FLIGHT_PRICE_PREDICTION

Submitted by:
Prachi NishantKumar Mehta

# ACKNOWLEDGMENT

Here, only one type of Dataset Train Dataset.
Refering online website for dataset, webscraping data from that and than analyse it and apply machine learning model.

## • Business Problem Framing

The Flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, duration of flights. In the proposed system a predictive model will be created by applying machine learning algorithms to the collected historical data of flights. Optimal timing for airline ticket purchasing

Pricing in the airline industry is often compared to a brain game between carriers and passengers where each party pursues the best rates. Carriers love selling tickets at the highest price possible — while still not losing consumers to competitors. Passengers are crazy about buying flights at the lowest cost available — while not missing the chance to get on board. All this makes flight prices fluctuant and hard to predict. But nothing is impossible for people armed with intellect and algorithms.

## • Conceptual Background of the Domain Problem

It is very difficult for the customer to purchase a flight ticket at the minimum price. For this several techniques are used to

obtain the day at which the price of air ticket will be minimum. Most of these techniques are using sophisticated artificial intelligence(AI) research is known as Machine Learning.

Air passengers (the buyers) are often looking for the best time period to purchase airfares to get as much saving as possible while airlines (the sellers) always try to maximize their revenues by revising different prices for the same service. The airline implements dynamic pricing for the flight ticket. Flight ticket prices change during the morning and evening time of the day. Also, it changes with the holidays or festival season. There are several different factors on which the price of the flight ticket depends. The price of an airline ticket is affected by a number of factors, such as flight distance, purchasing time, fuel price, etc. The sellers have all the necessary information (for example historical sale, market demand, customer profile, and behaviour) to make the decision whether to increase or decrease.

- **Review of Literature**

A lot of factors that affect the overall price of airline tickets, including the airline, the date of travel, source, destination, route, duration, and so on. Each provider seems to have its own unique set regulations and methods for determining pricing. Recent breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML) allow for the inference of such principles as well as the modelling of price volatility. This article is a study conducted on predicting flight prices. Utilizing two datasets for testing and training, this study analyses various machine learning methods for predicting flight prices.

- **Motivation for the Problem Undertake**

To begin, we need information on aircraft business and mass transit in order to develop the airline ticket pricing model just at market level. As a result, we have two datasets: training and testing. The training dataset contains 10,684 items with parameters such as airline, date of travel, source, destination, route, time of departure, estimated time of arrival, length, maximum stop, extra data, and price.

# Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

Load dataset into Jupyter Notebook by scraping data from online website.
Analyze the dataset using various operations:
Check the shape of both dataset, checking null-value, getting all information about dataset using info and describe method.
Then checking correlation of dataset with eachother.

- Data Sources and their formats

Out[10]:

| | Airline | Source | Destination | Duration | Total_Stop | Price |
|---|---|---|---|---|---|---|
| **0** | Fly Dubai | AMD | DXB | 3h 25m | Non Stop | 17,068 |
| **1** | Vistara | AMD | DXB | 11h 20m | 1 Stop | 17,950 |
| **2** | IndiGo | AMD | DXB | 3h 15m | Non Stop | 18,148 |
| **3** | IndiGo | AMD | DXB | 7h 00m | 1 Stop | 19,250 |
| **4** | SpiceJet | AMD | DXB | 3h 30m | Non Stop | 19,303 |
| **...** | ... | ... | ... | ... | ... | ... |
| **104** | Vistara | AMD | DXB | 13h 20m | 1 Stop | 90,931 |
| **105** | Qatar Airways | AMD | DXB | 18h 10m | 1 Stop | 1,00,453 |
| **106** | Vistara | AMD | AUH | 23h 55m | 1 Stop | 1,13,112 |
| **107** | Vistara | AMD | AUH | 16h 10m | 1 Stop | 1,15,265 |
| **108** | Vistara | AMD | AUH | 6h 35m | 1 Stop | 1,39,496 |

- Data Preprocessing Done

Data Preprocessing can be done using various mathematical operations
By replacing all number to numb, times into hour and minute each, differe all source and destination using encoding method.

- Data Inputs- Logic- Output Relationships


- Hardware and Software Requirements and Tools Used

Jupyter Notebook and other libraries used

## Model/s Development and Evaluation

- **Identification of possible problem-solving approaches (methods)**

Data analyze, Data Preprocessing and then after apply different model to that dataset.

- ## Testing of Identified Approaches (Algorithms)

ExtraTreesClassifier

LinearRegression

RandomForestRegressor

metrics

HyperParameterTuning

- ## Run and Evaluate selected models

ExtraTreesClassifier

```
: from sklearn.datasets import make_classification
  from sklearn.ensemble import ExtraTreesClassifier
  X, y = make_classification(n_samples=1000, n_features=20, n_informative=15, n_redundant=5, random_state=4)
  model = ExtraTreesClassifier()
```

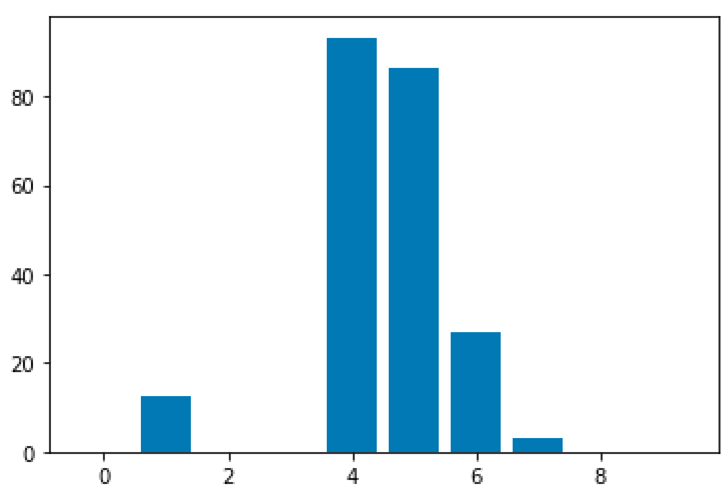
```
: ExtraTreesRegressor()
```

```
: ExtraTreesRegressor()
```

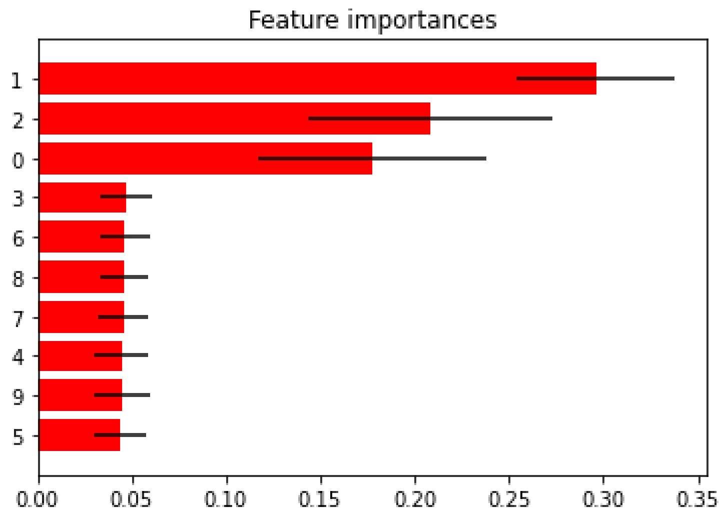
LinearRegression

```
from sklearn.datasets import make_regression
from sklearn.linear_model import LinearRegression
from matplotlib import pyplot
X, y = make_regression(n_samples=1000, n_features=10, n_informative=5, random_state=1)
# define the model
model = LinearRegression()
# fit the model
model.fit(X, y)
# get importance
importance = model.coef_
# summarize feature importance
for i,v in enumerate(importance):
 print('Feature: %0d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()
```

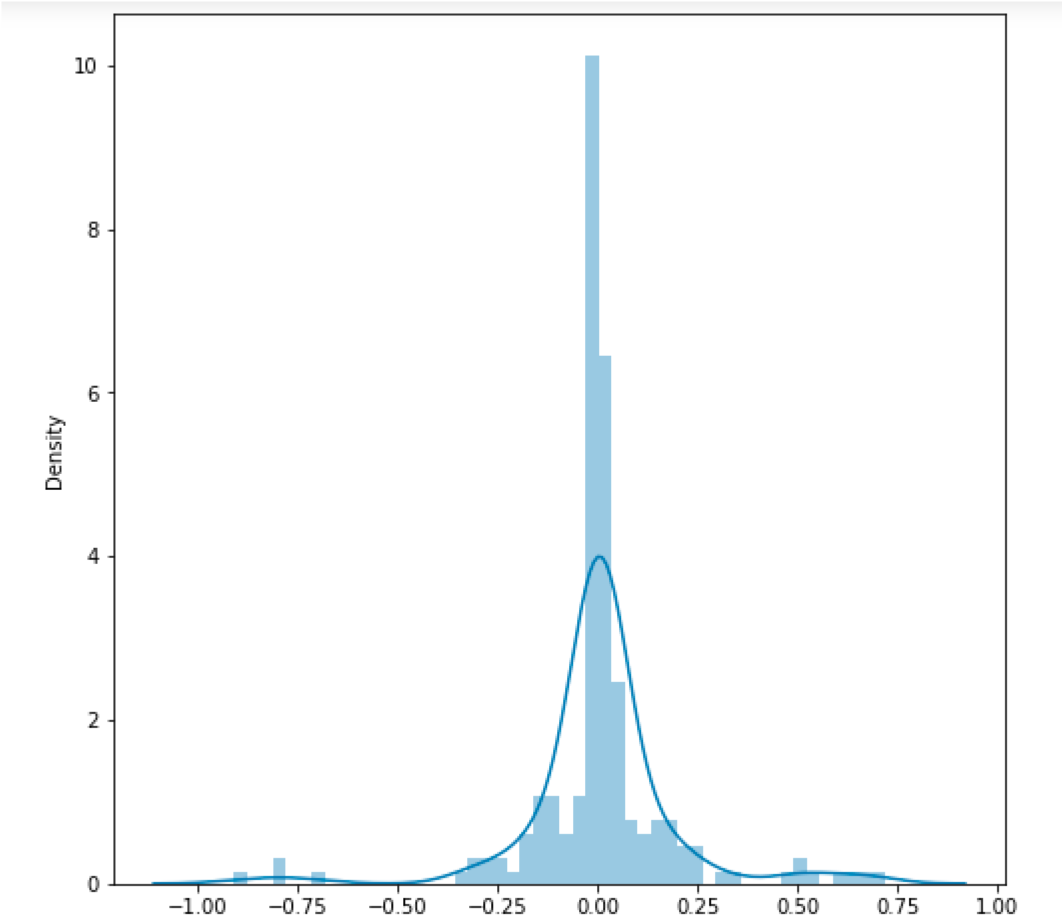
```
Feature: 0, Score: -0.00000
Feature: 1, Score: 12.44483
Feature: 2, Score: 0.00000
Feature: 3, Score: -0.00000
Feature: 4, Score: 93.32225
Feature: 5, Score: 86.50811
Feature: 6, Score: 26.74607
Feature: 7, Score: 3.28535
Feature: 8, Score: 0.00000
Feature: 9, Score: 0.00000
```



Feature Iimportances




Hyper-Parameter Tuning




- **Key Metrics for success in solving problem under consideration**

Here, we used different model for solving problem

- Visualizations

SNS Heatmap

- **Interpretation of the Results**

Result through accuracy of different models.