International Journal of Cyber Behavior, Psychology and Learning
Volume 13 • Issue 1

Deep Learning Models for Airport Demand Forecasting With Google Trends:
A Case Study of Madrid International Airports

Bahri Baran Koçak, Dicle University, Turkey*

ABSTRACT

Managers gain new insights into how operational benefits can be achieved. Forecasting problems for passenger flow in airports are gaining interest among marketing researchers, but comparison of stochastic optimisation methods via deep learning forecasts with search query data is not yet available in the aviation field. To fill this gap, the current study predicts the demand of Madrid airport demand with Google search query data using H2O deep learning method. The findings indicate that there is a long-term relationship between search queries and actual passenger demand. Besides, search queries "fly to madrid," and "flights to madrid spain" were found to be the cause of the actual domestic air passenger demand in Madrid. Also, to determine the best forecasting accuracy, stochastic gradient descent (SGD) optimisers were used. Specifically, findings indicate that Adam is a better optimiser increasing forecasting accuracy for Madrid airports.

KEYWORDS

Airport Demand Forecasting, Big Data Analytics, Consumer Search Behaviour, H2O Deep Learning, Stochastic Gradient Descent

1. INTRODUCTION

The global popularity of the Internet and the reach of digital information to the masses have enabled consumers to benefit from this technology and changed the way they seek information about the products desired to buy. Search engines such as Google, Bing and Yahoo have enabled people to get the information they need from the internet (Lai et al., 2017).

Theoretically, it can be thought that the Internet provides a large amount of information to its users with less time, effort, and cost. These conveniences can also be said to change the traditional information-seeking behavior of consumers, such as watching mass media or asking sales staff about their product or service (Peterson and Merino, 2003). In this context, to buy a new camera, to see the newest movies in theaters showtimes around, to search for air tickets or hotels, consumers can easily use internet search engines. Therefore, it can be said that it is possible to predict the collective search behavior by looking at the frequencies and time series of online search of activities such as retail, cinema, or travel (Goel et al., 2010). For example, as an important customer group of airline transport, tourists can use search engines to get air and traffic information and plan their routes as they wish

DOI: 10.4018/IJCBPL.324086 *Corresponding Author

(Fesenmaier et al., 2011; Li et al., 2017). In this respect, Google, one of the biggest search engines of today, has provided useful data about the search trends of the communities to the researchers with the "Google Trend" service (Dreher et al., 2018). Google Trends offers the daily or real-time search trends for a given region, as well as the frequency of searches for a given term from 2004. It can also display the search frequency of the search term by indexing it between 0-100 on the chart by filtering according to region, search category or search type (web, image, news, etc.)[1].

In the literature, most of the studies using Google Trends data have been performed for prediction. The first and most popular research among these studies was conducted by Ginsberg et al. (2009) to estimate the influenza virus activity in the US. Looking at the popular and the most cited studies related to demand forecasting with search query data, Pan et al. (2012) estimated the hotel room demand with search query data for a special tourist destination by applying autoregressive moving area (ARMA) models. Hand and Judge (2012) predicted cinema participation using the ARIMA method with Google Trends search information. Bangwayo-Skeete and Skeete (2015) estimated the demand for tourism with Google trend search data by comparing the Autoregressive Mixed-Data Sampling (MIDAS) method with other autoregressive models. In the prediction models, they showed that MIDAS method gives better results when Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) criteria are considered. Dimpfl and Jank (2016), using autoregressive models, determined investors' attention with daily search query data and estimated the stock market volatility with the help of these data. Rivera (2016) proposed a dynamic linear model (DLM) and predicted hotel registrations with search query volume. He compared the model with Seasonal Autoregressive Integrated Moving Average (SARIMA), Holt-Winters (HW) and seasonal naïve (SNAIVE) and found that DLM works better in long-term prediction. Önder and Gunter (2016) concluded that using autoregressive model, Google web and image search results in local language increased the predictive power of tourism demand. Li et al. (2017) predicted the tourism volume with web search queries related to various tourism words by applying Generalized Dynamic Factor Model (GDFM). In another study, Google web or image search volumes on country and city basis were estimated by comparing different models such as HW and Naïve (Önder, 2017). Park, Lee, and Song (2017) forecasted short-term tourists' entry on the basis of country with Google Trends data using ARIMA and SARIMA methods. Sun et al. (2019) predicted the number of tourists arrivals using machine learning methods and comparing the success of different search engines.

In the aviation indusrty, airport and airline planners need to predict demand and to improve forecast accuracy to escape uncertain economic climates and misinformed infrastructure investments (Suh and Reyerson, 2019). Also, handling airport capacity problems in Europe is one of the major objectives of extensive investment projects (Sismanidou and Tarradellas, 2017). In this sense, it is understood that the studies about forecasting aviation demand with web search queries are very limited. Accordingly, Kim and Shin (2016) have developed a model to predict short-term airline demand based on monthly passenger arrivals and weekly internet search queries. Shin et al. (2017) tested the relationship between international airport arrivals and search term volume data by Granger causality analysis, suggesting that search activities occur before the flights takes place and can be used for forecasting.

Lastly, demand forecasting studies for sectors such as tourism, entertainment and transportation have shown that web search query data increases predictive success. Furthermore, time series regression models, neural network methods and artificial intelligence algorithms are the main methods applied in these studies. However, no study, to my knowledge, predicted demand for the airports of a specific city with Google search queries using H20 deep learning method.

In this study, Madrid airport market demand was predicted with Google Trends search data of consumers using H2O deep learning method. Thus, the current work makes 2 contributions to the field:

1. To reflect demand for Madrid airports, the selection of search queries of passengers was made by using Google Ads Keyword Planner. The findings indicate that the planner recommends useful