A MACHINE LEARNING MODEL FOR AIR CRAFT TICKET PRICES

[1]G.Raj Kumar, [2]J.Praveen Kumar, [3]S.Sanjeeva Rao,[4]N.Niharika

Asst Professor Teegala Krishna Reddy Engineering College, Hyderabad, India.
rajkumar.gadda@tkrec.ac.in
Associate Professor Teegala Krishna Reddy Engineering College, Hyderabad, India.
praveenkrecit@gmail.com
Asst Professor Teegala Krishna Reddy Engineering College, Hyderabad, India.
ssanjeevarao@gmail.com
Asst Professor ,Stanley College of Engineering &Technology for Women Hyderabad, India.
niharikasanku@gmail.com

Abstract

Flight ticket fare is the most fluctuating data which varies every day. Depending on various factors that affect it directly or indirectly, we cannot say that price of the ticket stays the same or not. It is quite a tough task to predict the flight ticket fare. It may change throughout the week, month or sometimes days, but it can be predicted nearly accurate to actual flight ticket fare. The prime objective of our project "Flight fare prediction system" is to make a prediction of the flight ticket fare for the future flights. We build a machine learning model using a dataset of aircraft ticket prices for different carriers and different routes at different times of the year. The regression model which we have selected for our prediction is "Extreme Gradient Boosting".

DOI: 10.48047/ecb/2023.12.si12.159

INTRODUCTION

In computer science, machine learning (ML) refers to the study of algorithms that become better with time and utilisation of data. Models are built using training data and predictions or choices are made using this model without being trained to do so using machine learning algorithms. There are several uses for machine learning algorithms, such as fraud detection and email screening. Machine learning may be used in the 'Aviation business' to forecast flight pricing. Flight costs are affected by a variety of variables, including distance, flight length, number of stops, and more. The machine learning models are trained on this pattern to forecast the price of a flight in the future, automating the procedure and speeding up the process.

The airline industry is frequently considered as one of the most advanced when it comes to smart pricing techniques. Even seats next to one other on the same aircraft might alter in price dynamically. The price of a given flight might fluctuate as often as seven times a day. At the same time that passengers are seeking the lowest possible fares, airlines are always striving to preserve the greatest possible total income. Customers often pay more or the airline loses money when there is a mismatch between available seats and demand. With their extensive tools and capacities, airline businesses are able to regulate the price process. Customers, on the other hand, are becoming more strategic as a result of the emergence of several internet tools for comparing airfares offered by different airlines.

Demand prediction and price discrimination are two types of dynamic pricing models that have been developed by a large number of researchers. An airline firm may schedule flights and set prices for a route if it has an idea of how many people will be travelling on it. Current demand prediction models are primarily concerned with estimating an airline's market share and the demand for a certain aircraft or route. Due to the rapid expansion of the Internet and e-commerce, travellers may now quickly compare

the prices and schedules of many airlines from all over the globe. When a customer is happy with a price, they may purchase their tickets online using the official websites of the airline or agency. A variety of forecasting models have been developed to estimate the cost of flight in advance, in order to assist clients in finding the best deals. Customers received 'buy' or 'wait' signals based on a variety of data mining approaches and time series data analytics. Regression approaches such as Partial Least Square Regression or Linear Quantile Mixed Regression may also be used to build prediction models. A recurring theme in the United States of America. Aside from these few elements, additional airfare data may be gathered online to construct a model that can anticipate the price of a flight more accurately and help clients make smarter selections to obtain the best deal.

Because it's impossible to predict when all of an aircraft's seats have been taken, waiting to buy the lowest ticket might cause you to miss your departure. Flights have been more popular in recent years. Since prices fluctuate so often owing to many factors, airlines have a tough time keeping their costs stable. Our project “Flight Fare Prediction System” resolves this problem, and provide a facility where people will be able to predict the flight-ticket price before purchasing the ticket. By projecting what costs they can sustain, this is helpful to airlines. Using this information, users may better plan their travels and anticipate future flight costs.

1. LITERATURE SURVEY

2.1Airline ticket price and demand prediction: A survey

For the same aircraft, even in the same cabin, airline ticket costs may vary dramatically and drastically, even between seats that are just a few feet apart. As a consequence, passengers are hunting for the greatest offer, while airlines are striving to maximise their profit margin. Demand prediction and price

discrimination are two of the numerous computational methods airlines use to increase revenue. Various academics have produced models that anticipate the best time to buy a ticket and the cheapest ticket price in order to save money for customers. A review of consumer and airline prediction models is presented in this study. According to our evaluation, there are just a few features that can be utilised to develop models on both sides.

2.2 Based on Kalman Filtering, a Bayesian Approach to Flight Fare Prediction.

In current computer-aided solutions and applications, decision-making under ambiguity is a key challenge. Using Bayesian methods of prediction may be helpful in these kinds of investigations. Researchers in this research are making an attempt to predict airline ticket prices using the Kalman filter, a well-known Bayesian estimation approach. This approach is based on the linear Kalman Filter model. According to this method, it is possible to anticipate the cost of a flight based on previous pricing. To estimate the cost of a certain impending flight, an input matrix containing the observed data is used in conjunction with a linear model.

2.3 Predictive Analytics Platform for Airline Industry

As a result of these factors and others, air travel might be significantly affected. The buying patterns of passengers are thus vulnerable to fast changes rapidly. Because of this, it is difficult to build a strategy for accurately forecasting income for each route. We'll utilise a semi-supervised learning approach to get around this. Because we've sold tickets already, we're bringing in money, and we already have people lined up. Passenger data from earlier flights is also available to us. Much of the information is easily available, except for one unknown factor that may have a significant impact on travel patterns.

We will construct a best-fit model for predicting demand for flight OD level passengers based on historical data.

2.4 A technique for predicting airfare in poor countries

When it comes to purchasing plane tickets, customers (the buyers) are always on the lookout for the greatest deals, while airlines (the sellers) are constantly seeking for methods to increase their profits by raising the pricing of the same services. In order to decide whether or not to raise or lower tickets at various points before departure dates, All of the necessary information is in the hands of the vendors (for example, past sales, market demand, consumer profile and behaviour). Buyers, on the other hand, may only see a limited amount of information to help them decide whether or not to purchase airfare immediately. Our research in this article aims to develop a novel model that may assist buyers in anticipating airline fare patterns without relying on official airline data. Using publicly available ticket data, our findings indicate that the proposed model can estimate trends and actual price changes up to the departure dates, even if important metrics like the number of unsold seats on flights are missing.

2. Existing System

The existing system calculates only one side predicted fare. It uses random forest machine learning algorithm which has less accuracy , high error and overall speed is less compared to the new machine learning algorithms.

4. Problem Statement

Flight ticket costs are notoriously difficult to predict; although one day we may see one price, the next day we may see a completely different one.In order to resolve this issue, we've received data on the

pricing of airline tickets for various months and places, and our goal is to use this data to develop a model that can estimate travel costs based on numerous input variables.

5. Proposed system

The proposed approach is using machine learning algorithms and the one which we are using isExtraTree classifier algorithm. We are gathering our data from a site. The data is containing some of the details of Indian flight of a short duration. This project involves Feature engineering for processing the dataset (data) convert it into data frame. It's now time to normalise the data frame after processing. We've chosen Extreme Gradient Boosting as the regression model for our forecast. We are training model with the normalized data frame. After experimenting and tuning the hyper parameters we are obtaining predicted results and the accuracy.

6. Implementation

The project is heavily reliant on statistical and machine learning techniques. Python and R were used to build the models and automate the process.

1. Automated Script to Collect Historical Data

We require historical data in order to solve any prediction or classification challenge. Flight fares for each route must be obtained on a daily basis for this project. A python script was executed on a remote server to gather prices every day at a predetermined time and date.

2. Cleaning & Preparing Data

In order to make use of the data, we must first clean and prepare it. For each machine learning activity, this is both the most crucial and time-consuming step. Various statistical methods and logics were used, all of which were implemented using R's built-in packages.

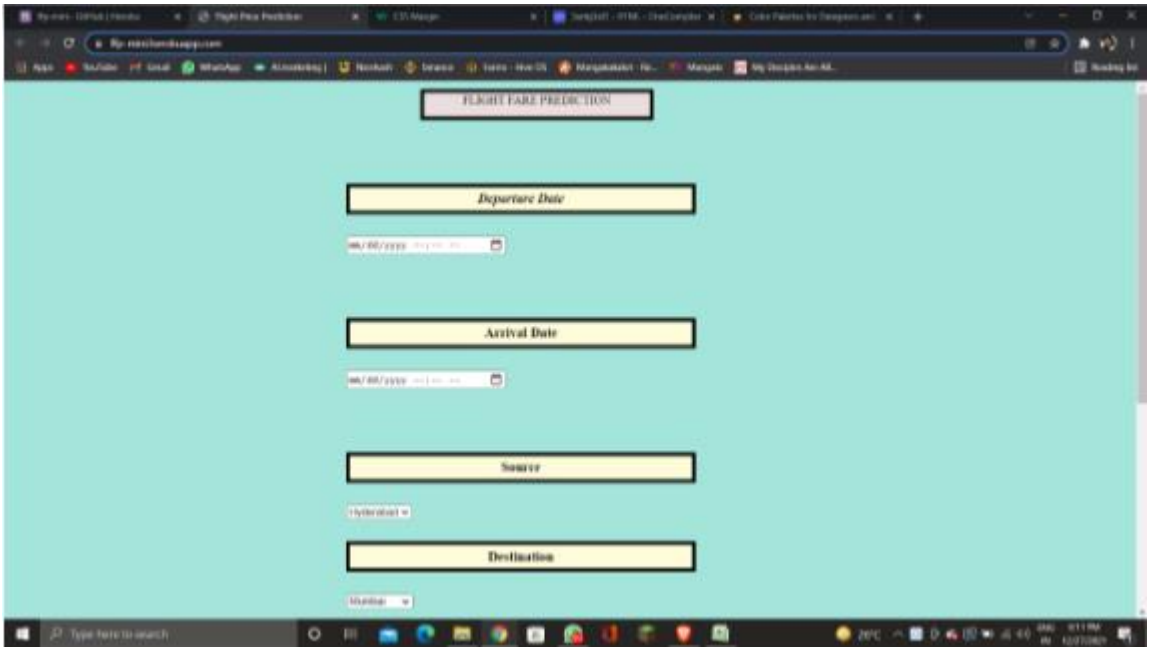
3. Analysing & Building Models

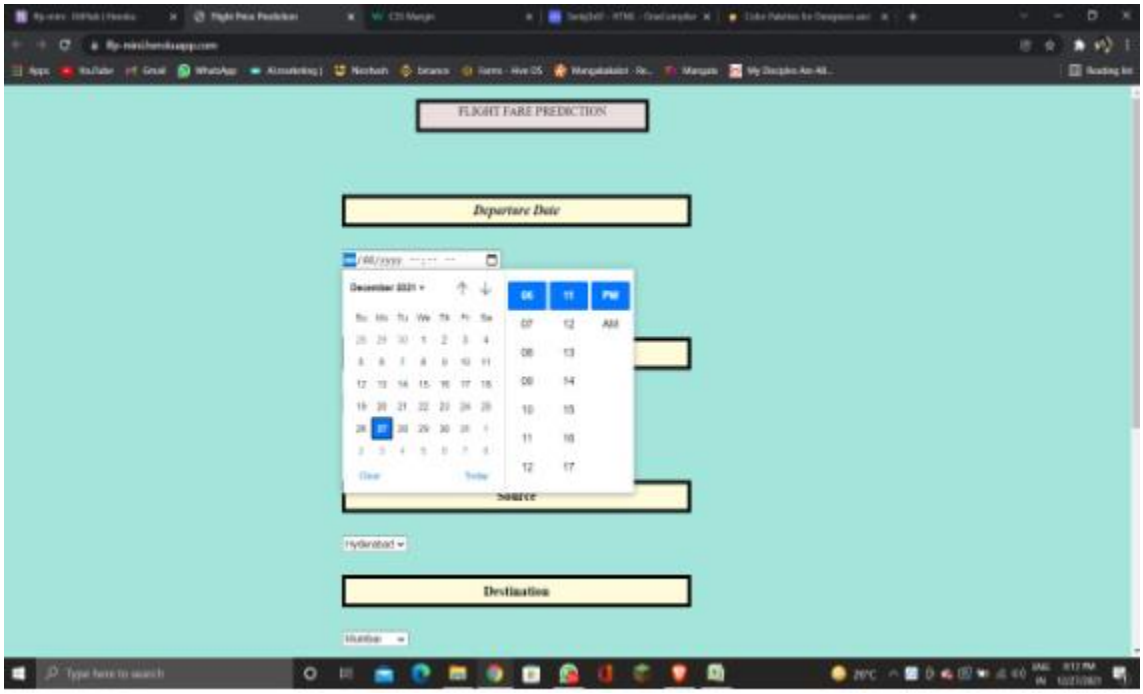
After preparing the data, several prediction and classification models are applied to the training set in order to reveal hidden patterns.

4. Merging Models & Accuracy Calculation

Now that we've developed a variety of models, it's time to put them to the test on our testing set and choose the most accurate measure to use. It's also common for many models to be combined to forecast a single target variable that is more accurate.

7. OUTPUT SCREENS


FLIGHT FARE PREDICTION
Departure Date
Arrival Date
Source
Destination


FLIGHT FARE PREDICTION
Departure Date
Source
Destination

8. CONCLUSION

An inexperienced traveller may be able to save money if this project is successfully performed since it gives data on airline pricing trends and an anticipated value for the price. In the end, this service can be implemented with a high degree of precision. There is a lot of room for development in this kind of service since the forecasted value isn't always correct. We may conclude the following based on the data gathered and exploratory data analysis: Flight fares fluctuate during the course of the year and around the holidays. Airlines fall into two categories: those that are cheap and those that are splurgey. Indian airlines Jet Airways and Air India are in the premium class, whereas Spicejet, AirAsia a, IndiGo, and Go Air are in the budget class. The trend on Vistara is more dispersed. Time of departure affects the cost of the flight, hence the timeslot employed in an analysis is critical. During the Christmas season, the cost of flying goes up. For all values of days to departure, the fare was higher during Diwali in our era. Our data is just a few months old at this point, and hence we haven't taken into account the

holidays. Traveling on a certain day of the week affects the cost of air travel. It is higher on weekends and Mondays, and lower during the rest of the week. Airlines sometimes conduct promotions that lead their prices to decrease dramatically. As a result of this, our mathematical models tend to be inaccurate. The Mumbai-Delhi flight fare fluctuates according on how many days remain till travel. This might be due to a combination of factors, including the large number of flights, strong demand, or even intense competition.

9. REFERENCES

[1] J. Santos Dominguez-Menchero, Javier Rivera and Emilio TorresManzanera"Optimal purchase timing in the airline market".

[2] K. Tziridis T. KalampokasG.Papakostas and K. Diamantaras"Airfare price prediction using machine learning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO .2017.8081365L. Li Y. Chen and Z. Li" Yawning detection for monitoring driver fatigue based on two cameras" Proc. 12th Int. IEEE Conf. Intell. Transp. Syst. pp. 1-6 Oct. 2009.

[3] SupriyaRajankar, Nehasakhrakar and Omprakashrajankar"Flight fare prediction using machine learning algorithms" International journal of Engineering Research and Technology (IJERT) June 2019.

[4] Tianyiwang, samiraPouyanfar, haimanTian and Yudong Tao "A Framework for airline price prediction: A machine learning approach"

[5] T. Janssen "A linear quantile mixed regression model for prediction of airline ticket prices"

[6] William Groves and Maria Gini"An agent for optimizing airline ticket purchasing" in proceedings of the 2013 international conference on autonomous agents and multi-agent systems.

[7] Wohlfarth, T.clemencon, S.Roueff"A Data mining approach to travel price forecasting" 10th international conference on machine learning Honolulu 2011.

[8] medium.com/analytics-vidhya/mae-mse-rmse-coefficient-ofdetermination- adjusted-r-squared-which-metric-is-bettercd0326a5697e article on performance metrics

[9] www.keboola.com/blog/random-forest-regression article on random forest

[10] https://towardsdatascience.com/machine-learning-basics-decisiontree-egression-1d73ea003fda article on decision tree regression