BRACT's

Vishwakarma Institute of Information Technology, Kondhwa(BK), Pune-48

Department of Computer Engineering

IT workshop (Python)

Mini Project

On

Flight Price Prediction

SY BTech Computer Engineering Academic

Year: 2022-23

Group Members:

Sr. No.	Gr. No.	Roll No.	Name of Student
1	22111231	222056	Sanghmitra Janrao
2	22220088	222072	Dhanashri Mahale
3	22220169	222075	Shruti Kamble
4	22220287	222078	Muskan Shaikh

Faculty Mentor :

Mrs. Shweta Tiwaskar

Index

Abstract

People who frequently travel through flight will have better knowledge on the best discount and right time to buy the ticket. For the business purpose many airline companies change prices according to the seasons or time duration. They will increase the price when people travel more. A lot of factors that affect the overall price of airline tickets, including the airline, the date of travel, source, destination, route, duration, and so on. Each provider seems to have its own unique set regulations and methods for determining pricing. As a result, having at least a general concept of the costs of the flights before beginning to organize the trip will undoubtedly assist a great number of people in saving both money and time. Features are taken from the chosen dataset, Utilizing three datasets for testing and training. The accumulated flight history will be used in our system to inform the development of a prediction model, this study analyses various machine learning methods for predicting flight prices. People will get an understanding from this system about the trends that prices follow, and it will also provide a forecast price value that they can refer to before purchasing flight tickets in order to save money. Customers may receive this kind of system or service if it is made available to them.

Introduction

Since the airline company's privatization, the airfare pricing scheme has evolved into a complex framework of sophisticated regulations including numerical simulations that determine airfare marketing strategies. Even though these principles are still mostly unknown, research has revealed that they are influenced by a range of circumstances. Using a machine learning model, this project intends to produce an application that will anticipate travel pricing for different flights. The user will get the expected values, and with that information, they may book their tickets appropriately. In the present marketplace, airlines attempt to control flight ticket pricing to maximize profits. There are a large number of frequent fliers who are aware of the optimal time to get inexpensive airline tickets. However, there are also a significant number of novice ticket buyers who fall victim to discount traps set by firms and wind up paying more than they should. The suggested solution may help clients save millions of rupees by providing performance measurements. Also, Passengers should be compatible with the airline companies to get adjusted for the increase and decrease of the price. Passengers or customers should make their own plans to get the best offers available on different airlines and travel through less price. Plane ticket prices change as time passes, pulling out the elements which creates the difference. Reporting the correlations and models which are used to price the flight tickets. Then, using that information, build the model which helps passengers to pull out the ticket to buy and predicting air ticket prices which progresses in the future. Duration, Arrival time, Price, Source, Destination and much more are the attributes used for flight price prediction.

Literature Survey

1. "Flight Price Prediction for Users by Machine Learning Techniques" This research paper, published by Pavithra Maria K , Anitha K L in the (**IARJSET**) International Advanced Research Journal in Science, Engineering and Technology.

   Presents the design and implementation of a Flight Price Prediction. In the preceding work on improving prediction models for airline prices by using Machine Learning (ML) techniques, the different exploration team has concentrated on various attributes and have trained the models on various kinds of Airlines. Specific trend is that they are trying to predict the price. Specifically, categorizing flight price with two divisions of elements helps the studied impact on mean price of the plane. Authors have examined the airline profit by applying pricing modes and have found that after a time duration of 70 days, categorical cases for a flight are observed as flight departure and the discount opportunities also tend to increase over time. Through the analysis we identify equal pricing techniques applied by the airline companies to positively manage the airline offers and demand to increase their business profit. Results shows airlines worry about the price changes according to the season in websites.

2. "Flight Ticket Price Predictor Using Python" This research paper, published by Vivekanand P. Thakare, Ankita Sanjay Murraya, Roshani Bandu Gawade, Mrunali Mukundrao Sawarkar, Trupti Khemraj Shende, Ujjwala Kamlesh Badole Department of Computer science & Engineering, GWCET, Nagpur, Maharashtra, India, published in the Advancement and Research in Instrumentation Engineering.

   For this project they have gathered information have gathered data from 18 routes all over India while the data of 4 routes were broadly used for the analysis due to the sheer volume of data collected over 4 months bring about 5.28 lakh data points each across the Mumbai Delhi and Delhi-Mumbai route and 1.05 lakh data points crosswise the Delhi Guwahati and Guwahati-Delhi route.

3. "Flight Fare Prediction System Using Machine Learning" This research paper, published by Mr. Parwaz Singh Sarao and Mr. Pushpendu Samanta Dept. of Computing

Technologies SRM Institute of Science and Technology Chennai, Tamil Nadu, India.sa9789@srmist.edu.in

Using a machine learning model, this project intends to produce an application that will anticipate travel pricing for different flights. The user will get the expected values, and with that information, they may book their tickets appropriately. The standard procedure for purchasing airline tickets is to do it quite a few days in advance of the departure date of the journey in order to avoid being subjected to the highest possible prices. On the other hand, there are a lot of people who have never booked tickets before, so they end up getting caught in the discount scams that businesses set up, and as a result, they wind up spending more money than they should have.

Software Requirements

Vscode:
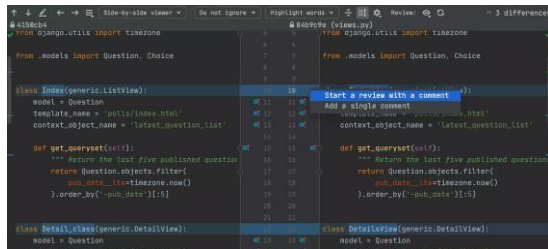
VS Code stands for Visual Studio Code. It is a free, open-source source-code editor developed by Microsoft. It is widely used by developers to write, debug and deploy code across a variety of programming languages such as Python, JavaScript, C++, and many more. VS Code offers a plethora of features like syntax highlighting, code completion, debugging, Git integration, and many more. It is available for Windows, macOS, and Linux operating systems.

Front End : Python

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation via the off-side rule. Python is dynamically typed and garbage-collected.

Python is often used as a support language for software developers, for build control and management, testing, and in many other ways. SCons for build control. Buildbot and Apache Gump for automated continuous compilation and testing. Roundup or Trac for bug tracking and project management.

Python is widely considered among the easiest programming languages for beginners to learn. If you're interested in learning a programming language, Python is a good place to start.

Methodology

First we collected all the flight data through the dataset from kaggel.com which is the source for this project.

The present study is divided into four distinct phases:
1) The first phase involves identifying the specific features of a flight that have an impact on air ticket prices. This step is important as it defines the problem that the study aims to address.
2) The second phase involves gathering enough data from various airlines to train and test the machine learning models that will be created.
3) In the third phase, the machine learning models that will be compared are chosen.
4) Finally, in the fourth phase, the machine learning models are evaluated.

Each of these steps will be discussed in more detail in the following :

Step-1 (Feature Selection) - In this step the most informative aspects of a flight that impact the pricing of the air tickets are chosen. This step is highly significant as it specifies the issue under solution.

For every journey the following characteristics were considered:
- Departure Date
- Arrival Date
- Source
- Destination
- Number of Stops
- Airlines

Step-2 (Machine learning Model Selection) - various ml models were chosen for the present investigation and subjected to the same dataset of flights.

The Machine learning technique used in this project is: ●
Linear Regression (LR).

Dataset Overview

To begin, we need information on aircraft business and mass transit in order to develop the airline ticket pricing model just at market level. Data on different websites is gathered with unique attributes to provide the best accuracy. The data is collected from website kaggle.com and the models are implemented using python.

As a result, we have one dataset: "Flight Price Prediction" within these dataset there are three different excel files :

I. Data_Train

II. Test_set

III. Sample_Submission

The Data_Train dataset contains 10,684 items with parameters such as Airlines, Date of journey, Source, Destination, Route, Time of Departure, Estimated Time of Arrival, Duration, Total Stops, Additional Information, and price.

The Testing dataset contains 2,672 items with the following attributes: Airlines, Date of journey, Source, Destination, Route, Time of Departure, Estimated Time of Arrival, Duration, Total Stops, Additional Information

The Sample_Submission dataset contains 2,672 items with the attribute as Price.

Dataset features description table :

Feature Name	Description
Airlines	As a result, this article will include all sorts of airlines such as Indigo, Jet Airways, Air India, among others.
Date of Journey	This column will inform us of the date the passenger's travel will begin.
Source	This column includes the names of the location from which the guest's journey will begin.
Destination	This column contains the name of the location where the passenger's journey will begin.
Route	This column includes the names of the location from where the customer's journey would begin.
Departure Time	The duration of a flight is the amount of time it takes to go from point A to point B.
Arrival Time	It will indicate how many spots the flight will stop over its journey.
Duration	The flight's endurance in hours.
Total Stops	The total number of breaks in the Voyage.
Additional Information	It will indicate whether a meal is included with the journey or not.

Experimental Setup

1. Data collection: Collected historical data on flight prices, including departure and arrival cities, dates, times, airlines, and other relevant information. Used flight fare dataset for this purpose.

2. Data preprocessing: Many characteristics in the datasets have the same data. Furthermore, the statistics given by airlines can contain inaccurate figures due to human mistake, payment processing error, and so on. As a result, a well-designed data pre-treatment pipeline is critical for generating reliable input data for the machine learning algorithms. We discovered that the variables 'Route' and 'Total Stops' contain extremely few missing values in the data. We have one date form variable named 'Date of Journey,' as well as time variables called 'Dep Time' and 'Arrival Time.' The 'Journey Day' and 'Journey Month' variables may be extracted from the 'Date of Journey' field. 'Voyage day' indicates the month in which the journey began. Similarly, we may extract 'Departure Hour' as well as 'Departure Minute' from the 'Dep Time' as well as 'Arrival Time' variables, as well as 'Arrival Hour' and 'Arrival Minute' from the 'Arrival Time' variables. This 'Duration' field also contains duration information. This variable combines time hours & minutes data. 'Duration hours' as well as 'Duration minutes' can be extracted individually from the 'Duration' variable.

3. Model selection: Choose an appropriate machine learning algorithm that can predict flight prices accurately. Commonly used algorithms for regression problems include linear regression, decision trees, random forests, and neural networks.

4. Feature Engineering: Here, we divide the characteristics and names first, then convert the hours to minutes. We are arranging the form of such a date of travel in our information in Date of Journey for easier preprocessing inside the model phase. Dep Time converts the departing date into clock time. Arrival Time is translated to minutes and hours by Arrival Time.

5. Feature Selection: To enhance model performance, a features extraction approach is used to investigate the level of influence of each information on the prediction outcome. The price section has been removed since it is no longer useful.

6. Modeling: We will now fit the image and forecast the results. We fit given data into different regression analysis in order to compare their efficiency and choose the right model.

Results & Discussion

Future Scope

- More routes can be added and the same analysis can be expanded to major airports and travel routes in India.
- The analysis can be done by increasing the data points and increasing the historical data used. That will train the model better giving better accuracies and more savings.
- More rules can be added in the Rule based learning based on our understanding of the industry, also incorporating the offer periods given by the airlines.
- Developing a more user friendly interface for various routes giving more flexibility to the users.
- This framework may be expanded in the future to also include airline tickets payment details, that can offer more detail about each area, such as timestamp of entry and exit, seat placement, covered auxiliary items, and so on.

<u>Conclusion</u>

In conclusion the "Flight Price Prediction" is built by using python. In the dataset, Machine Learning model that is Linear Regression is employed to forecast the dynamic flight pricing. This provides the expected airfare values to get a flight ticket at the lowest possible price. We used training data to train the training data and test data to test it. These records were used to extract a number of characteristics. Our suggested model can estimate the quarterly average flight price using attribute selection strategies.To the highest possible standard, much prior studies into flight price prediction using the large dataset depended on standard statistical approaches, which have their own limitations in terms of underlying issue estimates and hypotheses.

References

[1]"*Flight Price Prediction for Users by Machine Learning Techniques*" This research paper, published by Pavithra Maria K , Anitha K L published in the International Advanced Research Journal in Science, Engineering and Technology (**IARJSET**).

[2]"*Flight Ticket Price Predictor Using Python*" This research paper, published by Vivekanand P. Thakare, Ankita Sanjay Murraya, Roshani Bandu Gawade, Mrunali Mukundrao Sawarkar, Trupti Khemraj Shende, Ujjwala Kamlesh Badole published in the Advancement and Research in Instrumentation Engineering.

[3]"*Flight Price Prediction*" This research paper, published by Prithviraj Biswas, Rohan Chakraborty, Tathagata Mallik, Rohan Chakraborty, Sk Imran Uddin, Shreya Saha, Pallabi Das, Sourish Mitra Computer Science & Engineering, Guru Nanak Institute of Technology, Kolkata, published in the International Journal for Research in Applied Science & Engineering Technology (**IJRASET).**

[4]"*Flight Fare Prediction System Using Machine Learning*" This research paper, published by Mr. Parwaz Singh Sarao, Mr. Pushpendu Samanta Dept. of Computing Technologies SRM Institute of Science and Technology.

[5]"*Airfare price prediction using machine learning techniques"* This research paper, published by K. Tziridis T. Kalampokas G.Papakostas and K. Diamantaras in the European Signal Processing Conference (EUSIPCO).