iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** II **Month of publication:** February 2021

DOI: https://doi.org/10.22214/ijraset.2021.33115

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue II Feb 2021- Available at www.ijraset.com

A Machine Learning Approach to Predict Price of Airlines Tickets

Pranita Rajure[1], Samiksha Bankar[2], Harsimran Bakshi[3], Bhakti Patil [4]
[1, 2, 3]*Student, Computer Department, AISSMSCOE*
[4]*Faculty, Computer Department, AISSMSCOE*

Abstract: ***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 airfares at different times prior to the departure dates. On the other hand, the buyers are only able to access limited information only which is not enough to predict the airfare prices. Considering the features such as departure time, the number of days left for departure and time of the day it will give the best time to buy the ticket. Features are extracted from the collected data to apply Machine Learning (ML) models. Then using this information, we are intended to build a system that can help buyers whether to buy a ticket or not.***
Keywords: Airfare price, Airfare changes, Machine learning algorithm, predictive models

I. 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 [1]. Due to the high complexity of the pricing models applied by the airlines, it is very difficult to a customer to purchase an air ticket in the lowest price [2].

With the explosive growth of the internet and e-commerce, air passengers nowadays can check airfare and availability of any airlines around the world easily [4]. When satisfying with an airfare, these customers can purchase their desired tickets online through official airline or agent websites. Prediction of airfare provides buyers best time to buy tickets. So, we are using a machine learning approach to predict the flight ticket prices [6].

Machine learning is classified into supervised and unsupervised algorithms. The supervised algorithms consist of the pre-determined set of data which is provided for training of system, where the system predicts the results based on the previous training data. Whereas, in case of unsupervised algorithms the system tries to find the patterns from the example provided [3] [5].

We have used random forest algorithm which is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML [7]. It is based on the concept of **ensemble learning,** which is a process *of* combining multiple classifiers to solve a complex problem and to improve the performance of the model. With that said, random forests are a strong modelling technique and much more robust than a single decision tree. They aggregate many decision trees to limit over fitting as well as error due to bias and therefore yield useful results [8].

II. LITERATURE SURVEY

This paper, 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 and it uses Regression Tree Algorithm for prediction of airfare [1].

In this paper, Random Forest Algorithm is used for prediction they combined two public datasets (DB1B and T-100) and predicts quarterly average airfare price with an adjusted R squared score of 0.869 and it uses Random Forest Algorithm for prediction [2].

This paper not only predicts the airfare but also identify which airfare's features that have the strongest impacts on the airfare changes. It uses Random Forest and Multilayer Perceptron Algorithms [4].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue II Feb 2021- Available at www.ijraset.com

III. METHODOLOGY

A. Random Forest Algorithm

Random Forest Algorithm *can* be used for both Classification and Regression problems in ML. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

1) There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
2) The predictions from each tree must have very low correlations.

Below are some points that explain why we should use the Random Forest algorithm:

a) It takes less training time as compared to other algorithms.
b) It predicts output with high accuracy, even for the large dataset it runs efficiently.
c) It can also maintain accuracy when a large proportion of data is missing.

The below diagram explains the working of the Random Forest algorithm:

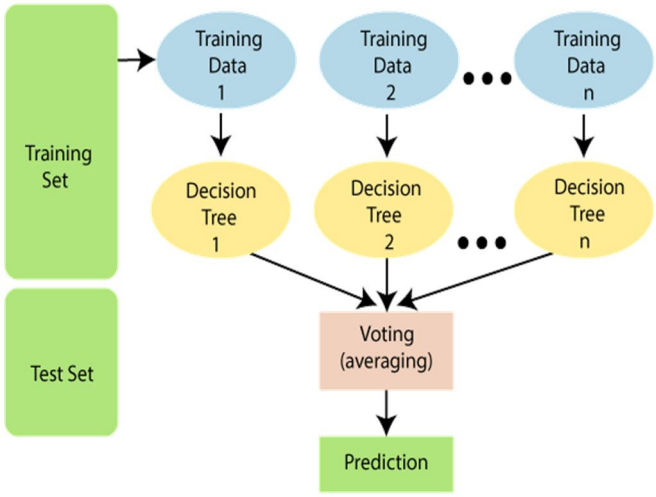


Fig 1 Random Forest Algorithm

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

- *Step-1:* Select random K data points from the training set.
- *Step-2:* Build the decision trees associated with the selected data points (Subsets).
- *Step-3:* Choose the number N for decision trees that you want to build.
- *Step-4:* Repeat Step 1 & 2.
- *Step-5:* For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue II Feb 2021- Available at www.ijraset.com

IV. PROPOSED WORK

The aim of the implementation is to predict the airfare using Random Forest (RF) ML Regression Algorithm. Early many researchers have proposed a model for predicting the tickets based on various machine learning algorithms like regression models, such as Linear Regression (LR), Support Vector Machines (SVMs), etc. We are using random forest algorithm because it gives a good accuracy compared to other ML algorithms.

The major goals of our new system are as follows:

1) To provide buyers best time to buy tickets.
2) To provide lowest price of tickets to buyers.
3) To design a system that can predict the flight fare so that a buyer can buy a ticket with the minimum fare.

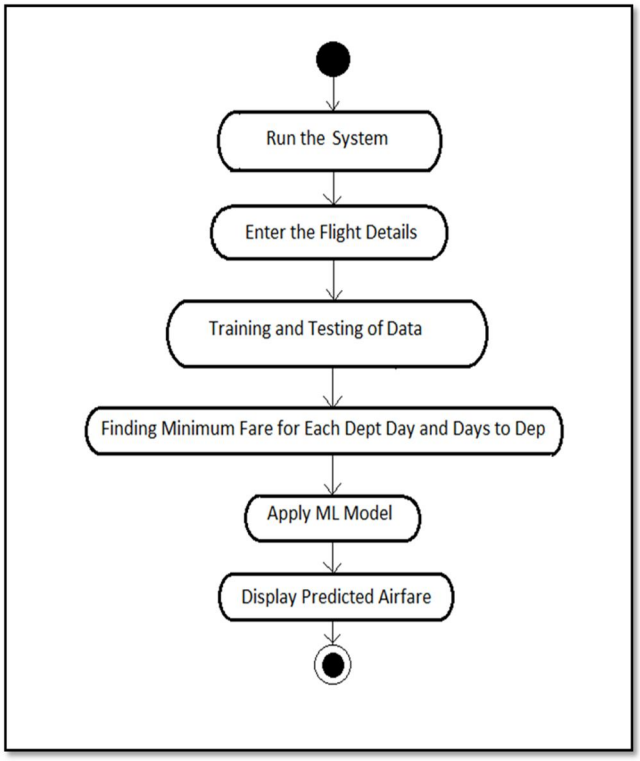


Fig 2 Proposed System Flow Diagram

A. Datasets

We are using the dataset which consist of flight tickets for various airlines between the months of March and June of 2019 and between various cities. Size of training set: 10683 records and Size of test set: 2671 records.

Following are the features which dataset contains:

1) *Airline:* The name of the airline.
2) *Date_of_Journey:* The date of the journey
3) *Source:* The source from which the service begins.
4) *Destination:* The destination where the service ends.
5) *Route:* The route taken by the flight to reach the destination.
6) *Dep_Time:* The time when the journey starts from the source.
7) *Arrival_Time:* Time of arrival at the destination.
8) *Duration:* Total duration of the flight.
9) *Total_Stops:* Total stops between the source and destination.
10) *Additional_Info:* Additional information about the flight
11) *Price:* The price of the ticket

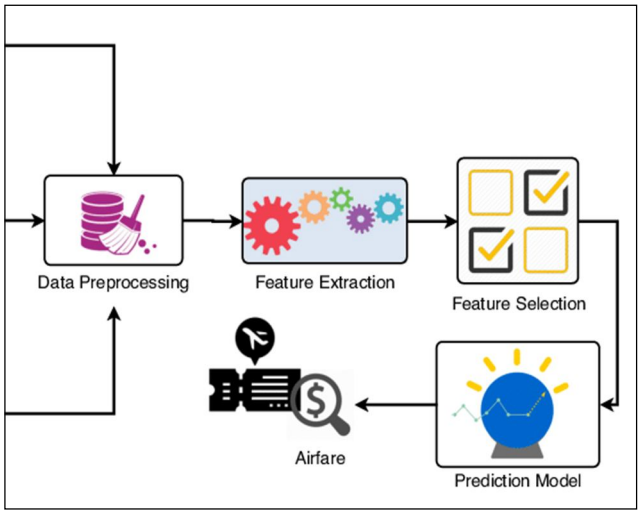


Fig 3 Architecture Diagram

B. Data Pre-Processing

Before building model, the data should be properly preprocessed and converted to quality, clean data even the resulting machine learning model will be of great quality.The data pre-processing includes three main parts that is data integration, data cleaning, data transformation. In data integration the data collected from various sources are integrated. In data cleaning process the data containing the null values, unnecessary rows with null values are being cleared. The data transformation includes the feature scaling, categorial data,etc to set the certain range of data.

A good data preprocessing in machine learning is the most important factor that can make a difference between a good model and a poor machine learning model. So, we need to do pre-process the data to get the perfect accuracy.

C. Feature Extraction

In this the features which are provided in data set are converted into the its easier form that is it means transforming raw data into a feature vector which helps in building the model. Making the data or the feature in its easier format would help to train the our model easily also it provides the a good accuracy rate.

D. Feature Selection

In the feature selection, we can find out the best feature which will contribute and have good relation with target variable. Some of the feature selection methods are Heatmap, FeatureImportances, selectKbest. In this we can find out which feature of data is independent and dependent, also to find out which feature cointains the most importance we can do this by using the Extra tree Regressor.

E. Fitting the Model

For fitting of the model we split dataset into train and test set in order to prediction. After splitting, you will train the model on the training set and perform the predictions on the test set. After training, check the accuracy using actual and predicted values. Thus, you can predict the price.

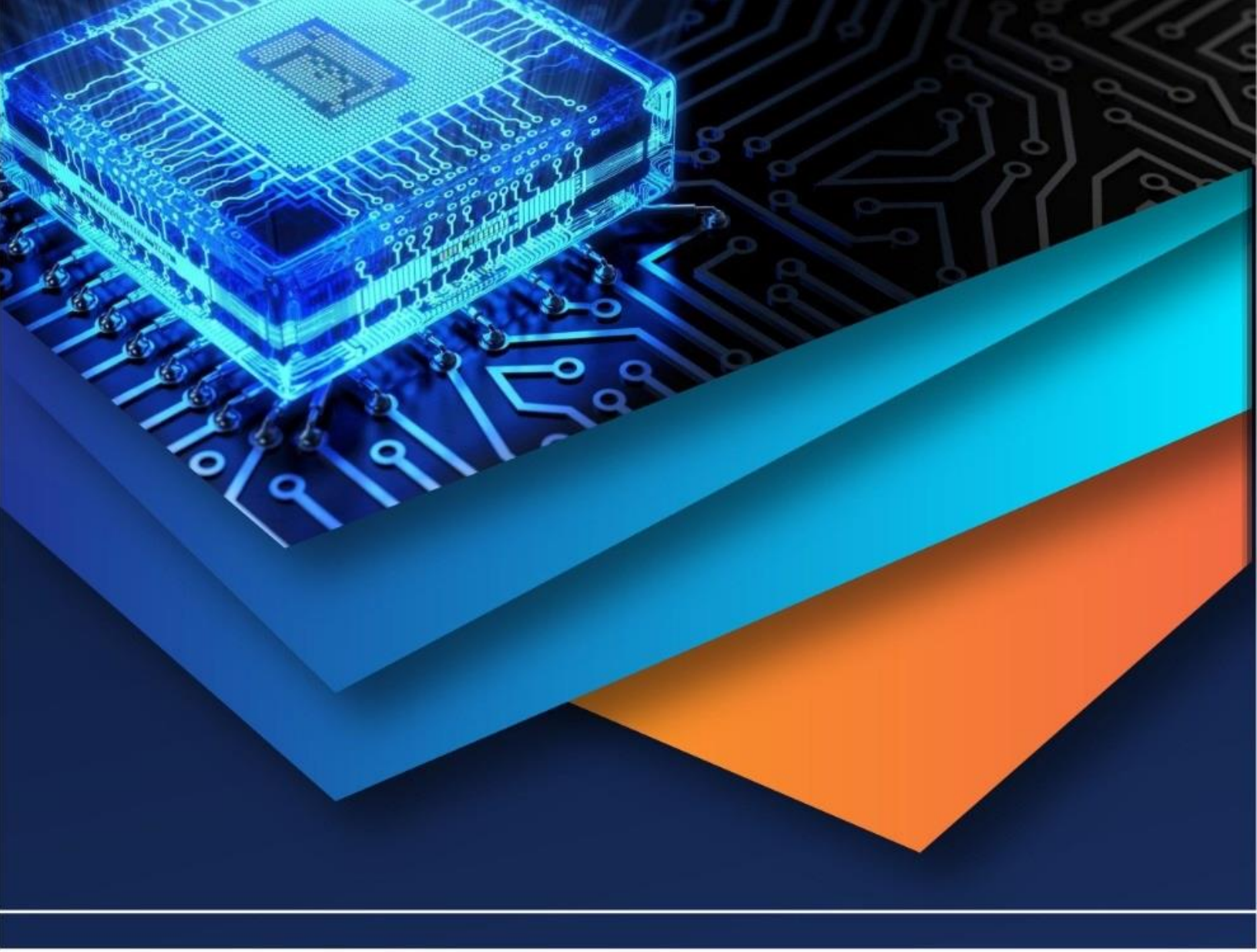
V. CONCLUSIONS

There are many systems which uses different machine learning algorithms such as Linear Regression (LR), Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors, etc for predicting the price for flight ticket. In this ML based system, we are using Random Forest Algorithm which gives more accuracy in predicting the airfare. Considering the features such as departure time, the number of days left for departure and time of the day it will give the best time to buy the ticket.

This system also helps the buyer to buy the flight ticket at lower price. It is easy to use and it gives more accuracy in prediction. It requires less time for prediction and it helps in reduction of over fitting. Travellers can save money if they choose to buy the ticket when its price is the lowest. It gives the customer the best time to buy a flight ticket for the desired destination and a period.

REFERENCES

[1] K. Tziridis, Th. Kalampokas, G.A. Papakostas, "Airfare prices prediction using machine learning techniques" IEEE,2017.
[2] Tianyi Wang, Samira Pouyanfar, Haiman Tian, Yudong Tao, Miguel Alonso Jr., Steven Luis and Shu-Ching Chen, "A Framework for Airfare Price Prediction: A Machine Learning Approach." IEEE(IRI),2019.
[3] Abhilash1, Ranjana Y2, Shilpa S3, Zubeda A Khan, "Survey on Air Price Prediction using Machine Learning Algorithms", IJIREEICE, 2019.
[4] Viet Hoang Vu; Quang Tran Minh; Phu H. Phung, "An airfare prediction model for developing markets",IEEE,2018
[5] T. Liu, J. Cao, Y. Tan, and Q. Xiao, "ACER: An adaptive context-aware ensemble regression model for airfare price prediction," in the international conference on progress in informatics and computing, 2017, pp. 312–317.
[6] Supriya Rajankar; Neha Sakharkar, **"**A Survey on Flight Pricing Prediction using Machine Learning", IJERT,2019
[7] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32, October 2001.
[8] Predicting_Flight_Prices_in_India, journel, researchgate publication, 33782141



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY