Flight Fare Prediction

Low level document (LLD)

Prashanth Kumar G V

8/14/21

Low Level Document

Document Version Control

Month	Version	Description	Author
July 21	V.0	Initial LLD	Prashanth Kumar GV
July 21	V.0	Initial LLD	Prashanth Kumar GV

Contents

1. INTRODUCTION

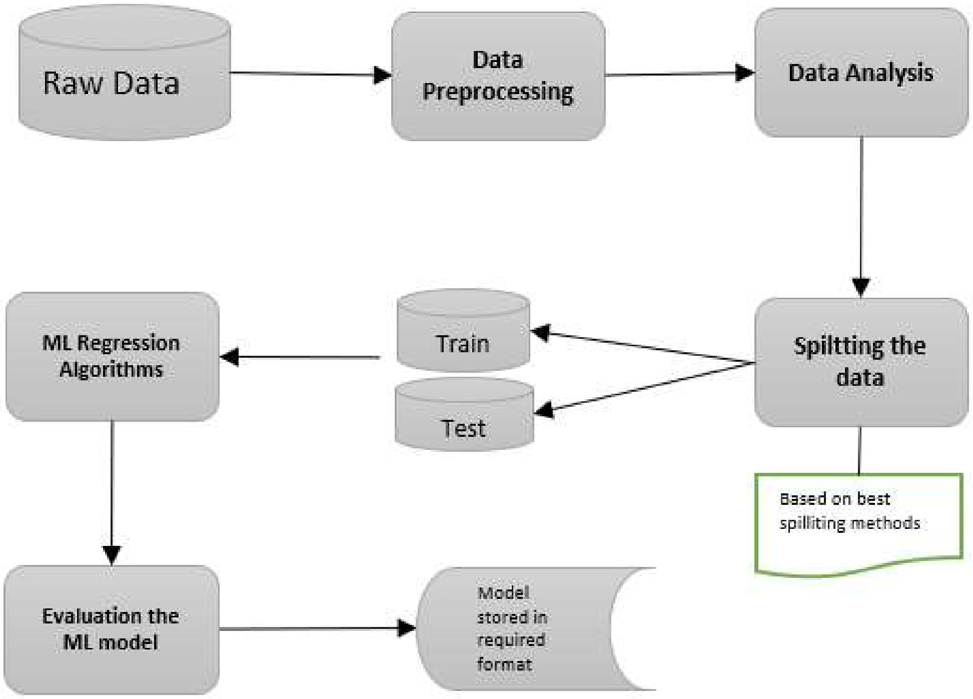
1.1. What is Low-Level design document?

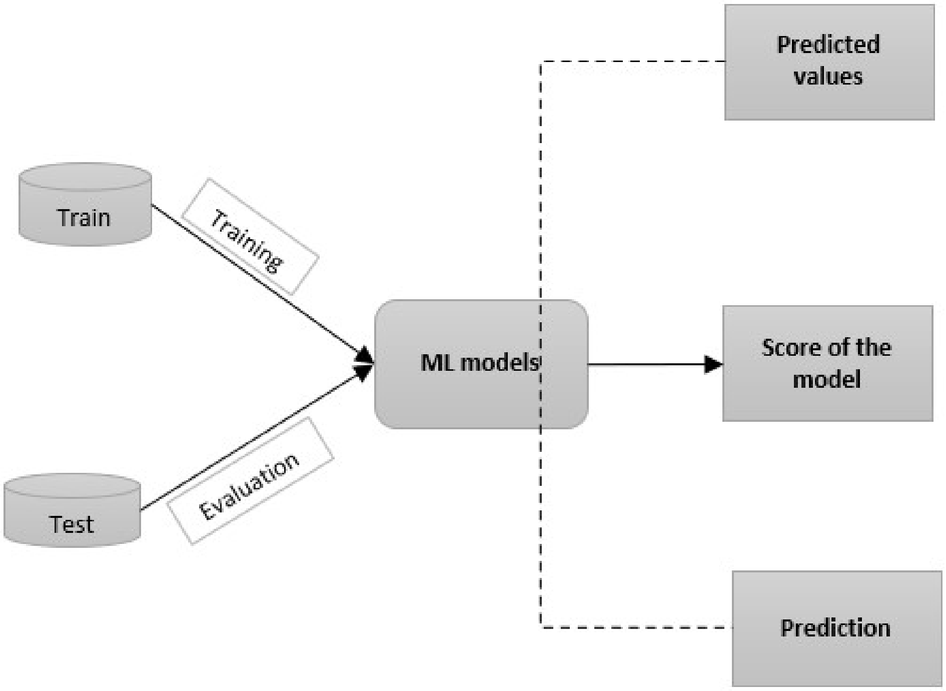
The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Food Recommendation System. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-bystep refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. ARCHITECTURE





2.1 Problem statement:

To predict flight fare charges of various airlines according to the 2019 dataset.

2.2 Data Overview:

It contains the train.xlsx excel format file contains the data related Airlines prices.

Data collection

Data is collected from Kaggle.com, raw data is in excel format.

Importing the data by using pandas' data frames into the python notebook to perfume the various operations and building useful insights.

Exploratory Data Analysis

Data visualization by using the matplotlib, seaborn libraries for analysing the data and exploring the data in different perspectives. And also, we find the useful insights related to data.

Data Pre-processing

Data Analysis for relation between the different features which finds the relations between the one feature to another feature to find the ground truth predictable.

Data Splitting

After analysing the data by visualization and various statistical parameters and methods we splitting the data into train and test because first we train the data to ML model then we test the Model to test data.

Production the Model

Building the various models, We train the model on various supervised regression machine learning algorithms.

Testing the Models

Testing the model by best models as mentioned as above.

Hyperparamter tuning the algorithms by Random search.

We got the best score by XGBoost and Random forest algorithms

Performance of model

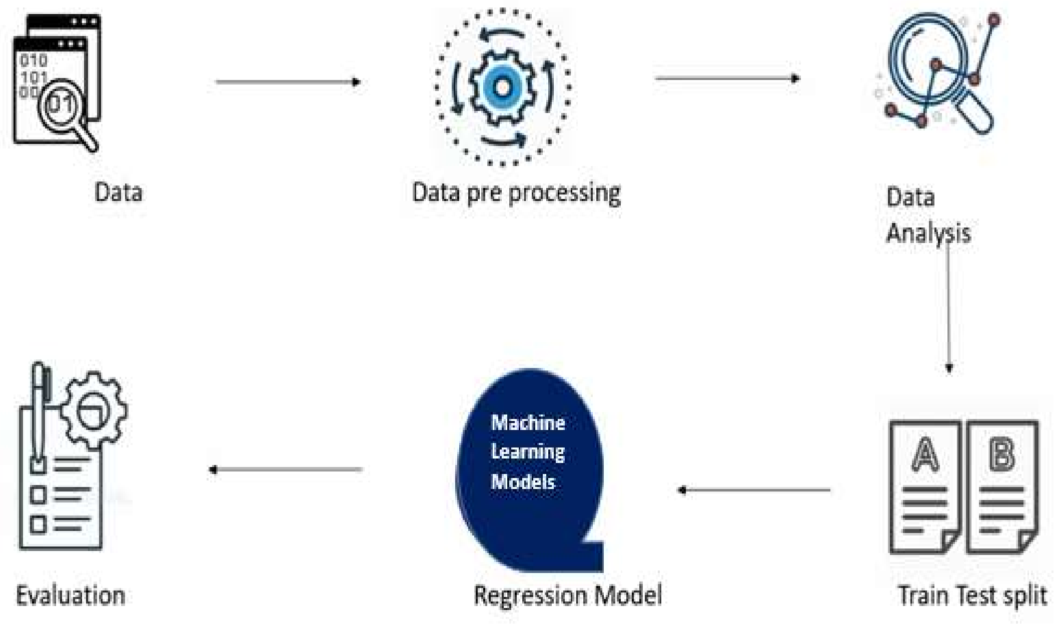
Evaluating the model by performance metrics to pick the best model accurately predicting the truth variables based on the train set.

After all performing the model we got the 0.843 score and R2 error of 0.857

Storing the Model

Store the model into standard format like JSON, PICKLE, SQL etc for the further process like building the web APIs or model deployments and testing.

3. WORK FLOW



References

1. Krish Naik youtube Classes