

FLIGHT FARE PREDICTION USING MACHINE LEARNING

## High Level Design (HLD)

## cheepurupalli yaswanth kumar

## november- december, 2021

iNeuron

# INTRODUCTION

The main objective of this project is to predict the flight price given certain factors that are proven to be effective on Flight Price. In Airline Business, flight Price is an important factor that drives customer base. Customers generally tend to go for budget friendly. So, airline company may use different factors to estimate flight price and adjust the pricing accordingly in real time. The use case of this application can be expanded to Flight Price Forecast and Flight Recommendation Applications.

# PROBLEM STATEMENT

As per the problem statement, we can make use of different machine learning regression algorithms to do the predictions. The data used for the algorithm was well structured in tabular format and proper data dictionary was available for the data. So, we proceeded with all the data processing and modeling activities.

# DATASET INFORMATION

# The dataset used for this project was extracted from Kaggle.

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

# Data Format: Excel

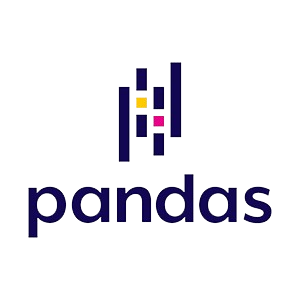
# 

# Columns Information:

* Airline: Name of the airline used for traveling
* Date\_of\_Journey: Date at which a person traveled
* Source: Starting location of flight
* Destination: Ending location of flight
* Route: This contains information on starting and ending location of the journey in the standard format used by airlines.
* Dep\_Time: Departure time of flight from starting location
* Arrival\_Time: Arrival time of flight at destination
* Duration: Duration of flight in hours/minutes
* Total\_Stops: Number of total stops flight took before landing at the destination.
* Additional\_Info: Shown any additional information about a flight
* Price: Price of the flight (**Target**)

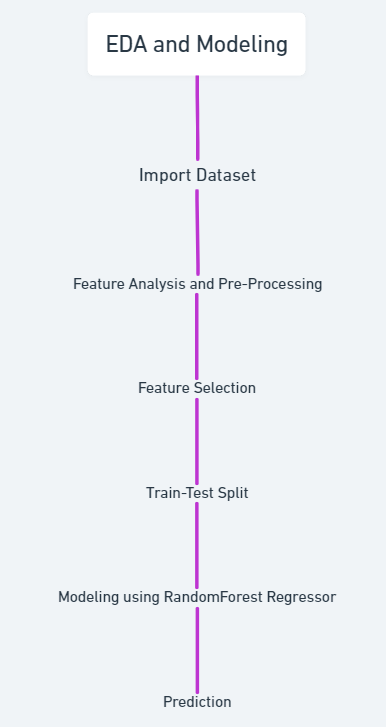
# TOOLS USED

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn are used to build the whole model.

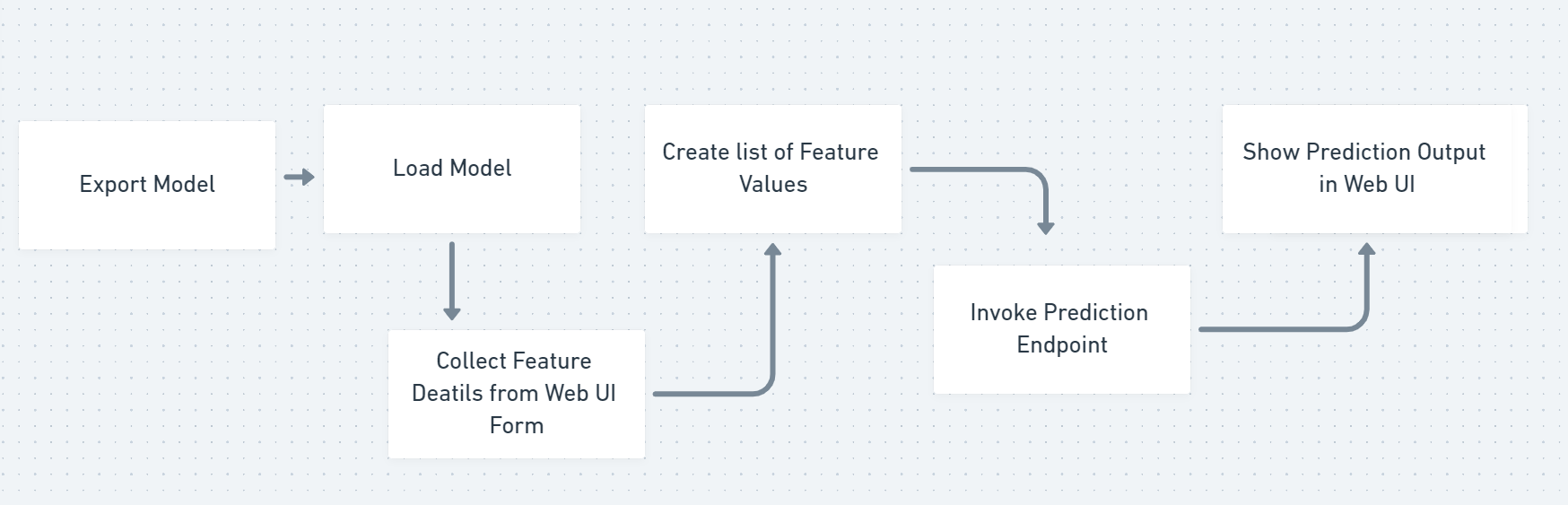


# DESIGN DETAILS

* 1. Process flow



* 1. Deployment process



# CONCLUSION

The project is designed in Streamlit. As we proceed with the Data pre-processing and cleaning and necessary data conversion, we may perform feature selection and elimination prior to model building and if necessary, post model building as well. We will make use ensemble technique for prediction preferably RandomForestRegressor. We will crosscheck the metrices for evaluating the generalized model prior to deployment.