CENTER FOR DEVELOPMENT OF ADVANCED COMPUTING

Airline Ticket Price Predication

Presented by:

Rohit S. Patil - 1338
Pratik M. Gajbhiye - 1342

Prashant Karhale
Center Coordinator

Akshay Tilekar (External Guide)

Introduction

- Now a days, the airline corporations are using complex strategies and methods to assign airline prices in a dynamic fashion. These strategies are taking into account several financial, marketing, commercial and social factors closely connected with the final airline prices.

- Due to the high complexity of the pricing models applied by the airlines, it is very difficult a customer to purchase an airline ticket in the lowest price, since the price changes dynamically.

- For this reason, several techniques able to provide the right time to the buyer to purchase an air ticket by predicting the airline price, have been proposed recently. The majority of these methods are making use of sophisticated prediction models from the computational intelligence research field known as Machine Learning (ML).

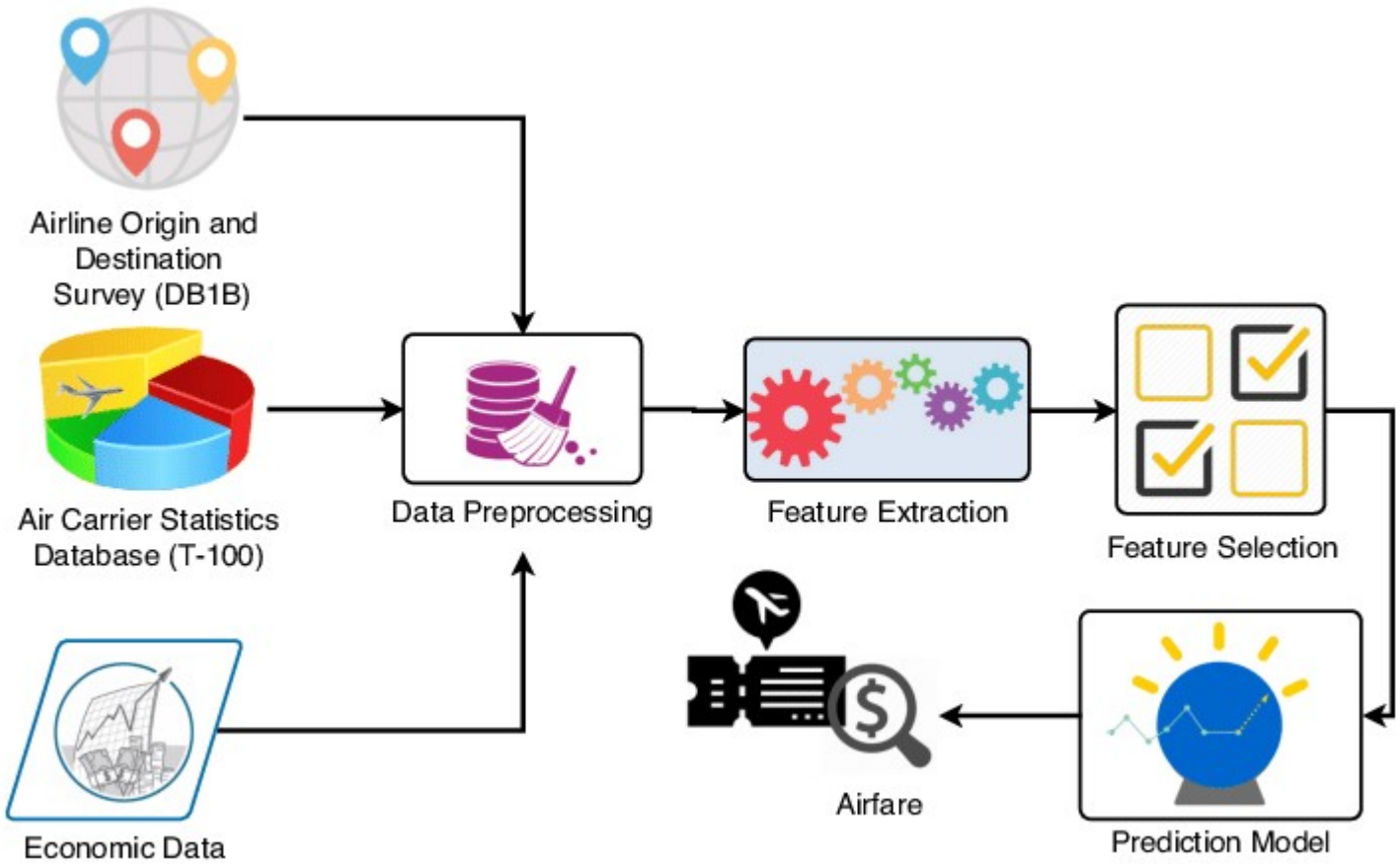
Machine Learning

- Machine Learning is one of the most hot research topics in computer science and engineering, which is applicable in many disciplines. It provides a collection of algorithms, methods and tools able to embody some kind of intelligence to machines.

- The power of ML is the provided modeling tools, which are able to be trained, via a learning procedure, with a set of data describing a certain problem and to respond to similar unseen data with a common way.

- One of the reasons that ML has attracted scientists from several disciplines is its ability to provide human-like intelligence to machines as the amount of data used during learning increases. However, the increase of the training data needs parallel implementations of the ML algorithms using specialized software and/or hardware platforms.

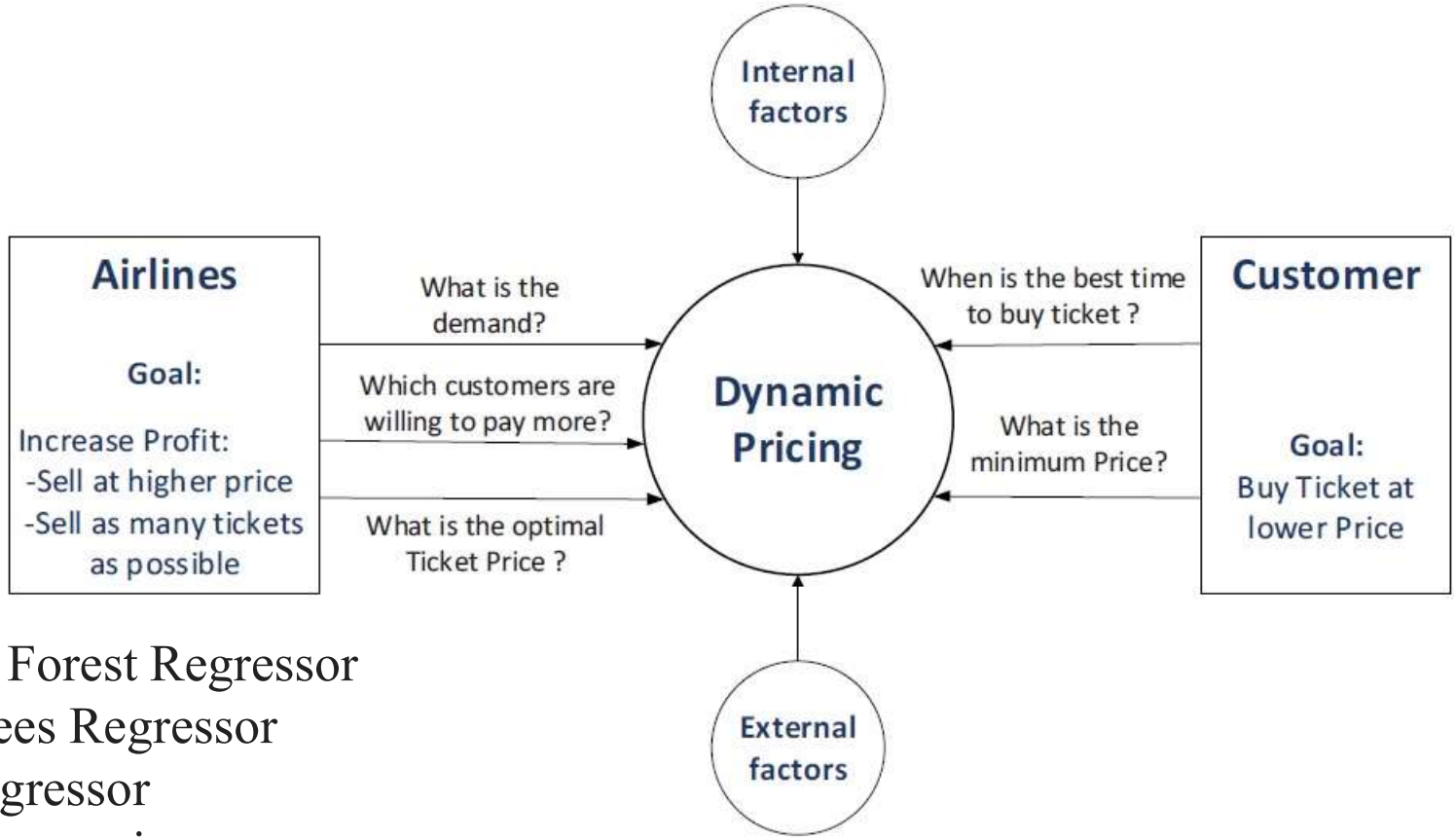
Proposed Methodology



Dataset

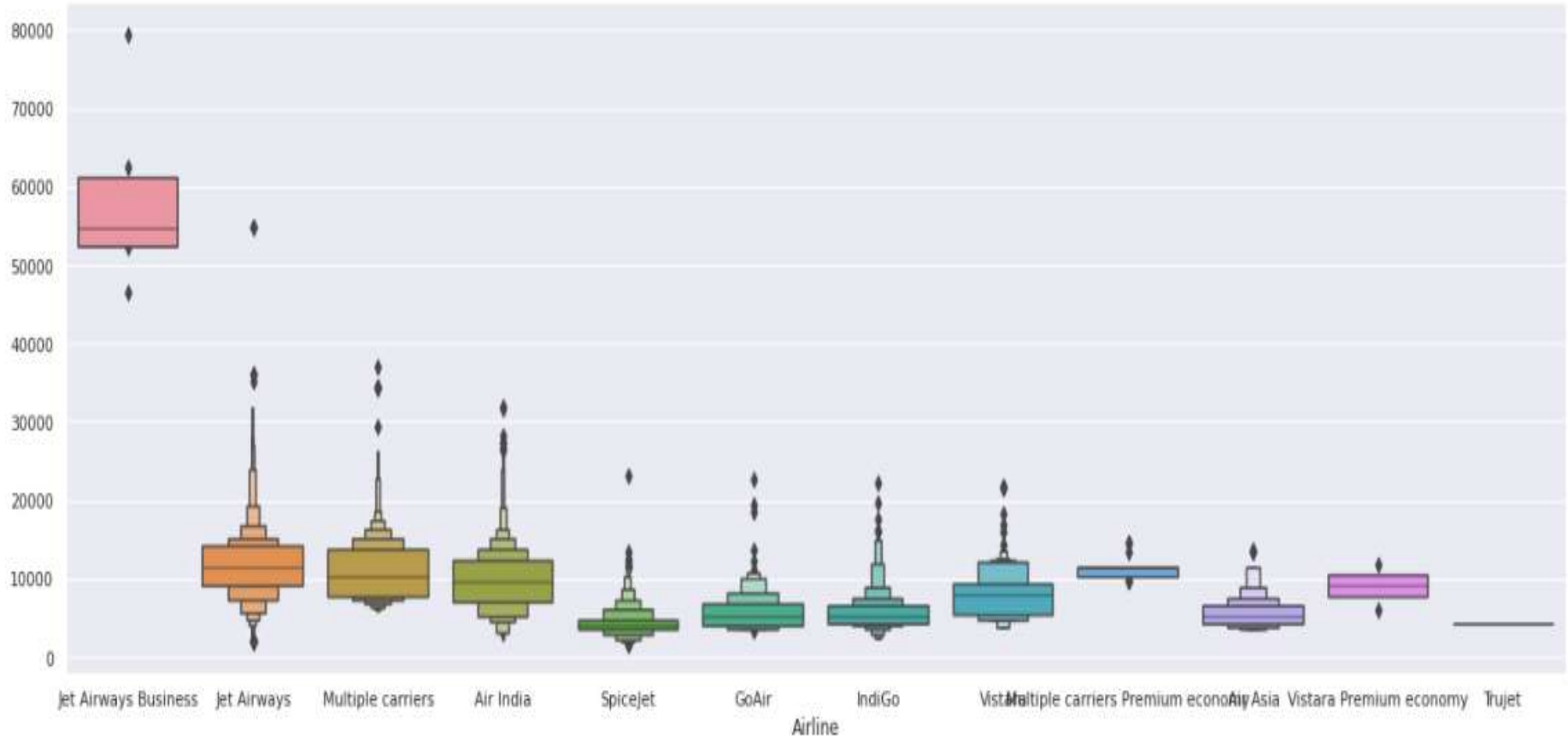
	A	B	C	D	E	F	G	H	I
1	**Airline**	**Date_of_Journey**	**Source**	**Destination**	**Route**	**Dep_Time**	**rrival_Tim**	**Duration**	**Total_Stops**
2	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 M	2h 50m	non-stop
3	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR	05:50	13:15	7h 25m	2 stops
4	Jet Airway	9/06/2019	Delhi	Cochin	DEL → LKO	09:25	04:25 10 Ju	19h	2 stops
5	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NA	18:05	23:30	5h 25m	1 stop
6	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NA	16:50	21:35	4h 45m	1 stop
7	SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BL	09:00	11:25	2h 25m	non-stop
8	Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BO	18:55	10:25 13 M	15h 30m	1 stop
9	Jet Airway	01/03/2019	Banglore	New Delhi	BLR → BO	08:00	05:05 02 M	21h 5m	1 stop
10	Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BO	08:55	10:25 13 M	25h 30m	1 stop
11	Multiple ca	27/05/2019	Delhi	Cochin	DEL → BO	11:25	19:15	7h 50m	1 stop
12	Air India	1/06/2019	Delhi	Cochin	DEL → BLR	09:45	23:00	13h 15m	1 stop
13	IndiGo	18/04/2019	Kolkata	Banglore	CCU → BL	20:20	22:55	2h 35m	non-stop
14	Air India	24/06/2019	Chennai	Kolkata	MAA → CC	11:40	13:55	2h 15m	non-stop
15	Jet Airway	9/05/2019	Kolkata	Banglore	CCU → BO	21:10	09:20 10 M	12h 10m	1 stop
16	IndiGo	24/04/2019	Kolkata	Banglore	CCU → BL	17:15	19:50	2h 35m	non-stop
17	Air India	3/03/2019	Delhi	Cochin	DEL → AM	16:40	19:15 04 M	26h 35m	2 stops
18	SpiceJet	15/04/2019	Delhi	Cochin	DEL → PN	08:45	13:15	4h 30m	1 stop
19	Jet Airway	12/06/2019	Delhi	Cochin	DEL → BO	14:00	12:35 13 Ju	22h 35m	1 stop
20	Air India	12/06/2019	Delhi	Cochin	DEL → CC	20:15	19:15 13 Ju	23h	2 stops
21	Jet Airway	27/05/2019	Delhi	Cochin	DEL → BO	16:00	12:35 28 M	20h 35m	1 stop
22	GoAir	6/03/2019	Delhi	Cochin	DEL → BO	14:10	19:20	5h 10m	1 stop
23	Air India	21/03/2019	Banglore	New Delhi	BLR → CO	22:00	13:20 19 M	15h 20m	1 stop

Main Steps to build a Model

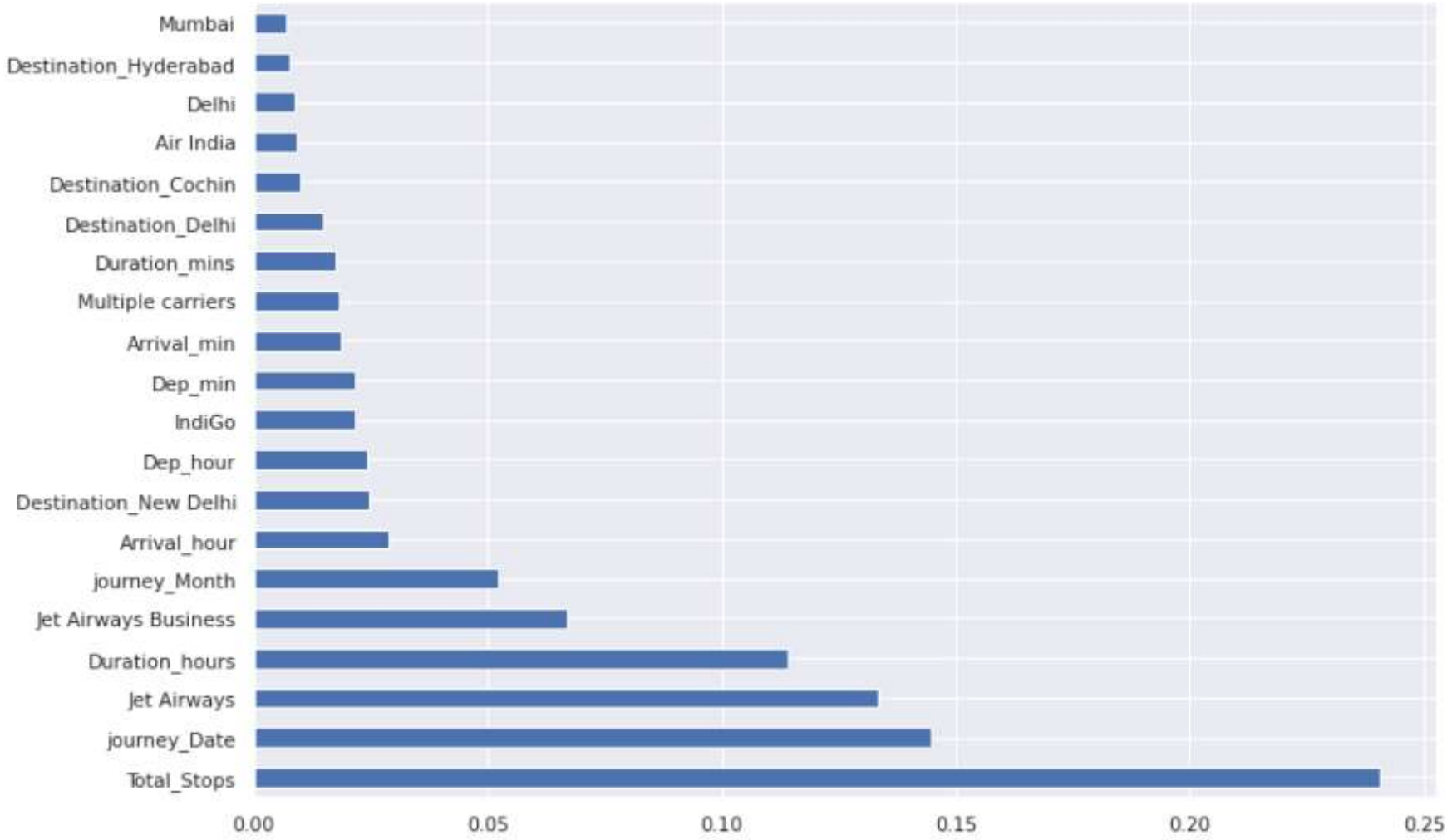


- Random Forest Regressor
- Extra Trees Regressor
- XGB Regressor
- Linear Regression

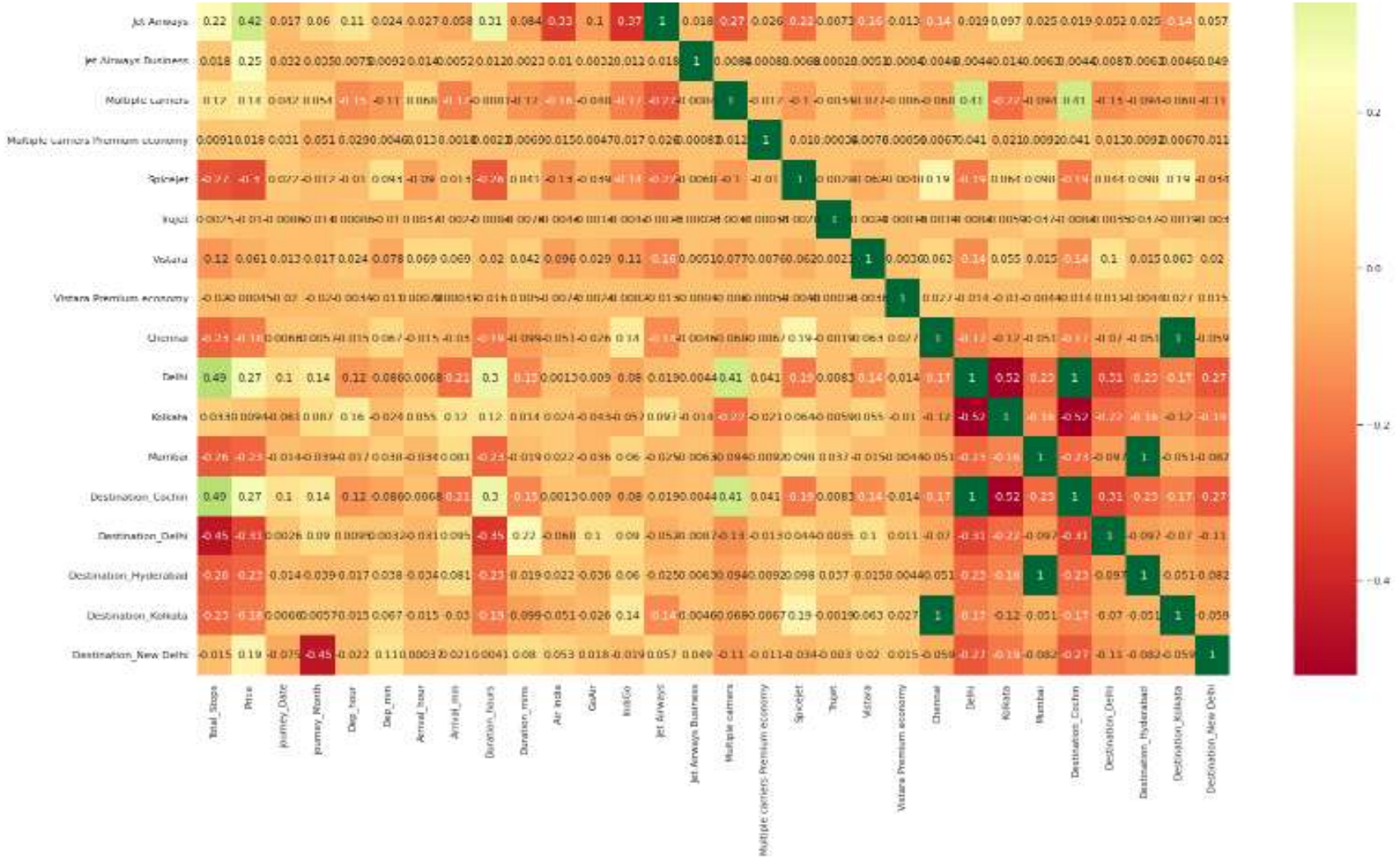
Exploratory Data Analysis (EDA)



Exploratory Data Analysis (EDA)



Exploratory Data Analysis (EDA)



Accuracy

Random Forest Regressor

```
1 metrics.r2_score(y_test,y_pred)
```

```
0.8271912137161707
```

Extra Trees Regressor

```
1 metrics.r2_score(y_test,y_pred1)
```

```
0.7799302479623234
```

XGB Regressor

```
1 metrics.r2_score(y_test,y_pred4)
```

```
0.7488919240810953
```

Linear Regression

```
1 metrics.r2_score(y_test,y_pred2)
```

```
0.5888639023134974
```

Conclusion :

- We presented a survey of airline ticket price prediction models classifying them into customer side and airline side models based on their designed goals.
- We also discussed about the strength and weaknesses of the this work.
- Our results showed us that few issues regarding performance, dataset, usage of dynamic external features such as social media data and search engine query needs to be thoroughly investigated.
- Hence, for this we suggest a machine learning and social media data based prediction model as a promising approach.

Future Aspects

- One future directions that has great potential to improve the airline ticket price prediction is to use the latest and advanced ML techniques in conjunction with valuable social media-based data.
- Various features from social media can possibly be useful to forecast airlines passenger demands and ticket prices.
- For example, sentiment analysis of different Twitter hashtags can give the idea of any event happening at the flight origin or destination city that can improve the prediction of ticket prices .

Reference

Dataset - https://www.kaggle.com/vinayshaw/airfare-price-prediction/data

Libraries - https://pandas.pydata.org/pandas-docs/stable/index.html

https://scikit-learn.org/stable/modules/ensemble.html

https://www.tensorflow.org/api_docs/python/tf/keras

Blogs - https://www.analyticsvidhya.com/machine-learning/

https://stackoverflow.com/questions/tagged/machine-learning

Thank You