Detail Project Report
Flight Price Prediction

Revision Number : 1.2
Last Date of Revision : 28/06/2023

Document Version Control

Date	Version	Description	Author
17/06/2023	1.0	Abstract Introduction General Description	ANISH MAURYA
21/06/2023	1.1	Technical Requirements Data Requirements Data Pre-Processing Design Flow	ANISH MAURYA
28/06/2023	1.2	Data from user and its validation Rendering the results Deployment Conclusion	ANISH MAURYA

Contents

iNeuron

Abstract

Flight price prediction is a challenging task that involves forecasting the prices of airline tickets based on various factors such as historical data, seasonality, demand, and other external factors. Accurate price prediction can be beneficial for travelers, airlines, and travel agencies in making informed decisions related to ticket purchasing, revenue management, and pricing strategies.

This project aims to develop a flight price prediction model using machine learning techniques. The dataset used for training and testing the model consists of historical flight data, including features such as departure and arrival locations, date and time of travel, airline carrier, and other relevant information. The dataset is preprocessed to handle missing values, outliers, and feature engineering to extract useful insights.

Different machine learning algorithms such as regression, time series analysis, and ensemble methods are implemented and evaluated to identify the most suitable approach for flight price prediction. The models are trained on a subset of the dataset and evaluated using appropriate evaluation metrics such as Accuracy Score

In addition to traditional machine learning techniques, advanced methods like deep learning and neural networks may also be explored to capture complex patterns and dependencies in the data. Transfer learning and feature selection techniques can be employed to enhance the model's performance and reduce computational complexity.

The trained flight price prediction model can be integrated into a user-friendly application or website where travelers can input their desired travel details and receive accurate price estimates. The model's performance can be continuously monitored and refined using additional data to improve its accuracy over time.

Overall, this project aims to contribute to the field of travel and tourism by developing a reliable and accurate flight price prediction model that can assist travelers and industry professionals in making informed decisions related to flight bookings and pricing strategies.

1. Introduction

1.1 Why this DPR Document?

The main purpose of this DPR documentation is to add the necessary details of the project and provide the description of the machine learning model and the written code. This also provides the detailed description on how the entire project has been designed end to end.

Keypoints:

- Describes the design flow
- Implementations

- Software requirements
- Architecture of the project
- Non-functional attributes like
- Reusability
- Portability
- Resource Utilization

2. General Description

2.1 Problem Perspective

The Flight Price Prediction is a machine learning model that helps companies to predict how many fare of between to city based on some input data

2.2 Problem Statement

The main goal of this model is to predict flight price based on the input features

2.3 Proposed Solution

To solve the problem, we have created a user interface for taking the input from the user To predict flight price using our trained ML Model after processing the input and at last the predicted value from the model is communicated to the user.

3. Technical Requirements

As technical requirements, we don't need any specialized hardware for visualization of the application. The user should have a device that has the access to the web and the fundamental understanding of providing the input. And for the backend, we need a server to run all the required packages to process the input and predict the desired output

3.1 Tools Used

Python programming language and frameworks such as numpy, pandas, scikit-learn, Vs Code and Azure are used to build the whole model.

- VS Code is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn, and Plotly are used.
- Azure is used for deployment of the model
- Front end development is done using Azure
- Python is used for backend development
- Github is used as version control system

4. Data Requirements

The Data requirements totally supportted the matter statement and also the data is available on the Kaggle in (Excel) file format.

4.1 Data Collection

The data for this project is collected from the Kaggle , the URL for the same is https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

4.2 Data Description

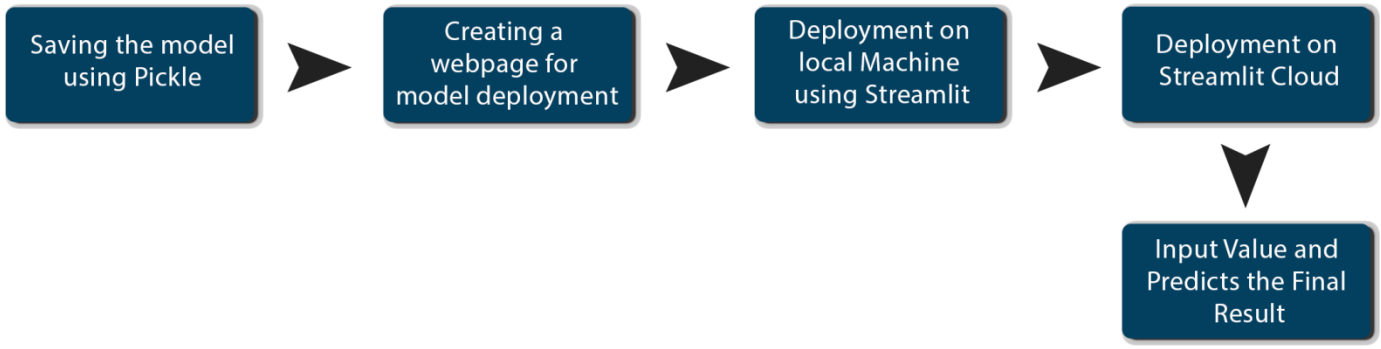
The flight price dataset used in this project contains historical data related to airline flights and their corresponding prices. The dataset includes various features that are used for predicting flight prices. Here is a description of the common features found in a flight price dataset. Row 10683 and Columns 11 columns name is Airline name, Date of journey, Source, Destination, Stops, Dep_time , Arrival_time, Price.

4.3 Data Pre-processsing

- Checked the datatype of features in dataset using df.info()
- Checked for Null values, because the null values can affect the accuracy of the model.
- Perform Label Encoding for the features that have categorical data.
- Checked the distribution of the features to interpret its importance.
  Now, the data is prepared to train a machine learning model.

5. Design Flow

5.1 Deployment Process



5.2 Data Description

In Logging, at each time an error or an exception occurs, the event is logged into the system log file with reason and timestamp. This helps the developer to debug the system bugs and rectify the error.

5.3 Data from user

The data from the user is retrieved from the created Azure web app.

5.4 Data Validation

- The data provided by the user is then being processed by appliction.py file and validated.
- The validated data is then sent to the prepared model for the prediction.

5.5 Rendering the results

The data sent for the prediction is then rendered to the web page.

6. Deployment

The tested model is then deployed to Azure. So, users can access the project from any internet device.

7. Conclusion

Flight price prediction is a complex task that requires the integration of various factors and machine learning techniques. In this project, we developed a flight price prediction model using historical flight data and machine learning algorithms. The dataset was preprocessed and features were engineered to extract useful insights.

By developing a reliable and accurate flight price prediction model, this project contributes to the field of travel and tourism. It empowers travelers and industry professionals with valuable insights for making informed decisions regarding flight bookings and pricing strategies.

8. Frequently Asked Questions (FAQs)

Q1) What's the source of data?

The data for training is accessible to all at Kaggle , the URL for the same is https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

Q2) What are the type of the data ?

The data was the combination of numerical and categorical values.

Q3) What's the complete flow you followed in this Project ?

Refer Page no 6 for better understanding

Q4) After the file validation what you do with incompatible file or files which didn't pass the validation ?

Files like these are moved to the Achieve Folder and a list of these files has been shared with the client and we removed the bad data folder.

Q5) How logs are managed ?

We are using different logs as per the steps that we follow in validation and modelling like Data validation log, Data Insertion, Model Training log, Prediction log etc.

Q6) What techniques were you using for data pre-processing?

- Removing unwanted attributes.
- Visualizing relation of independent variables with each other and output variable
- Checking and changing distribution of continuous values.
- Removing Outliers
- Cleaning data and imputing if null values are present.
- Converting categorical data into numeric values using one hot encode

Q7) How training was done or what models were used ?

- Before dividing the data in training and validation set, we performed pre-processing over the data set and made the final dataset.
- As per the dataset training and validation data were divided.
- Algorithms like Random Forest Regression were used based on the recall, final model was used on the dataset and we saved that model.

Q8) How prediction was done?

The testing files are shared by the client. We performed the same life cycle on the provided dataset. Then, on the basis of dataset, model is loaded and prediction is performed . In the end we get the accumulated data of predictions.

Q9) What are the different stages of deployment ?

- First, the scripts are stored on Github as a storage interface.
- The model is first tested in the local environment.
- After successful testing, it is deployed on Azure cloud.