FLIGHT PRICE PREDICTION USING IBM WATSON

A UG PROJECT PHASE-1 REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Submitted by

EMME ANIL	**19UK1A05N7**
SAMALA HIMAVANTHSAI	**20UK1A0522**
SHAIK SHAREEF	**20UK1A0523**

Under the esteemed guidance of

Mr. A.ASHOK KUMAR

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
VAAGDEVI ENGINEERING COLLEGE
(Affiliated to JNTUH, Hyderabad)
Bollikunta, Warangal – 506005

2019– 2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
VAAGDEVI ENGINEERING COLLEGE
BOLLIKUNTA, WARANGAL – 506005 2019 – 2023

CERTIFICATE OF COMPLETION UG

PROJECT PHASE-1

This is to certify that the UG Project Phase-1 entitled "**FLIGHT PRICE PREDICTION USING IBM WATSON**'' is being submitted by ***E.ANIL (H.NO:19UK1A05N7), S.HIMAVANTHSAI (H.NO:20UK1A0522), S.SHAREEF (H.NO:20UK1A0523),***in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University** Hyderabad during the academic year **2022-23**, is a record of work carried out by them under the guidance and supervision.

Project Guide
Mr. A.ASHOK KUMAR
(Assistant Professor)

Head of the Department
Dr. R. Naveen Kumar
(Professor)

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr.P.PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.R.NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **Mr. A.ASHOK KUMAR** Assistant professor, Department of CSE for her constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

E.ANIL **(19UK1A05N7)**
S.HIMAVANTHSAI **(20UK1A0522)**
S.SHAREEF **(20UK1A0523)**

CONTENTS

INTRODUCTION

1.1 OVERVIEW

Any individual who has booked a flight ticket previously knows how dynamically costs change. Aircraft uses advanced strategies called Revenue Management to execute a distinctive valuing strategy [1]. The least expensive accessible ticket changes over a period the cost of a ticket might be high or low. This valuing method naturally modifies the toll as per the time like morning, afternoon or night. Cost may likewise change with the seasons like winter, summer and celebration seasons.

1.2 PURPOSE

The extreme goal of the carrier is to build its income yet on the opposite side purchaser is searching at the least expensive cost. Purchasers generally endeavor to purchase the ticket in advance to the takeoff day. Since they trust that airfare will be most likely high when the date of buying a ticket is closer to the takeoff date, yet it is not generally true. Purchaser may finish up with the paying more than they ought to for a similar seat. A report says India's affable aeronautics industry is on a high development movement. India is the third-biggest avionics showcase in 2020 and the biggest by 2030. Indian air traffic is normal to cross the quantity of 100 million travelers by 2017, whereas there were just 81 million passengers in 2015. Agreeing to Google, the expression "Cheap Air Tickets" is most sought in India. At the point when the white collar class of India is presented to air travel, buyers searching at modest costs. The rate of flight tickets at the least cost is continuously expanding.

LITERATURE SURVEY

2.1 EXISTING PROBLEM:

It is very difficult for the customer to purchase a flight ticket at the minimum price. For this several techniques are used to obtain the day at which the price of air ticket will be minimum. Most of these techniques are using sophisticated artificial intelligence(AI) research is known as Machine Learning. Utilizing AI models, [2] connected PLSR(Partial Least Square Regression) model to acquire the greatest presentation to get the least cost of aircraft ticket buying, having 75.3% precision. Janssen [3] presented a direct quantile blended relapse model to anticipate air ticket costs for cheap tickets numerous prior days takeoff. Ren, Yuan, and Yang [4], contemplated the exhibition of Linear Regression (77.06% precision), Naive Bayes (73.06% exactness, Soft max Regression (76.84% precision) and SVM (80.6% exactness) models in anticipating air ticket costs. Papadakis [5] anticipated that the cost of the ticket drop later on, by accepting the issue as a grouping issue with the assistance of Ripple Down Rule Learner (74.5 % exactness.), Logistic Regression with 69.9% precision and Linear SVM with the (69.4% exactness) Machine Learning models. Gini and Groves[2] took the Partial Least Square Regression(PLSR) for developing a model of predicting the best purchase time for flight tickets. The data was collected from major travel journey booking websites from 22 February 2011 to 23 June 2011. Additional data were also collected and are used to check the comparisons of the performances of the final model. Janssen [3] built up an expectation model utilizing the Linear Quantile Blended Regression strategy for San Francisco to New York course with existing every day airfares given by www.infare.com. The model utilized two highlights including the number of days left until the takeoff date and whether the flight date is at the end of the week or weekday.

2.2 PROPOSED SOLUTION:

The model predicts airfare well for the days that are a long way from the takeoff date, anyway for a considerable length of time close the takeoff date, the expectation isn't compelling. Wohlfarth [15] proposed a ticket buying time enhancement model dependent on an extraordinary pre-preparing step known as macked point processors and information mining systems (arrangement and bunching) and measurable investigation strategy. This system is proposed to change over heterogeneous value arrangement information into added value arrangement direction that can be bolstered to unsupervised grouping calculation. The value direction is bunched into gathering dependent on comparative estimating conduct. Advancement model gauge the value change designs. A tree based order calculation used to choose the best coordinating group and afterward comparing the advancement model.

THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM

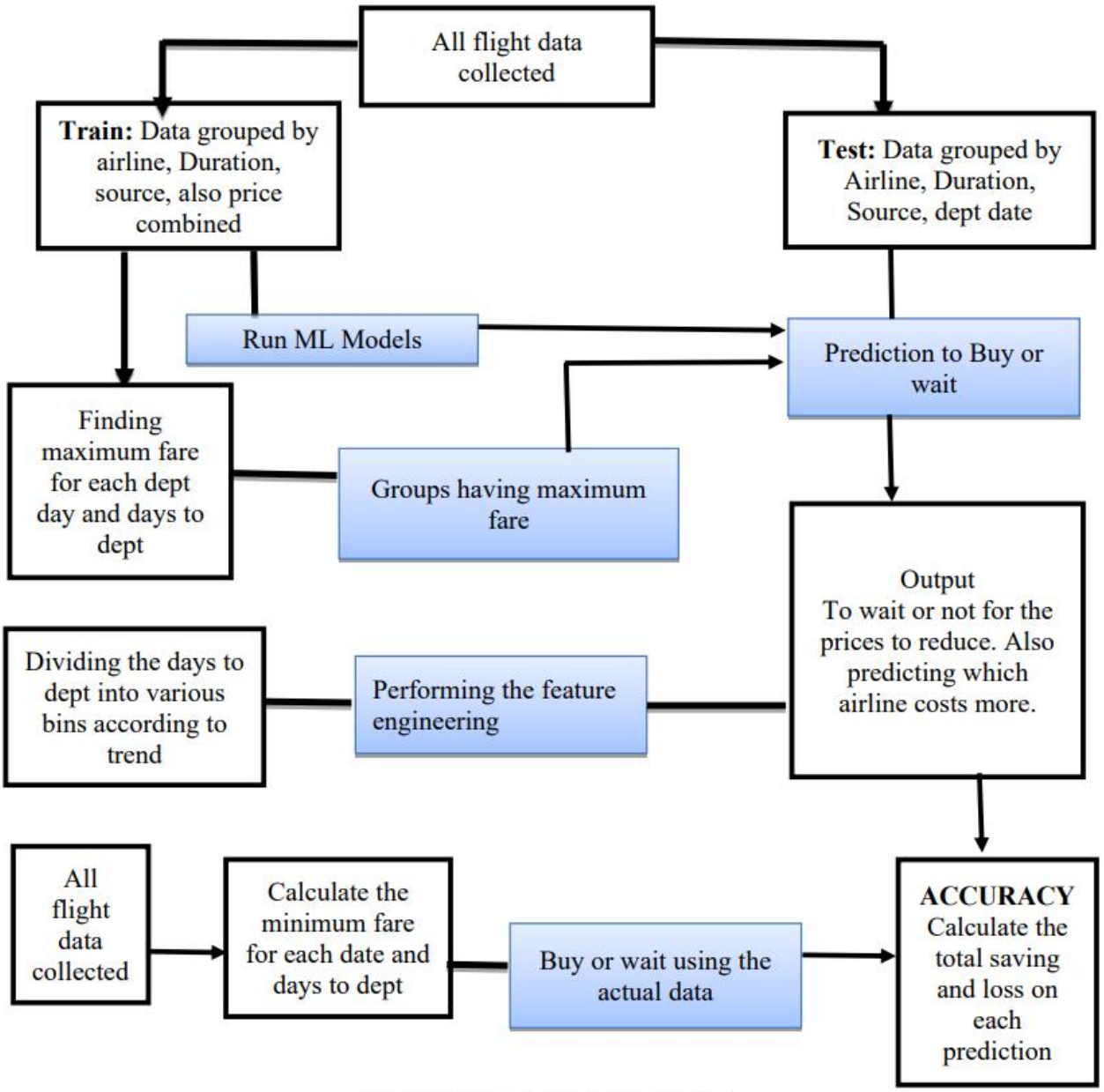


Fig. 1 Workflow for Flight Price Prediction

3.2 HARDWARE / SOFTWARE DESIGNING

The hardware required for the development of this

project is: Processor : Intel CoreTM i5-

9300H

Processor speed :

2.4GHz RAM Size :

8 GB 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,JAVASCRIPT

Programming : PYTHON

Cloud Computing Service : IBM Cloud Services

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 mainly used for data manipulation.

NumPy: This python library is used for numerical analysis.

Matplotlib and Seaborn: Both are the data visualization library used for plotting graph which will help us for understanding the data.

csr_matrix() :A dense matrix stored in a NumPy array can be converted into a sparse matrix using the CSR representation by calling the csr_matrix() function.

Train_test_split: used for splitting data arrays into training data and for testing data.

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

Reading the Dataset

For this project, we make use of single dataset(data_train). 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. Let's load a .csv data file into 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 directly read it, without any path. After the next Steps we made following bellow:

1.Datavisualization

2.Collabrative and

filtering

3.Creating the Model

4.Test and save the model

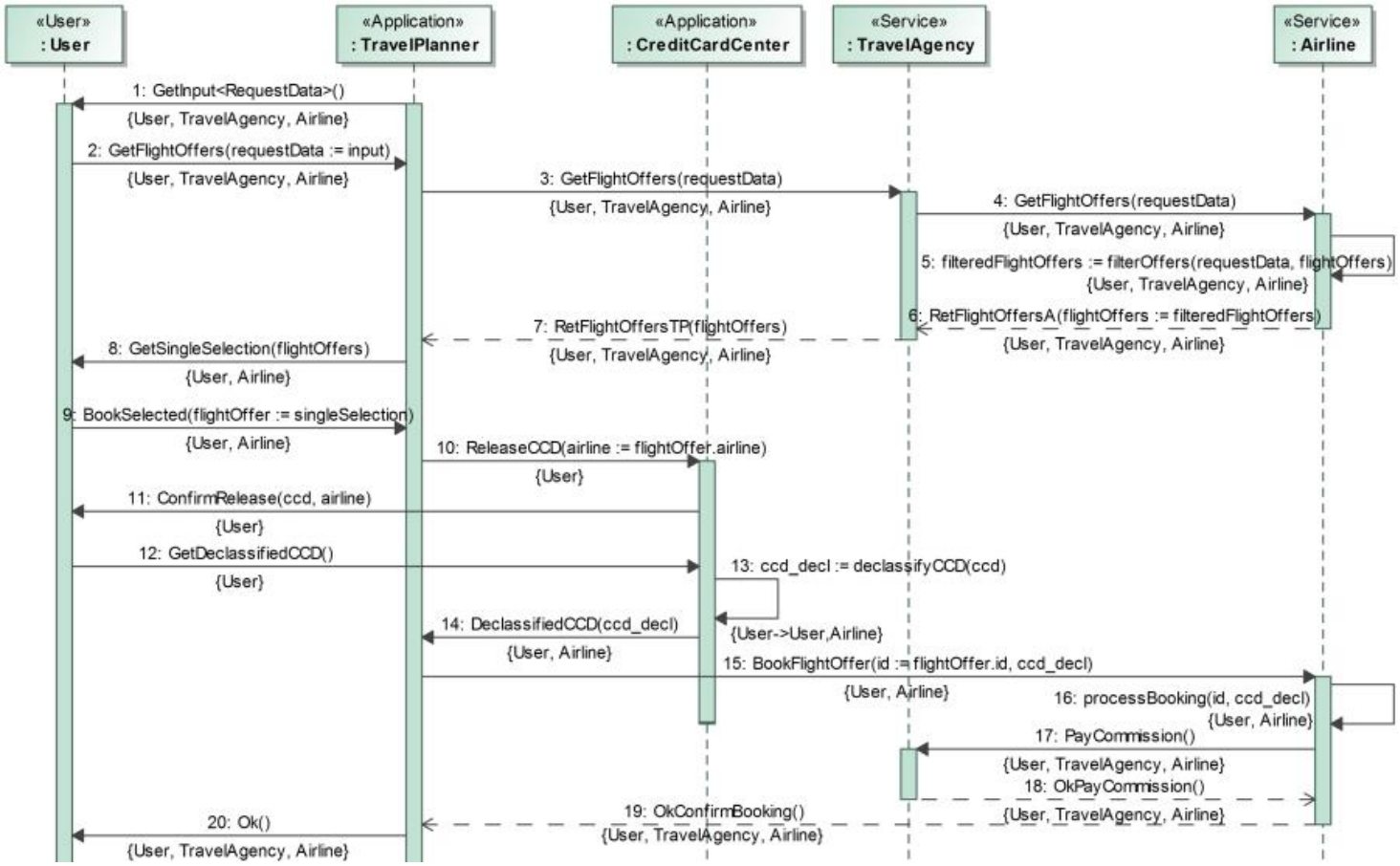
5.Build Python Code

6.Build HTML Code

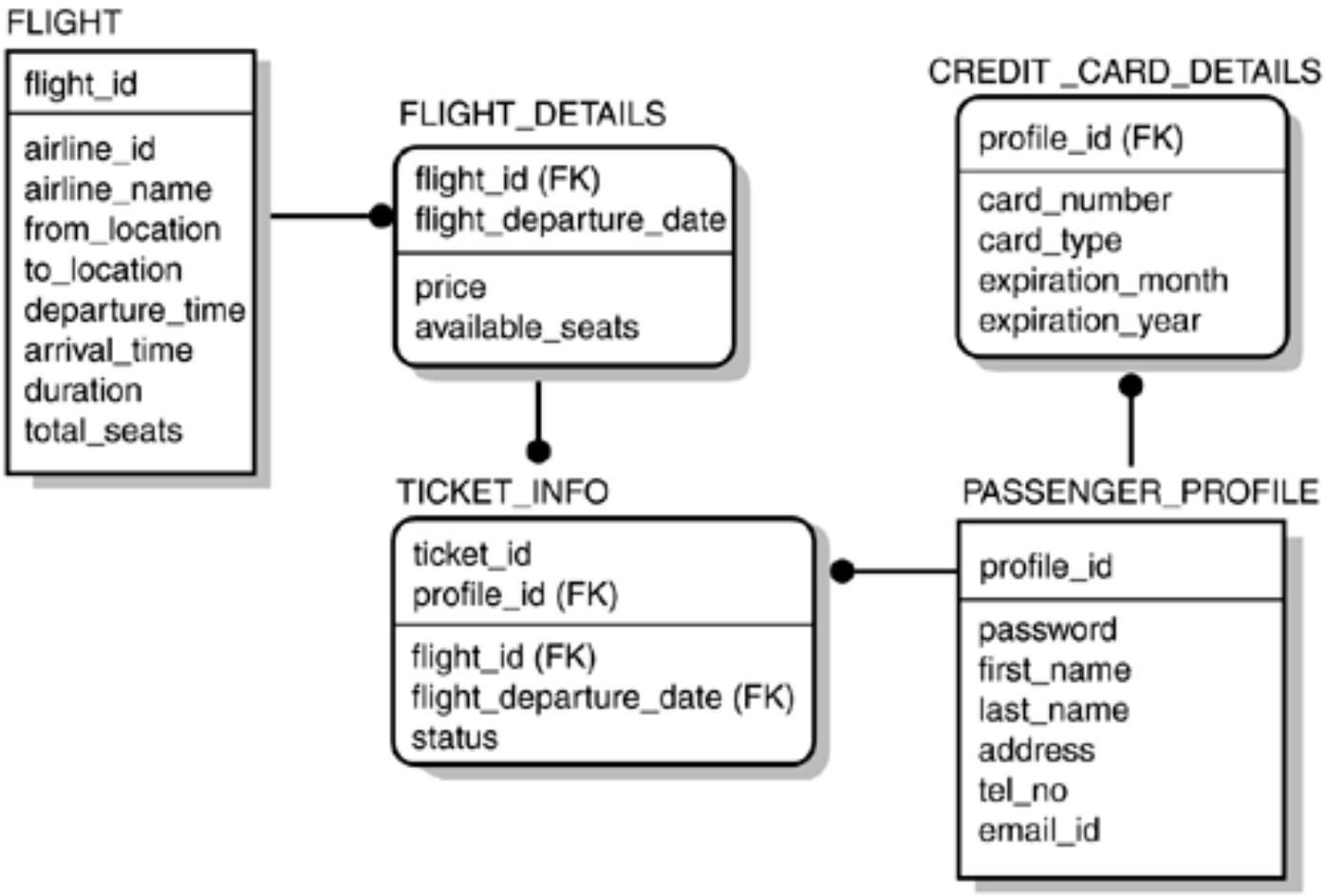
7.Run the Application

We are the following above sections we did and investigate it.

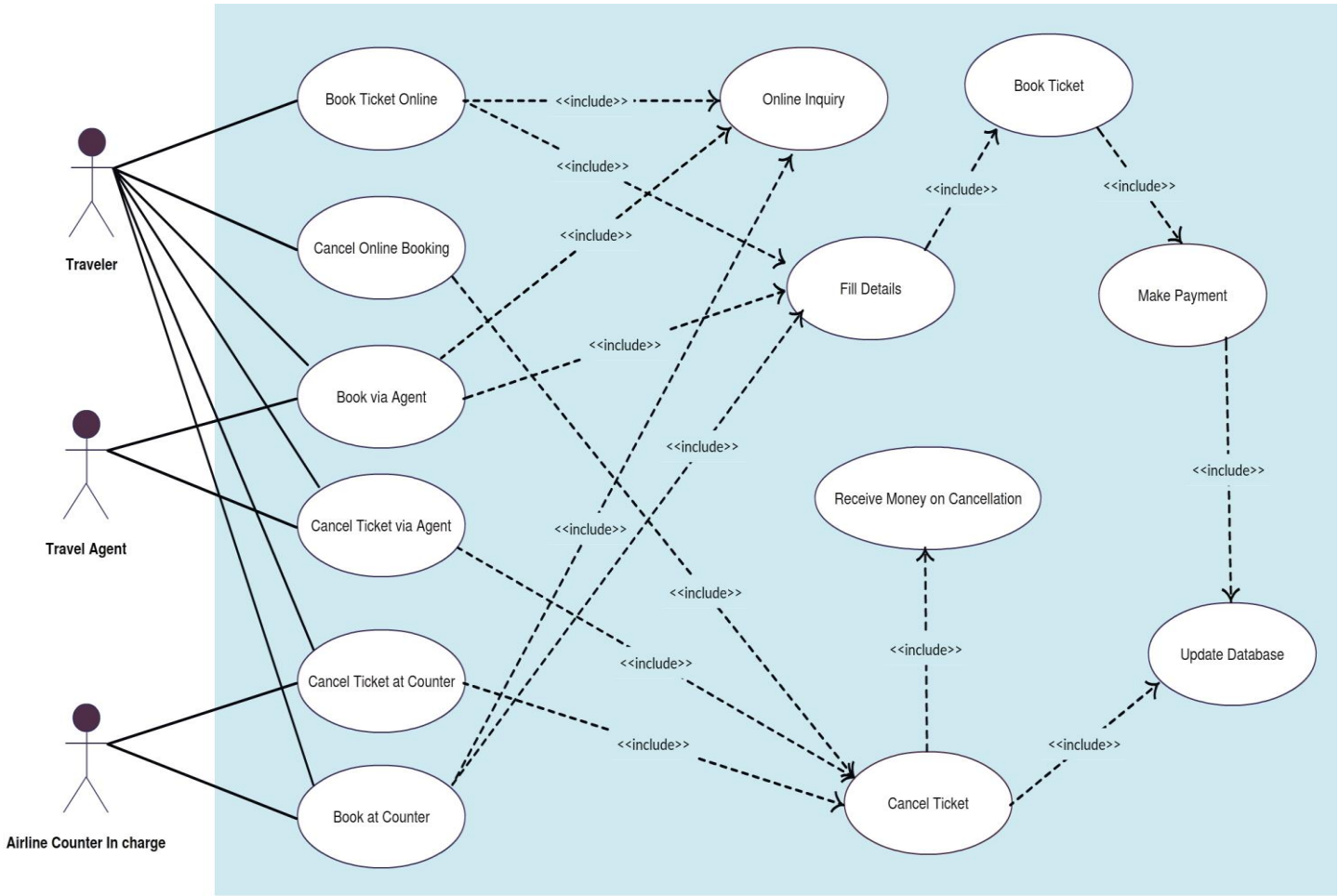
SEQUENCE DIAGRAM FOR FLIGHT PRICE



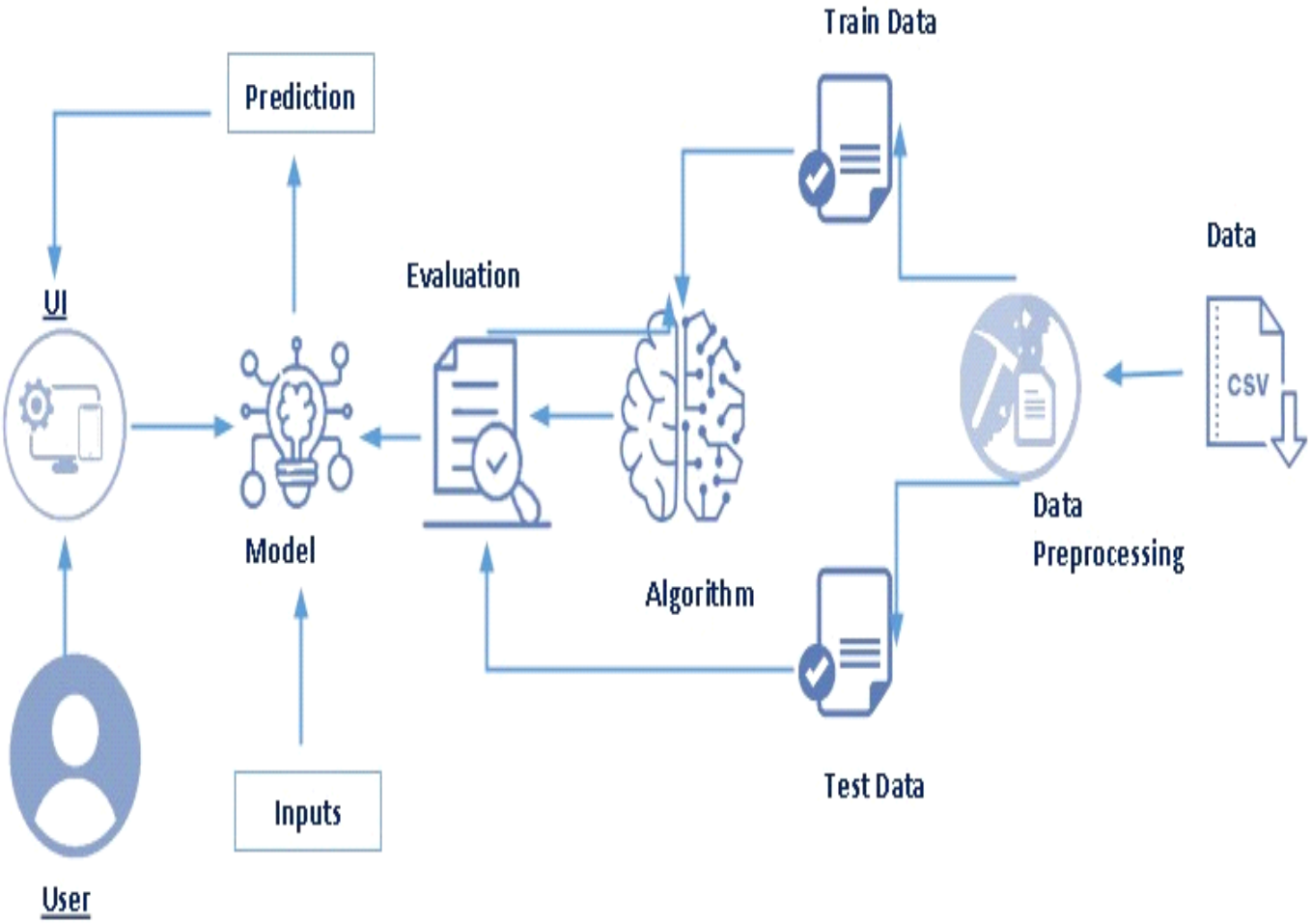
CLASS DIAGRAM FOR FLIGHT PRICE



USECASE DIAGRAM FOR FLIGHT PRICE



FLOWCHART


Train Data
Prediction
Data
Evaluation
UI
CSV
Model
Algorithm
Data
Preprocessing
Test Data
Inputs
User

Project Flow:

- User interacts with the UI (User Interface) to upload the input features.
- Uploaded features/input is analyzed by the model which is integrated
- Once model analyses the uploaded inputs, the prediction is showcased on the UI.

1. **Data Collection.**

    - Collect the dataset or Create the dataset.

2. **Data Pre- processing.**

    - **Import the Libraries:**
      Using the import keyword, you can import certain library functions of the entire library at once.

    - **Importing the dataset:**
      Upload data from external sources and combine it with data you collect via analytics.

    - **Exploratory Data Analysis:**
      It is an approaching of analysing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

    - **Data Visualization:**
      The representation of information in the form of a chart, diagram, picture, etc.

3. **Collaborating Filtering**

    - **Dataset:**
      A collection of related sets of information that is composed of separated elements but can be manipulated as a unit by a computer.

    - **Creating the Model:**
      Model involves making a representation of something.

    - **Predicting the results:**
      To say that an event or action will happen in the future, especially as a result of

knowledge or experience.

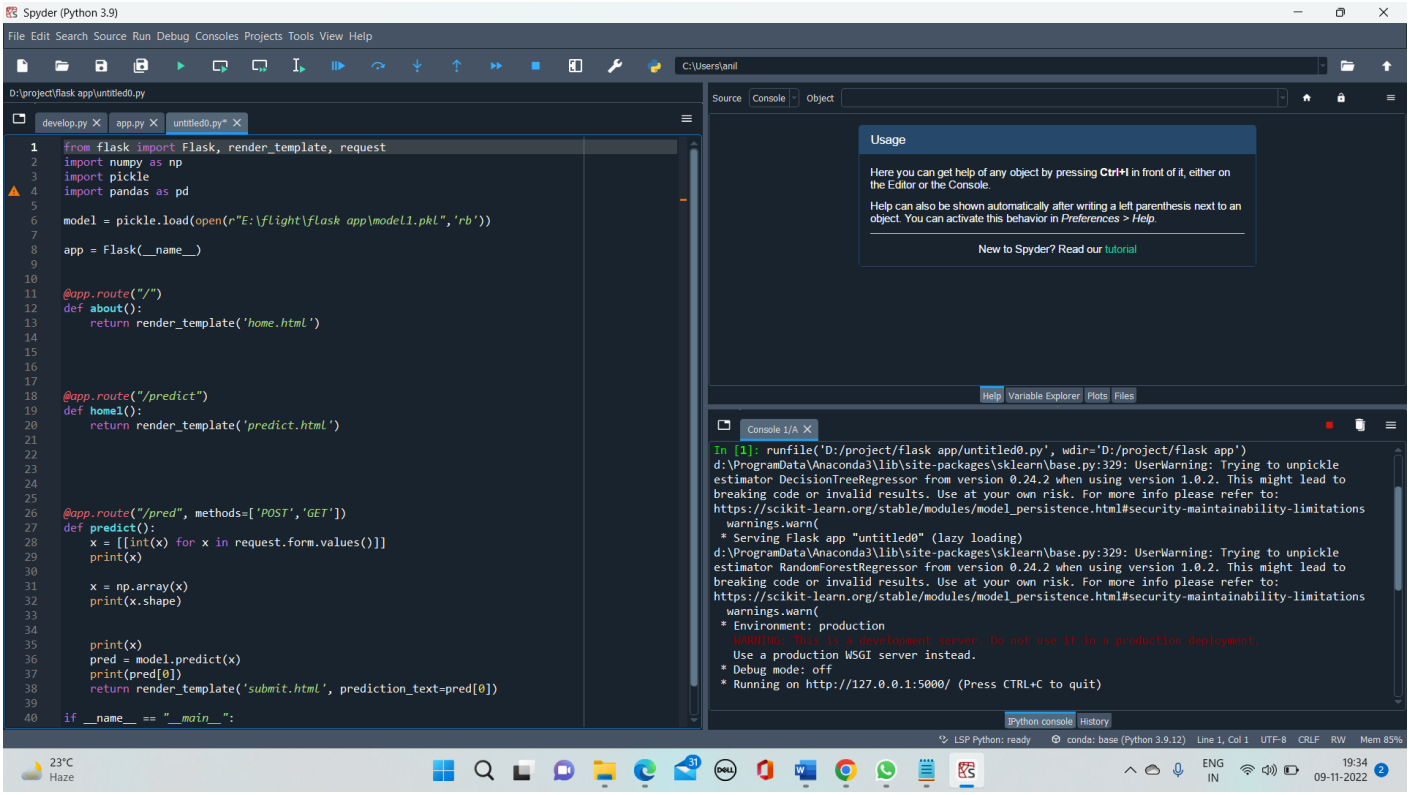
- Saving our model and dataset

4. Application Building

- **Create an HTML file:**
  Hyper Text Markup Language is the extension for web pages created for display in browsers.

- **Build a Python Code:**
  The most basic building block of any programming language is the concept of a variable, name and a place in memory that we reserve for a variable.

RESULTS:

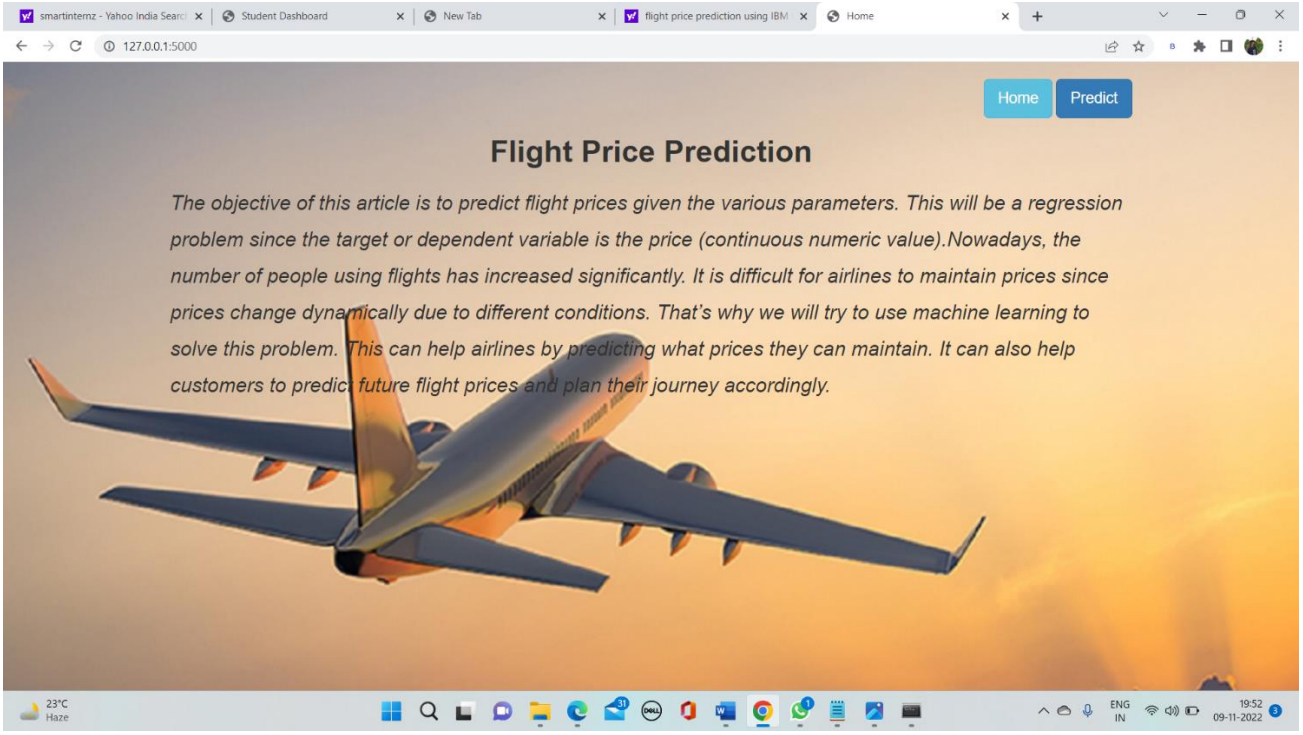

Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\anil
D:\project\flask app\untitled0.py
develop.py
app.py
untitled0.py*
from flask import Flask, render_template, request
import numpy as np
import pickle
import pandas as pd
model = pickle.load(open(r"E:\flight\flask app\model1.pkl",'rb'))
app = Flask(__name__)
@app.route("/")
def about():
return render_template('home.html')
@app.route("/predict")
def home1():
return render_template('predict.html')
@app.route("/pred", methods=['POST','GET'])
def predict():
x = [[int(x) for x in request.form.values()]]
print(x)
x = np.array(x)
print(x.shape)
print(x)
pred = model.predict(x)
print(pred[0])
return render_template('submit.html', prediction_text=pred[0])
if __name__ == "__main__":
Source
Console
Object
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.
New to Spyder? Read our tutorial
Help
Variable Explorer
Plots
Files
Console 1/A
In [1]: runfile('D:/project/flask app/untitled0.py', wdir='D:/project/flask app')
d:\ProgramData\Anaconda3\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator DecisionTreeRegressor from version 0.24.2 when using version 1.0.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/modules/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Serving Flask app "untitled0" (lazy loading)
d:\ProgramData\Anaconda3\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator RandomForestRegressor from version 0.24.2 when using version 1.0.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/modules/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Debug mode: off
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
IPython console
History
LSP Python: ready
conda: base (Python 3.9.12)
Line 1, Col 1
UTF-8
CRLF
RW
Mem 85%
23°C
Haze
ENG
IN
19:34
09-11-2022

For this project create three HTML files namely

- home.html
- predict.html
- submit.html

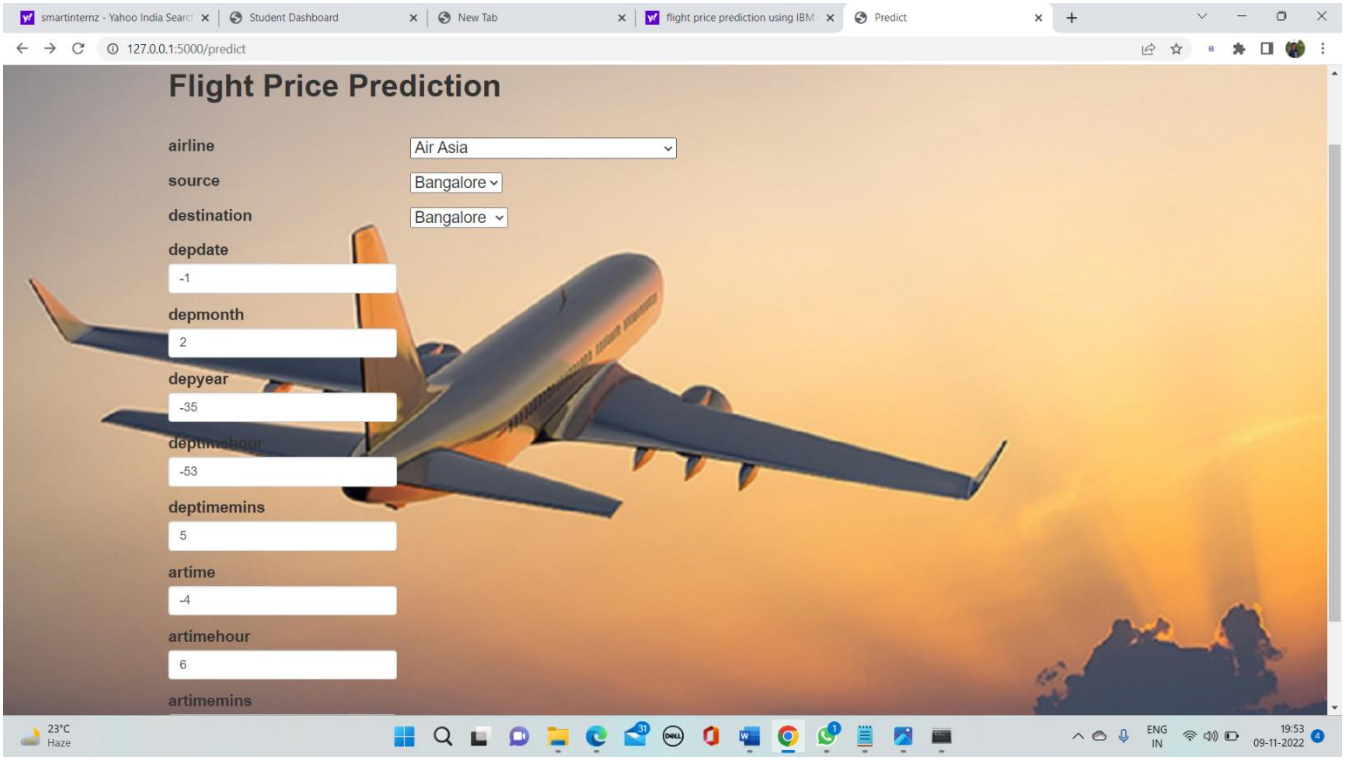
and save them in templates folder.

Let's see how our home.html page looks like



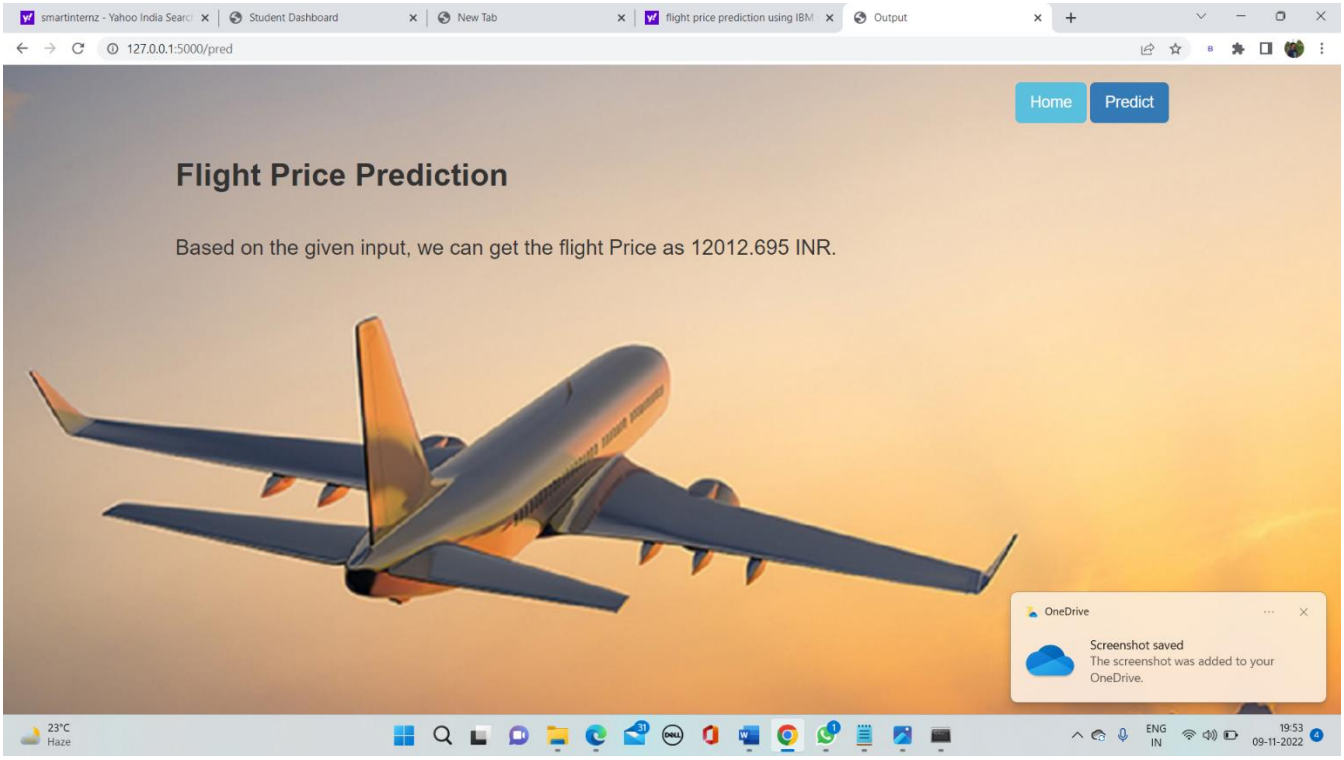
Now when you click on predict button from top right corner you will get redirected to predict.html

Lets look how our predict.html file looks like:



Now when you click on submit button from left bottom corner you will get redirected to submit.html

Lets look how our submit.html file looks like



ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- This feature to attract more visitors looking for the best rates.
- The technology to forecast rates of competitors and adjust their strategies accordingly.

DISADVANTAGES

Problems:

- The time of the flight also leads to fluctuations in the flight ticket prices.
- Depending upon the season the prices of the flight tickets also see a drastic spike.
- Travel type determines if you get a direct flight with less travel time.

APPLICATIONS

- This software has been installed into online booking platforms and mobile apps, giving potential passengers hintd and tips about when lower airfares will be released.
- The flight price prediction that could save travellers money ,time and stress.

CONCLUSION AND FUTURESCOPE

Evaluating the algorithmic rule, a dataset is collected, pre-processed, performed data modelling and studied a value difference for the number of restricted days by the passengers for travelling. Machine Learning algorithms with square measure for forecasting the accurate fare of airlines and it gives accurate value of plane price ticket at limited and highest value. Information is collected from Kaggle websites that sell the flight tickets therefore restricting data which are often accessed. The results obtained by the random forest and decision tree algorithm has better accuracy, but best accuracy is predicted by decision tree algorithm as shown is the above analysis. Accuracy of the model is also forecasted by the R-squared value. In Upcoming days when huge amount of information is accessed as in detailed information in the dataset, the expected results in future are highly correct. For further research anyone desire to expand upon it ought to request different sources of historical data or be a lot of organized in collection knowledge manually over amount of your time to boot, a lot of different combination of plane are going to be traversed. There is whole possibility that planes differ their execution ideas consisting characteristics of the plane. At last, it is curious to match our model accuracy with that of the business models accuracy offered nowadays.

BIBILOGRAPHY

[1] Bachis, E., & Piga, C. A. (2011). “Low-cost airlines and online price dispersion. International Journal of Industrial Organization”, 29(6), 655– 667. doi:10.1016/j.ijindorg.2011.02.006
[2] Groves, W. and Gini, M., 2021. “A Regression Model For Predicting Optimal Purchase Timing For Airline Tickets.”. Available at:
[3] T. Wohlfarth, S. Clemencon, F. Roueff and X. Casellato, "A Data-Mining Approach to Travel Price Forecasting," doi: 10.1109/ICMLA.2011.11.
[4] T. Wang et al., "A Framework for Airfare Price Prediction: A Machine Learning Approach," doi: 10.1109/IRI.2019.00041.
[5] V. H. Vu, Q. T. Minh and P. H. Phung, "An airfare prediction model for developing markets," doi: 10.1109/ICOIN.2018.8343221.
[6] K. Tziridis, T. Kalampokas, G. A. Papakostas and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," doi: 10.23919/EUSIPCO.2017.8081365.
[7] Y. Chen, J. Cao, S. Feng and Y. Tan, "An ensemble learning based approach for building airfare forecast service," doi: 10.1109/BigData.2015.7363846
[8] T. Liu, J. Cao, Y. Tan and Q. Xiao, "ACER: An adaptive context-aware ensemble regression model for airfare price prediction," doi: 10.1109/PIC.2017.835956

APPENDIX

PYTHON SOURCE FILE OF FLASK APP:

```
from flask import Flask,
render_template, request
import numpy as np
import pickle
import pandas as pd

model =
pickle.load(open(r"E:\flight\flask
app\model1.pkl",'rb'))

app = Flask(__name__)



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




@app.route("/predict")
def home1():
    return
render_template('predict.html')
```

```
@app.route("/pred",
methods=['POST','GET'])
def predict():
    x = [[int(x) for x in
request.form.values()]]
    print(x)

    x = np.array(x)
    print(x.shape)



    print(x)
    pred = model.predict(x)
    print(pred[0])
    return
render_template('submit.html',
prediction_text=pred[0])

if __name__ == "__main__":
    app.run(debug=False)
```