International Conference - Innovation-2021-Innovation-2021
International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN : 2456-3307 (www.ijsrcseit.com)
Volume 8, Issue 3, May-June-2021

A Random Forest Regression Approach to Predict Flight Fare

Komal Kalane[1], Shivam Ghorpade[1], Omkar Jawale[1], Abhishek Jaiswal[1], Snehal More[1], Prof. Monika Dangore[2]
[1]Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Pune, Maharashtra, India
[2]Assistant professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Pune, Maharashtra, India

ABSTRACT

This paper deals with the problem of flight prices prediction. The aviation industry keeps changing the flight prices. Prices on last minute airfare can be highly volatile so customer try to book ticket in advance. To estimate the minimum flight price, data for a specific air route has been collected including the features like departure date, arrival date, source, destination and airways. Features are extracted from the gathered data to apply Machine Learning (ML) Models. Machine Learning regression methods are used to predict the price at the given time. The training set is used to train the algorithm for accurate prediction and this will help to decide a specific airline as per budget.

Keywords : Machine Learning Algorithm, Prediction Model, Flight Price, Regression.

I. INTRODUCTION

The aviation industry is using complex strategies and methodologies to assign flight prices these days, in a dynamic fashion. These strategies are taking into account several commercial, financial and marketing factors closely connected with the final flight prices.

The high complexity of the pricing models applied by the airlines is the major difficulty faced by customers while purchasing the ticket, it is very difficult for customer to purchase an air ticket in the lowest price, since the price changes rapidly.

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

The potential of machine learning is the provided modelling tools, which are able to be trained, via a learning policy, with a set of data describing a certain problem and to counter to similar unrevealed data with a common way.

For anticipating the flight ticket prices, numerous 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 Random Forest Algorithm. Using python library scikit learn these models have been accomplished. The parameters like R-square, MAE and MSE are appraised to demonstrate the performance of these models.

II. METHODS AND MODEL

A. Methods

a. Collection of Data-

Python script is used on remote server for data collection, which provide output as excel record. The document contains data with features and its details. Output accumulated from the site contains number of parameters for each flight. The gathered data is two years old.

Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	litional_I	Price
IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
Jet Airway	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302
SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLR	09:00	11:25	2h 25m	non-stop	No info	3873
Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight	11087
Jet Airway	01/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight	11087
Multiple c	27/05/2019	Delhi	Cochin	DEL → BOM → COK	11:25	19:15	7h 50m	1 stop	No info	8625
Air India	1/06/2019	Delhi	Cochin	DEL → BLR → COK	09:45	23:00	13h 15m	1 stop	No info	8907
IndiGo	18/04/2019	Kolkata	Banglore	CCU → BLR	20:20	22:55	2h 35m	non-stop	No info	4174
Air India	24/06/2019	Chennai	Kolkata	MAA → CCU	11:40	13:55	2h 15m	non-stop	No info	4667
Jet Airway	9/05/2019	Kolkata	Banglore	CCU → BOM → BLR	21:10	09:20 10 May	12h 10m	1 stop	In-flight	9663
IndiGo	24/04/2019	Kolkata	Banglore	CCU → BLR	17:15	19:50	2h 35m	non-stop	No info	4804
Air India	3/03/2019	Delhi	Cochin	DEL → AMD → BOM → COK	16:40	19:15 04 Mar	26h 35m	2 stops	No info	14011
SpiceJet	15/04/2019	Delhi	Cochin	DEL → PNQ → COK	08:45	13:15	4h 30m	1 stop	No info	5830
Jet Airway	12/06/2019	Delhi	Cochin	DEL → BOM → COK	14:00	12:35 13 Jun	22h 35m	1 stop	In-flight	10262
Air India	12/06/2019	Delhi	Cochin	DEL → CCU → BOM → COK	20:15	19:15 13 Jun	23h	2 stops	No info	13381

Fig 1: Collected Data

b. Cleaning data-

The collected data needs to be clean, it is important to clean the data according to the model requirements. All the unnecessary data need to remove to achieve desire output. There are various techniques used for cleaning and pre-processing the data.

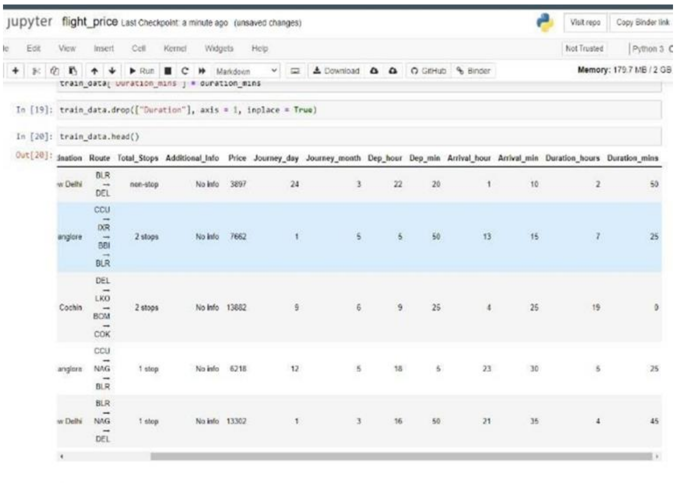

Fig 2: Data cleaning and processing

c. Analyzing Data-

The inspecting of data is the most key aspect of this project. Data preparation is followed by analyzing the data, unwrapping the invisible trends and then applying various machine learning models.

d. Machine Learning Algorithms-

To develop the model for the flight price prediction, many standard machine learning algorithms are evaluated. Those algorithms are as follows: K- Nearest neighbors, Multilayer Perceptron, Support Vector Machine (SVM), Linear regression, Decision tree, Gradient Boosting and Random Forest Algorithm. All these models are implemented in the scikit learn to evaluate the performance of this model, certain parameters are considered, the parameters are as follows: Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared value.

B. Models

Different models which we have tried:

*LinearRegression- 2779.0455708889144
*ElasticNet- 3379.6819876610443
*Lasso- 2759.449381312224
*Ridge- 2710.8476127741037
*KNeighborsRegressor- 3249.005561971264
*DecisionTreeRegressor- 2017.530360334335
*RandomForestRegressor- 1662.7359733973055
*SVR- 4246.460099935076
*AdaBoostRegressor- 3135.985374101527
*GradientBoostingRegressor- 1904.7364927923986
*ExtraTreeRegressor- 2432.1393735590073
*HuberRegressor- 3108.870789540331
*XGBRegressor- 1603.7426369307445
*BayesianRidge- 2773.275561516677

XGBRegressor, RandomForestRegressor and GradientBoostingRegressor gave the lowest RMSE so we have chosen these model and

Model	Savings (In Lakhs)	Loss (In Lakhs)	Profit per Transaction (In Rs.)	Accuracy
Decision Trees	4.7	1.3	140	73.0%
Gradient Boosting	5.5	2.2	145	73.0%
Logistic regression	6	1.8	177	76.0%
Random Forest	5.8	1.8	180	77.8%
Trend Based Model	7	2.2	210	81.8%

Fig 3: Comparison between models

III. RESULTS AND DISCUSSION

As discussed in Section II, we used 10 thousand samples to train the data and test the classifiers with the suggested 80-20 split. For the decision tree algorithm, we parameterized the depth of the tree for better accuracy. In order to analyze the impact of the used features to the prediction accuracy of the models, the same experiment is repeated several times by leaving out some features, one at a time. Several techniques used to clean and pre-processing the data to achieve desire output.

The correlation between independent and dependent attribute as shown in figure below: -

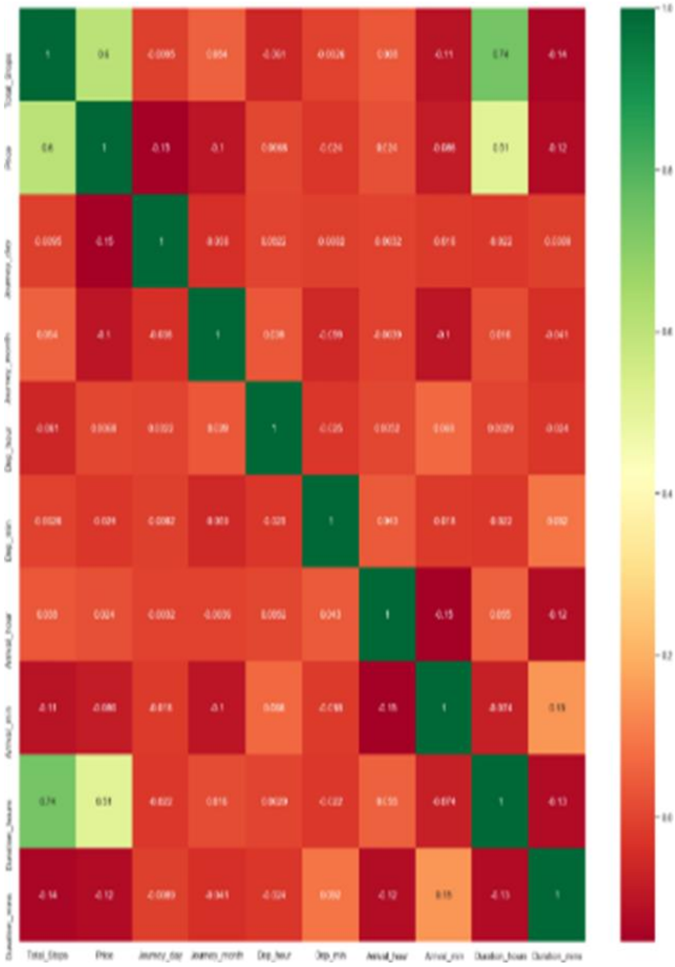

fig 4: Co-relation between independent and dependent attribute

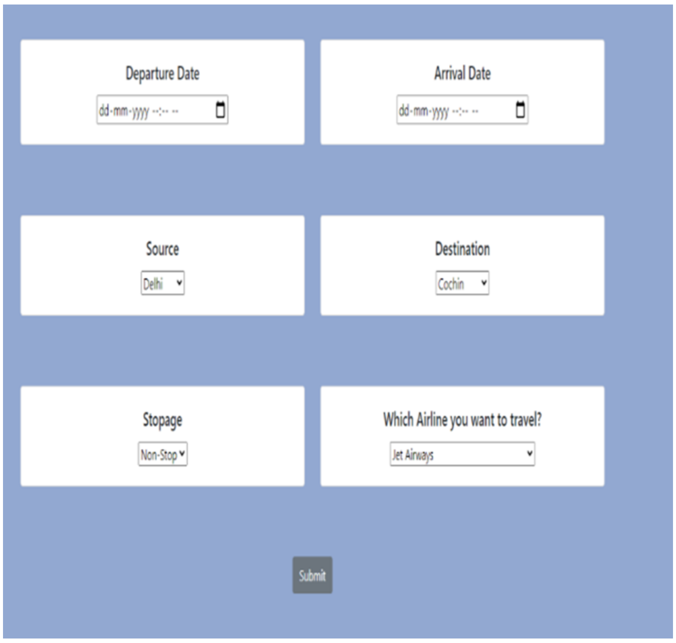

Fig 5: UI

The above figure shows the user interface, user need to fill the data according to user requirements to get the approximate flight price for the trip.

IV. CONCLUSION

The information gathered remote server by using python script and showed that it is feasible to predict prices for flights based on previous air fare data. The experimental results show that machine learning models are a satisfactory tool for predicting flight prices. Other important factors in flight prices prediction are the data collection and feature selection from which we drew some useful conclusions. From the experiments we wind up which features influence the flight price prediction at most.

In the future, this work could be extended to predict the flight prices for the entire aviation industry. Additional experiments on larger airfare data sets are essential, but this initial work highlights the potential of Machine Learning models to guide consumers to make an airfare purchase in the best market period.

V. REFERENCES

[1]. K. Tziridis, K.I. Diamantaras,"Airfare Prices Prediction Using machine Learning Technique",European signal processing conference 2017

[2]. M. Papadakis, "Predicting Airfare Prices," 2014.

[3]. A. Mottini and R. Acuna-Agost, "Deep choice model using pointer networks for airline itinerary prediction," in the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017

[4]. 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 Analysis. In: 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, 2013, pp. 440–444.

[5]. Chieh-Hua Wen, Po-Hung ChenPassenger booking timing for low-cost airlines: a continuous logit approach J. Air Transport Manage., 64 (2017), pp. 91-99

[6]. Estimating dynamic demand for airlines Econ. Lett., 124 (1) (2014), pp. 26-29

[7]. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," Journal of Machine Learning Research, vol. 12, no. Oct, pp. 2825–2830, 2011

[8]. A. Ng, "Cs229: Machine learning lecture notes," Standford University Lecture, 2011.

[9]. L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984

[10]. Dominguez-Menchero, J.Santo, Reviera, "optimal purchase timing in airline markets" ,2014

[11]. Abhilash, Ranjana,shilpa and Zubeda'"Survey on Air Price Prediction using Machine Learning Algorithm", IJIREEICE 2019

[12]. L. Breiman, ―Random forests, ‖ Machine Learning, vol. 45, pp. 532, 2001

[13]. S.B. Kotsiantis, "Decision trees: a recent overview," Artificial Intelligence Review, vol. 39, no. 4, pp. 261-283, 2013.