



# **Walchand College of Engineering, Sangli.**

**(An Autonomous Institute)**

**Department  
of  
Computer Science and Engineering**

**Mini-Project Synopsis  
on**

***Flight Price Prediction***

by

Mahendra Gharge(2018BTECS00033)  
Shreyash Marwade (2018BTECS00107)  
Aditya Zade (2018BTECS00111)

Under the Guidance  
of  
**Prof. N. L. Mudegol**  
**Guide**

**Dr. M. A. Shah**  
HOD

Computer Sci. & Engg. Dept,  
WCE, Sangli

2020-



### **1) 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 data scientists, we are going to prove that given the right data anything can be predicted. Here you will be provided with prices of flight tickets for various airlines between the months of March and June of 2019 and between various cities.

### **2) Abstract:**

Travelling is one of the most entertaining things that everybody wants to avoid city crowds. Going to another island with a unique nature brings a new perspective about new things. The Indian, one of the most entertaining cities with its uniqueness, brings a lot of wonderful islands. Based on the Outlook of the Indian travel Industry, the tourism industry contributes around \$98 billion in 2018 and is expected to arise in the next period. This trend brings India as one of the immense potential countries because of the country's rich cultural and geographical diversity. These trends also bring uncertainty about the flight ticket price. It can be hard to guess the flight ticket price when we check it today compared to the other day. The tourists who want to visit a new place in India should know the ticket price in order to get the cheapest and certain ticket price with their needs. This gap brings the idea to make a prediction about the flight tickets in order to make the tourists easier to book their tickets with their needs.

### **3) Customer Domain:**

Our customer domain consists of frequent travelers, explorers and the general public.

### **4) Objectives:**

1. Learning the fundamentals of Data Science, Machine Learning and different models used for classification.
2. Collecting and cleaning a dataset and feature engineering it.
3. Designing a machine learning model for prediction of flight with maximum accuracy.
4. Uploading the project on GitHub for version controlling and open sourcing.

## 5) Outcomes:

1. Getting familiar with the concepts of Data Science, Machine Learning.
2. A fairly accurate and reliable machine learning model that can be used for flight price prediction
3. Providing a time-efficient way for the estimation of airfare.

## 6) Project Potentials:

This is a student project made for educational and learning purposes only.

## 7) Project plan (Schedule- Gantt chart (LA wise), work distribution, budget (if any))

### Phase 1:

- Learning the fundamentals of Data Science, Machine Learning and different models used for classification.
- Collecting and cleaning a dataset and feature engineering it.

### Phase 2:

- Data featurizing and building machine learning model for given data.

### Phase 3:

- Performance optimization of model for better accuracy of prediction.

### Phase 4:

- Deployment of model and open sourcing on Github.

**Work distribution:** Our project is divided into 3 parts-

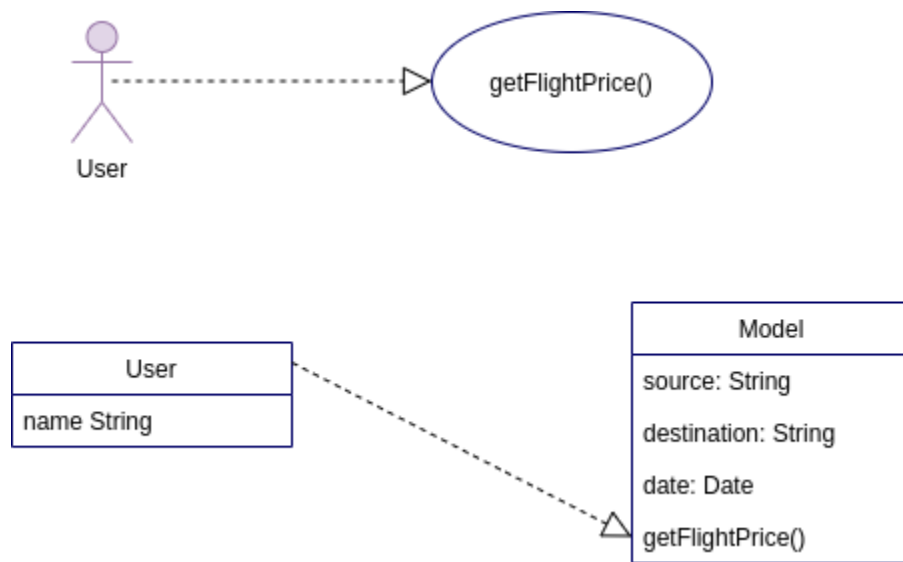
- Learning fundamentals of Data Science , Machine learning and various model used in its classification
- Data featurizing and building machine learning model for given data.
- Training and testing of data and working on accuracy of each model and selecting further.

Budget- Nill.

## 8) Methodology :

1. **Tech-Stack** : Machine Learning will be used for training and testing the model..
2. **Algorithms**: Linear Regression, Random forest Regression, Decision tree Regression, KNN Regression.
3. **Techniques**:
  - One hot encoding to ensure working with numerical data.
  - Choosing the best algorithm based upon the accuracy of the model .

## 9) UML diagrams :



## 10) Test cases :

- We will be selecting models with a minimum accuracy.
- We will be testing our model with features like bar graph histogram, r-squared, confusion matrix.
- We will testing and comparing different ml models on our dataset by their R-square number. The models are-
  1. Linear Regression
  2. Random forest Regression,
  3. Decision tree Regression
  4. KNN Regression.

## 11) References

- a) Viet Hoang Vu, Quang Tran Minh and Phu H. Phung, "An Airfare Prediction Model for Developing Markets", IEEE paper 2018.
- b) S.B. Kotsiantis, —Decision trees: a recent overview, Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.
- c) L. Breiman, —Random forests, Machine Learning, vol. 45, pp. 5- 32, 2001.
- d) Wohlfarth, T. Clemencon, S. Roueff, —A Data mining approach to travel price forecasting, 10 th international conference on machine learning Honolulu 2011.
- e) <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cad=rja&uact=8&ved=2ahUKEwjw37S2-pHwAhWSbn0KHfPCDE4QFjAAegQIAxAD&url=http%3A%2F%2Fwww.ijstr.org%2Ffinal-print%2Fdec2019%2FPredicting-The-Price-Of-A-Flight-Ticket-With-The-Use-Of-Machine-Learning-Algorithms.pdf&usg=AOvVaw0wY6-JKLbCUFFwktmFsALS>