# FLIGHT FARE PREDICTION WEB APP

# A PROJECT REPORT

***Submitted by***


**Yash Sharma (19MIM10068)**

**Mayank Samadhiya (19MIM10077)**

**Shivam Bajpeyi (19MIM10082)**

**Himanshu (19MIM10085)**


*in partial fulfillment for the award of the degree*

**MASTER OF TECHNOLOGY(Integrated)**

*in*

**COMPUTER SCIENCE AND ENGINEERING**

*Specialization in*

Artificial intelligence and machine learning



**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**VIT BHOPAL UNIVERSITY**

**KOTHRI KALAN, SEHORE**

**MADHYA PRADESH - 466114**


APR 2021

# VIT BHOPAL UNIVERSITY, KOTHRI KALAN, SEHORE MADHYA PRADESH – 466114

## BONAFIDE CERTIFICATE

Certified that this project report titled **"Flight Fare Prediction Webapp"** is the bonafide work of "**Yash Sharma (19MIM10068), Mayank Samadhiya (19MIM10077), Shivam Bajpeyi (19MIM10082), Himanshu (19MIM10085)"** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**PROGRAM CHAIR**
Dr. Pandimurugan Vellaisamy,
Assistant Professor
School of AI & ML division
VIT BHOPAL UNIVERSITY

**PROJECT GUIDE**
Dr. Paras Jain
Assistant Professor (Senior)
School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY

The Project Exhibition II Examination is held on _____

# ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr. Nageswara Guptha M, Head AI Division for much of his valuable support and encouragement in carrying out this work.

I would like to thank my internal guide Dr. Paras jain, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Aeronautical Science, who extended directly or indirectly all support.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

# LIST OF ABBREVIATIONS

| ABBREVIATION NO. | ABBREVIA-TIONS | FULL FORM |
|---|---|---|
| 1 | SVM | Support vector machine |
| 2 | AI | Artificial Intelligence |
| 3 | KNN | k-nearest neighbors |
| 4 | ML | Machine Learning |
| 5 | EDA | Exploratory Data Analysis |
| 6 | PLSR | Partial Least Square Regression |
| 7 | MSE | Mean Square Error |
| 8 | RMSE | Root Mean Square Error |

# LIST OF FIGURES


| FIGURE NO. | TITLE | PAGE NO. |
|---|---|---|
| 1 | Loading the data | 19 |
| 2 | EDA (Data Visualization and encoding) | 20 |
| 3 | Feature Selection(correlation) | 21 |
| 4 | Feature Selection(Feature Importance) | 21 |
| 5 | Model training and evaluation | 22 |
| 6 | WebApp Interface (Input) | 23 |
| 7 | WebApp Output (Output) | 23 |
| 8 | Scatter plot between y_pred and y_test | 26 |
| 9 | Line plot of difference between actual price and predict price | 26 |

# ABSTRACT

Someone who purchases flight tickets frequently would be able to predict the right time to procure a ticket to obtain the best deal. Many airlines change ticket prices for their revenue management. The airline may increase the prices when the demand is to be expected to increase the capacity. To estimate the minimum airfare, data for a specific air route has been collected including the features like departure time, arrival time, source, destination, route, total stops and airlines over a specific period. Features are extracted from the collected data to apply Machine Learning (ML) models. In this project, we used the random forest algorithm to predict the prices at the given time, airline, source, and destination. The derived experimental results reveal that the ML models are able to handle this regression problem with 81.51% accuracy, for a certain type of flight features.

# TABLE OF CONTENTS


| CHAPTER NO. | TITLE | PAGE NO. |
|---|---|---|
| | List of Abbreviations | **4** |
| | List of Figures | **5** |
| | Abstract | **6** |
| 1 | **INTRODUCTION**<br>1.1 Introduction<br>1.2 Motivation for the work<br>1.3 About Introduction to the project including techniques<br>1.4 Problem Statement<br>1.5 Objective of the work<br>1.6 Organization of the thesis<br>1.7 Summary | **9** |
| 2 | **LITERATURE SURVEY**<br>2.1 Introduction<br>2.2 Core area of the project<br>2.3 Existing Algorithms<br>2.3.1 Linear Regression<br>2.3.2 Decision Tree<br>2.3.3 Support Vector Machine (SVM)<br>2.3.4 k-Nearest-Neighbours (KNN)<br>2.3.6 Bagging Regression<br>2.4 Research issues/observations from literature Survey<br>2.5 Summary | **12** |

| | | |
|---|---|---|
| 3 | **SYSTEM ANALYSIS**<br>3.1 Introduction<br>3.2 Disadvantages/Limitations in the existing system<br>3.3 Proposed System<br>3.4 Summary | **17** |
| 4 | **SYSTEM DESIGN AND IMPLEMENTATION**<br>4.1 Introduction<br>4.2 Data Preprocessing<br>4.3 Exploratory Data Analysis (EDA)<br>4.4 Feature Selection<br>4.5 Model Training and Evaluation<br>4.6 Building WebApp<br>4.7 Summary | **19** |
| 5 | **PERFORMANCE ANALYSIS**<br>5.1 Introduction<br>5.2 Performance Measures (Table/text)<br>5.3 Performance Analysis(Graphs/Charts)<br>5.4 Summary | **25** |
| 6 | **FUTURE ENHANCEMENT AND CONCLUSION**<br>6.1 Introduction<br>6.2 Limitation/Constraints of the System<br>6.3 Future Enhancements<br>6.4 Conclusion | **28** |
| 7 | **REFERENCES** | **30** |

# 1. INTRODUCTION

## 1.1 Introduction

Nowadays, the airline corporations are using complex strategies and methods to assign airfare prices in a dynamic fashion. These strategies are taking into account several financial, marketing, commercial and social factors closely connected with the final airfare prices. Due to the high complexity of the pricing models applied by the airlines, it is very difficult for a customer to purchase an air ticket at 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 airfare 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). All the aforementioned works applied only a small number of ML models, with emphasis to some classical ones, to predict the airfare prices of airlines worldwide. However, to the authors' best knowledge, the performance of the state of the art ML models to this problem is still unexplored.

## 1.2 Motivation of the work

Predicting flight prices without having proper ideas about a particular airline company is near to impossible, especially when you want to book any flight real quick. Using a Machine Learning approach it becomes quite easy as the model predicts how much you need to spend on your flight expenses by getting rid of all unnecessary calculations and brain scratching thoughts. From the user's point of view this somehow proves to be a profitable project model as users can select from a wide range of airline companies whichever of their choice is and decide budget for their ongoing journey without spending anything extra.

## 1.3 About Introduction to the project including techniques

Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. We have datasets for different different flights so a simple platform like an app or website can help easily to the customers. By keeping in mind all these things we thought that we can make a simple user-friendly web app for computing expected future prices of several airlines so that the customer will get to know which flight is financially profitable to them.

## 1.4 Problem statement

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, it will be a different story. We might have often heard travelers saying that flight ticket prices are so unpredictable. As being AI students, we are gonna prove that given the right data anything can be predicted. Here in this project we will be provided with prices of flight tickets for various airlines

## 1.5 Objective of our work

Being an AI student, we thought why not do a detailed analysis on the data collected from the flights and understand how the whole process works. Once we get some insights through analysis, we can develop a simple, easy to use web app which will help customers to find cheaper flight tickets.

## 1.6 Organization of the thesis

Chapter 1: Introduction

Chapter 2: Literature Survey

Chapter 3: Project Procedure

Chapter 4: Work done

Chapter 5: Observation

Chapter 6: Result & Conclusion

Chapter 7: Recommendation for future work

Chapter 8: References

## 1.7 Summary

Dataset related to our work gives the information about the highs and lows in the airfares according to the days, weekends and time of the day that is morning, evening and night also the machine learning models in the computational intelligence fields that are evaluated before on different datasets are studied. Their accuracy and performances are evaluated and compared in order to get better results. For the prediction of the ticket prices perfectly different prediction models are tested for the better prediction accuracy. As the pricing models of the company are developed in order to maximize the revenue management. So to get results with maximum accuracy regression analysis is used. From the studies , the features that influence the prices of the ticket are to be considered. In future the details about the number of available seats can improve the performance of the model.

# 2. LITERATURE SURVEY

## 2.1 Introduction

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 the air ticket will be minimum. The majority of these systems are utilizing the modern computerized system known as Machine Learning.

## 2.2 Core area of the project

We are providing a platform to the travellers to compare the predicted fare of different airlines. We will take inputs from the user using our webapp and according to these inputs like date, destination, and duration, etc. the predicted price will be calculated and shown to the user and hence the user can easily compare the price of several different flights and get the cheaper fare.

## 2.3 Existing algorithms

For predicting the flight ticket prices, many algorithms are introduced in machine learning. The algorithms are: Support Vector Machine (SVM), Linear regression, K-Nearest neighbours, Decision tree, Multilayer Perceptron, Gradient Boosting and others also. Using python library scikit learn these models have been implemented. The parameters like R-square, MAE and MSE are considered to verify the performance of these models:

### 2.3.1 Linear Regression

To determine the correlation between two continuous variables, simple linear regression analysis is used. One of the two variables is the predictor variable of which value is to be found. It gives

the statistical relationship not the deterministic relationship between two variables. Linear regression algorithm gives the best fit line to the given data for which the prediction error is minimum. Gradient descent and cost function are the two major factors to understand linear regression. The equation for linear regression is :

$$y(pred) = b0+b1 * x \quad \_\_\_\_(1)$$

The value of coefficients b1 and b0 are chosen so that the error value is as small as possible. The square of predicted and actual value difference gives the error. To deal with the negative values, the mean square error is taken (MSE). Here b0 gives the positive or negative relationship between the x and y, whereas b1 is called bias. The accuracy of the regression problem is measured in terms of R-squared, MAE, and MSE.

### 2.3.2 Decision Tree

This tree count isolates the collected information into small subsets, at a comparative same time makes it persistent. The last results show the tree having the decision centres, likewise, the leaf centres. This decision centre point may contain two branches at any rate. At first think about the whole informational index as root. Feature regards are kicked out of the chance. In case the characteristics are relentless then they have to be discretized before structuring the model. In view of estimation property records are corrected recursively. Information Gain and Gini index are two essential properties in the decision of tree computation. Information Gain is defined as the amount change in entropy. Higher entropy indicates more effectiveness of the substance. Thus the entropy is a proportion of vulnerability of arbitrary variables. Gini Index measures how regularly an arbitrarily picked component would be falsely recognized. It implies a characteristic with a lower Gini index ought to be liked. For a Regression tree, cost capacity can be a basic squared condition.

### 2.3.3 Support Vector Machine (SVM)

The SVM is supervised ML algorithm is used for classification and regression analysis. It takes a large time to process so usually applied to the small datasets. It finds the Hyperplane that separates the feature into different parts. It gives an optimal hyperplane which classifies the different domains. The data points which are nearest to the hyperplane are called support vector points and distance between the vector planes and these points are called as margins.

The proposed work has exploited SVM for regression analysis. The performance depends on kernel function selection as a nonparametric technique. Linear, Radial Basis Function and Polynomial are the kernels of support vector machine algorithms.

### 2.3.4 K-Nearest Neighbours(KNN)

In k-nearest neighbour regression analysis, the output is mean of its k nearest neighbours. Like SVM this is also a nonparametric method. Considering few values, results are computed to achieve the best value. KNN is a supervised classification algorithm that can also be used as a regressor. It assigns a new data point to the class. It is non-parametric because it does not take any assumptions. It calculates the distance between every training example and a new data point. To compute this distance following distance calculation methods are used:

- Euclidean Distance
- Manhattan distance
- Hamming distance

K- entries in the dataset are picked by the model that are close to the new data point.

### 2.3.5 Bagging Regression

Tree The drawback of decision trees is, it has a large bias with simple trees and large variance with complex trees. Bagging comes from Bootstrap aggregating, a method of selecting a random number of data from a dataset with replacement. It is mostly used to reduce the variance of the tree. From literature it is clear that maximum accuracy is achieved by gradient boosting and random forest methods.

## 2.4 Research issues/observations from literature Survey

It is hard for the client to buy an air ticket at the most reduced cost. For this few procedures are explored to determine time and date to grab air tickets with minimum fare rate. The majority of these systems are utilizing the modern computerized system known as Machine Learning. To determine the ideal purchase time for flight tickets Gini and Groves exploited Partial Least Square Regression(PLSR) for building up a model. The information was gathered from major travel adventure booking sites from 22 February 2011 to 23 June 2011. Extra information was additionally gathered and are utilized to check the correlations of the exhibitions of the last model. Janssen implemented a desired model using the Linear Quantile Blended Regression methodology for San Francisco–New York course. Two features such as number of days for departure and whether departure is on weekend or weekday are considered to develop the model. The model guesses airfare well in advance from the departure date. But the model isn't convincing in a situation for an extensive time allotment, it closes the departure date. Wohlfarth proposed a ticket purchasing time improvement model subject to a significant pre-processing known as macked point processors, data mining frameworks ( course of action and grouping) and quantifiable examination system. This framework is proposed to change various added value arrangements into included added value arrangement heading which can support solo gathering estimation. This value heading is packed into a gettogether reliant on near evaluating conduct. Headway models measure the value change plans. A tree-based analysis used to pick the best

planning gathering and a short time later looking at the progression model. An investigation by Dominguez-Menchero suggests the perfect purchase timing reliant on a nonparametric isotonic backslide technique for a specific course, carriers, and time frame. The model provides the most acceptable number of days before buying the flight ticket. The model considers two types of variables such as the entry and its date of obtainment.

## 2.5 Summary

We gathered airfare data from a specific Greek airline corporation (Aegean Airlines) from the web and showed that it is feasible to predict prices for flights based on historical fare data. The experimental results show that ML models are a satisfactory tool for predicting airfare prices. Other important factors in airfare prediction are the data collection and feature selection from which we drew some useful conclusions. From the experiments we concluded which features influence the airfare prediction at most. Apart from the features selected, there are other features that could improve the prediction accuracy. In the future, this work could be extended to predict the airfare prices for the entire flight map of the airline. Additional experiments on larger airfare data sets are essential, but this initial pilot study highlights the potential of Machine Learning models to guide consumers to make an airfare purchase in the best market period.

# 3. SYSTEM ANALYSIS

## 3.1 Introduction

Machine Learning algorithms are applied on the dataset to predict the dynamic fare of flights. This gives the predicted values of flight fare to get a flight ticket at minimum cost. The values of R-squared obtained from the algorithm give the accuracy of the model.

## 3.2 Disadvantages/Limitations in the existing system

From the very beginning people try to book flights at cheaper cost so for doing that they follow some of these facts-

- Generally, flights after 6 pm are cheaper so people prefer to book flights at that time.
- Airfares usually go up on Friday and start declining on Monday or Tuesday whereas Wednesday and Thursday are the cheapest days to book airline tickets.
- According to a survey, to book a ticket just 60 days before the flight is considered to be good because in such a situation, we will get the ticket very cheap, the more they delay, the more expensive the ticket will be.
- People also use some private companies' websites or app to buy the tickets for their journey but they don’t know, Is that price cheaper or not?

## 3.3 Proposed System

- We are providing the platform to the travellers to compare the predicted fate of different airlines according to the date , duration and destination , etc. so that they can get the cheaper flight tickets.
- Do feature engineering and data preprocessing for the factors involved in fare prediction.
- Analyze the data and bring some meaningful insights using exploratory data analysis (EDA).
- Build a user friendly **'Web App'** that can predict airfare of various airlines to get flights at very low.

## 3.4 Summary

- The trend of flight prices vary over.
- The airfare varies depending on the Time of departure , by making the time slot used in analysis an important parameter.
- Airfare varies according to the day of the week of travel.
- There are a few times when an offer is run by an airline because of which the prices drop suddenly. These are difficult to incorporate in our mathematical models, and hence lead to error.

# 4. SYSTEM DESIGN AND IMPLEMENTATION

## 4.1 Introduction

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, it will be a different story. We might have often heard travelers saying that flight ticket prices are so unpredictable. As data scientists, we are gonna prove that given the right data anything can be predicted.

## 4.2 Data Preprocessing

In this part we have basically checked the data, how the data is and other descriptions related to data like max, min values, mean, etc. We have also checked whether the dataset contains null values or not and then we have removed the null values.

```
In [2]:  1 train_data = pd.read_excel(r"Data_Train.xlsx")
         2 train_data.shape

Out[2]: (10683, 11)

In [3]:  1 pd.set_option('display.max_columns', None)

In [4]:  1 train_data.head()
```

Out[4]:

| | Airline | Date_of_Journey | Source | Destination | Route | Dep_Time | Arrival_Time | Duration | Total_Stops | Additional_Info | Price |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | IndiGo | 24/03/2019 | Banglore | New Delhi | BLR → DEL | 22:20 | 01:10 22 Mar | 2h 50m | non-stop | No info | 3897 |
| 1 | Air India | 1/05/2019 | Kolkata | Banglore | CCU → IXR → BBI → BLR | 05:50 | 13:15 | 7h 25m | 2 stops | No info | 7662 |
| 2 | Jet Airways | 9/06/2019 | Delhi | Cochin | DEL → LKO → BOM → COK | 09:25 | 04:25 10 Jun | 19h | 2 stops | No info | 13882 |
| 3 | IndiGo | 12/05/2019 | Kolkata | Banglore | CCU → NAG → BLR | 18:05 | 23:30 | 5h 25m | 1 stop | No info | 6218 |
| 4 | IndiGo | 01/03/2019 | Banglore | New Delhi | BLR → NAG → DEL | 16:50 | 21:35 | 4h 45m | 1 stop | No info | 13302 |

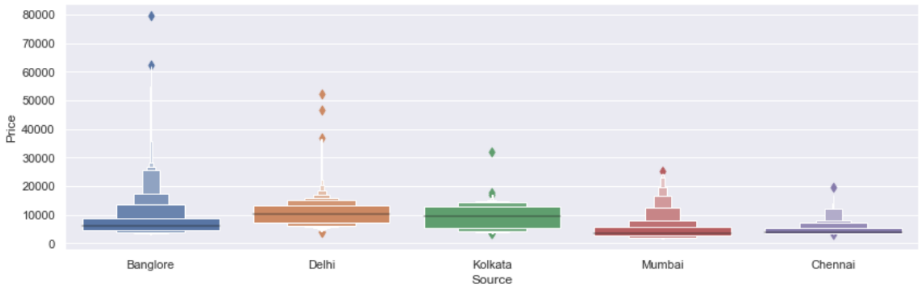
**Figure 1: loading the data**

## 4.3 Exploratory Data Analysis (EDA)

The main purpose of EDA is to look at data before making any assumptions. In this part we have handled categorical data by one hot encoding and label encoding for nominal and ordinal data respectively as well as we have done data visualization to read the pattern of the data.

In [25]:
```
# Source vs Price

sns.catplot(y = "Price", x = "Source", data = train_data.sort_values("Price", ascending = False),
            kind="boxen", height = 4, aspect = 3)
plt.show()
```




In [26]:
```
# As Source is Nominal Categorical data we will perform OneHotEncoding

Source = train_data[["Source"]]

Source = pd.get_dummies(Source, drop_first= True)

Source.head()
```

Out[26]:

| | Source_Chennai | Source_Delhi | Source_Kolkata | Source_Mumbai |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 1 | 0 |
| 4 | 0 | 0 | 0 | 0 |

**Figure 2: EDA (Data Visualization and encoding)**

## 4.4 Feature Selection

To find out the important features for our model that have good relation with the target variable, we are using heatmap, feature importance, selectKbest methods.

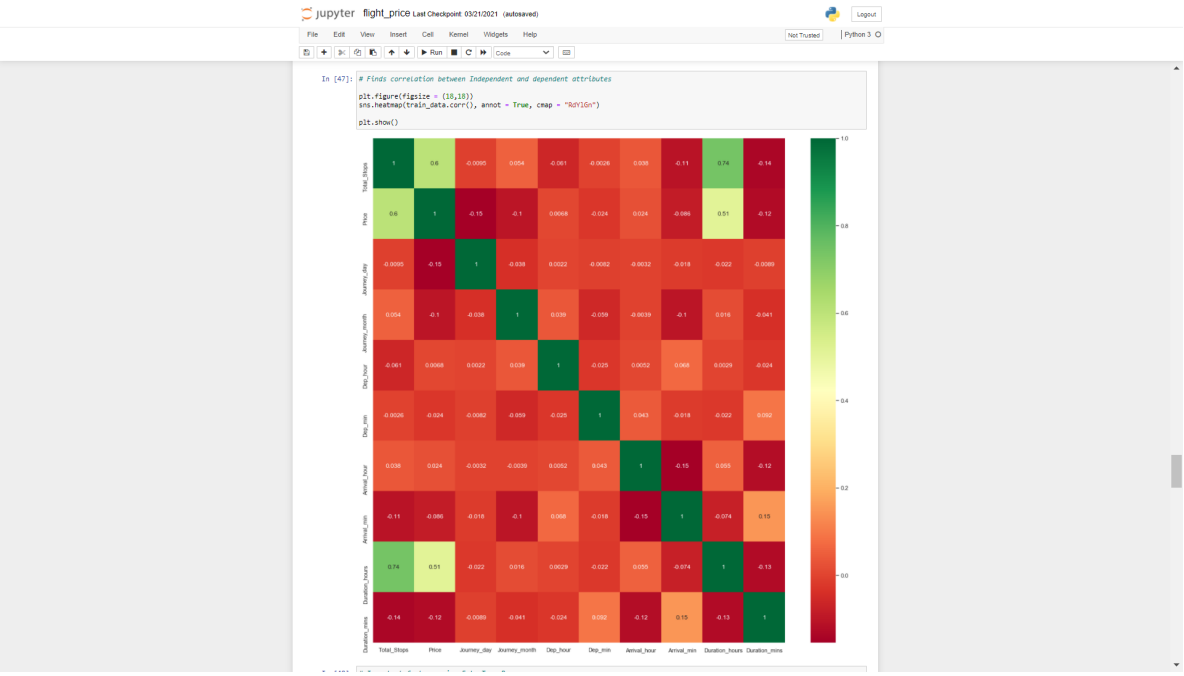



**Figure 3: Feature Selection(correlation)**

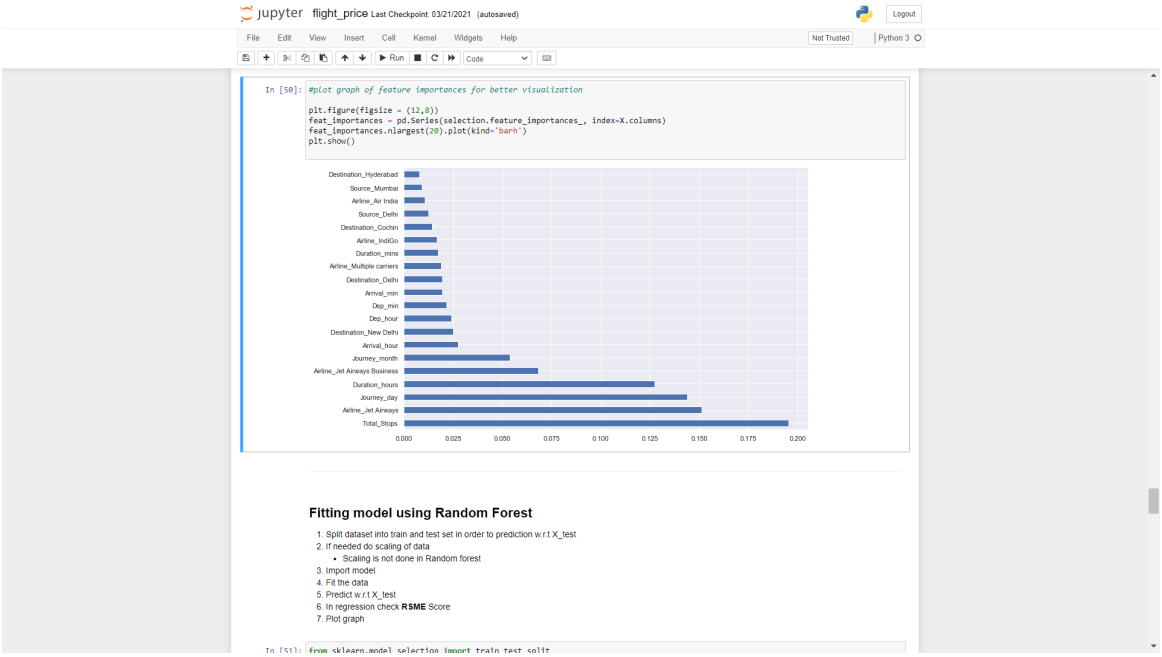
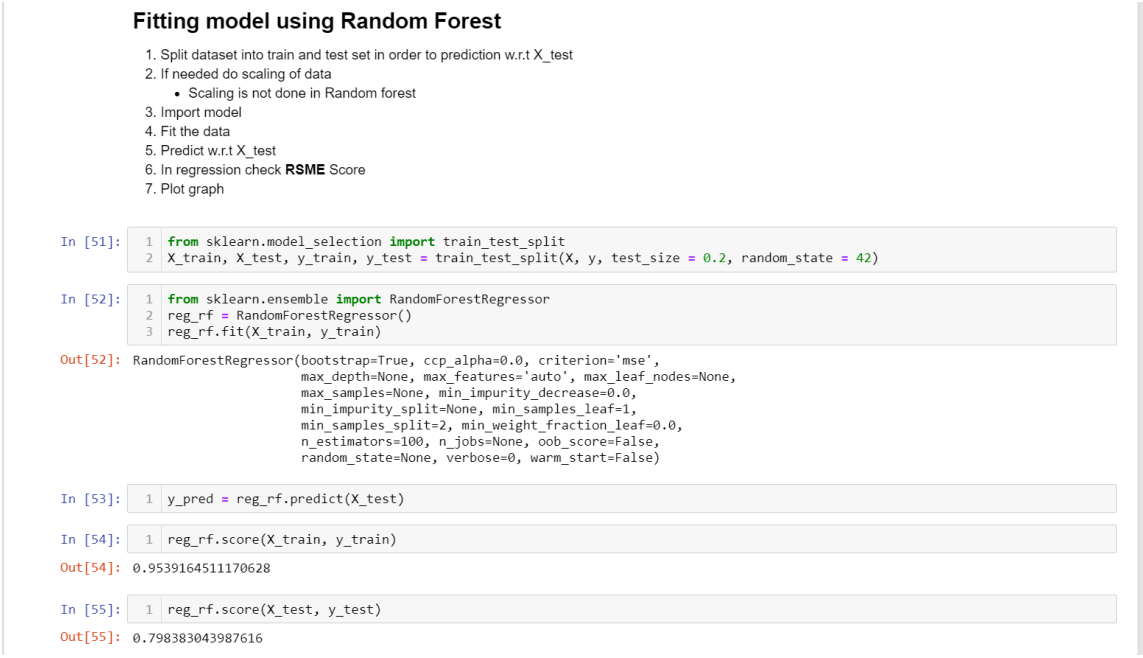



**Figure 4: Feature Selection(Feature Importance)**

## 4.5 Model Training and Evaluation

For training the model, we have split the dataset into train and test, then fit the model , and evaluated its performance.



**Fitting model using Random Forest**

1. Split dataset into train and test set in order to prediction w.r.t X_test
2. If needed do scaling of data
   - Scaling is not done in Random forest
3. Import model
4. Fit the data
5. Predict w.r.t X_test
6. In regression check **RSME** Score
7. Plot graph

```
In [51]:
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 42)

In [52]:
1 from sklearn.ensemble import RandomForestRegressor
2 reg_rf = RandomForestRegressor()
3 reg_rf.fit(X_train, y_train)

Out[52]: RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                               max_depth=None, max_features='auto', max_leaf_nodes=None,
                               max_samples=None, min_impurity_decrease=0.0,
                               min_impurity_split=None, min_samples_leaf=1,
                               min_samples_split=2, min_weight_fraction_leaf=0.0,
                               n_estimators=100, n_jobs=None, oob_score=False,
                               random_state=None, verbose=0, warm_start=False)

In [53]:
1 y_pred = reg_rf.predict(X_test)

In [54]:
1 reg_rf.score(X_train, y_train)

Out[54]: 0.9539164511170628

In [55]:
1 reg_rf.score(X_test, y_test)

Out[55]: 0.798383043987616
```

**Figure 5: Model training and evaluation**

## 4.6 Building WebApp

For webapp development, we are using the flask framework and will make a user-friendly web app.

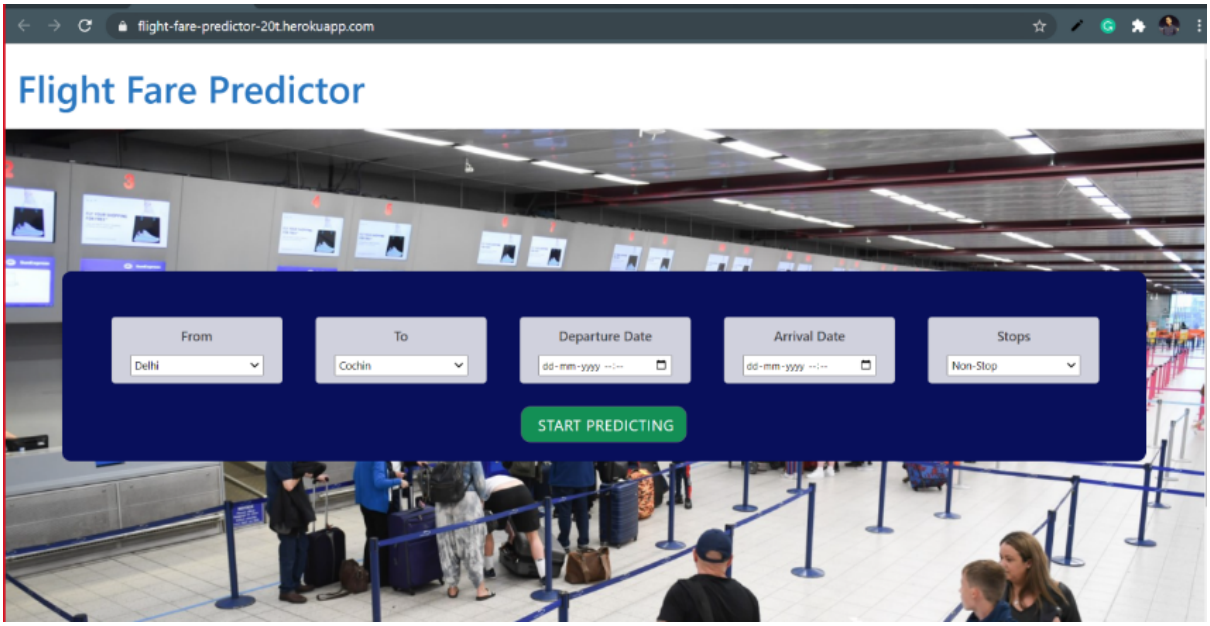



**Figure 6: WebApp Interface (Input)**

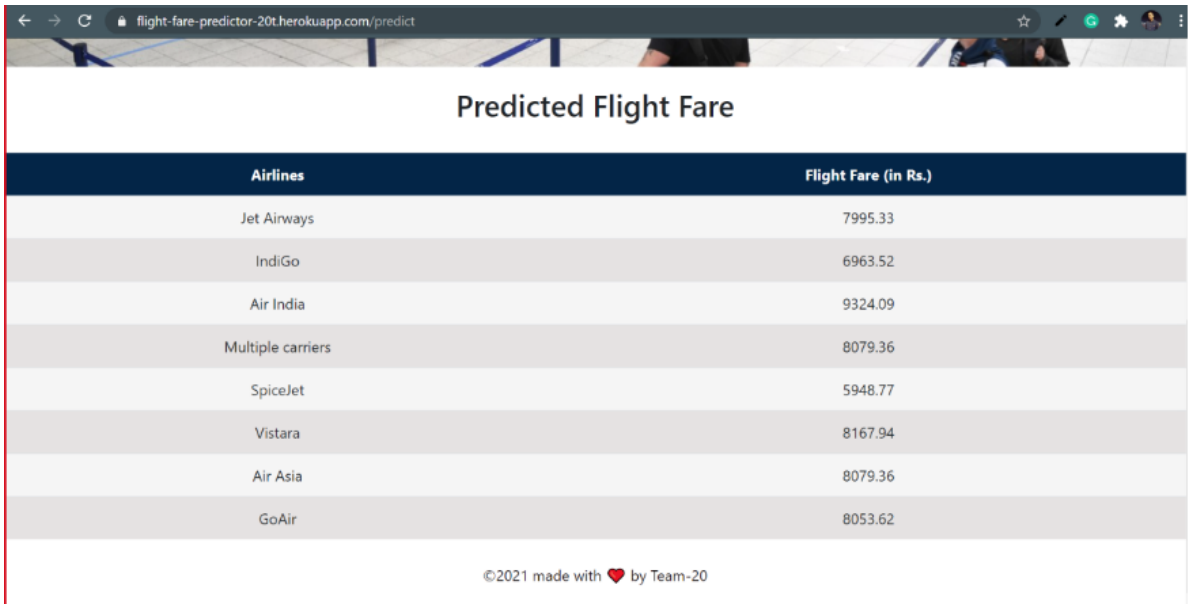



**Figure 7: WebApp Interface (Output)**

## 4.7 Summary

In this type of problem Feature Engineering is the most crucial thing . You can see how we have handled the categorical and numerical data and also how we build different ML models on the same dataset . We also check the RMSE score of each model so that we can understand how it should perform in our test dataset . At last You can also further improve the Model by Running different parameters which are being used in the model .

# 5. PERFORMANCE ANALYSIS

## 5.1 Introduction

After building the machine learning model, the most important part is to analyse the performance of that model. For that purpose we have used line plot and scatter plot for graphical visualization and compared the accuracy between predicted value and actual value. As well as we have calculated the mean absolute error and root mean squared error.

## 5.2 Performance Measure

For flight fare prediction, We have made our ML model using a random forest algorithm and calculated the accuracy of the model which is 81.51%. We have also found the mean absolute error and root mean squared error which are 1155.30 and 1950.95 respectively.

## 5.3 Performance Analysis

To analyse performance of the model, we have plotted the scatter plot and check the relationship between predicted and actual value. This graph shows the highly positive correlation between those values. Then, we also plotted the line plot of difference between actual and predicted value. The figure[8] shows the scatter plot and the figure[9] shows the line plot.

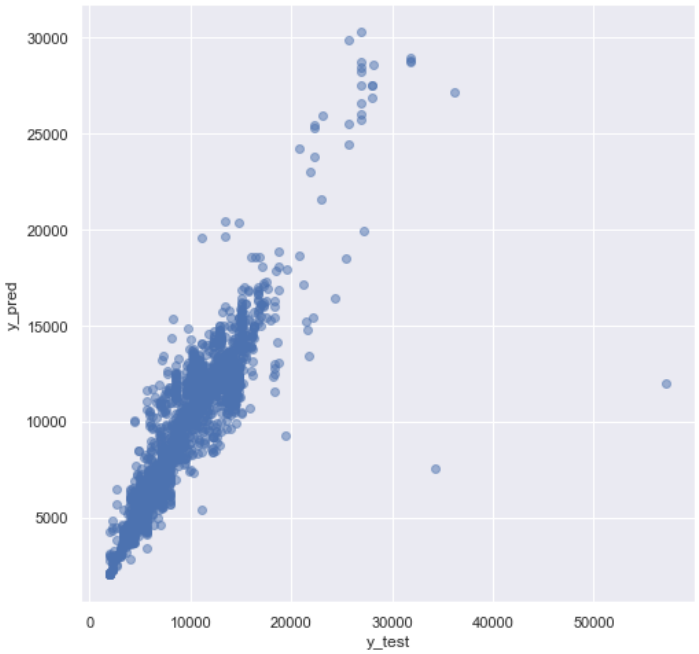



**Figure 8: Scatter plot between y_pred and y_test**

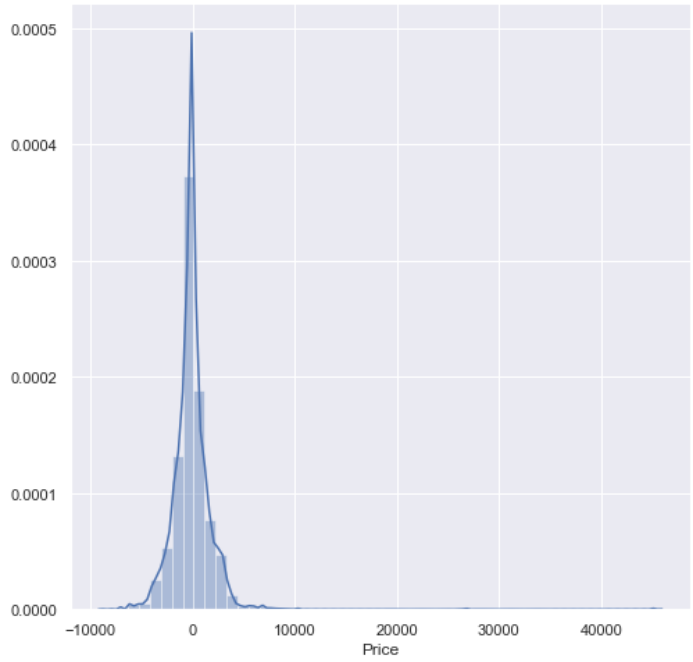



**Figure 9: Line plot of difference between actual price and predict price**

## 5.4 Summary

The accuracy of the price prediction model using random forest algorithm is 81.17% which is greater than the accuracy of other models that we have studied in the research papers. The scatter plot between predicted and actual value is also positively correlated and the line plot of difference between these values is also normally distributed.

# 6. FUTURE ENHANCEMENT AND CONCLUSION

## 6.1 Introduction

In this type of problem Feature Engineering is the most crucial thing . You can see how we have handled the categorical and numerical data and also how we build different ML models on the same dataset . We also check the RMSE score of each model so that we can understand how it should perform in our test dataset . At last You can also further improve the Model by running different parameters which are being used in the model

## 6.2 Limitation/Constraints of the System

- User can not be predict more than one airlines prize at a single time
- Data is limited so user can not predict from more cities to more cities airlines fare prize
- User can only predict the Airlines fare prize not book the tickets from our webapp
- It can not be predict the actual value of airlines fare prize

## 6.3 Future Enhancements

- More routes can be added and the same analysis can be expanded to major airports and travel routes in India.
- The analysis can be done by increasing the data points and increasing the historical data used.That will train the model better giving better accuracies and more savings.
- More rules can be added in the Rule based learning based on our understanding of the industry,also incorporating the offer periods given by the airlines.
- Developing a more user friendly interface for various routes giving more flexibility to the user.

## 6.4 Conclusion

- Flight prices almost always remain constant or increase between the major cities
- Tourist routes and routes that offer services involving Tier-2 cities of the country have uneven trends related to the increase and decrease of airline ticket prices
- The model in the worst case almost breaks even with the profits and losses, and in most cases saves an average of about Rs. 200 per transaction when predicting to wait.
- Routes with data collected over a longer duration of time tend to facilitate much more accurate predictions in the model and thus lead to higher average savings

# 7. REFERENCES


- Vu, V.H., Minh, Q.T., Phung, P.H., 2018. An airfare prediction model for developing markets. In: 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 765–770.

- Chen, Y., Cao, J., Feng, S., Tan, Y., 2015. An ensemble learning-based approach for building airfare forecast service. In: 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, 2015, pp. 964-969.

- Liu, T., Cao, J., Tan, Y., Xiao, Q., 2017. ACER: An adaptive context-aware ensemble regression model for airfare price prediction. In: 2017 International Conference on Progress in Informatics and Computing (PIC), Nanjing, 2017, pp. 312–317.

- K. Tziridis, T. Kalampokas, G.A. Papakostas, K.I. Diamantaras: Airfare prices prediction using machine learning techniques, 25th European Signal Processing Conference (EUSIPCO): Kos, 2017 (2017), pp. 1036-1039