request.form[‘destination’]
if(destination == “map”):
        destination1,destination2,destination3,destination4,destinatio
n5 = 0,0,0,0,1
 if(destination == “dtw”):
        destination1,destination2,destination3,destination4,destinatio
n5 = 1,0,0,0,0
if(destination == “jfk”):
        destination1,destination2,destination3,destination4,destinatio
n5 = 0,0,1,0,0
if(destination == “sea”):
        destination1,destination2,destination3,destination4,destinatio
n5 = 0,1,0,0,0
if(destination == “alt”):
        destination1,destination2,destination3,destination4,destinatio
n5 = 0,0,0,1,0
dept = request. form[‘dept’]
arrtime = request. form[‘arrtime’]
actdept = request. form[‘actdept’]
dept15=int (dept)-int [actdept’]
total=[(name,month,dayofmonth,dayofweek,origin1,origin2,origin3
,origin4,origin5,destination1,destination2,destination3,destination4
,destination5,dept,arrtime,actdept,dept15)]
y_pred = model.predict(total)
print(y_pred)
if(y_pred==[0.]);
ans=”The  flight will be on time”
else:
ans=”The flight will be delayed”
return render_template(“index.html,showcase = ans)
```

Here the route for prediction is given and necessary steps are performed in order to get the predicted output.

Main Function:

```
if _ name_ == ‘__ main __’:
   app. run(debug = True)
```

OUTPUT:


127.0.0.1:5000
Prediction of Flight Delay
Enter the Flight Number : 1399
Month : 2
Day of Month : 4
Day of Week : 5
origin JFK
destination SEA
Scheduled Departure Time : 1
Scheduled Arrival Time : 22
Actual Departure Time : 1
SUBMIT


localhost:5000/prediction
Prediction of Flight Delay
Enter the Flight Number :
Month :
Day of Month :
Day of Week :
origin MSP
destination MSP
Scheduled Departure Time :
Scheduled Arrival Time :
Actual Departure Time :
SUBMIT
The Flight will be on time

ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

Predicting flight delays can improve airline operations and passenger satisfaction, which will result in a positive impact on the economy.

DISADVANTAGES:

Flight delays not only irritate air passengers and disrupt their schedules but also cause a decrease in efficiency. Ad increase in capital cost, reallocation of flight crews and aircraft, and additional crew expenses.

CONCLUSION:

This paper presented the need to develop a system to predict the delay in flights along with its methodology. The paper gives details about the range of different methodology that is used or can be used to find out the delay in flights. As flight delay cost a lot to the airlines as well as passangers in financial and environmental terms, flight delay is a the talk of the hour.Flight delay causes surging of prices by costing a lot on operational purpose They may increase prices to customers and operational prices to airlines. As the outcome is directly associated with the passanger and the airlines which inturn is liked to another set of airline and pasaangers it is very crucial to get real time delay for each player within the air transport system. hence there is a requirement to develop a system to predict

the delay in flights to scale back monetory loss and for the higher and smooth operation. Classification or reggrerssion ways are often accustomed determine the delay which includes Feed forward network, Neural Network, Random Forrest, decision tress, Naïve Bayes Classification Tree, Regression Tree, etc. As seen from the articles and papers these methodologies offer virtually identical accuracy however we want an algorithmic rule that is good with real world prediction and analysis and thus: naïveBayes. except being smart with real time prediction algorithmic rule that considers or assumes independence among predictors that makes the system scalable as other independent attribute may be superimposed up to the algorithmic rule for computation of the delay. the expected delay can thus facilitate the ground employees for creating correct and smooth operation plans and therefore the data if sent to the passengers will profit the airlines also because the passengers.

FUTURE SCOPE :

Further supportive study is required to correlate all the problem, scope and method for getting most accurate result. Although weather conditions are the major reasons for flight delay, other unprecedented events such as major calamities , natural or man-made can cause major delay in flight.

Reference:

[1] M. Abdel-Aty, C. Lee, Y. Bai, X. Li, and M. Michalak. Detecting periodic patterns of arrival delay. Journal of Air Transport Management, 13(6):355–361, Nov. 2007. ISSN 0969-6997. 11 preprint

[2] K. F. Abdelghany, S. S. Shah, S. Raina, and A. F. Abdelghany. A model for projecting flight delays during irregular operation conditions. Journal of Air Transport Management, 10(6):385–394, Nov. 2004. ISSN 0969-6997.

[3] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba. Analysis of the potential for delay propagation in passenger airline networks. Journal of Air Transport Management, 14(5):221–236, Sept. 2008. ISSN 0969-6997.