Airline Choice:
A Comparison of Classifiers in Traditional Analysis vs Decision Trees

Archana Shrivastava, Amity School of Business, Amity University Uttarpradesh, Noida, India

https://orcid.org/0000-0001-7955-3844

P. James Daniel Paul, Ernst and Young LLP, Bengaluru, India

J.K. Sharma, Amity School of Business, Amity University Uttarpradesh, Noida, India

ABSTRACT

Widespread use of e-commerce in the airline industry is generating data at unprecedented scale, thus rendering it amenable to decision analysis. Classification accuracy is one of the key factors in forecasting and in the decision sciences. The traditional classification analysis was carried out by several methods such as ANOVA, Logit, Probit. However, for decision analysis algorithms and decision trees have emerged for classification analysis. The objective of the article is to analyze the airline choice data using the traditional ANOVA and compare them with the decision trees and different algorithms.

KEYWORDS

Airlines Choice, Airlines Price Determination, Comparison of Classifiers, Decision Sciences, Decision Trees, Traditional Analysis

INTRODUCTION

Data driven mathematical models are used for forecasting and decisions as large volume of data is being collected in a range of application areas (Jagdish et al., 2014). Widespread use of E-commerce is inherently generating data at unprecedented scale, thereby rendering it amenable to decision analysis (Chong et al., 2017).

Classification accuracy is one of the key factors in forecasting and in the decision sciences (Hofman et al., 2017). The traditional classification analysis was conducted using various methods such as ANOVA, Logit, Probit, etc. However, for the decision analysis the market analysts started using number of tools such as decision trees and various other algorithms.

According to CAPA - Centre for Aviation, airlines are largely e-commerce driven. As a result, massive volume of data is being collected that is further used by airlines to study buyer behavior, facilitating them in making decisions on pricing, time of departure etc. Researchers use number of techniques to analyze the data. We have gathered information from makemytrip.com in order to compare results between ANOVA and decision tree.

The objective of this paper is to compare the airline decisions using traditional ANOVA and compare them with decision trees and different algorithms.

Authors have examined if advance booking, departure and arrival time, number of stops, hours of journey have an impact on the price of a ticket.

The research analyses classification accuracy of the traditional ANOVA regression and the decision trees like the MSP tree and the Rep tree. While in ANOVA most of the variables indicated significance of impact of price on choice of airlines. In Regression analysis the departure time was found to be the critical factor in the choice. Similarly, in MSP Tree also the number of days ahead and departure time emerged as a significant factor. However, in the case of the Rep Tree, the number of days ahead was a clear factor. This indicates that the MSP tree is an intermediate solution providing the best of the regression results and the tree results.

As for as the business implications are concerned, Air France has emerged as the second most preferred airlines that provides seven choices in the third category of price range (40,000 to 60,000). It also provides one choice in the top category of price range (60001 and more). Alitalia, Etihad Airways, Jet Airways and Lufthansa are the third most preferred airlines. In the third category (40,000 to 60,000 INR) category Air France and Jet Airways provide the highest number of options. In the top price category Lufthansa offers the highest options. Rep tree offers three levels of choices to the customers. At the first level there are Air France and the Lufthansa. At the second level prices range from 20001 to 40000 there are Ethihad, Al Italia and Qatar Airways offer most choices with total number of choices 53. At the third level there are Air India, Turkish and Jet Air, Finn Air. Rep tree shows that Air France and Lufthansa are chosen based on their departure time. Ethihad, Qatar Airways and Al Italia are chosen based on number of stops, arrival time and duration of the journey. Air India is considered if departure time, no. of stops, arrival time and duration of the journey suit the travelers.

The research reveals the variations in results obtained using various tools and gives an insight on how consumers make choices, what are the relevant factors that play a role in choosing an airline and on what factors airlines determines their prices.

METHODOLOGY

This paper uses the data from the popular multiband ticket-booking site called www.makemytrip.com. The search for the price of tickets between London and New Delhi were 30 days before the travel date and 15 days before the travel date. 98 different price combinations were obtained and then the data was analyzed. But before the analysis of the data the literature on similar studies were collected and analyzed. The route Delhi to London was chosen arbitrarily. We have used Word Cloud, SPSS, Tableau, and WEKA for analysis and visualization.

REVIEW OF LITERATURE

We reviewed recent research articles on online buying of air tickets on issues ranging from price dispersion to whether price plays a central role in purchase decision of air tickets. It was observed that a lot of studies have been conducted on consumer perceptions and experience of buying air ticket on Internet. Research has also been conducted on price dispersion and market structure of airlines. We reviewed the methodologies used in the articles to understand the variables and statistical tools used by different researchers.

Bing Pan and Daniel R. Fesenmaier (2007) in their article "Travel Information Search on the Internet: A Preliminary Analysis", used 15 subjects from a major university located in the Midwest, United States during August of 2002 for their study. The subjects were asked to plan a trip to a specified US destination (San Diego, California) on the Internet using computers located in the National Laboratory for Tourism and ecommerce, University of Illinois at Urbana-Champaign. The data was collected using questionnaire, voice recorder, online camcorder, Internet monitor

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/airline-choice/246027

Related Content