Application Of Machine Learning Algorithms To Predict Flight Delays

Team-42

Done By:

T.Aarthika

B.harshitha

Chanpreet kaur chawla

P.Ramya sree sai

1.INTRODUCTION

1.1OVERVIEW

The primary goal of our project is to predict airline delays caused by various factors. Flight delays lead to negative impacts, mainly economical for commuters, airline industries and airport authorities. Furthermore, in the domain of sustainability, it can even cause environmental harm by the rise in fuel consumption and gas emissions. Hence, these factors indicate how necessary and relevant it has become to predict the delays no matter the wide-range of airline meshes. To carry out the predictive analysis, which encompasses a range of statistical techniques from supervised machine learning and, that studies current and historical data to make predictions or just analyze about the future delays. Moreover, apart from the assessment related to the passengers, delay prediction analysis will also help in important decision-making procedures necessary for every pivotal player in the air transportation system.

1.2 PURPOSE

Due to the highly dynamic environments of the industry, relying only on historical datasets of flight delays may not be sufficient and applicable to forecast the future of flights. The purpose of this project is to study the flight delays from a new angle by utilising data generated from the previous records. Our primary goal is to improve the understanding of the roots and signs of flight delays as well as discovering related factors.

We adopt predictive modelling to study the factors that contribute to flight delays and to predict the flight delays in the future. The results of our work show a high correlation among the developed features.

In particular, our current prediction model achieves 82.77% in accuracy.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

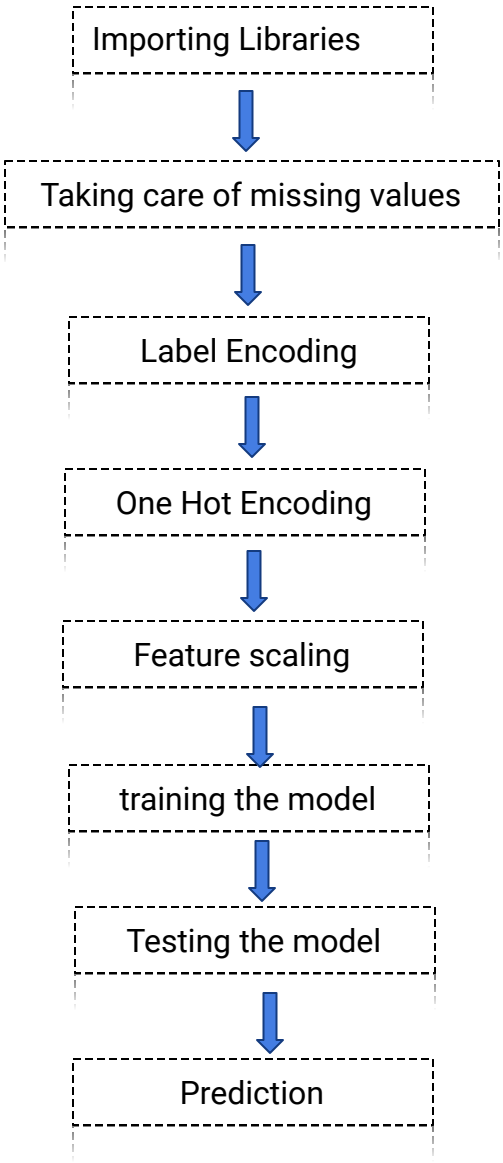
This stream of literature concerns arrival time estimation. Airlines use aircraft performance models together with parametric or physics-based trajectory models to calculate flight time. Often these tools are incapable of considering external influences such as weather and airport congestion. Even though there are more advanced methods to predict an aircraft's arrival time, airlines still heavily rely on simple models due to a lack of data availability and real time data integration . Proactive delay management is closely linked to arrival time predictions. Researchers have focused on using classification models to predict if an aircraft is delayed. Especially the detection of periodic and reoccurring delays have been investigated. Early work in the field of delay propagation was done by Peterson

2.2 PROPOSED SOLUTION

Using machine learning model, we tend predict the 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 considered to be delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifier with logistic regression for various figures of merit. Finally, we integrate it to web based application

3. THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 HARDWARE / SOFTWARE DESIGNING

This project has HTML page for user interface, where they can give the flight details they are going to travel . The web application gives the output of whether the flight lands on time or it has a delay of more than 15 minutes or less than 15 minutes. Background is build with python program where the inputs from HTML page is taken and converted to required format for inputs. The output is generated and it is showcased on HTML page.

4. EXPERIMENTAL INVESTIGATION

4.1 PROJECT IDEA

Over the last twenty years, air travel has been increasingly preferred among travellers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in the air traffic and on the ground. Increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. Our Idea mainly focuses on building a model that can predict the flight delays accurately in order to optimize the flight operations and minimize the delays.

4.2 BACKGROUND RESEARCH

These flight delays are responsible for large economic and environmental losses. 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.
These delays also contribute to growing concerns of fuel emissions and their negative impact on health, there is an active research in the aviation industry for finding techniques to predict flight delays accurately.

4.3 COMPOSING A HYPOTHESIS

The inputs to the algorithm must be :

1. Flight number
2. Date
3. Origin airport of the flight
4. Destination airport of the flight
5. Scheduled departure time
6. Actual departure time
7. Scheduled arrival time
8. Actual arrival time

The output of the algorithm must be:

1. Flights that have a delay of more than 15 minutes

4.4 DESIGNING OF THE EXPERMENT:

_First step is to collect the data then we have to analyze the data, apply suitable classification algorithm, train and test the data and build a model.
Next , then the model will be integrated to web based application.

4.5 COLLECTION OF DATA

The data for our experiment is collected from Kaggle which precisely has all the input and the output data required.

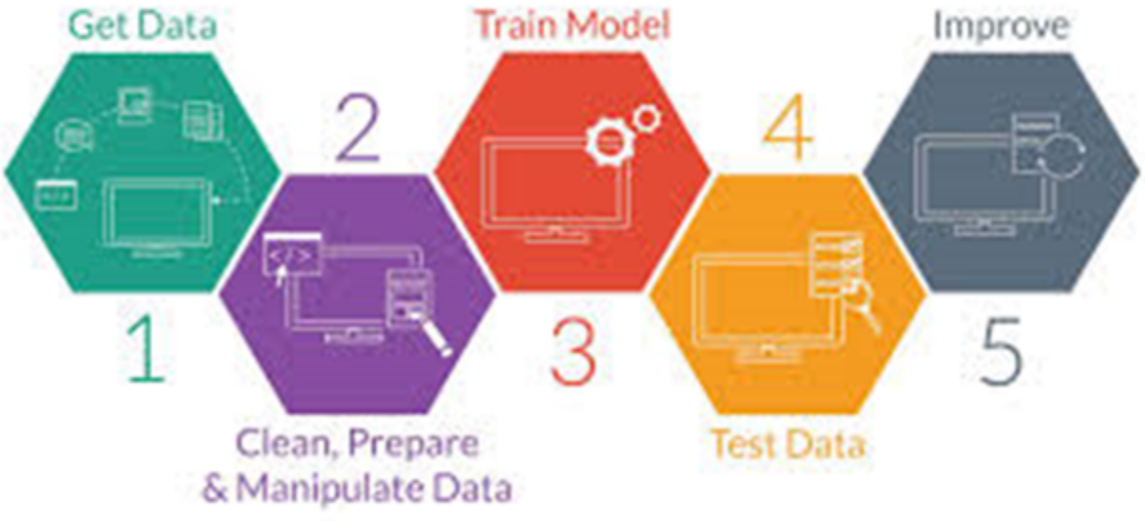
Column	Description
YEAR	Year that the flight took place
QUARTER	Quarter that the flight took place (1-4)
MONTH	Month that the flight took place (1-12)
DAY_OF_MONTH	Day of the month that the flight took place (1-31)
DAY_OF_WEEK	Day of the week that the flight took place (1=Monday, 2=Tuesday, etc.)
UNIQUE_CARRIER	Airline carrier code (e.g., DL)
TAIL_NUM	Aircraft tail number
FL_NUM	Flight number
ORIGIN_AIRPORT_ID	ID of the airport of origin
ORIGIN	Origin airport code (ATL, DFW, SEA, etc.)
DEST_AIRPORT_ID	ID of the destination airport
DEST	Destination airport code (ATL, DFW, SEA, etc.)
CRS_DEP_TIME	Scheduled departure time
DEP_TIME	Actual departure time
DEP_DELAY	Number of minutes departure was delayed
DEP_DEL15	0=Departure delayed less than 15 minutes, 1=Departure delayed 15 minutes or more
CRS_ARR_TIME	Scheduled arrival time
ARR_TIME	Actual arrival time
ARR_DELAY	Number of minutes flight arrived late
ARR_DEL15	0=Arrived less than 15 minutes late, 1=Arrived 15 minutes or more late

CANCELLED	0=Flight was not cancelled, 1=Flight was cancelled
DIVERTED	0=Flight was not diverted, 1=Flight was diverted
CRS_ELAPSED_TIME	Scheduled flight time in minutes
ACTUAL_ELAPSED_TIME	Actual flight time in minutes
DISTANCE	Distance traveled in miles

4.6 ANALYSIS OF THE DATA

Using machine learning model, we predict the flight arrival delays. We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifier with logistic regression for various figures of merit.

5. FLOWCHART



6. RESULT

The algorithm tells whether or not there is a delay in the particular flight.

7. Advantages and disadvantages.

The application predicts the flight delays and gives the output.So accordingly passengers can plan their schedules.Because these delays are responsible for large economic and environmental losses. 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 .

8. Applications

1. It is useful for the passengers and also by aviation industry
2. It also helps us to know the flights that are not reliable during the time of some emergence.

9. Conclusion

In this project we successfully applied machine learning algorithm to predict flight delays with the help of Decision tree classifier. The proposed project to predict the flight delay has a accuracy of 84.12%. This project is not only useful for passengers but is useful for aviation industry. It also portrays a negative reputation of the airlines but decrease their reliability. The project predicts flight delays based on previous available data. This project can be used as a prototype for the real time dataset. This project has also showed the importance of Regression Analysis in Machine Learning.

10. Future Scope

The future scope of this project involves application of more advanced and novel pre-processing techniques, sampling algorithms and Machine Learning Hybrid Models tuned with Grid search for achieving better model performance. This project further used to know about taxi-delay, train-delay and many more

11. Bibliography

https://www.researchgate.net/
https://www.kaggle.com/adveros/flight-delay-eda-exploratory-data-analysis/data
https://ieeexplore.ieee.org/document/8903554
https://mapr.com/webinars/predicting-flight-delays-with-apache-spark-machine-learning/

SOURCE CODE

```
from flask import Flask,render_template,request

import pickle
import numpy as np

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

app = Flask(__name__)

@app.route('/')
def home():
    return render_template("index.html")



@app.route('/prediction',methods =['POST'])
def login():
    name = request.form['name']
    month = request.form['month']
    dayofmonth = request.form['dayofmonth']
    dayofweek = request.form['dayofweek']
    origin = request.form['origin']
    if(origin == "msp"):
        origin1,origin2,origin3,origin4,orgin5 = 0,0,0,0,1
    if(origin == "dtw"):
        origin1,origin2,origin3,origin4,orgin5 = 1,0,0,0,0
    if(origin == "jfk"):
         origin1,origin2,origin3,origin4,orgin5 = 0,0,1,0,0
```

```
    if(origin == "sea"):
        origin1,origin2,origin3,origin4,orgin5 = 0,1,0,0,0
    if(origin == "alt"):
        origin1,origin2,origin3,origin4,orgin5 = 0,0,0,1,0


    destination = request.form['destination']
    if(destination == "msp"):
       destination1,destination2,destination3,destination4,destination5 = 0,0,0,0,1
    if(destination == "dtw"):
       destination1,destination2,destination3,destination4,destination5 = 1,0,0,0,0
    if(destination == "jfk"):
        destination1,destination2,destination3,destination4,destination5 = 0,0,1,0,0
    if(destination == "sea"):
        destination1,destination2,destination3,destination4,destination5 = 0,1,0,0,0
    if(destination == "atl"):
        destination1,destination2,destination3,destination4,destination5 = 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,orgin5,destination1,destination2,
destination3,destination4,destination5,int(arrtime),int(dept15)]]
    #print(total)
    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)


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