FLIGHT TICKET PRICE PREDICTION USING MACHINE LEARNING

Md. Ashraf*1, K. Sajitha*2, M. Vamshi*3, Mr. A. Ramesh*4

*1,2,3,4Department Of Computer Science And Engineering, ACE Engineering College,

Hyderabad, Telangana, India.

ABSTRACT

As domestic air travel is getting more popular now a days in India with various air ticket booking channels coming up online, travelers are trying to understand how these airline companies make decisions regarding ticket prices over time. Due to high complexity of pricing models applied by the airlines, it is very difficult for a customer to purchase an flight ticket at highest price, since price changes dynamically. The majority of making use of complicated prediction models from the computational intelligence research field known as Machine Learning. In this machine learning project, there is only one module namely, User. User can log in with valid credentials in order to access web application. A traveler can access this module to get the future price prediction of individual airlines. By considering certain features like departure time, arrival time, number of days for departure and based on the time it gives the best time to purchase a ticket.

Keywords: Machine Learning, Random Forest Regression.

I.   INTRODUCTION

The flight ticket buying system is to purchase a ticket many days prior to flight take-off so as to stay away from the effect of the most extreme charge. All organizations have the privilege and opportunity to change its ticket costs at any time. Explorer can set aside cash by booking a ticket at the least costs. People who had travelled by flight frequently are aware of price fluctuations. The airlines use complex policies of Revenue Management for execution of distinctive evaluating systems. The evaluating system as a result changes the charge depending on time, season, and festive days to change the header or footer on successive pages. The ultimate aim of airways is to earn profit whereas the customer searches for the minimum rate. Customers usually try to buy the ticket well in advance of departure date so as to avoid hike in airfare over specific period of time

II.   METHODOLOGY

Following steps were performed while building the system.

A. Data Collection

Both the training and testing data sets have been extracted from Kaggle data repository. They contain categorical as well as nominal data related to the Indian Airlines from the year 2019. The data set provides vital information about some impacting features to predict the fare of a flight - such as the places of departures and arrivals, time of departure and arrivals, the route of the flight, the number of halts during the journey and the price of the ticket depending on those features.

B. Data Cleaning

The null values present in the training data set where removed. A few columns which were of no use for the feature selection process were deleted from the data set. The columns of attributes having the categorical data were dropped from the data set after the new columns containing the numerical values extracted from the pre-processed data were stored for the prediction. Thus, the training data set suitable for use was obtained and it had the following attribute columns.

C. Data Pre-processing

While pre-processing the data, we converted the date of journey, departure time and the arrival time from string data type to date-time object and extracted the numeric values from them; the month-date numeric value from the date of journey attribute and hour-minute numeric value from the departure time and arrival time attributes respectively. Later, we have implemented the 'One hot encoding' method for the nominal categorical data and the label encoding method for ordinal categorical data present in both the training as well as the

testing data set. 'One hot encoding' is a process of converting the categorical data variables into numerical values thus making it suitable to use while implementing machine learning algorithms. One hot encoding method was applied to nominal categorical data attributes such as the 'source', the 'destination' and the 'airline company' chosen by the user. 'Label encoding' helps us convert the labels into numeric values in order to make the data set suitable for use. Label encoding method was applied to the nominal categorical data attributes such as the 'total number of halts in the journey'. The columns were re- arranged at the last step.

D. Model generation

The model has been generated using the Random Forest Regression.

E. Predicting Fare Price

The user input fields will be provided on a web page developed using the flask framework. The web page body was built using HTML5 and the same was styled using CSS3. After the user fills all the required input fields and submits the form, the data will be sent to the generate random forest regression model and the predicted value of the ticket price will be displayed.
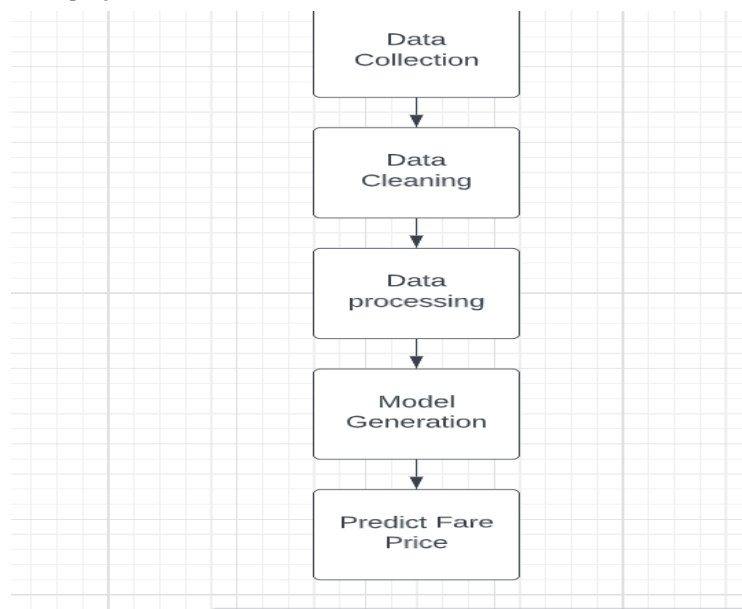

Fig 1: Methodology

III. MODELING AND ANALYSIS

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 data set and takes the average to improve the predictive accuracy of that data set. 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 data set, 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 data set 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 data set it runs efficiently.

e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022   Impact Factor- 6.752   www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

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:
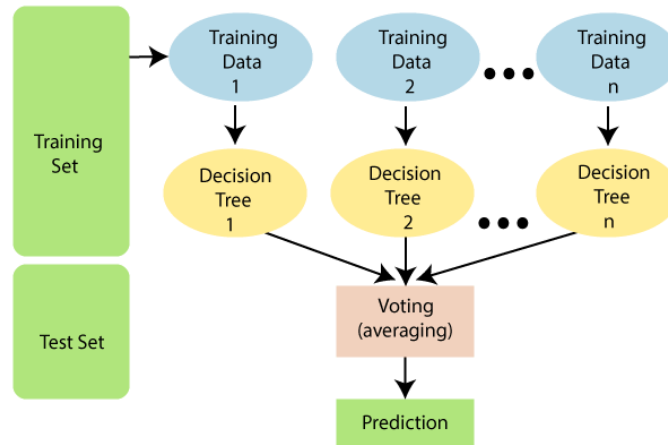
Fig 2: Random Forest Algorithm

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.

Working process can be explained in the below steps:

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
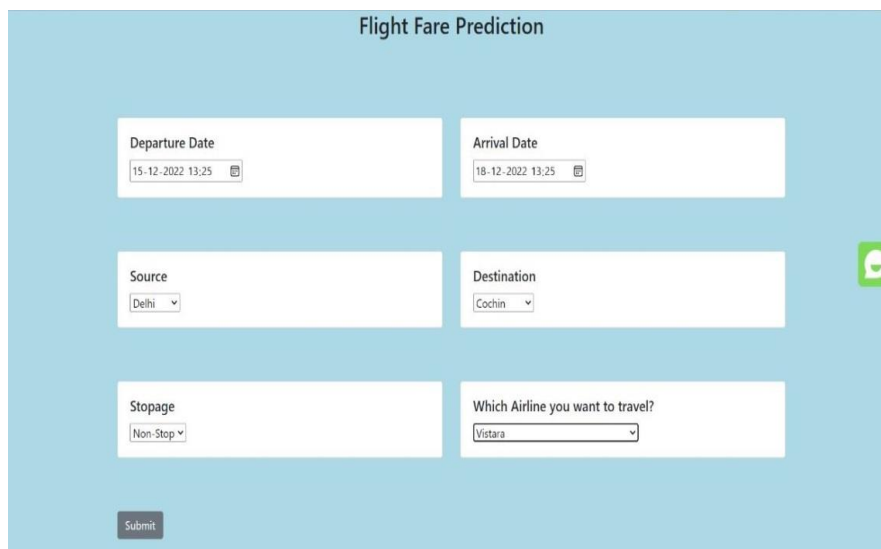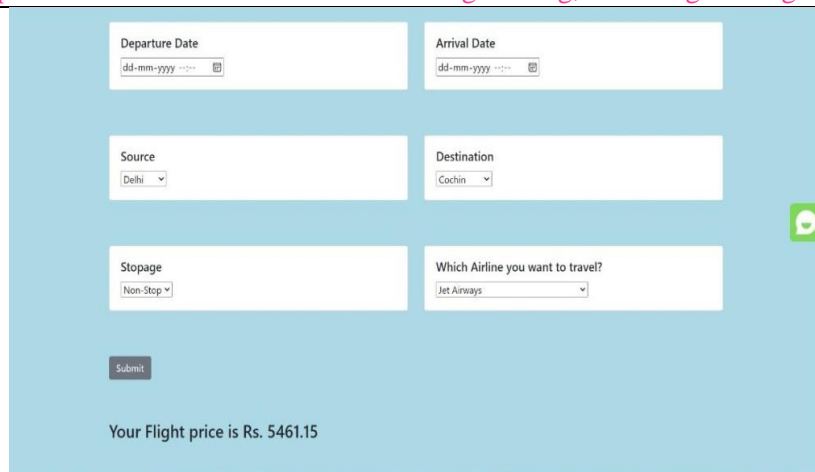
IV.   RESULTS AND DISCUSSION

Fig 3: Taking user input departure date, arrival date, source, destination, stoppage, airline

Fig 4: Prediction of Flight ticket price

V.   CONCLUSION

Machine Learning algorithms are applied on the data set to predict the dynamic fare of flights. For the prediction of the ticket prices perfectly different prediction models are tested for the better prediction accuracy. With the help of our project the travelers can find out the right time to buy their tickets at the lowest cost and also can plan accordingly. So to get result with maximum accuracy regression analysis is used. 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.

ACKNOWLEDGEMENTS

VI.   REFERENCES

[1]   Sharma, L., Carpenter, M. (eds.): Analysis of machine learning techniques for airfare prediction. In: Computer Vision and Internet of Things: Technologies and Applications, 1st edn. Chapman and Hall/CRC (2022)

[2]   Khandelwal, K., Sawarkar, A., Hira, S.: A novel approach for fare prediction using machine learning techniques. Int. J. Next Gener. Comput. Suppl. 12(5), 602–609 (2021). 8p..

[3]   Arjun, K.P., Achuthshankar, A., Soumya, M.K., Sreenarayanan, N.M., Priya, V.V., Faby, K.A.: PROvacy: protecting image privacy in social networking sites using reversible data hiding. In: 2016 10th International Conference on Intelligent Systems and Control (ISCO), pp. 1–4, January 2016

[4]   Gupta, J., Singh, I., Arjun, K.P.: Artificial Intelligence for Blockchain I, Blockchain, Internet of Things, and Artificial Intelligence, vol. 6. CRC Press, April 2021

[5]   Achuthshankar, A., Achuthshankar, A., Arjun, K., Sreenarayanan, N.: Encryption of reversible data hiding for better visibility and high security. Procedia Technol. 25, 216–223 (2016)

[6]   Champawat, J.S., Arora, U., Vijaya, K.: Indian flight fare prediction: a proposal. Int. J. Adv. Technol. Eng. Sci. 9(3) (2021)

[7]   Tian, H., Presa-Reyes, M., Tao, Y., et al.: Data analytics for air travel data: a survey and new perspectives. ACM Comput. Surv. 54(8), 1–35 (2022)

[8]   Rajankar, S., Sakhrakar, N., Rajankar, O.: Flight fare prediction using machine learning algorithms. Int. J. Eng. Res. Technol. (IJERT) (2019).