DETAILED PROJECT REPORT (DPR)

FLIGHT FARE PREDICTION

Last Date of Revision – 20/12/2022

Kalpagam . S

Contents

ABSTRACT

In the airline industry, Airlines love selling tickets at the highest price possible — while still not losing consumers to competitors. Passengers are crazy about buying flights at the lowest cost available — while not missing the chance to get on board. All this makes flight prices fluctuant and hard to predict. Therefore, having some basic idea of the flight fares before planning the trip will surely help many people save money and time. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. This system will give people the idea about the trends that prices follow and also provide a predicted price value which they can refer to before booking their flight tickets to save money.

1. INTRODUCTION

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.

1.1 Problem Statement

The objective of this project is to predict flight prices given the various parameters. Data used in this project is publicly available at Kaggle.
Dataset link:

https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh

This will be a regression problem since the target or dependent variable is the price (continuous numeric value).

1.2 Problem Solution

This project aims to develop an application which will predict the flight prices for various flights using a machine learning model. The user will get the predicted values and with its reference the user can decide to book their tickets

accordingly. In the current day scenario flight companies try to manipulate the flight ticket prices to maximize their profits. There are many people who travel regularly through flights. But there are also many people who are inexperienced in booking tickets and end up falling into discount traps made by the companies where actually they end up spending more than they should have. The proposed system can help save millions of rupees of customers by providing them the information to book tickets at the right time.

2. TECHNICAL REQUIREMENTS TO RUN THIS APP

There are not any hardware needs needed for victimization of this application, the user should have an interactive device that has access to the web and should have the fundamental understanding of providing the input.

3. STEPS INVOLVED IN CREATING THIS PROJECT

3.1 Importing Python libraries and loading the dataset

The necessary python libraries such as
- Pandas
- Numpy
- Matplotlib
- Seaborn
- Scikit-learn

are imported and the dataset in the form of CSV file is imported.

3.2 Data Inspection

In this step the size and shape of the dataset is inspected. The appropriate data types for each column in the data frame and summary is checked for numerical features.

3.3 Exploratory Data Analysis (EDA)

Visualizations are done using Matplotlib and Seaborn libraries to get some insights about the dataset.

3.4 Data pre-processing

This is done to transform raw data into features that can be used for creating a predictive model using Machine learning.
Steps involved in pre-processing are:
- Converting the string data type into the desired data type.
- Extracting the Date Column separately as Day, Month, Year.
- Hours and minutes are extracted from Departure Time.
- Checking for null values and filling them.
- Performing Label Encoding to convert categorical values into numerical values.

3.5 Feature Selection

This step helps in finding out the best feature which will contribute and have good relation with the target variable (Price). Heat map is used here to find correlation between features.

3.6 Building ML models

After selecting the features which are more correlated to price the next step involves applying machine algorithms and creating a model. As our dataset consists of labeled data, we use Supervised machine learning algorithms. We will apply Regression Algorithms to solve this use case as our dataset contains continuous values in the target column. Regression models are used to describe relationships between dependent and independent variables. The machine learning algorithms used in this project are:

- K neighbors regressor
- Decision Tree regressor
- Random Forest Regressor

3.7 Performance Metrics

Performance metrics are statistical models which will be used to compare the accuracy of the machine learning models trained by different algorithms. The sklearn.metrics module will be used to implement the functions to measure the

errors from each model using the regression metrics. Following metrics will be used to check the error measure of each model.

- R2_score
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)

After training and building the models, accuracies are checked for the 3 models and the model with maximum accuracy is selected. The **Random Forest Regressor model** turned out to be the most accurate one out of the 3 models.

3.8 Hyperparameter tuning

The accuracy is improved by doing Hyperparameter tuning. Random Search method is used to find the best parameters. After hypertuning, the accuracy is found to increase.

3.9 Saving the model to a pickle file

The model is then saved to a pickle file for future use.

3.10 Creating flask app

Now a web application is designed where the user can input all the attribute values and the data will be given to the model, based on the training given to the model, the model will predict the flight ticket fares.

The back-end is created using Flask Framework where API end-points such as GET and POST will be created to perform operations related to fetching and displaying data on the front-end of the application.

3.11 Creating front end using HTML and CSS.

The front-end of the application will be created using HTML, CSS, Bootstrap framework where users will have the functionality of entering their flight data.

3.12 UI Integration

The data from the user is retrieved from the created HTML web page.
This data will be sent to the back-end service where the model will predict the output according to the provided data. The predicted value is rendered to the front-end and displayed.

3.13 App Deployment on Render

The project is then deployed on Render which is a Cloud platform.

Link to access the app: https://test3-lr78.onrender.com

4. TOOLS USED

- Python is used as the programming language and frameworks like pandas, scikit-learn for building the model.
- Jupyter notebook and Visual Studio Code is used as IDE.
- Seaborn, Matplotlib are used for visualization.
- HTML, CSS are used for front end development.
- Flask is used as a backend.
- Render is used for Deployment.

5. DESIGN FLOW

This project implements the machine learning life cycle to create a basic web application which will predict the flight prices by applying machine learning

algorithms to historical flight data. The below diagram shows the steps and design process followed in implementing the model.
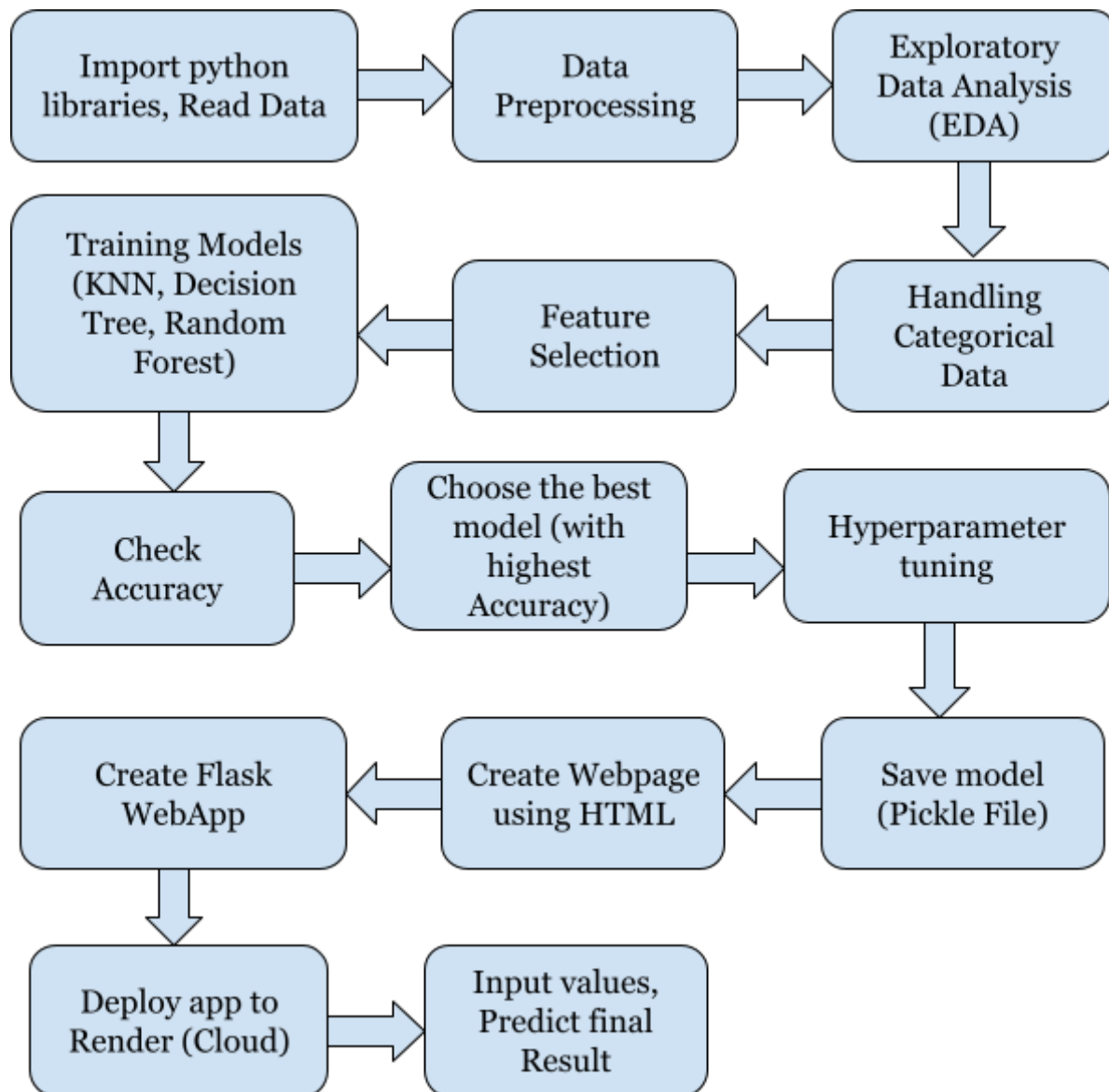
6. Conclusion

Flight Prediction Service in the aviation industry can help customers in booking tickets. This system will give an idea about the price trends and also provide a predicted price value which travelers can refer to before booking their flight tickets to save money.