HIGH LEVEL DESIGN

FLIGHT FARE PREDICTION

(Domain- Aviation)

Revision number 2.0

Last date of Revision 16/Sept/20201

By

Staffin T Shabu

Document Version Control

Date issued	Version Description	Author
13/08/2021	Initial HLD — V1.0	Staffin
16/09/2021	Deployment — V1.1	Staffin

INDEX

ABSTRACT

Travelling through flights has become an integral part of today's lifestyle as more and more people are opting for faster travelling options. The flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, duration of flights. various occasions such as vacations or festive season. 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. This kind of system or service can be provided to the customers by flight booking companies which will help the customers to book their tickets accordingly.

1. INTRODUCTION

This project aims to develop an application which will predict the flight prices for various flights using 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 and so they have an idea about the best time to book cheap tickets. But there are also many people who are inexperienced in booking tickets and end up falling in 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 proving them the information to book tickets at the right time. The proposed problem statement is "Flight Fare prediction system".

1.1 Scope

Currently, there are many fields where prediction-based services are used such as stock price predictor tools used by stockbrokers and service like Zestimate which gives the estimated value of house prices. Therefore, there is requirement for service like this in the aviation industry which can help the customers in booking tickets. There are many research works that have been done on this using various techniques and more research is needed to improve the accuracy of the prediction by using different algorithms. More accurate data with better features can also be used to get more accurate results.

1.2 Definitions

ML- Machine Learning
Dataset- Collection of Data needed for training and Testing
Visual Studio- IDE used

2. General Description

2.1. Product Perspective

This project aims to provide a solution to those who want to book cheap tickets. Using high end ML algorithm this application will help passengers to plan their travel in advance.

2.2 Problem Statement

Flight ticket prices are highly volatile. The price is very much dependent on peak hours, festival time and various other parameters. It is very difficult to predict the ticket fare. An efficient solution is vital in order for travellers to schedule their travel in a cost efficient manner.
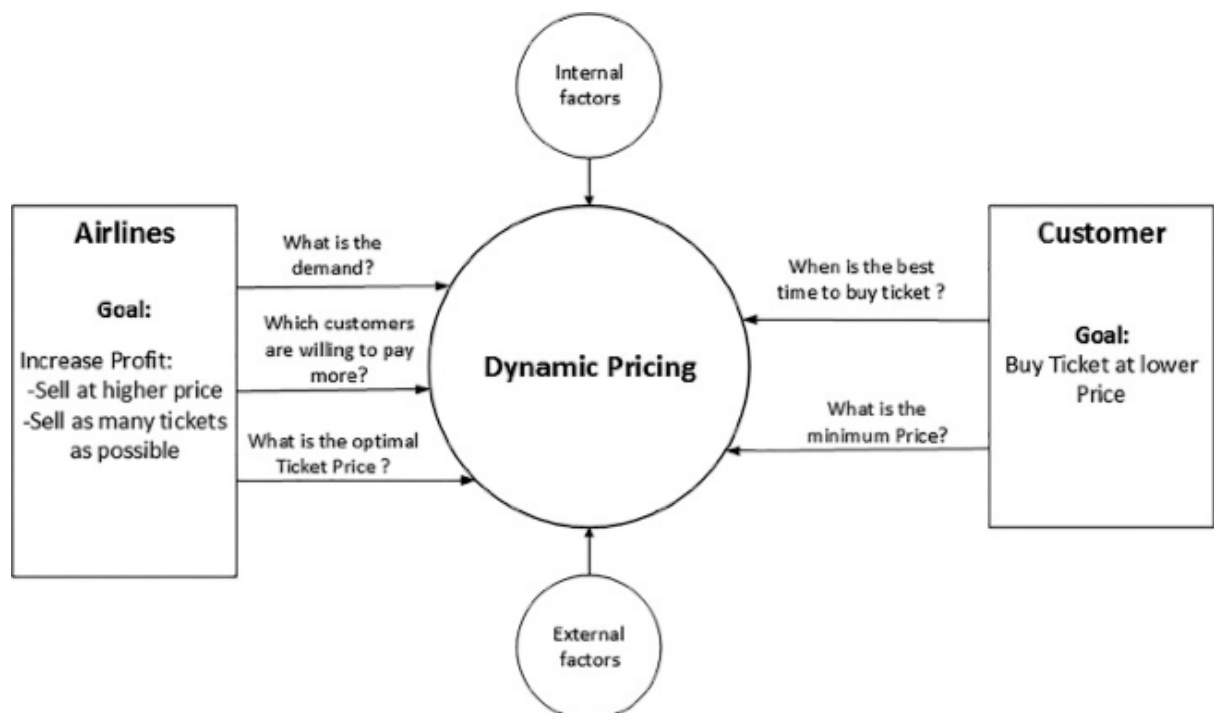
2.3 Proposed Solution

For this project, we have implemented the machine learning life cycle to create a basic web application which will predict the flight prices by applying machine learning algorithm to historical flight data using python libraries like Pandas, NumPy, Matplotlib, seaborn and sklearn.

Data selection is the first step where historical data of flight is gathered for the model to predict prices. Our dataset consists of more than 10,000 records of data related to flights and its prices. Some of the features of the dataset are source, destination, departure date, departure time, number of stops, arrival time, prices, and few more. In the exploratory data analysis step, we cleaned the dataset by removing the duplicate values and null values. If these values are not removed it would affect the accuracy of the model. We gained further information such as distribution of data. Next step is data pre-processing where we observed that most of the data was present in string format. Data from each feature is extracted such as day and month is extracted from date of journey in integer format, hours and minutes is extracted from departure time. Features such as source and destination needed to be converted into values as they were of categorical type. For this One hot-encoding and label encoding techniques are used to convert categorical values to model identifiable values. Feature selection step is involved in selecting important features that are more correlated to the price. There are some features such as extra information and route which are unnecessary features which may affect the accuracy of the model and therefore, they need to be removed before getting our model ready for prediction. After selecting the features which are more correlated to price the next step involves applying machine algorithm and creating a model. As our dataset consist of labelled data, we will be using supervised machine learning algorithms. We will be using random forest Algorithm.

Random Forest

Random Forest Random Forest is an ensemble learning technique where training model uses multiple learning algorithms and then combine individual results to get a final predicted result. Under ensemble learning random forest falls into bagging category where random number of features and records will average value of the predicted values if considered as the output of the random forest model.
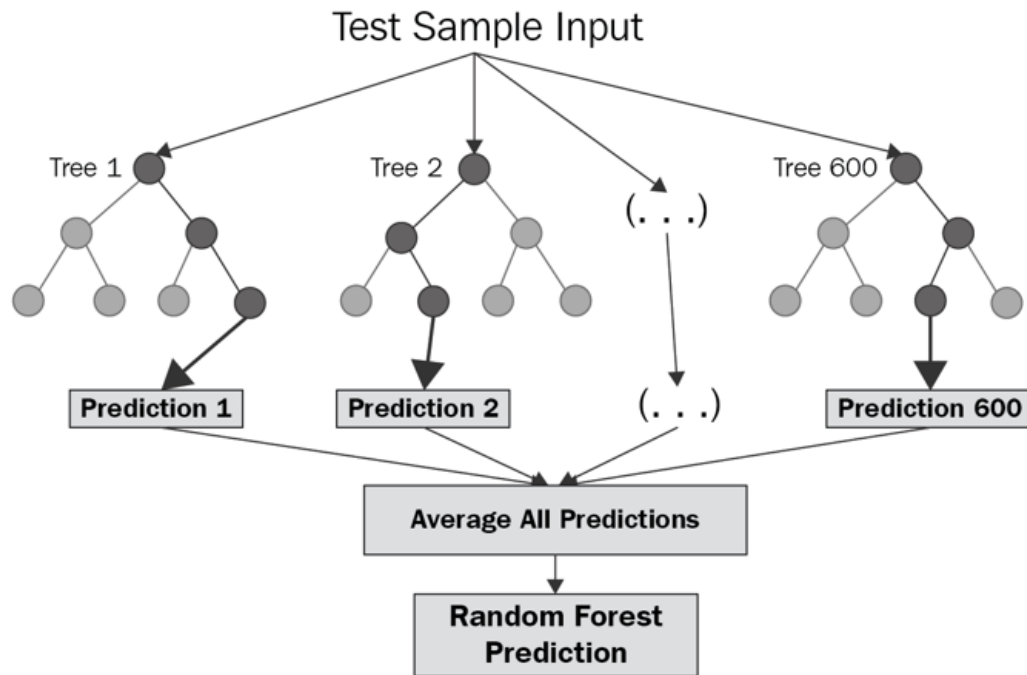
2.4 Further Improvements

We can add further features that would allow the passengers to book directly from the application. This will add more convenience to the passengers.

2.5 Technical requirements

1. Active internet connection with minimum 1mb/sec
2. Minimum 4 Gb Ram
3. Windows/Linux/Mac os
4. Intel Pentium or higher.

2.6 Data Requitements.

The dataset consists of more than 10,000 records of data and 10 columns related to flights and its prices. Some of the features of the dataset are source, destination, departure date, departure time, number of stops, arrival time, prices, and few more.

2.7 Tools used

- Python is used for bac end development
- Numpy, Seaborn, Plotly are used for data analysis and vuisulization
- Flask is used for deployment.
- Front end development is done using HTML/CSS
- Git is used as version control system

2.8 Constrains

The designed application should have a user-friendly interface and should be hassle free experience for the users. The accuracy of the prediction should be highly reliable.
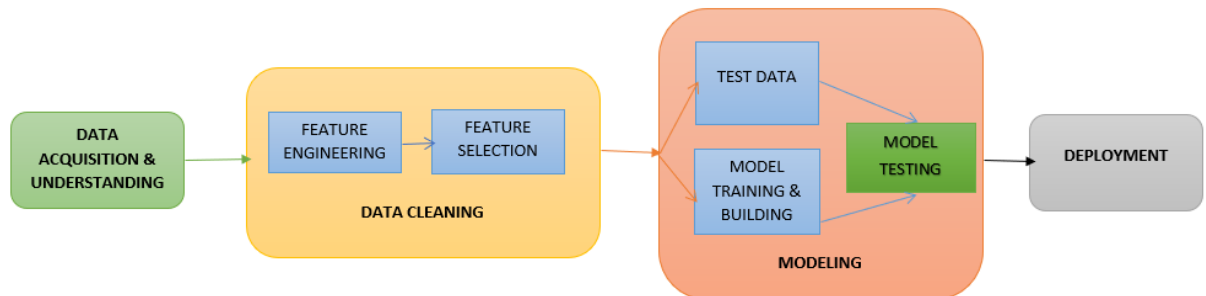
2.9 Assumptions

It assumes that there is minimal or no multicollinearity among the independent variables. It usually requires a large sample size to predict properly. It assumes the observations to be independent of each other.
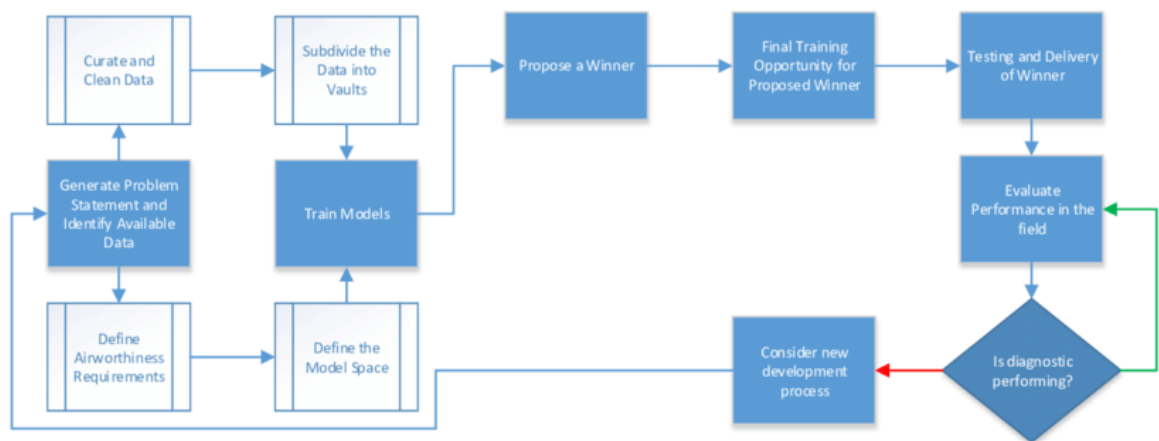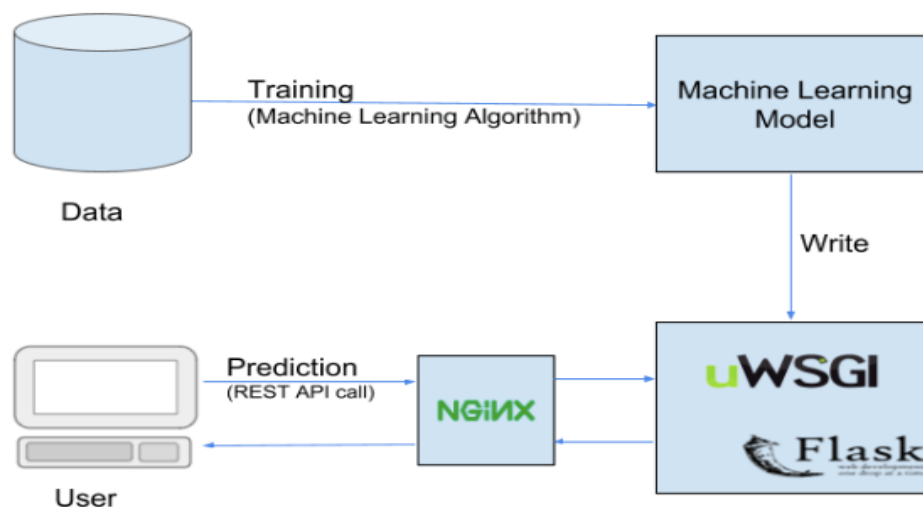
3. Design Details

3.1 Process Flow

In order to predict the flight fare we will use ML model.

3.1.1 Model training and Evaluation

3.1.2 Deployment Process

3.2 Error handling

In case if any error is encountered proper prompt messages will be displayed to the user. An error will be defined as anything that falls outside the normal and intended usage.

4. Performance

The flight fare prediction system is designed to help the passengers to find the dates when flight tickets are cheaper. The prediction should be accurate enough so that it doesn't cause any sort of inconvenience to the user in the future. Proper model training is necessary for accurate prediction.

4.1 Reusability

The code is written in such a way that it can be reused for further implementation or to integrate with other projects.

4.2 Application Compatibility

Various components of this project uses python as an interface between them. Each component is designated with a particular task to be executed.

4.3 Resource Utilization

The application is expected to use minimal system resources to provide a reliable experience.

4.4 Deployment

The project is locally deployed using Flask.

5.Conclusion

A proper implementation of this project can result in saving money of inexperienced people by providing them the information related to trends that flight prices follow and also give them a predicted value of the price which they use to decide whether to book ticket now or later. In conclusion this type of service can be implemented with good accuracy of prediction. As the predicted value is not fully accurate there is huge scope for improvement of these kind of service.