****

**NetFare Manual Project Work 2020**

|  |
| --- |
|  |

**Project Code : P23**

**Group :G4**

**Submitted Date:2 July 2020**

**Place :Bangalore, India**

**Project Time :(26 May 2020 – 5 July 2020)**

|  |
| --- |
| **Mentors** |

Designing of the project and the conceptualization of the activities carried out by the following members:

* Munumun
* Varun

|  |
| --- |
| **Group Members** |

Development of the project and the deployment of the Model carried out by the following members:

* Raja Ram Chaudhary
* Umesha B.K
* Gowthami R.

|  |
| --- |
| **FAQ** |

* **What types of problems in provided dataset?**

There are Linear and Time Series Forecasting problems in dataset. We solve by creating predictive Linear Model.

* **What is the problem that we are tried to solve?**

Passengers want to fly in cheaper rate but companies want to generate valuable revenue at same time. Due to this, there has fluctuations in price of Air over time. This is the problem to solve by building a predictive model so that both parties may survive by exploring historical airfare data.

* **How you solve the question?**

Diagnosed data quality:

* If there has a problem with data quality.
* The data has corrected.
* Explored data to understand the data and find scenarios for performing the analysis.
* Identifed dependency over many endogenous variables.
* Compared different models to evaluate best model.
* **What data scientist tools used for solution?**
* Jupyter Notebook-6.0.3
* Python-3.6.10
* iPython-7.13.0
* Numpy-1.18.4
* Pandas-1.0.3
* Matplotlib-3.2.1
* Seaborn-0.10.1
* Scipy-1.4.1
* Scikit-learn-0.22

|  |
| --- |
| **Methodology** |

**Dataset:** air\_fair.csv

* InvoiceDate converted into datetime format
* Extracted of Date, Year, Month, Day, Hour & Minute from InvoiceDate columns
* Droped InvoiceDate column after extraction
* Show total counts of records in different Years
* Each column has 278466 rows of data, all columns containing missing data:
* "Year","Month","Day","Hour","Minute": 2 missing data
* "ProductType": 2 missing data
* "ItineraryType": 32777 missing data
* "NetFare": 60890 missing data
* Deal with missing data
* "Year","Month","Day","Hour","Minute": 2 missing data, drop them , two observations among large dataset are not so big deals
* "ProductType": 2 missing data, drop them , two observations among large dataset are not so big deals
* "ItineraryType": 32777 missing data, drop them , they are totally empty, we have huge data
* "NetFare": 60890 missing data, drop whole rows, they are totally empty

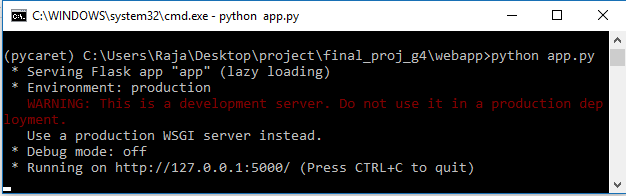
#### Identified Outliers with Interquartile Range (IQR)

* Created dummies variables
* Correlation heatmap of dataset

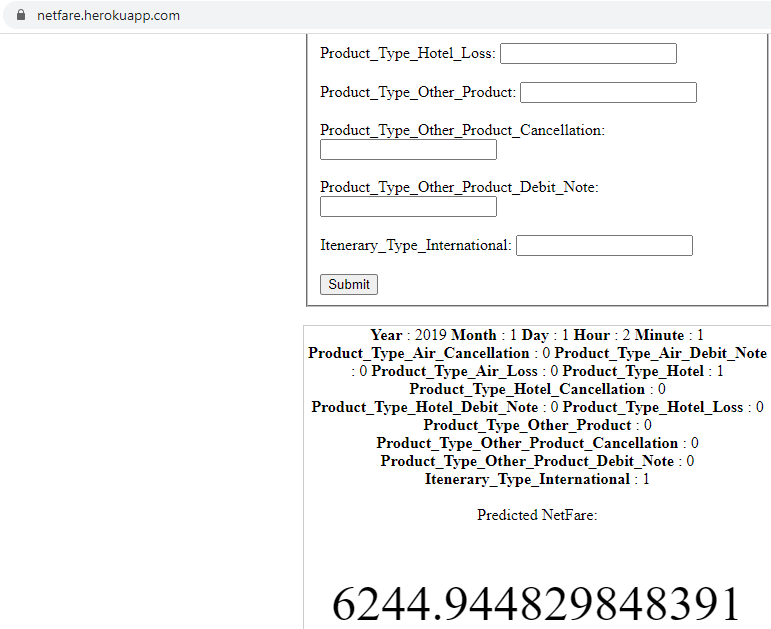
### Fit or Train the Model

* Pickle Model
* Prediction Model

**Pickle:** 30june\_lr\_model.pickle

* Flask webserver 
* Heroku webview

Link: [netfare.herokuapp.com](http://netfare.herokuapp.com)



|  |
| --- |
| **Summary** |

Predictive Model shows good prediction.