International Journal of Advanced Trends in Computer Science and Engineering

Flight Ticket Price Predicting With the Use of Machine Learning

Jibin Joseph[1], Abhijith P[2], Aryasree S[3] , Jinsu Anna Joseph[4], Meghana Sara Oommen[5], Abin T Abraham[6]

[1] PG Student, Department of Computer Applications –Saintgits College of Engineering, Kottayam, Kerala, India, jibinjoseph0077@gmail.com

[2] PG Student, Department of Computer Applications -Saintgits College of Engineering, Kottayam, Kerala, India, abhijthp41997@gmail.com

[3] PG Student, Department of Computer Applications -Saintgits College of Engineering, Kottayam, Kerala, India,aryasreepadom@gmail.com

[4] PG Student, Department of Computer Applications -Saintgits College of Engineering, Kottayam, Kerala, India, jinsuannajoseph@gmail.com

[5] PG Student, Department of Computer Applications -Saintgits College of Engineering, Kottayam, Kerala, India,meghanasaraoommen77@gmail.com

[6] Assistant Professor, Department of Computer Applications - Saintgits College of Engineering, Kottayam Kerala, India, abin.t@saintgits.org

ABSTRACT

As domestic air travel in India is becoming increasingly popular with different air ticket booking channels coming online these days, passengers are trying to understand how these airline companies make decisions over time about ticket prices. Therefore, many methods are ready to provide the proper time to do so. The customer who buys an air ticket by estimating the price of the airfare is recently proposed. The majority of these strategies make use of sophisticated Computational Intelligence Prediction Models an area of science known as Machine Learning (ML). This paper highlights the parameters and also includes the guidelines that are important for project work to be developed that is indicated above.

Key words: Computational intelligence, Machine learning

1. INTRODUCTION

These days, domestic air travel is becoming more and more common in India. Travelers are trying to learn how these airline companies make choices over time about ticket prices with multiple air travel booking outlets coming online. For a passenger, it is a time-consuming method to search websites for deals and offers. The cost can therefore depend on various variables. This venture uses AI to show the types of flight tickets after some time to estimate the costs. Both organizations have the right and the ability at any time to change their ticket prices. By reserving a ticket at the lowest cost, explorer can set aside money. People who have traveled by flight are also aware of the variations in costs. Complex revenue control policies are used by airlines for the introduction of distinctive assessment schemes. As a result, the appraisal scheme adjusts the fee to adjust the header or footer on successive pages based on time, season, and festive days. The ultimate goal of the airways is to achieve profit, while the customer is looking for the minimum cost. Usually consumers try to book the ticket well in advance of the departure date to prevent airfare hikes as the date gets closer. But that's not the truth, really. By giving more than they should for the same seat, the customer can finish up.

2. LITERATURE SURVEY

It is difficult for a customer to receive a low-cost airline ticket. For this, a few procedures are investigated in order to assess the best time and date to buy low-cost airline tickets. The majority of these systems make use of Machine Learning, a modern computerized method. Gini and Groves [1] used Partial Least Square Regression (PLSR) to build a model to decide the best time to buy a flight ticket. From February 22nd to June 23rd, 2011, data was gathered from major adventure travel booking sites. Extra data was collected as well, which was used to verify the similarities between the previous model's exhibitions. Janssen[2] used

the Linear Quantile Blended Regression technique to create a desire model for the San Francisco to New York course, where www.infare.com provides daily airfares. The model is designed using two features: the number of days for departure and whether the departure is on a weekend or weekday. The model forecasts airfare months ahead of time. However, in a scenario involving a long time commitment, the model fails to persuade, and the departure date is pushed back. Wohlfarth[3] suggested a ticket buying time improvement model involving macked point processors, data mining systems (course of action and grouping), and a quantifiable inspection scheme. This structure is proposed as a way to convert different added value arrangements into included added value arrangement headings that can help with solo gathering estimation. This value heading is jammed into a get-together based on close assessing actions. The value shift proposals are measured using the Headway model. A tree-based analysis was used to choose the best planning gathering and then look at the progression model a short time later. Dominguez Menchero[4] proposes the perfect purchase timing for a particular course, carriers, and time frame using a nonparametric isotonic backslide technique. The definition of the model. The model determines the most appropriate number of days before purchasing a plane ticket. The model takes into account two types of variables: entry and date of acquisition.

3. DATA COLLECTION

The most critical aspect of this project is the accumulation of knowledge. To prepare the models, the distinct wellsprings of the data on various sites are used. Sites provide information on the different firms, hours, aircraft, and charges. For data scratching, various sources from API's to customer travel sites are available. Details on the various sources and criteria that are obtained is discussed in this chapter. Here we collected dataset from Kaggle.com[5] and site and the models are implemented using python.

3.1 Data Collection

The document contains the data with features and its details. Choosing the features needed for the estimation of the predicted flight price is an important prospect. The site's output contains the number of parameters for each flight: but not all are needed, so only the accompanying components are required.

- Airway company
- Date of Journey
- Date of Arrival
- Date of Departure
- Time of Arrival
- Total number of stops
- Place of Destination/Arrival
- TotalFare

Table 1: Collected dataset

Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	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 → IXF	05:50	13:15	7h 25m	2 stops	No info	7662
Jet Airway	9/06/2019	Delhi	Cochin	DEL → LKC	09:25	04:25 10 Jun	19h	2 stops	No info	13882
IndiGo	12/05/2019	Kolkata	Banglore	CCU → NA	18:05	23:30	5h 25m	1 stop	No info	6218
IndiGo	01/03/2019	Banglore	New Delhi	BLR → NA	16:50	21:35	4h 45m	1 stop	No info	13302
SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLI	09:00	11:25	2h 25m	non-stop	No info	3873
Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BOI	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight meal nc	11087
Jet Airway	01/03/2019	Banglore	New Delhi	BLR → BOI	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
Jet Airway	12/03/2019	Banglore	New Delhi	BLR → BOI	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight meal nc	11087
Multiple c	27/05/2019	Delhi	Cochin	DEL → BO	11:25	19:15	7h 50m	1 stop	No info	8625
Air India	1/06/2019	Delhi	Cochin	DEL → BLF	09:45	23:00	13h 15m	1 stop	No info	8907
IndiGo	18/04/2019	Kolkata	Banglore	CCU → BLI	20:20	22:55	2h 35m	non-stop	No info	4174
Air India	24/06/2019	Chennai	Kolkata	MAA → C(	11:40	13:55	2h 15m	non-stop	No info	4667
Jet Airway	9/05/2019	Kolkata	Banglore	CCU → BO	21:10	09:20 10 May	12h 10m	1 stop	In-flight meal nc	9663
IndiGo	24/04/2019	Kolkata	Banglore	CCU → BLI	17:15	19:50	2h 35m	non-stop	No info	4804
Air India	3/03/2019	Delhi	Cochin	DEL → AM	16:40	19:15 04 Mar	26h 35m	2 stops	No info	14011
SpiceJet	15/04/2019	Delhi	Cochin	DEL → PN	08:45	13:15	4h 30m	1 stop	No info	5830
Jet Airway	12/06/2019	Delhi	Cochin	DEL → BO	14:00	12:35 13 Jun	22h 35m	1 stop	In-flight meal nc	10262
Air India	12/06/2019	Delhi	Cochin	DEL → CCI	20:15	19:15 13 Jun	23h	2 stops	No info	13381
Jet Airway	27/05/2019	Delhi	Cochin	DEL → BO	16:00	12:35 28 May	20h 35m	1 stop	In-flight meal nc	12898

The original dataset obtained from kaggle.com is shown in Table I. It is essentially raw information containing all the characteristics. For many routes, this information has been obtained.

3.2 Cleaning and preparing Data

All the information collected took a great deal of effort, but it should have been perfect and ready after the accumulation of information, as shown by the model prerequisites. Including copies and null attributes, all unnecessary data is removed. This breakthrough is the most critical and time. Different mathematical methods and logics clean and customize knowledge in python. For example, the extracting date and time.

3.3 Analyzing the data

Data preparation is monitored by breaking down the data, exposing the hidden patterns and applying various regression models afterwards. Similarly, from the existing features, a few features can be calculated. Flight days can be given by measuring the difference between the date of the flight and the date of collection of the details. In addition, the flight date, whether on a festive day or a weekday or a weekend, is significant. The flights scheduled during the weekends instinctively cost more than the flights on weekdays. In addition, time plays a major role.

4. MACHINE LEARNING MODEL PERFORMANCE

In machine learning, several algorithms are applied to forecast the prices of flight tickets. The algorithms are: Linear regression, Decision tree, K-Nearest neighbors, and Random Forest Algorithm. These models have been implemented using the python library Sklearn. The parameters like MAE and MSE, RMSE are considered to check the efficiency of these models.

4.1 Linear Regression

Simple linear regression analysis is used to determine the association between two continuous variables. The indicator variable of what importance is to be found is one of the two variables. It is not the deterministic relationship between two variables that gives the statistical relationship. The linear regression algorithm gives the given data the best fit line for which the prediction error is limited. The two main factors for understanding linear regression are gradient descent and cost function. The equation for linear regression is:

$$y(pred) = b0 + b1 * x$$

The value of coefficients b1 and b0 is selected such that the value of the error is as minimal as possible. The error is given by the square of the expected and actual value difference. The mean square error is taken to deal with the negative values (MSE). The positive or negative relationship between x and y is given by b0 here, while b1 is called bias. Regression problem accuracy is calculated in terms of R-squared, MAE, and MSE.

4.2 Decision Tree

This tree count isolates the data obtained into small subsets, rendering it permanent at a comparable time. The new findings show the tree with the decision centers, and the leaf centers as well. At any rate, this decision-center point will contain two branches. Think of the entire knowledge index as a root at first. Function aspects are kicked out of the opportunity. If the characteristics are relentless, then before structuring the model, they have to be discretized. In view of estimation property records are corrected recursively. In the decision of tree computation, Knowledge Gain and Gini index are two basic properties. Information Gain is characterized as the change in entropy in quantity. Higher entropy suggests the substance's greater efficacy. Therefore, entropy is a proportion of an arbitrary variable's susceptibility. The Gini Index tests how to falsely identify an arbitrarily chosen component on a regular basis. This suggests that a feature with a lower Gini index should be liked. For Regression tree, cost capacity can be a basic squared condition:

$$E = \sum (y - \hat{y})^2$$

Where y is the actual value from the dataset and y cap is predicted value. Have a class with the maximum sum of the expected value obtained by a split function called the gain of knowledge. If the class is kept dividing and dividing at the

leaf node without any condition, the algorithm will be truly massive, slow and over-fitted. To stop this, a minimum count on the training example on the leaf node is assigned.

5.EXPERIMENTAL RESULTS

In our project we had implemented various Machine Learning Algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression and compared the accuracy of results based on our test data set. Based on the various accuracy levels we find that Random Forest Regression gives the highest accuracy i.e. 81%. Therefore we selected Random Forest Regression and created User Interface based on it.

Table 2: Regressions with its accuracy

Algorithms	Accuracy
Linear Regression	0.62
Decision Tree Regression	0.65
Random Forest Regression	0.81

6.LIMITATION OF SYSTEM

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. So the collected train data should be accurate if not it may result in wrong prediction. And also it is necessary to update the train data time to time for best results.

7. CONCLUSION

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. 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. From the studies, the feature that influences the price ticket are to be considered. In future the details about number of available seats can improve the performance of the model.

REFERENCES

[1] W. Groves and M. Gini, "An agent for optimizing airline ticket purchasing," 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), St. Paul, MN, May 06 - 10, 2013, pp. 1341-1342.

[2] T. Janssen, "A linear quantile mixed regression model for prediction of airline ticket prices," Bachelor Thesis, Radboud University, 2014. [3] M. Papadakis, "Predicting Airfare Prices," 2014.

[3] Wohlfarth, T. Clemencon, S.Roueff, "A Dat mining approach to travel price forecasting", 1 0 th international conference on machine learning Honolulu 2011.

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

[5] Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioner.Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.