

Flight Fare Prediction DPR

INDEX

1. ABSTRACT
2. INTRODUCTION
3. DESCRIPTION
 - 3.1 Problem perspective
 - 3.2 Problem statement
 - 3.3 Problem solution
 - 3.4 Improvements
4. PROJECT REQUIREMENTS
 - 4.1 Tools used
 - 4.2 Hardware
5. Data requirements
 - 5.1 Data gathering
 - 5.2 Data description
 - 5.3 Import data into database
 - 5.4 Export data from database
6. Data Pre-processing
7. Design
 - 7.1 Modelling
 - 7.2 UI integration
 - 7.3 Modelling process
 - 7.4 Deployment process
8. Data from User
9. Data Validation
10. Results
11. Deployment
- Conclusion

1. Abstract

In the present day contemporary scenario, Airlines have become popular and affordable domestically at least, and these flight fare is a factor people consider to travel. All these fares aspect varies according to season, weather condition, availability of staff and so on. So here we are going to build ML model that predicts the price of an Airline which would be useful prior to bookings.

2. INTRODUCTION

Basically, the main of this detailed project report is to add necessary details of the project and provide information about the ML model and the code base.

Key points include:

- Describe the design flow
- Implementation
- Software requirements
- Project architecture
- Non functional attributes

3. Description

- Problem Perspective
 - The Flight Fare prediction is a ML project which helps to predict the flight prices for a given particular date.
- Problem Statement
 - We read the dataset and come to an assumption about the given dependent variable i.e, nothing but the prices of flight.
- Proposed Solution
 - The ML model will be capable of predicting the price of flight when certain parameters like date, boarding place, destination are provided.

- Improvements

We can further improve or enhance the model by hyper parameter tuning it and with some additional features like most preferred flights by people and recommending it .

4. Technical Requirements

Currently, all that is needed for this project is a set of libraries and an IDE for development. High end Hardware isn't required that much.

Tools used:

- Python 3.9.12(Conda env) for programming
- Pandas, sklearn, numpy for ML model.
- Jupyter and VS code as IDE
- HTML and CSS as frontend UI.
- Flask for app development
- GitHub for version control

Hardware used:

- CPU – AMD Ryzen 7
- GPU – AMD raedon
- RAM – 8 GB

5. Data Requirements

Data Gathering:

Dataset for this project was gathered from Kaggle datasets where most of the datasets are available for ML projects.

Data Description:

The Dataset contains close to 10600 records of information that has Airway carriers, routes, Source of boarding, destination, and additional info. Some of the columns has less or no information that's useful. It has train and test data in separate file.

FLIGHT FARE PREDICTION

| Data_Train [Protect] | | | | | | | | | | | |
|---|-------------|-----------------|----------|-------------|-----------|----------|--------------|----------|-------------|-----------------|-------|
| File Home Insert Page Layout Formulas Data Review View Help Power Pivot Tell me what you want | | | | | | | | | | | |
| A1 Airline | | | | | | | | | | | |
| | A | B | C | D | E | F | G | H | I | J | K |
| 1 | Airline | Date_of_Journey | Source | Destination | Route | Dep_Time | Arrival_Time | Duration | Total_Stops | Additional_Info | Price |
| 2 | IndiGo | 24/03/201 | Banglore | New Delhi | BLR → DEL | 22:20 | 01:10 22 M | 2h 50m | non-stop | No info | 3897 |
| 3 | Air India | 1/05/2019 | Kolkata | Banglore | CCU → IXF | 05:50 | 13:15 | 7h 25m | 2 stops | No info | 7662 |
| 4 | Jet Airway | 9/06/2019 | Delhi | Cochin | DEL → LKO | 09:25 | 04:25 10 J | 19h | 2 stops | No info | 13882 |
| 5 | IndiGo | 12/05/201 | Kolkata | Banglore | CCU → NA | 18:05 | 23:30 | 5h 25m | 1 stop | No info | 6218 |
| 6 | IndiGo | 01/03/201 | Banglore | New Delhi | BLR → NA | 16:50 | 21:35 | 4h 45m | 1 stop | No info | 13302 |
| 7 | SpiceJet | 24/06/201 | Kolkata | Banglore | CCU → BL | 09:00 | 11:25 | 2h 25m | non-stop | No info | 3873 |
| 8 | Jet Airway | 12/03/201 | Banglore | New Delhi | BLR → BO | 18:55 | 10:25 13 M | 15h 30m | 1 stop | In-flight m | 11087 |
| 9 | Jet Airway | 01/03/201 | Banglore | New Delhi | BLR → BO | 08:00 | 05:05 02 M | 21h 5m | 1 stop | No info | 22270 |
| 10 | Jet Airway | 12/03/201 | Banglore | New Delhi | BLR → BO | 08:55 | 10:25 13 M | 25h 30m | 1 stop | In-flight m | 11087 |
| 11 | Multiple ca | 27/05/201 | Delhi | Cochin | DEL → BO | 11:25 | 19:15 | 7h 50m | 1 stop | No info | 8625 |
| 12 | Air India | 1/06/2019 | Delhi | Cochin | DEL → BLF | 09:45 | 23:00 | 13h 15m | 1 stop | No info | 8907 |
| 13 | IndiGo | 18/04/201 | Kolkata | Banglore | CCU → BL | 20:20 | 22:55 | 2h 35m | non-stop | No info | 4174 |
| 14 | Air India | 24/06/201 | Chennai | Kolkata | MAA → CC | 11:40 | 13:55 | 2h 15m | non-stop | No info | 4667 |
| 15 | Jet Airway | 9/05/2019 | Kolkata | Banglore | CCU → BG | 21:10 | 09:20 10 M | 12h 10m | 1 stop | In-flight m | 9663 |
| 16 | IndiGo | 24/04/201 | Kolkata | Banglore | CCU → BL | 17:15 | 19:50 | 2h 35m | non-stop | No info | 4804 |
| 17 | Air India | 3/03/2019 | Delhi | Cochin | DEL → AM | 16:40 | 19:15 04 M | 26h 35m | 2 stops | No info | 14011 |
| 18 | SpiceJet | 15/04/201 | Delhi | Cochin | DEL → PNI | 08:45 | 13:15 | 4h 30m | 1 stop | No info | 5830 |
| 19 | Jet Airway | 12/06/201 | Delhi | Cochin | DEL → BO | 14:00 | 12:35 13 J | 22h 35m | 1 stop | In-flight m | 10262 |
| 20 | Air India | 12/06/201 | Delhi | Cochin | DEL → CC | 20:15 | 19:15 13 J | 23h | 2 stops | No info | 13381 |
| 21 | Jet Airway | 27/05/201 | Delhi | Cochin | DEL → BO | 16:00 | 12:35 28 M | 20h 35m | 1 stop | In-flight m | 12898 |
| 22 | GoAir | 6/03/2019 | Delhi | Cochin | DEL → BO | 14:10 | 19:20 | 5h 10m | 1 stop | No info | 19495 |
| 23 | Air India | 21/03/201 | Banglore | New Delhi | BLR → CO | 22:00 | 13:20 19 M | 15h 20m | 1 stop | No info | 6955 |
| 24 | IndiGo | 3/04/2019 | Banglore | Delhi | BLR → DEL | 04:00 | 06:50 | 2h 50m | non-stop | No info | 3943 |
| 25 | IndiGo | 1/05/2019 | Banglore | Delhi | BLR → DEL | 18:55 | 21:50 | 2h 55m | non-stop | No info | 4823 |
| 26 | Jet Airway | 6/05/2019 | Kolkata | Banglore | CCU → BG | 18:55 | 08:15 07 M | 13h 20m | 1 stop | In-flight m | 7757 |
| 27 | Jet Airway | 9/06/2019 | Delhi | Cochin | DEL → IDR | 21:25 | 12:35 10 J | 15h 10m | 2 stops | No info | 13292 |
| 28 | IndiGo | 1/06/2019 | Delhi | Cochin | DEL → LKO | 21:50 | 03:35 02 J | 5h 45m | 1 stop | No info | 8238 |
| 29 | GoAir | 15/05/201 | Delhi | Cochin | DEL → BO | 07:00 | 12:55 | 5h 55m | 1 stop | No info | 7682 |
| 30 | Vistara | 18/06/201 | Banglore | Delhi | BLR → DEL | 09:45 | 12:35 | 2h 50m | non-stop | No info | 4668 |
| 31 | Vistara | 15/06/201 | Chennai | Kolkata | MAA → CC | 07:05 | 09:20 | 2h 15m | non-stop | No info | 3687 |
| 32 | Vistara | 18/06/201 | Chennai | Kolkata | MAA → CC | 07:05 | 09:20 | 2h 15m | non-stop | No info | 3687 |
| 33 | Air India | 1/05/2019 | Kolkata | Banglore | CCU → GA | 09:50 | 23:15 | 13h 25m | 2 stops | No info | 13227 |
| 34 | IndiGo | 6/04/2019 | Banglore | Delhi | BLR → DEL | 04:00 | 06:50 | 2h 50m | non-stop | No info | 4422 |

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|------------|-----------|----------|-------------|-----------|----------|--------------|----------|-------------|-----------------------------|---|---|
| 1 | Airline | e_of_Jour | Source | Destination | Route | Dep_Time | Arrival_Time | Duration | Total_Stops | Additional_Info | | |
| 2 | Jet Airway | 6/06/2019 | Delhi | Cochin | DEL → BO | 17:30 | 04:25 07 J | 10h 55m | 1 stop | No info | | |
| 3 | IndiGo | 12/05/201 | Kolkata | Banglore | CCU → M | 06:20 | 10:20 | 4h | 1 stop | No info | | |
| 4 | Jet Airway | 21/05/201 | Delhi | Cochin | DEL → BO | 19:15 | 19:00 22 M | 23h 45m | 1 stop | In-flight meal not included | | |
| 5 | Multiple c | 21/05/201 | Delhi | Cochin | DEL → BO | 08:00 | 21:00 | 13h | 1 stop | No info | | |
| 6 | Air Asia | 24/06/201 | Banglore | Delhi | BLR → DE | 23:55 | 02:45 25 J | 2h 50m | non-stop | No info | | |
| 7 | Jet Airway | 12/06/201 | Delhi | Cochin | DEL → BO | 18:15 | 12:35 13 J | 18h 20m | 1 stop | In-flight meal not included | | |
| 8 | Air India | 12/03/201 | Banglore | New Delhi | BLR → TR | 07:30 | 22:35 | 15h 5m | 1 stop | No info | | |
| 9 | IndiGo | 1/05/2019 | Kolkata | Banglore | CCU → HY | 15:15 | 20:30 | 5h 15m | 1 stop | No info | | |
| 10 | IndiGo | 15/03/201 | Kolkata | Banglore | CCU → BL | 10:10 | 12:55 | 2h 45m | non-stop | No info | | |
| 11 | Jet Airway | 18/05/201 | Kolkata | Banglore | CCU → BO | 16:30 | 22:35 | 6h 5m | 1 stop | No info | | |
| 12 | Jet Airway | 21/03/201 | Delhi | Cochin | DEL → MA | 13:55 | 18:50 22 M | 28h 55m | 2 stops | In-flight meal not included | | |
| 13 | IndiGo | 15/06/201 | Delhi | Cochin | DEL → HY | 06:50 | 16:10 | 9h 20m | 1 stop | No info | | |
| 14 | Multiple c | 15/05/201 | Delhi | Cochin | DEL → BO | 09:00 | 19:15 | 10h 15m | 1 stop | No info | | |
| 15 | Jet Airway | 12/03/201 | Banglore | New Delhi | BLR → BO | 05:45 | 10:25 | 4h 40m | 1 stop | No info | | |
| 16 | Jet Airway | 3/06/2019 | Delhi | Cochin | DEL → BO | 19:15 | 12:35 04 J | 17h 20m | 1 stop | In-flight meal not included | | |
| 17 | Jet Airway | 06/03/201 | Banglore | New Delhi | BLR → BO | 21:25 | 08:15 07 M | 10h 50m | 1 stop | No info | | |
| 18 | Multiple c | 6/06/2019 | Delhi | Cochin | DEL → HY | 13:15 | 22:30 | 9h 15m | 1 stop | No info | | |
| 19 | Vistara | 24/03/201 | Kolkata | Banglore | CCU → DE | 09:55 | 22:10 | 12h 15m | 1 stop | No info | | |
| 20 | Jet Airway | 12/06/201 | Delhi | Cochin | DEL → BO | 19:15 | 04:25 13 J | 9h 10m | 1 stop | In-flight meal not included | | |
| 21 | Jet Airway | 12/03/201 | Banglore | New Delhi | BLR → BO | 22:55 | 08:15 13 M | 9h 20m | 1 stop | No info | | |
| 22 | IndiGo | 6/03/2019 | Delhi | Cochin | DEL → BO | 10:45 | 01:35 07 M | 14h 50m | 1 stop | No info | | |
| 23 | Jet Airway | 9/05/2019 | Kolkata | Banglore | CCU → BO | 20:00 | 10:05 10 M | 14h 5m | 1 stop | In-flight meal not included | | |
| 24 | Jet Airway | 18/03/201 | Banglore | New Delhi | BLR → BO | 21:25 | 09:00 16 M | 11h 35m | 1 stop | In-flight meal not included | | |
| 25 | Jet Airway | 9/05/2019 | Delhi | Cochin | DEL → JAI | 05:30 | 19:00 | 13h 30m | 2 stops | In-flight meal not included | | |
| 26 | Air India | 6/04/2019 | Banglore | Delhi | BLR → DE | 21:05 | 23:55 | 2h 50m | non-stop | No info | | |
| 27 | Jet Airway | 21/03/201 | Delhi | Cochin | DEL → BO | 16:00 | 04:25 22 M | 12h 25m | 1 stop | In-flight meal not included | | |
| 28 | IndiGo | 15/05/201 | Kolkata | Banglore | CCU → BL | 15:15 | 17:45 | 2h 30m | non-stop | No info | | |
| 29 | Jet Airway | 21/05/201 | Delhi | Cochin | DEL → BO | 17:30 | 19:00 22 M | 25h 30m | 1 stop | No info | | |
| 30 | Jet Airway | 6/06/2019 | Delhi | Cochin | DEL → BO | 09:00 | 04:25 07 J | 19h 25m | 1 stop | In-flight meal not included | | |
| 31 | Multiple c | 1/06/2019 | Delhi | Cochin | DEL → BO | 14:00 | 21:00 | 7h | 1 stop | No info | | |
| 32 | Air Asia | 3/03/2019 | Kolkata | Banglore | CCU → BL | 22:50 | 01:20 04 M | 2h 30m | non-stop | No info | | |

Import Data into Database:

We create an API for the upload of data into the Cassandra Database, steps performed are:

- Connection is made with database
- Created a database with name flight fare
- Cqlsh command is given for creating data table with required parameters.
- Cqlsh command is written for uploading the dataset into the table.

6. Data PreProcessing

For the Data preprocessing the following steps are performed:

- Check for the null values if any in the dataset

- Drop those null values if present
- Performed One-Hot encoding(did not perform label encoding for particular reasons mentioned in jupyter file).
- Scaling needs to be done.(But did not do for the current model since we used Random forest regressor).

7. Design

- Modelling:

The processed data is visualised and arrived at some insights. We use different Machine Learning algorithms like Linear Regression, XGBoost regressor, Random Forest Regressor and so on. Finally, we use Random forest Regressor as the final model.

- UI integration:

We have used HTML and CSS and integrated it with the flask app that's in a name app.py.

- Model process and deployment:

The detailed process is explained below in the steps as done with the project.

- Importing the necessary Libraries
- Read the data. (get insights on it).
- Check for null values. (Drop or use imputation)
- EDA
- Standardize the data set
- Encode the dataset (Label or One-Hot encoding upon requirement)
- Preprocessing
- Model Building
- Deploy using the Flask library
- Start the app (command is "python app.py")
- Input from the user
- Get the Result.

8. Data from User:

Data from the user is received by the front end HTML and CSS page that's used to take the input.

9. Data Validation

The data collected as input is then sent as a request to the flask app and validated to predict the future prices.

10. Results

The validated data is sent for prediction

11. Deployment

The tested project and the built ML model is deployed on

Conclusion:

This project is built on a regression mechanism that uses past prices to predict future price.