JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Airfare PREDICTION USING MACHINE LEARNING

[1]Mr. Manoj Kumar S , [2] Brijesh S, [3] Nikhil M, [4] Narasimha Raju R, [5] Rayyaan Mohiaddin

[1]Assistant Professor, [2,3,4,5] Undergraduates
[1]Department of Computer Science and Engineering
[1]K S Institute of Technology, Bangalore, India

Abstract: A number of factors influence the cost of an airline ticket, including flight distance, purchasing time, fuel price, and so on. Each carrier has its own set of proprietary rules and algorithms for determining pricing. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) allow for the inference of such rules and the modelling of price variation. This paper proposes a novel application based on two public air transportation data sources: the Airline Origin and Destination Survey (DB1B) and the Air Carrier Statistics database.
The proposed framework combines the two databases, as well as macroeconomic data, and employs machine learning algorithms to model the quarterly average ticket price based on various origin and destination pairs, referred to as the market segment. On the testing dataset, the framework achieves a high prediction accuracy of 0.869 adjusted R squared.

Price Prediction, Data Analysis, Random Forest Regressor, Testing, and Training are all index terms

I. INTRODUCTION

Since the airline industry's deregulation, airfare pricing strategy has evolved into a complex structure of sophisticated rules and mathematical models that drive airfare pricing strategies. Despite the fact that these rules are still largely unknown, studies have revealed that they are influenced by a variety of factors. Traditional variables such as distance, while still important, are no longer the sole determinants of pricing strategy. Economic, marketing, and societal trends have all played a growing role in determining airfare prices.Airline Ticket prices for the same flight fluctuate dramatically and significantly from day to day. It is extremely difficult for a customer to purchase an airline ticket at the lowest possible price because the price changes dynamically. Customers can also use septimate analysis (reviews) to help them make decisions about which airlines to fly by analysing the opinions of other customers. The goal of this research is to better understand the factors that influence airfare and to develop and fine-tune models that can predict airfare well in advance. The ultimate goal of the airlines is to make a profit, while the customer seeks the lowest possible cost.Typically, customers try to book their tickets well in advance of the departure date to avoid price increases as the date approaches. But, in reality, that is not the case. The customer can save money by paying more than they should for the same seat.

LITERATURE SURVEY

[1] A survey on machine learning-based flight pricing prediction.
Supriya Rajankar and Neha Sakharkar
The goal of this paper was to find a machine learning model that could more accurately forecast the fare of Indian flights. After experimenting with various models, it was discovered that the Random Forest algorithm predicted the outcome with the highest accuracy. In terms of results, the article outperforms models, and it aspires to improve in the future.

[2] A PROPOSAL FOR INDIAN FLIGHT FARE PREDICTION
Udhhav Arora, Jaywrat Singh Champawat, and Dr. K. Vijaya
Several academics have used various Machine Learning techniques to develop a model that predicts ticket prices more accurately. s Researchers have used a variety of regression models to estimate accurate airline fares, including Support Vector Machines (SVM), Linear Regression (LR), Decision Trees, Random Forests, and others.

[3]Survey on Machine Learning Algorithms for Air Price Prediction
Zubeda A Khan, Abhilash, Ranjana Y, Shilpa S, and Zubeda A Khan are the authors of this paper.
Random Forests are a versatile tool for solving regression and classification problems, including multiclass classification; they provide an internal estimate of generalisation error, removing the need for cross validation. They can be fine-tuned, but with the

default tuning parameters, they frequently perform admirably. Random Forests have been successfully applied in a wide range of applications and are widely used in a wide range of disciplines.

[4] Using Machine Learning to Predict Flight Ticket Prices
Jibin Joseph, Abhijith P, Aryasree S, Jinsu Anna Joseph, Meghana Sara Oommen, and Abin T Abraham are abbreviations for Jibin Joseph, Abhijith P, Aryasree S, Jinsu Anna Joseph, Meghana Sara Oommen, and Abin T Abraham.
A few techniques are investigated in order to determine the best time and date to purchase low-cost airline tickets. They used Partial Least Square Regression (PLSR) to create a model for determining the best time to buy a plane ticket. The vast majority of these systems make use of Machine Learning, a modern computerised technology.

[5] Using Machine Learning Algorithms to Predict the Cost of a Flight Ticket.
Neha Sakharkar, Supriya Rajankar, and Omprakash Rajankar
These few processes are investigated in order to determine the best time and date to buy low-cost airline tickets. The vast majority of these systems rely on Machine Learning, a modern computerised technology. Gini and Groves estimated the best time to buy a plane ticket using partial least square regression (PLSR).

[6] Using Machine Learning to Predict Airfares
Yash Shah, Sonia Relan, Vinal Raja, and Janhavi Vaki are among those who have appeared in the film.
We use logistic regression, a classification procedure, to assign observations to a discrete set of classes. Logistic regression modifies the data, as opposed to linear regression, which produces continuous numbers. The sigmoid function returns a probability value that can be transferred to two or more discrete classes.

II. DATASET

The project's data source is, which contains detailed information on flight bookings made via the Amadeus system. The dataset includes booking information such as airlines, Date of Journey, source, destination, route, departure time, Arrival Time, Duration, Total Stop, and Price.

Sl.no	attributes	Datatypes
1.	Airline	Object
2.	Date_of_Journey	Integer
3.	Source	Object
4.	Destination	Integer
5.	Route	Object
6.	Dep_Time	Float
7.	Arrival Time	Float
8.	Duration	Integer
9.	Total Stops	string
10.	Additional Info	Object
11.	Price	Float

Fig 1: Description of dataset

III. METHODOLOGY

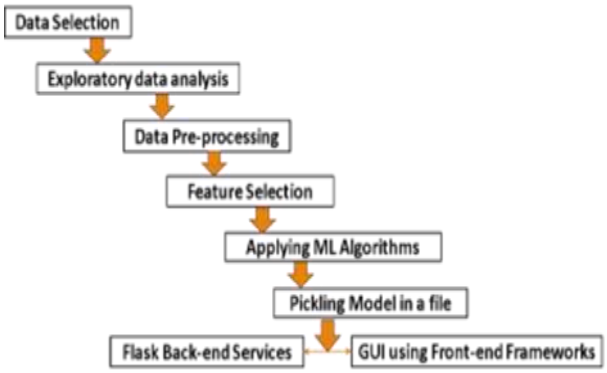

Fig 2: Methodology used for our proposed Airfare Prediction Using Machine Learning

1. **Dataset:**
   Data on aircraft traffic is required for the airline ticket price model at the market segment level.
   as well as the number of passengers in each market segment Data from 2018 is being used to train and evaluate.the suggested model Kaggle, a free platform for data scientists and machine learning enthusiasts, provided the data.

2. **Exploratory data analysis:**
   The dataset's attributes are all present. Using library math, we examine how the data is related to one another. lib, drawing, histogram scatter plot, and many other options are available. This plot shows how the independent variables and dependent variables were chosen.

3. **Data selection:**
   The first stage is to collect historical flight data in order to train the algorithm to predict pricing. Our dataset contains over 10,000 pieces of data related to flights and their pricing. The dataset includes the following elements: source, destination, departure date, departure time, number of stops, arrival time, prices, and a few others.
   We cleaned the dataset in the exploratory data analysis step by removing duplicate and null values. If these values are not removed, the model's accuracy will suffer. We obtained additional information, such as data distribution.

4. **Feature selection**
   The feature selection process includes selecting significant qualities that are more closely related to pricing. Some components, such as additional information and a route, are superfluous and may jeopardise the accuracy of the models chosen and distributed to the group. Following the selection of attributes that are more closely related to price, the next stage is to create a model using a machine algorithm. Because our dataset contains labelled data, we will employ supervised machine learning techniques. Because our dataset contains continuous values in the feature, we will also use support vector machine, random forest, and decision algorithms.

5. **Applying ML algorithm:**
   In our project, we will test various algorithms to determine which one is the most efficient. We will employ the support vector machine, the random forest algorithm, and the decision tree algorithm. We examined the above algorithms and discovered that the random forest algorithm is more efficient and significant.

IMPLEMENTATION

Support vector machine (SVM): Machine learning requires predicting and classifying data, and we use a variety of machine learning methods to do so depending on the dataset.

To address classification and regression issues, a linear model known as the Support Vector Machine (SVM) can be used. It is useful in a variety of situations and can solve both linear and nonlinear problems. SVM's basic idea is that it generates a line or hyperplane that divides data into classes.

I'll give a high-level overview of SVMs in this blog post. I'll go over SVM theory, how they can be used to separate non-linearly separable datasets, and a quick example. SVM implementation in Python. In subsequent articles, I'll dissect the algorithm and look at the underlying arithmetic.

Random algorithms: The supervised learning method is used by Random Forest, a well-known machine learning algorithm. It can be used in machine learning for both classification and regression problems. It is based on ensemble learning, a technique for combining multiple classifiers to solve a complex problem and improve the model's performance.

"Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to boost the dataset's projected accuracy," according to the name. Rather than relying on a single decision tree, the random forest collects forecasts from each tree and predicts the final output based on the majority votes of projections. The larger the font, the better.

Decision tree: The Decision Tree is a supervised learning technique that can be used to solve classification and regression problems. It is most commonly used to solve classification problems. Internal nodes represent dataset properties, branches decision rules, and each leaf node represents the conclusion.
A Decision Tree has two nodes: the Decision Node and the Leaf Node. Decision nodes are used to make decisions and have many branches, whereas Leaf nodes are the results of such decisions and have no additional branches.
A dataset's properties are used to make decisions or run tests. Depending on the circumstances, there is an issue or a decision to be made. It is called a decision tree because it functions similarly to a decision tree. It is a graphical representation of all the potential solutions to a problem or decision based on a set of criteria. It's called a decision tree because, like a tree, it begins with a root node and grows into a tree-like structure with many branches.

6. **Picking model in a file**: The random technique was used to choose the model for the project because it prevents overfitting in decision trees and increases accuracy. It detects missing values in data automatically. It does not require data normalisation because it uses a rule-based approach**.**

7. **Flash Back-end services** Machine learning approaches are not complete without model deployment, as the model must be made available to the general public or clients. The model can be used in a variety of ways, including through a website built with flask, html, and CSS. Finally, the model is trained, its performance is evaluated, and the best model is deployed via a website.

V. RESULTS AND DISCUSSION

This paper proposed developing a machine learning model that predicts the fare of Indian flights with greater accuracy. Working with various models, it was discovered that the Random Forest algorithm had the highest accuracy in predicting the output. The paper outperforms previously examined models and aims to improve in the future.

VI. ACKNOWLEDGEMENT

We would like to express our heartfelt appreciation and gratitude to Mr. Manoj Kumar S for his insightful suggestions and helpful advice throughout the project's planning and development. We would also like to thank all of the professors, staff, and management at KSIT for their ongoing encouragement and support.

REFERENCES

1. Machine Learning-Assisted Flight Ticket Price Prediction ISSN 2278-3091 Volume 10, No.2, March - April 2021. Jibin Joseph1, Abhijith P2, Aryasree S3, Jinsu Anna Joseph4, Meghana Sara Oommen5, Abin T Abraham6
2. A Machine Learning Approach to Airfare Price Prediction Tianyi Wang, Samira Pouyanfar, Haiman Tian, Yudong Tao, Miguel Alonso Jr., Steven Luis, and Shu-Ching Chen.
3. Prediction of Airline Prices Using Machine Learning Algorithms Abhilash1, Ranjana Y2, Shilpa S3, Zubeda A Khan4 are the authors. ISSN 2321-2004 (Online) Vol. 7, Issue 2, February 2019 ISSN (Print) 2321-5526 IJIREEICE copyright DOI 10.17148/IJIREEICE.2019.7203 14.
4. Prediction of airline ticket prices and demand: A survey Juhar Ahmed Abdella, NM Zaki, Khaled Shuaib, and Fahad Khan King Saud University Journal of Computer and Information Sciences 33 (2021) 375–391.
5. Machine Learning Techniques for Predicting Airfare Prices Based on Reviews Ruchi K. Oza, Aakanksha V. Jain, Aanal S.Raval ISSN: 2319-8753 ISSN: 2347-6710 Vol. 9, Issue 3, March 2020
6. A Survey on Machine Learning for Flight Pricing Prediction, International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV8IS060672 Vol. 8 Issue 06, June-2019 Supriya Rajankar, Neha Sakharkar
7. System for Predicting Flight Fares May 19, 2021: Vinod Kimbhaune, Harshil Donga, Asutosh Trivedi, Sonam Mahajan, and Viraj Mahajan
8. 8. INDIAN FLIGHT FARE PREDICTION: A SUGGESTION Volume 8, Issue 3, March 2021, Jaywrat Singh Champawat1, Udhhav Arora2, Dr. K. Vijaya3, www.ijates.com