

High Level Design (HLD)

Flight Fare Prediction

Abstract

The airline industry has always been the preferred mode of transport for those who would like to reach their destination quickly. The pricing of these airline tickets is considered to be a tug of war between the airline carriers and the passengers where each side wants the best rates. Passengers prefer to travel by flights that offer them the best rate at the lowest fare available while the airlines would like to bank in more money for their tickets. Flight prices are also affected by seasons, particular occasions when passengers are expected to travel more. This makes it harder to guess the flight prices. This project provides a method to predict flight fares using machine learning models like Linear regression, Random Forest, Decision tree, Extra trees regression, Gradient boosting and Ridge regression.

Contents

1. Introduction.....	1
2. Description.....	1
2.1 Problem Perspective.....	1
2.2 Problem Statement	1
2.3 Proposed solution.....	1
2.4 Data Requirements.....	1
2.5 Tools used	2
3. Design details.....	2
3.1 Methodology	2
3.2 Event log	2
4. Performance	2
4.1 Reusability	3
4.2 Application compatibility	3
4.3 Deployment.....	3
5. Conclusion	3

1. Introduction

The High-Level-Design (HLD) helps to add the necessary detail to the project description so that the model is suitable for coding. The document can avoid contradictions prior to coding and can be referred to as a manual

The HLD underlines the different design aspects, the user interface being used, performance requirements, design features and the architecture of the project

2. Description

2.1 Problem Perspective

The flight fare prediction system is a machine learning based regression model that will help us predict the fare of a flight based on the several inputs related to the travel.

2.2 Problem Statement

To create a solution to find the ticket price of the user's preferred flight using machine learning models which will make it easy for users to plan their trips

2.3 Proposed solution

The proposed solution is to build various machine learning models like Linear regression, Random Forest, Decision tree, Extra trees regression, Gradient boosting and Ridge regression and find which model is able to best predict the fare of the flight given a set of data.

2.4 Data Requirements

Data for the task depends solely on the problem statement. The dataset contains details pertaining to the flight like the preferred airline, date of travel, number of stops, the source and destination of travel.

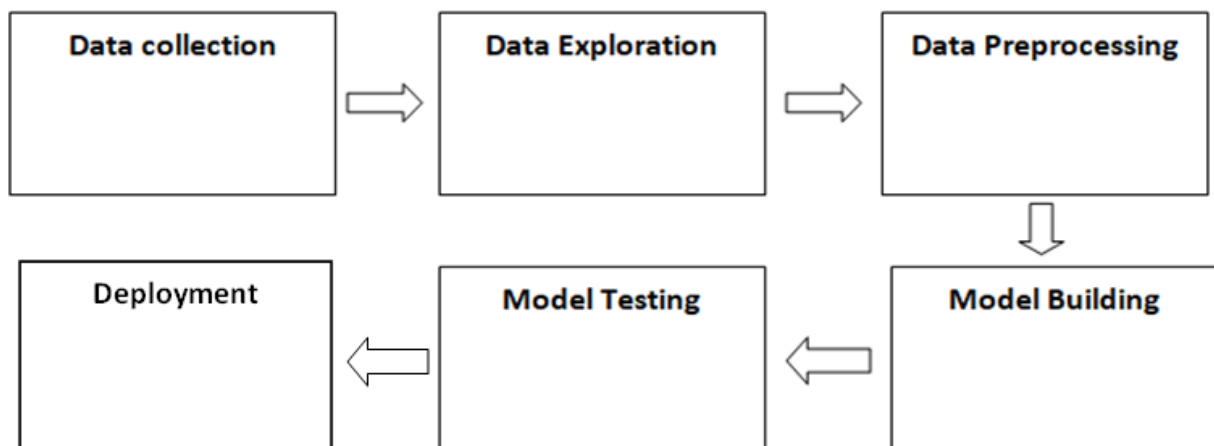
The dataset used for the task is in CSV format and consists of numerical and categorical data types

2.5 Tools used

Python is the main programming language used for the project and uses different libraries like pandas for reading the file and other tasks, numpy for mathematical operations, seaborn and matplotlib for visualization, sklearn for building machine learning models. Jupyter notebook is used as the IDE

3. Design details

3.1 Methodology



3.2 Event log

The system logs all the steps so that the user will be aware of the processes working internally. The type of logging chosen for each depends upon the type of operation that is carried out like info, critical, debug or warning logging. Logging helps to debug issues, so it is a mandatory step

4. Performance

The machine learning based predictor is used for prediction of flight fares given a set of variables. It can help to make informed decisions and take necessary actions. Model retraining can also help to improve performance of prediction

4.1 Reusability

The code used for the project can be easily reused with no errors.

4.2 Application compatibility

The whole project will be done using Python as the main interface and each component and library in it is used to perform various tasks related to the project and ensure the proper completion of the project.

4.3 Deployment

The best machine learning model will be deployed as an API using FastAPI.

5. Conclusion

The different models used in the project will be used to predict the fare of the flight based on a number of underlying factors. The goal is to identify the model which is able to make predictions well.